THE UNIVERSITY OF BRITISH COLUMBIA

CPSC 425: Computer Vision

Image Credit: Devi Parikh

Lecture 1: Introduction and Course Logistics

Course **logistics**

Instructor: Matthew Brown

About **me …**

I have been working in **Computer Vision** for the last 20 years

Research Scientist
2015—present
Google

Associate Professor
2011–2015
University of Bath

CTO
2009–2015
Cloudburst Research

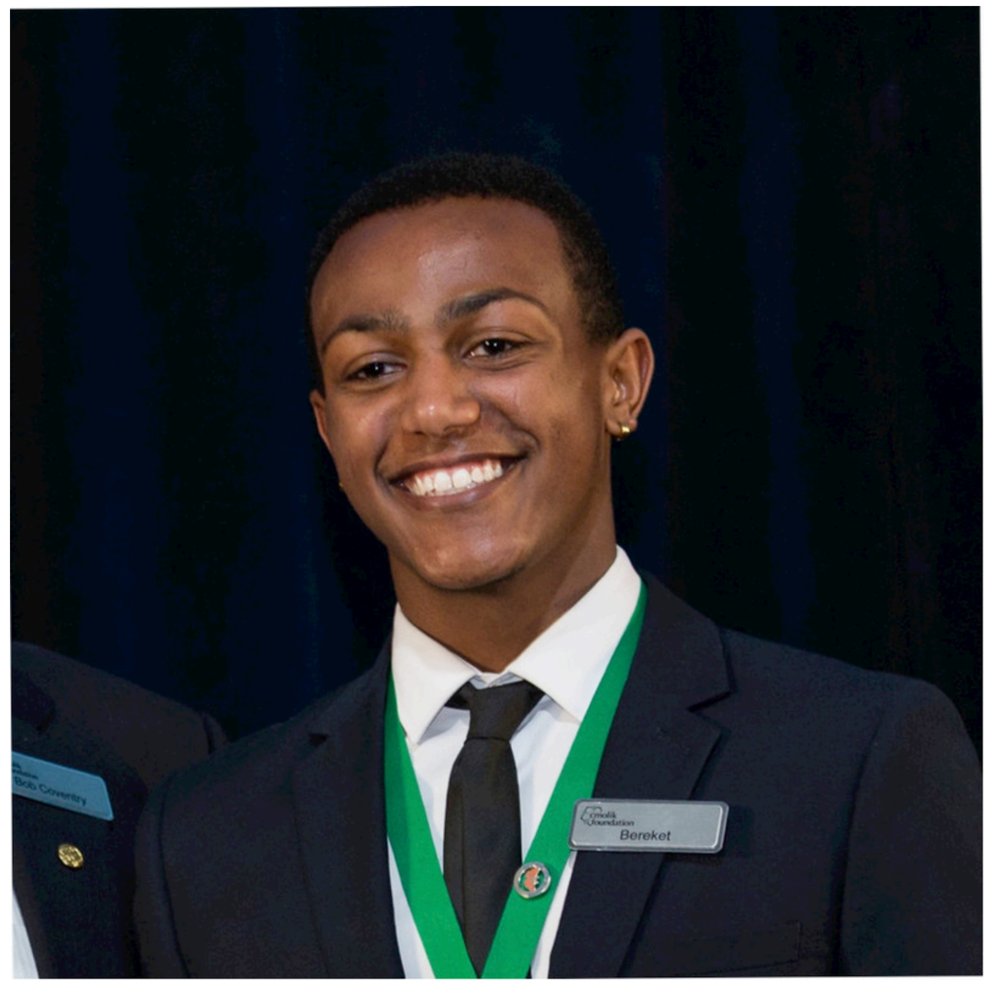
Postdoctoral Researcher
2006–2011
Microsoft Research
EPFL

PhD
2001–2005
UBC The University of British Columbia

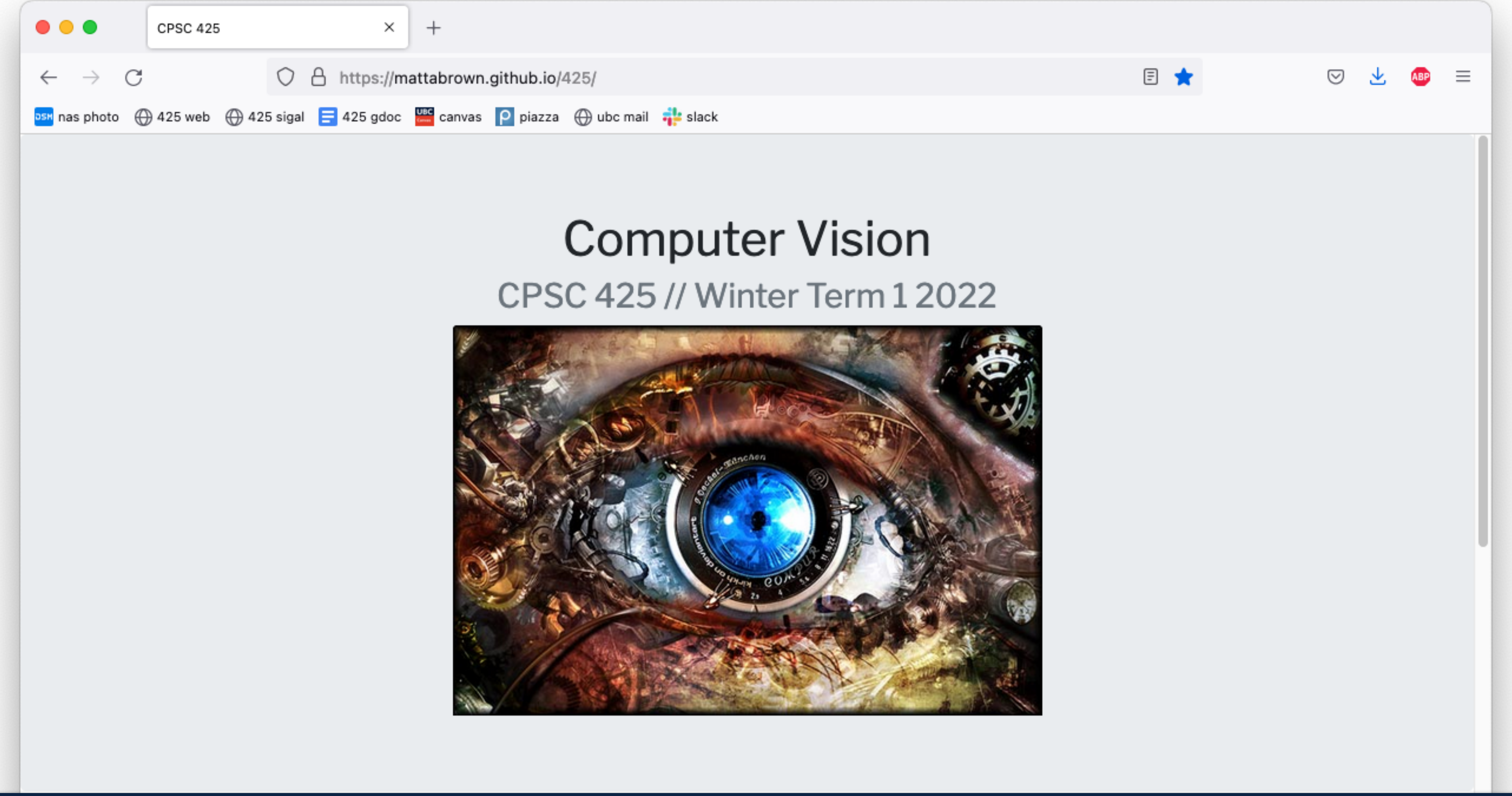
Course **logistics**

Instructor: Matthew Brown

TAs: Bereket Guta, Rayat Hossain, +1 TBC



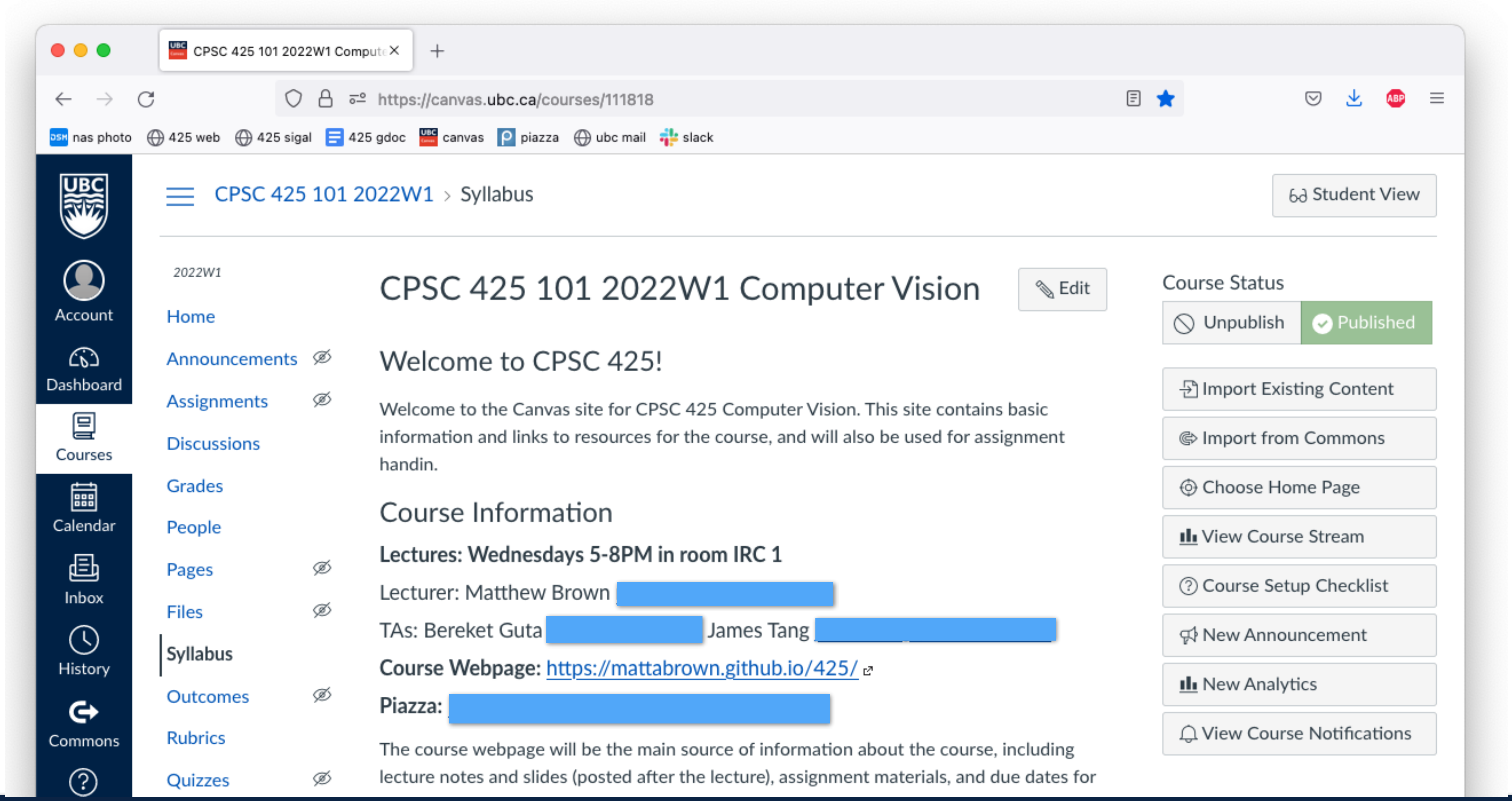
Course Webpage



- Schedule, Assignments
- Lecture Slides and Notes
- Course Information (public)

https://mattabrown.github.io/425

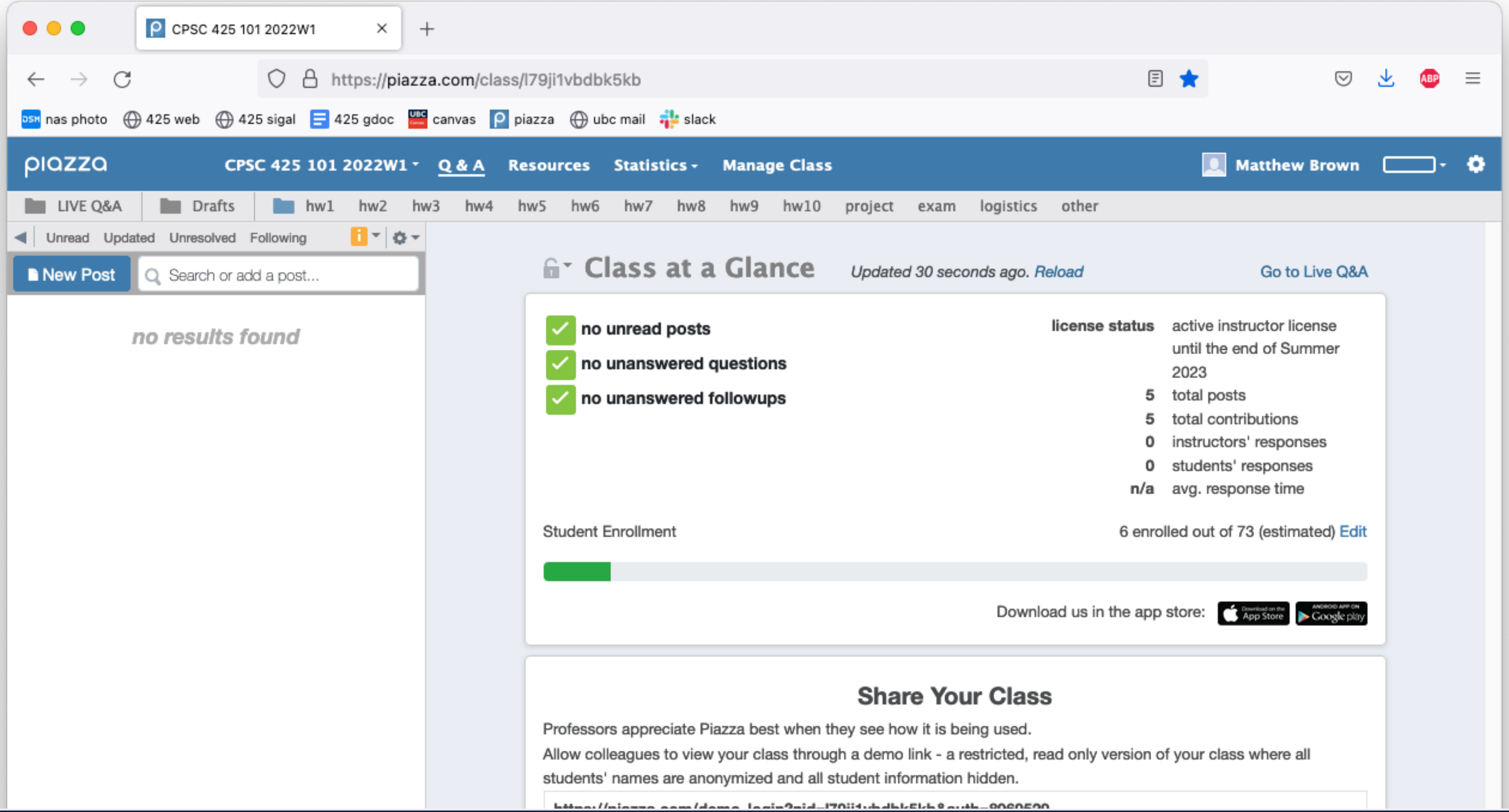
Canvas



- Assignment hand-in
- Course Information (private)
- Piazza link

https://canvas.ubc.ca/courses/123448

Piazza



- Discussions and Q+A
- Confused? Likely someone else has the same question as you!
- Lectures, Technical Issues, Assignments …

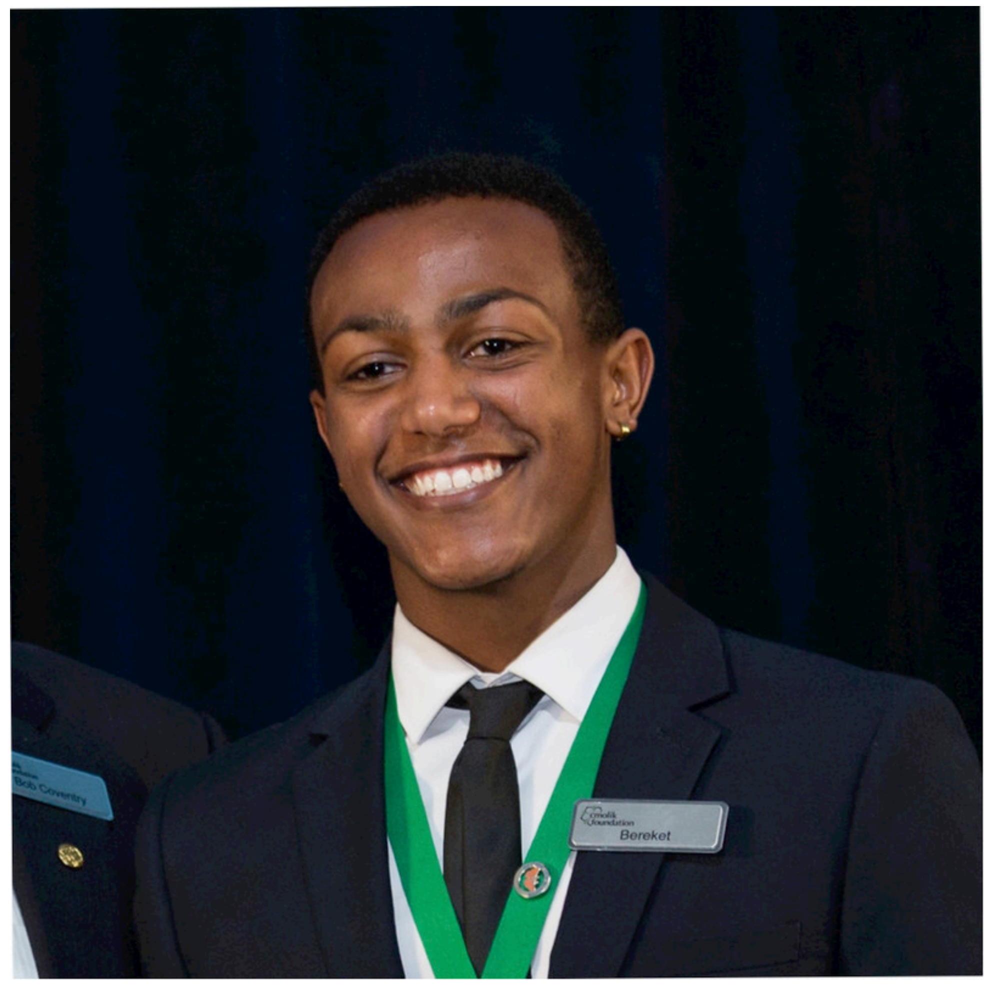
Link in Canvas

Office Hours

Instructor: Matthew Brown

Fri 11-12, Zoom

TAs: Bereket Guta, Rayat Hossain, +1 TBC



Mon 3-4, In Person

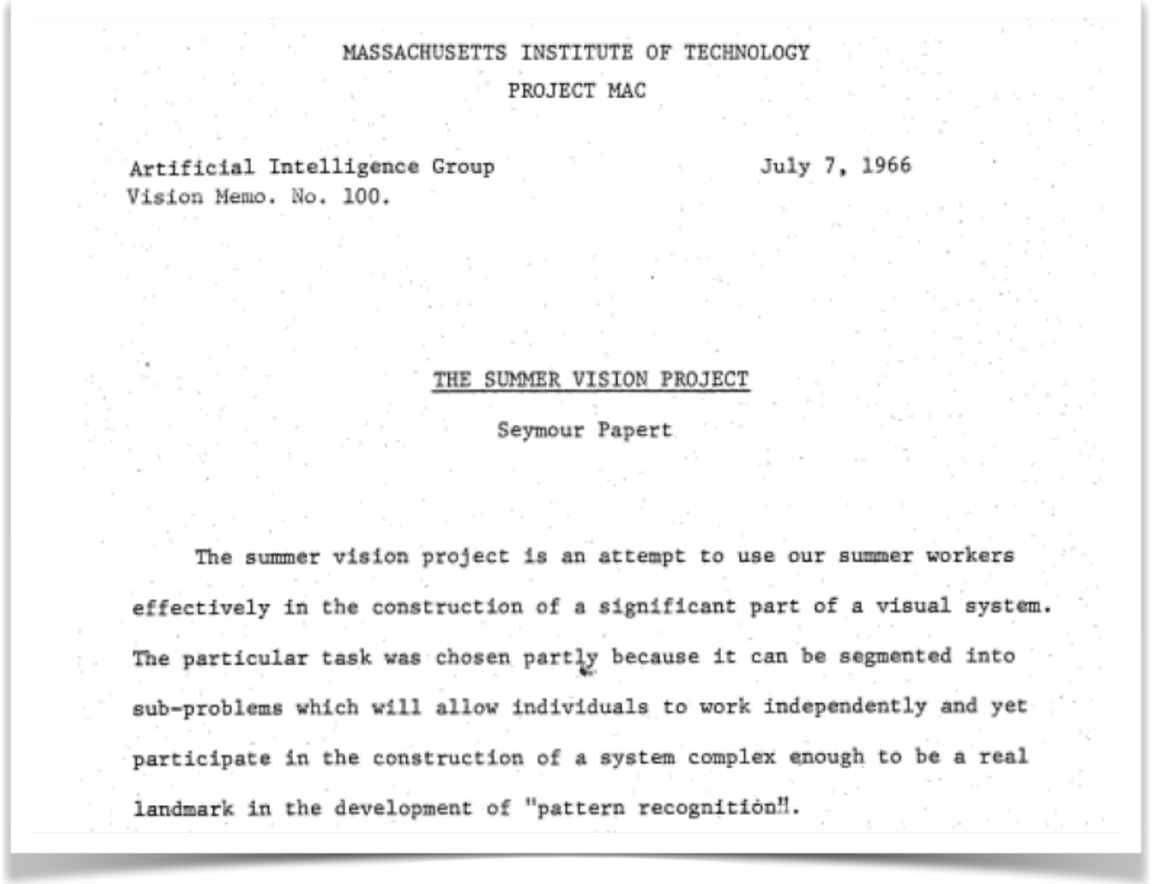
Wed 3-4, In Person

See Canvas for Links and Locations

What is **Computer Vision**?

Image Credit: https://www.deviantart.com/infinitecreations/art/BioMech-Eye-168367549

Computer vision … the beginning …



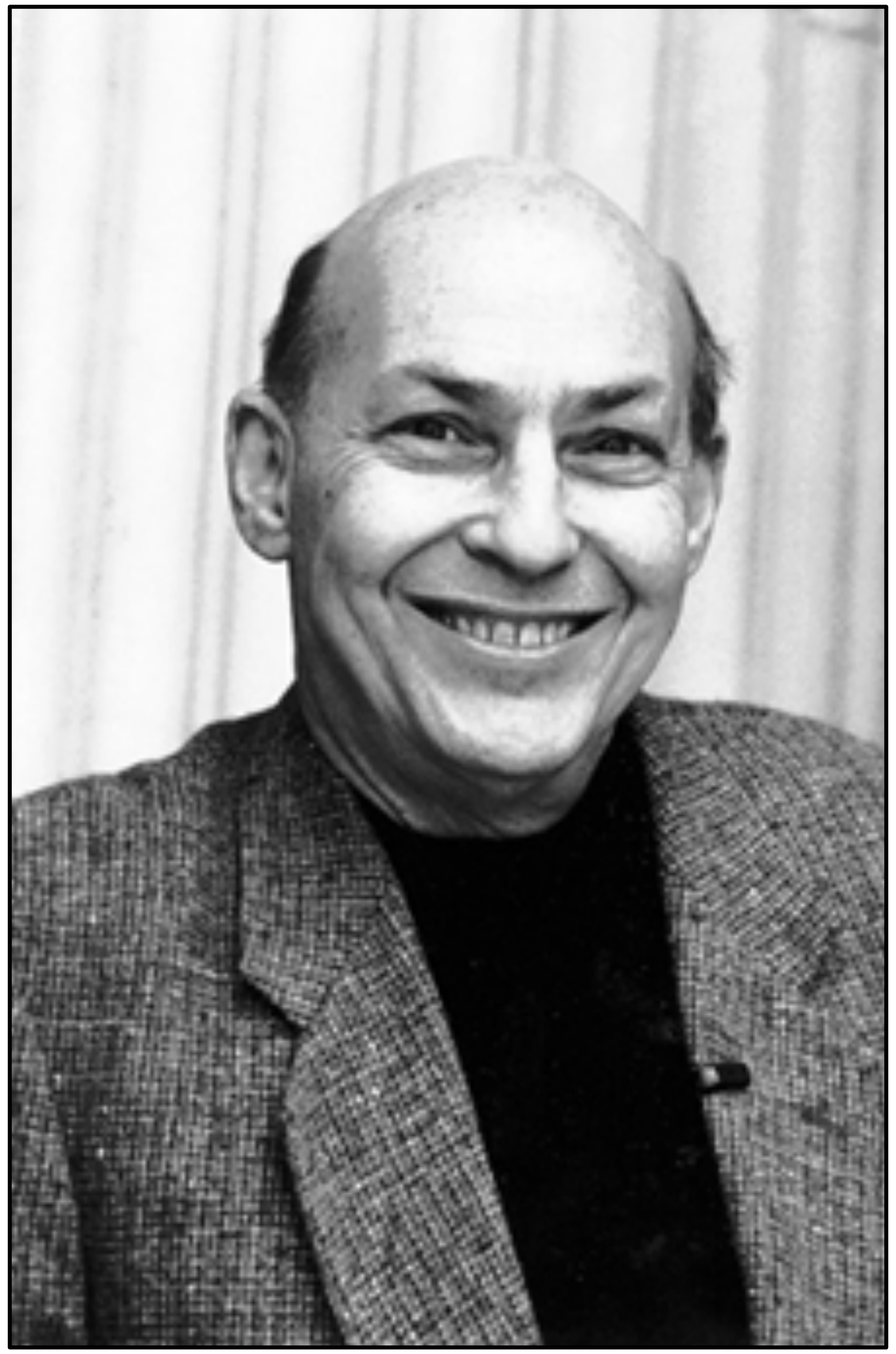
MASSACHUSETTS INSTITUTE OF TECHNOLOGY
PROJECT MAC

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".



The Summer Vision Project

"spend the summer linking a camera to a computer and getting the computer to describe what it saw"

- Marvin Minsky (1966), MIT
  Turing Award (1969)

… >50 years later

Slide Credit: Devi Parikh (GA Tech)

Computer vision … the beginning …

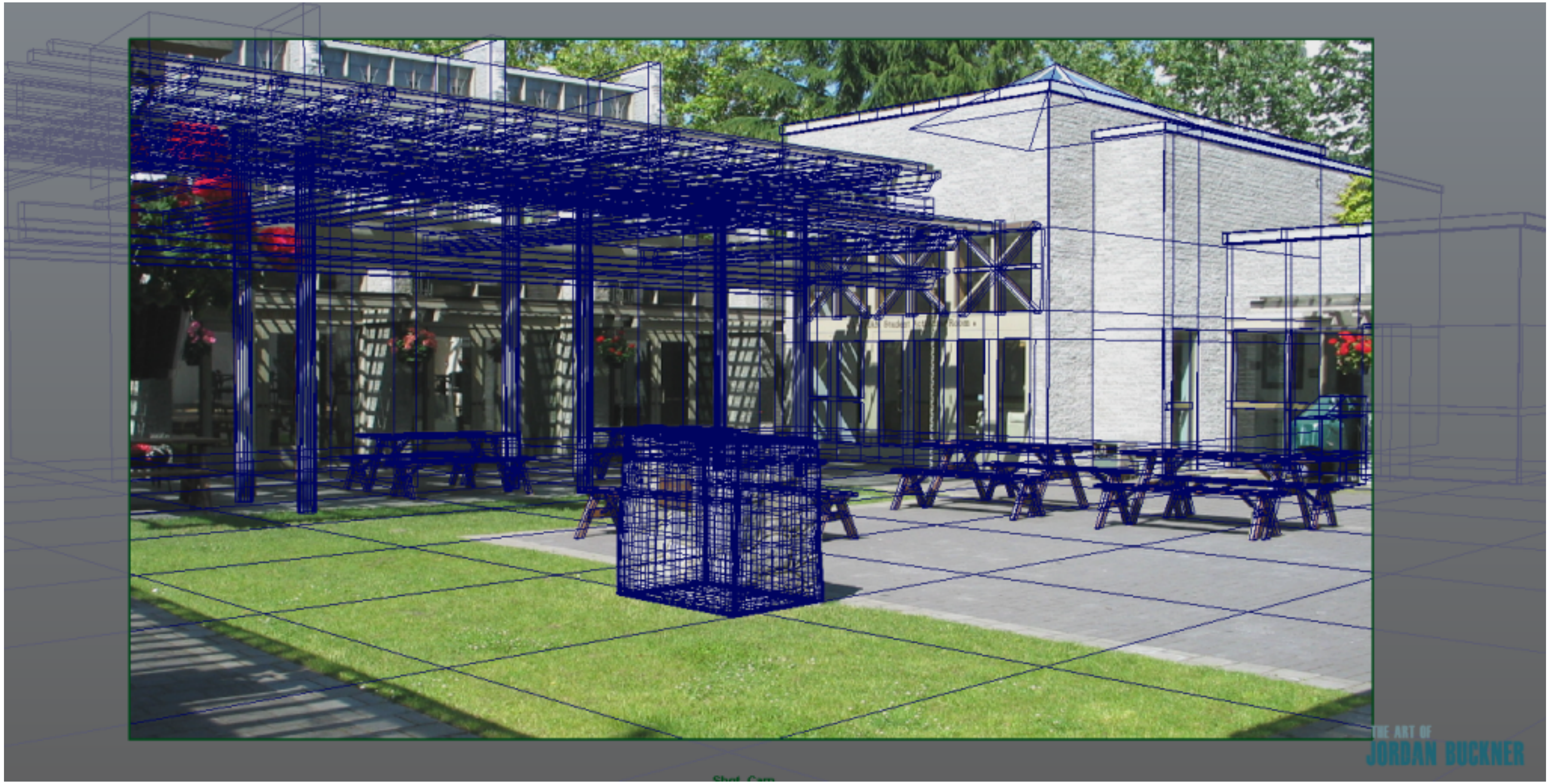
Gerald Sussman, MIT

"You'll notice that **Sussman** never worked in vision again!" – Berthold Horn

Slide Credit: Devi Parikh (GA Tech)

Definitions of Computer Vision #1

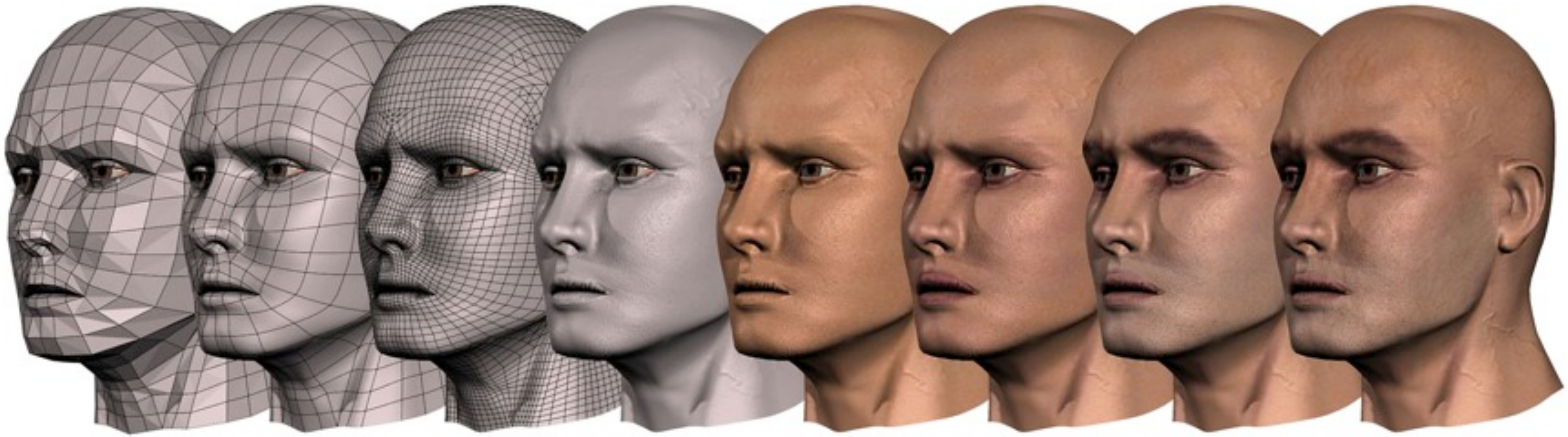
“Inverse Computer Graphics”

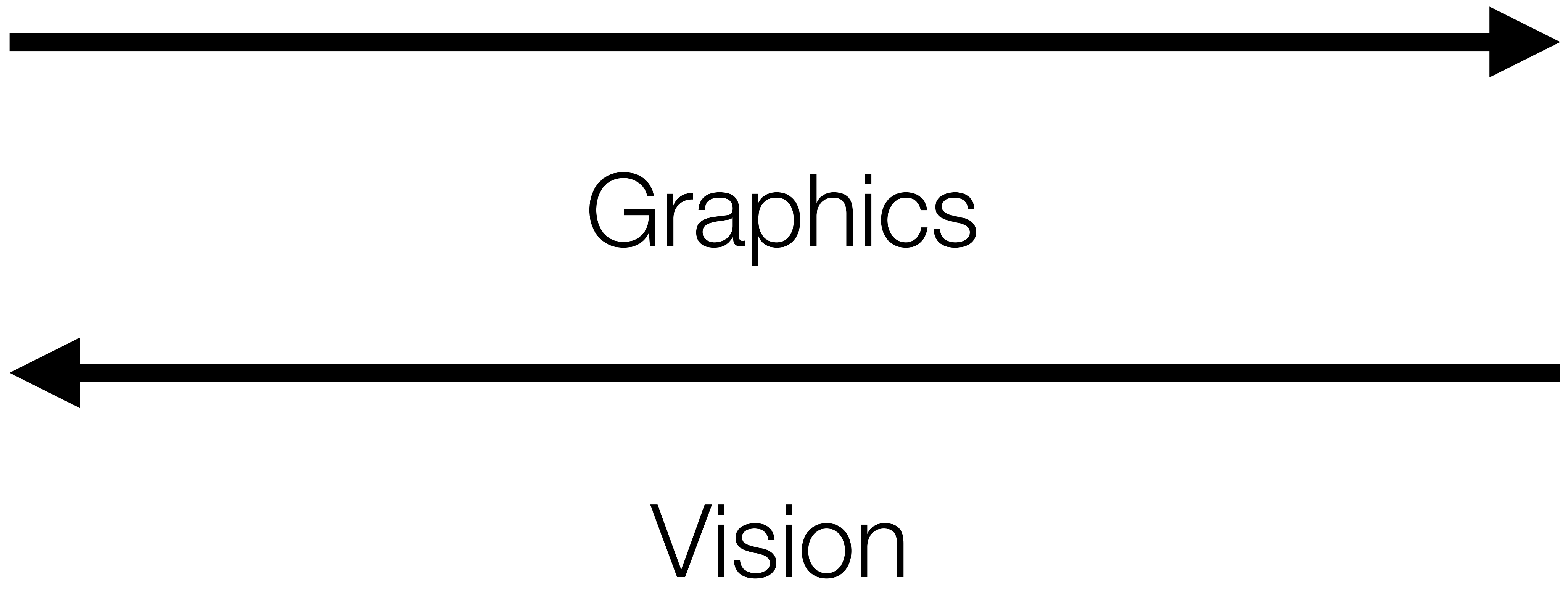


[J. Buckner]

Definitions of Computer Vision #1

"Inverse Computer Graphics"

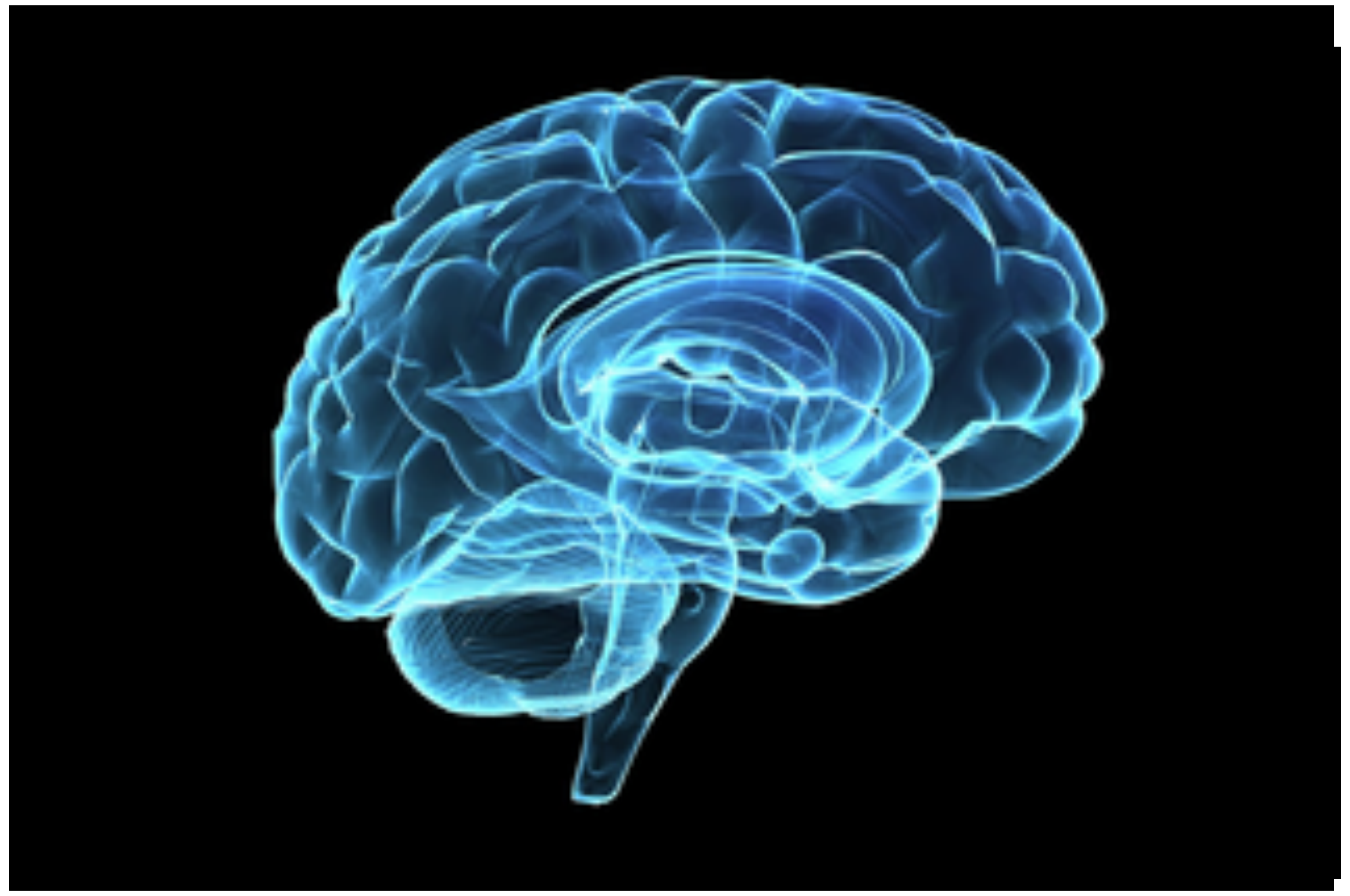




Definitions of Computer Vision #2

“Replicate Human Vision”

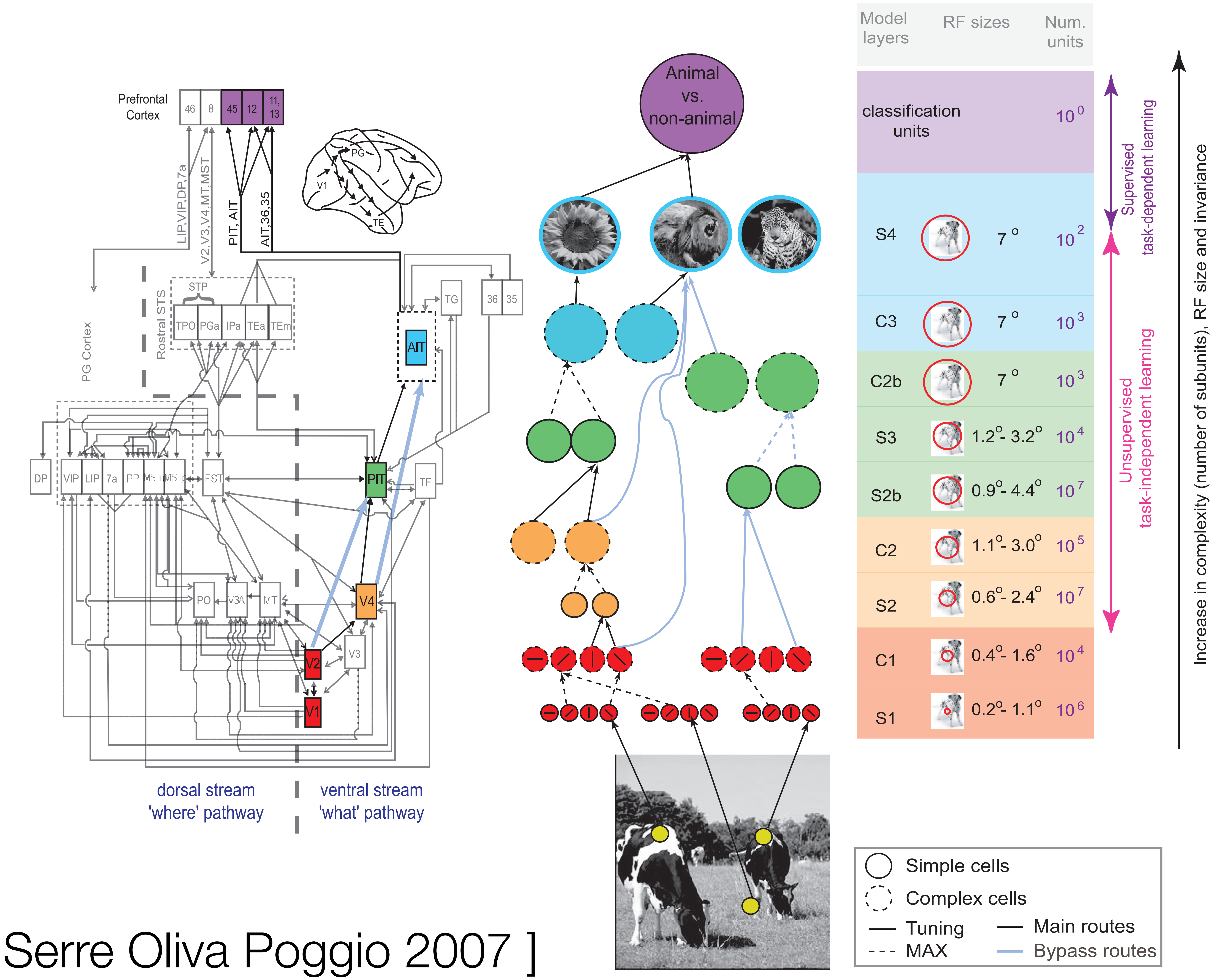
=



=

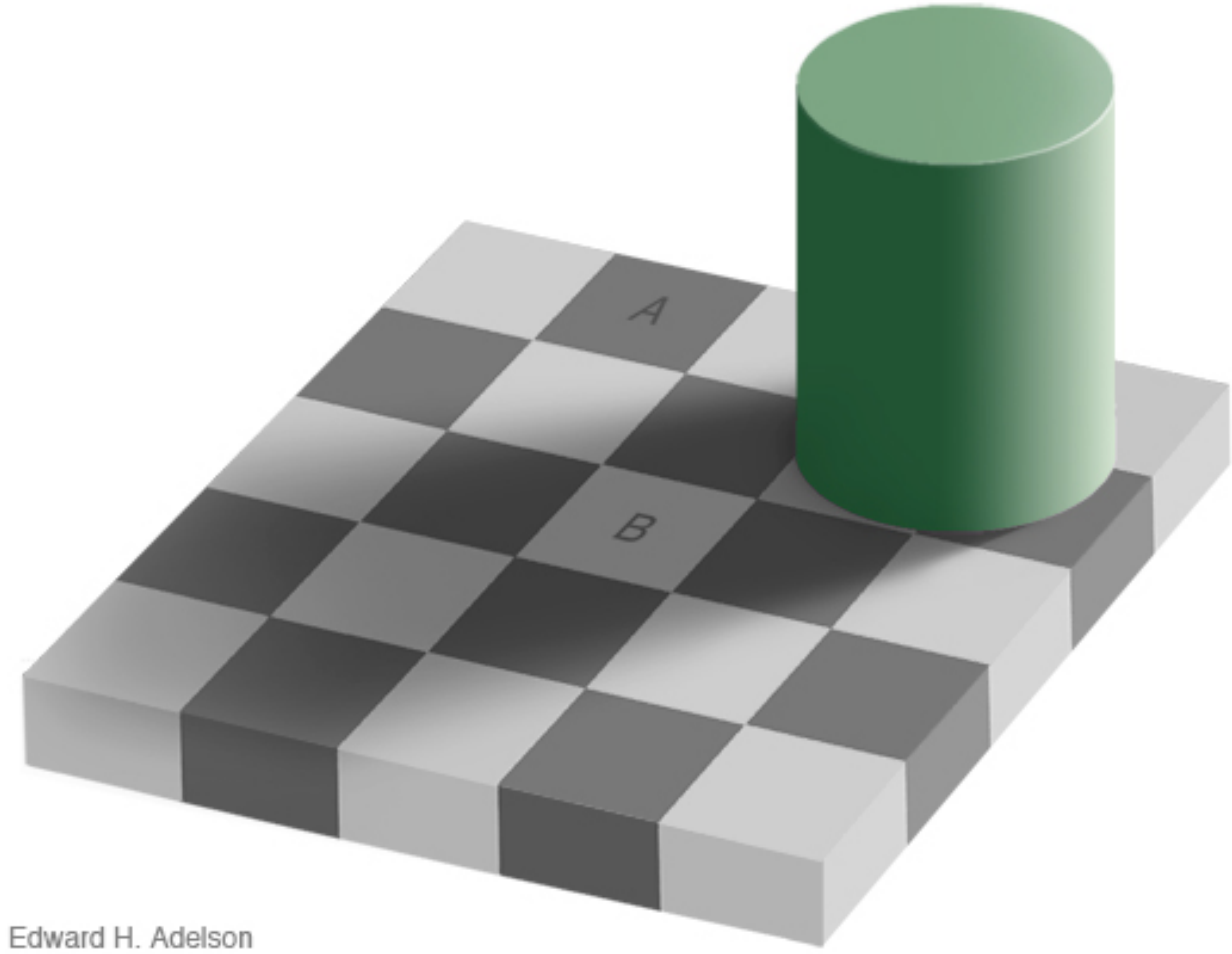
Definitions of Computer Vision #2

“Replicate Human Vision”

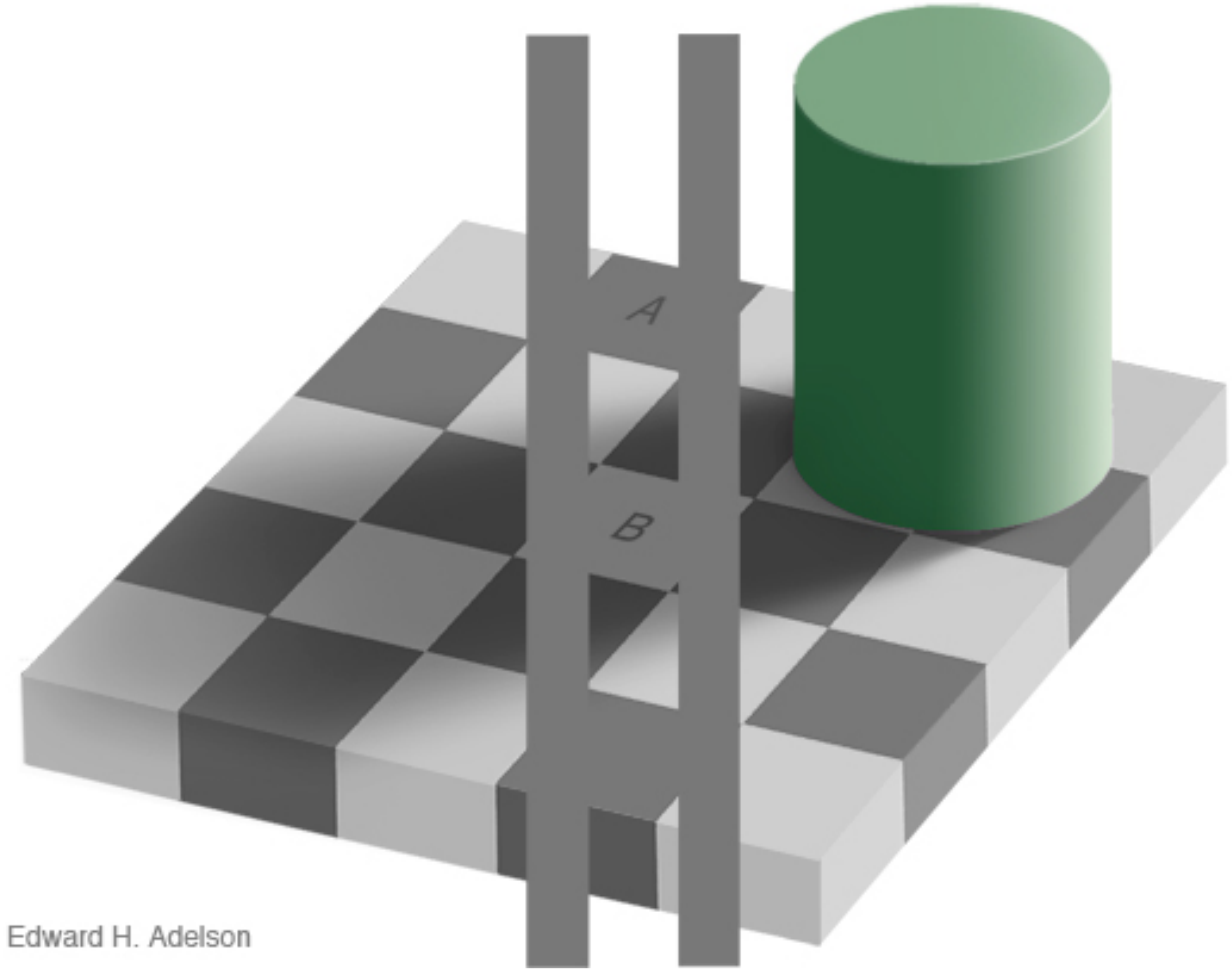


[Serre Oliva Poggio 2007]

Can computers **match (or beat)** human vision?

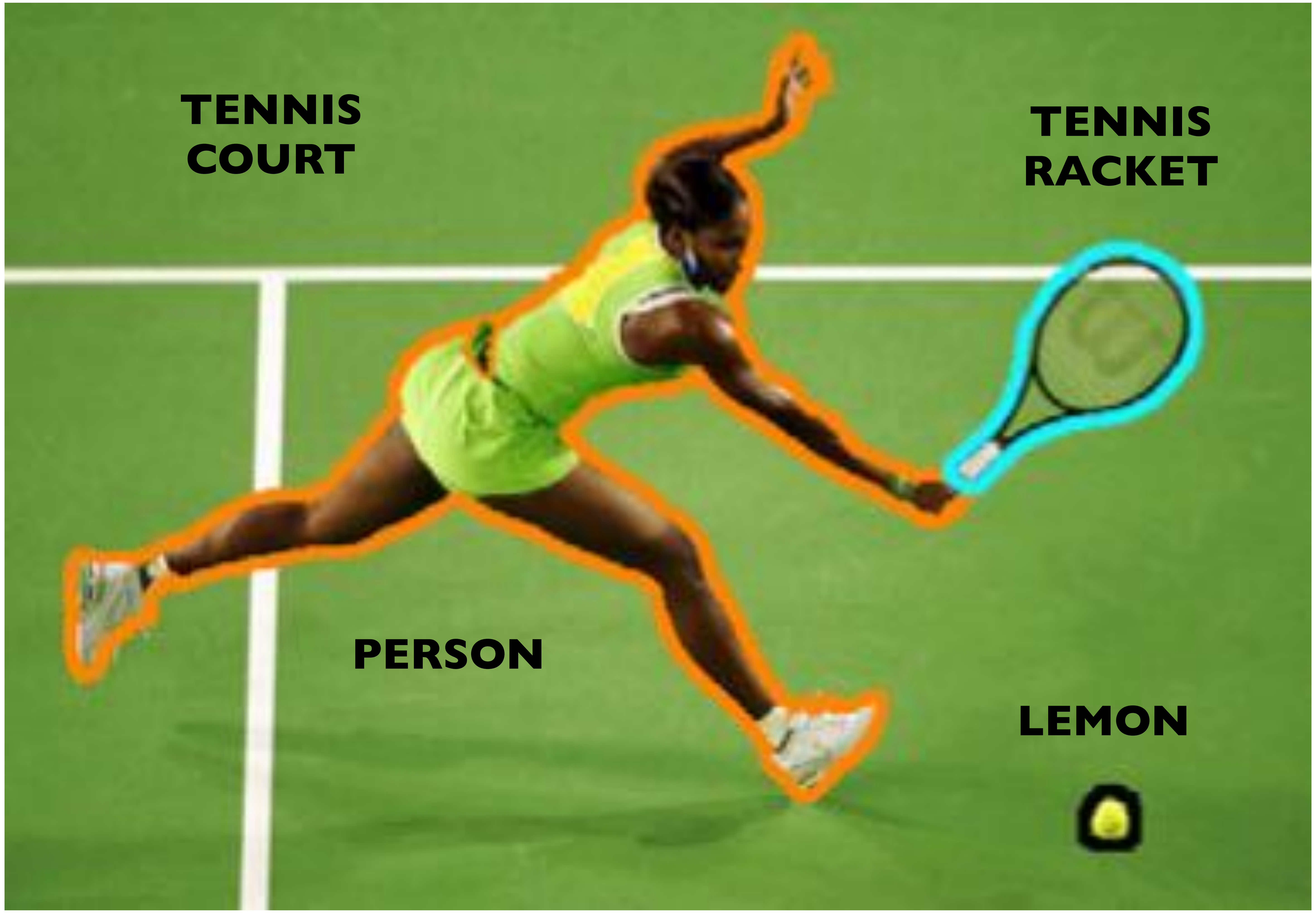


Can computers **match (or beat)** human vision?



Definitions of Computer Vision #3

“Image/Video Understanding”

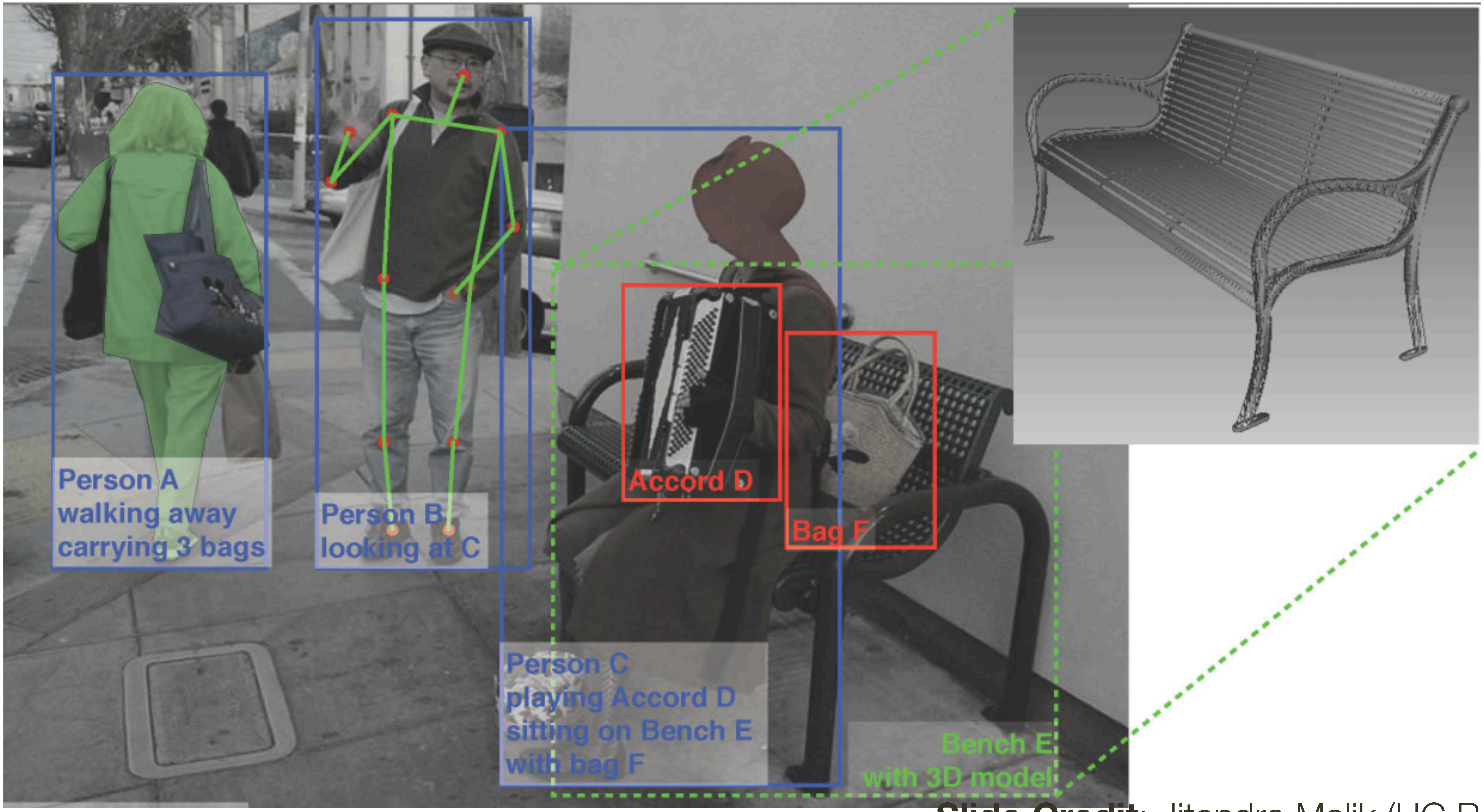


[Rabinovich, Galleguillos, Wiewiora, Belongie 2007]

What do **you** see?

Slide Credit: Jitendra Malik (UC Berkeley)

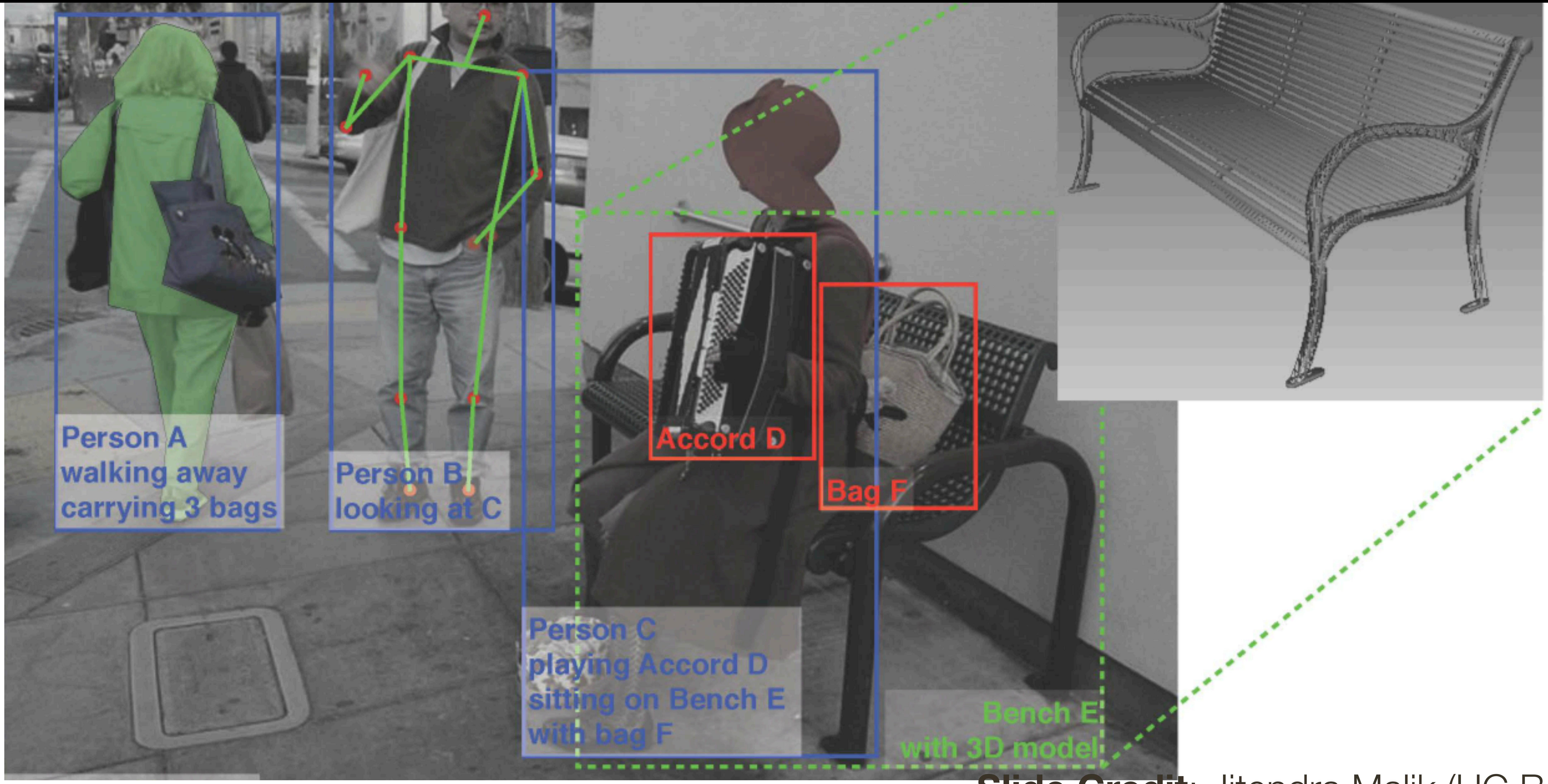
What we would like **computer to infer**?



Slide Credit: Jitendra Malik (UC Berkeley)

What we would like **computer to infer**?

Will person B put some money into person C's cup?



Slide Credit: Jitendra Malik (UC Berkeley)

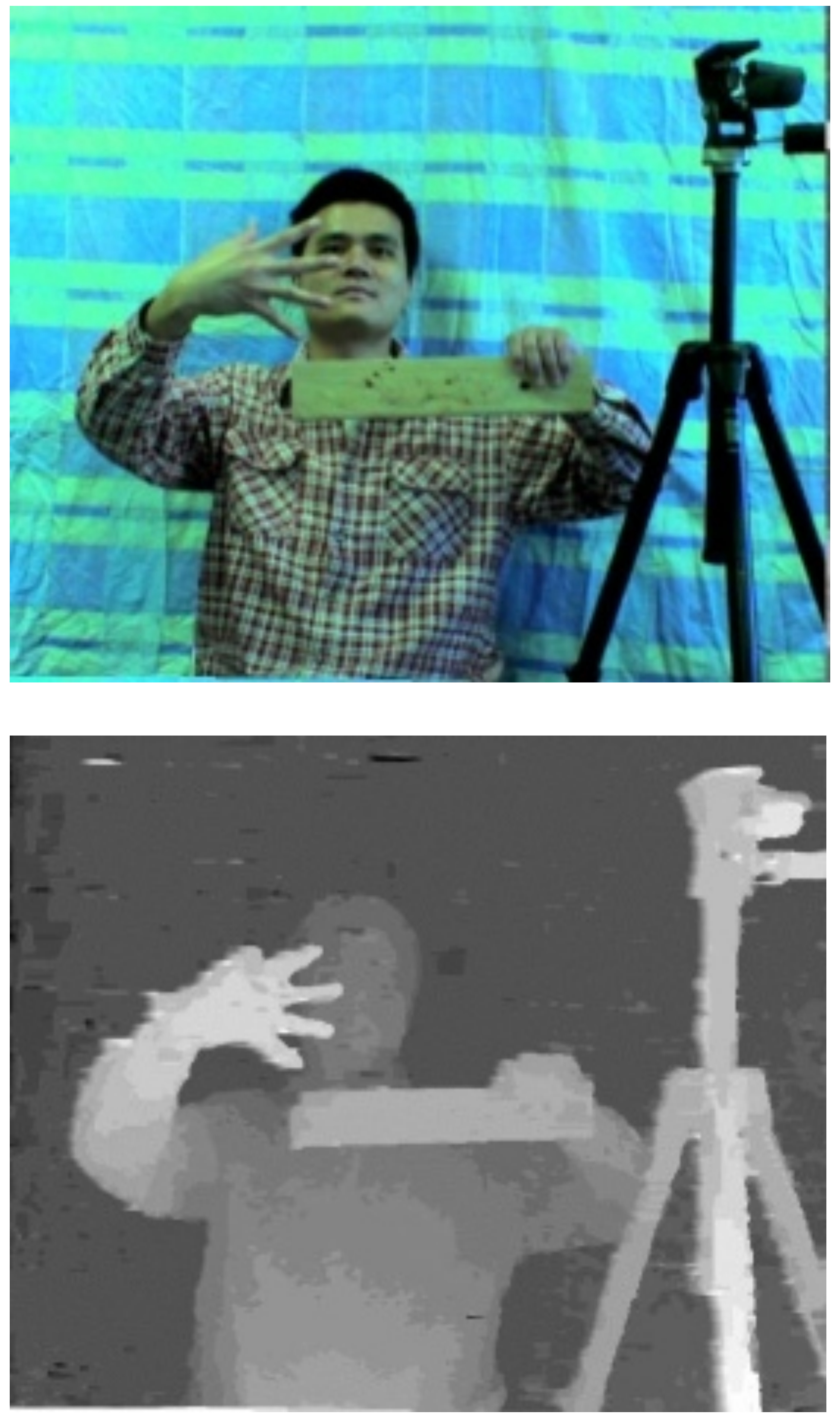
Computer **Vision Problems**

1. Computing properties of the 3D world from visual data (***measurement***)

Slide Credit: Kristen Grauman (UT Austin)

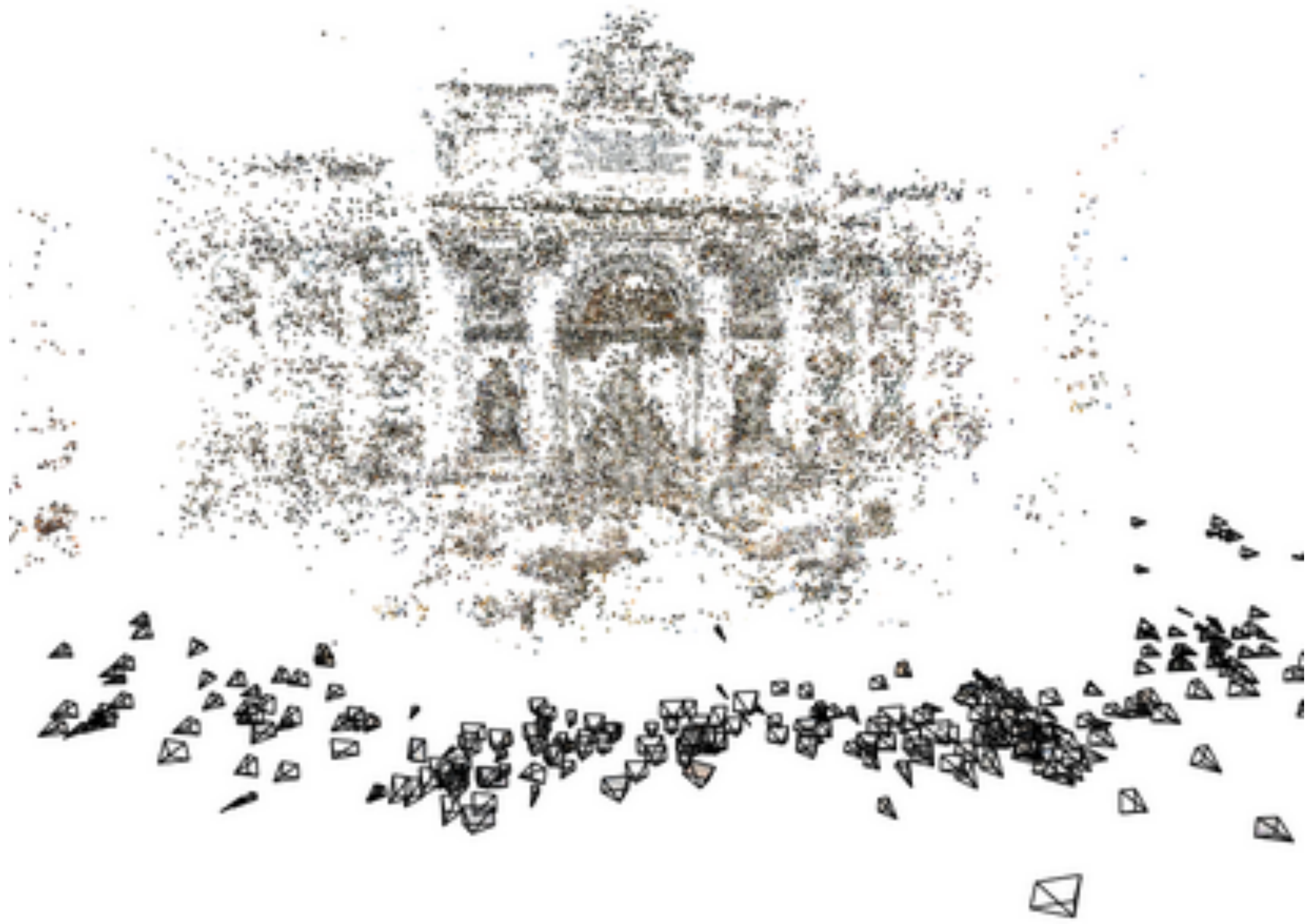
1. Vision for **Measurement**

Real-time stereo



Wang et al.

Structure from motion



Snavely et al.

Tracking

Demirdjian et al.

Slide Credit: Kristen Grauman (UT Austin)

Computer **Vision Problems**

1. Computing properties of the 3D world from visual data (***measurement***)

Ill-posed problem: real world is much more complex than what we can measure in images: 3D -> 2D

It is (literally) impossible to invert the image formation process

Slide Credit: Kristen Grauman (UT Austin)

Computer **Vision Problems**

1. Computing properties of the 3D world from visual data (***measurement***)

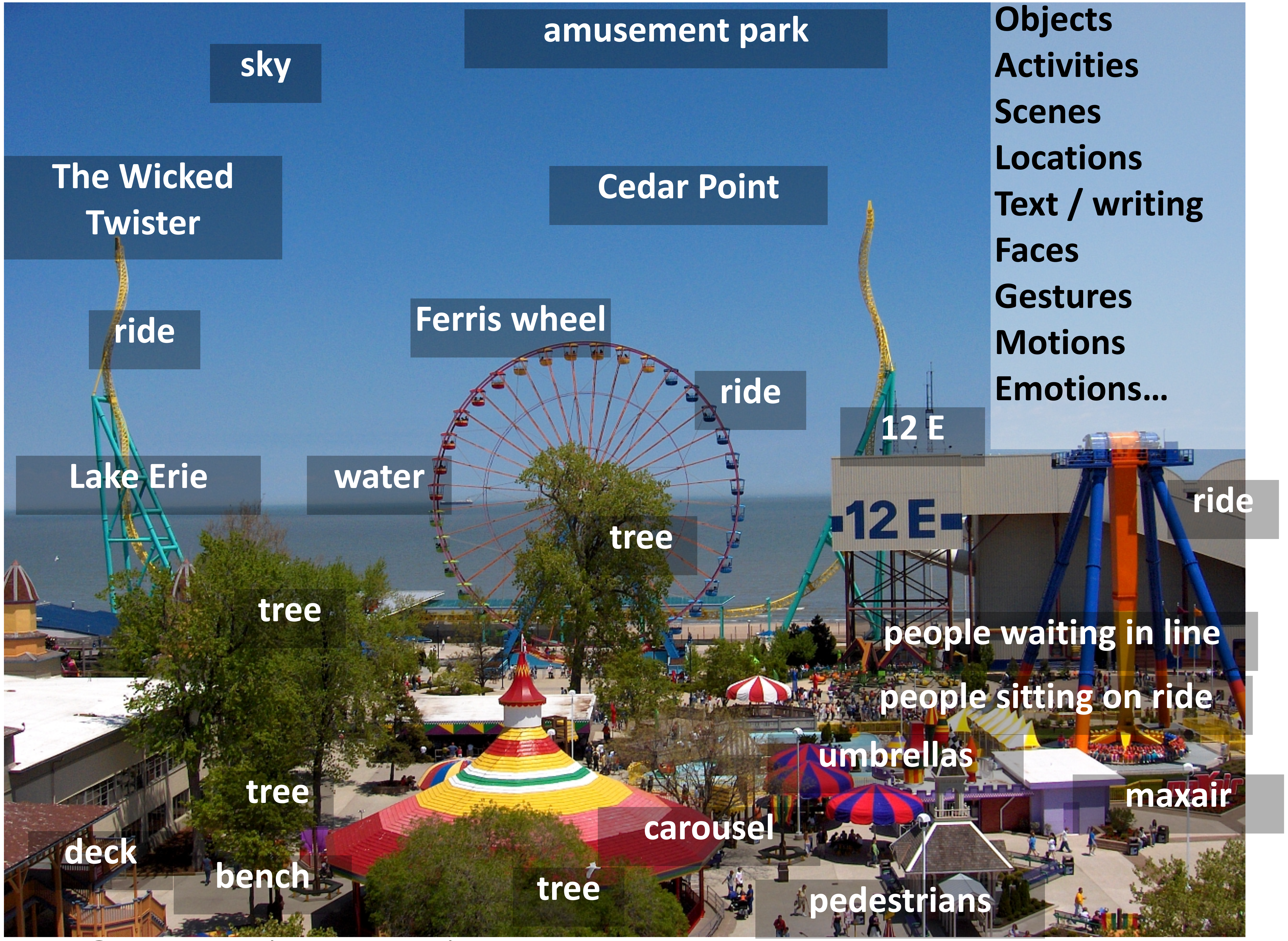
2. Algorithms and representations to allow a machine to recognize objects, people, scenes, and activities (***perception and interpretation***)

Slide Credit: Kristen Grauman (UT Austin)

2. Vision for **Perception and Interpretation**

Slide Credit: Kristen Grauman (UT Austin)

2. Vision for **Perception and Interpretation**



Slide Credit: Kristen Grauman (UT Austin)

Computer **Vision Problems**

1. Computing properties of the 3D world from visual data (***measurement***)

2. Algorithms and representations to allow a machine to recognize objects, people, scenes, and activities (***perception and interpretation***)

It is computationally intensive / expensive

Slide Credit: Kristen Grauman (UT Austin)

2. Vision for **Perception and Interpretation**

~ 55% of **cerebral cortex** in humans (13 billion neurons) are devoted to vision

more human brain devoted to vision than anything else

Computer **Vision Problems**

1. Computing properties of the 3D world from visual data (***measurement***)

2. Algorithms and representations to allow a machine to recognize objects, people, scenes, and activities (***perception and interpretation***)

It is computationally intensive / expensive

We do not (fully) understand the processing mechanisms involved

Slide Credit: Kristen Grauman (UT Austin)

Computer **Vision Problems**

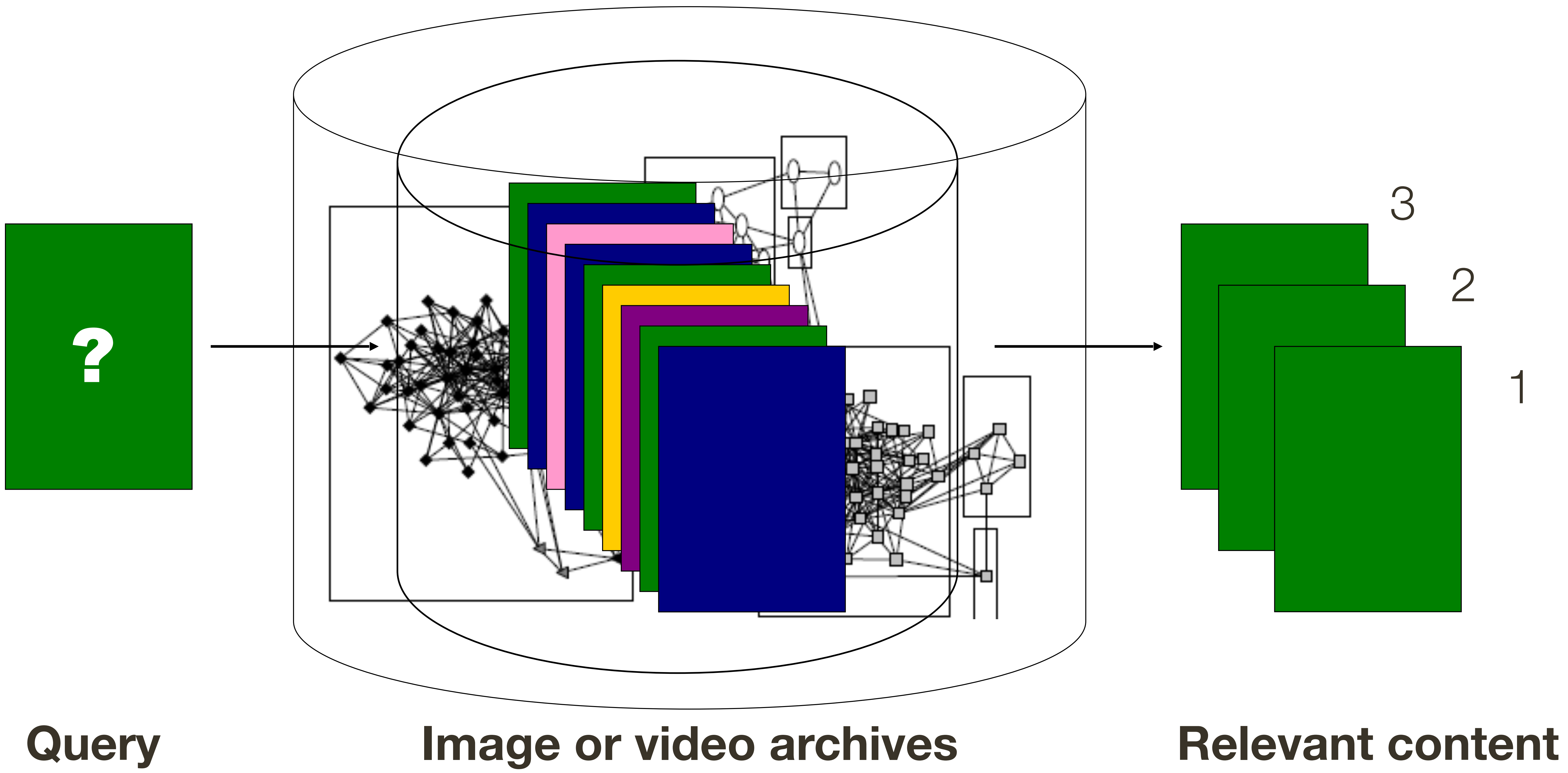
1. Computing properties of the 3D world from visual data (***measurement***)

2. Algorithms and representations to allow a machine to recognize objects, people, scenes, and activities (***perception and interpretation***)

3. Algorithms to mine, search, and interact with visual data (***search and organization***)

Slide Credit: Kristen Grauman (UT Austin)

3. **Search and Organization**



Slide Credit: Kristen Grauman (UT Austin)

Computer **Vision Problems**

1. Computing properties of the 3D world from visual data (***measurement***)

2. Algorithms and representations to allow a machine to recognize objects, people, scenes, and activities (***perception and interpretation***)

3. Algorithms to mine, search, and interact with visual data (***search and organization***)

Scale is enormous, explosion of visual content

Slide Credit: Kristen Grauman (UT Austin)

3. **Search and Organization**

*from iStock by GettyImages

Snapchat	WhatsApp	Facebook
31.7 Million / hour	**29.2 Million** / hour	**14.6 Million** / hour

Instagram	Flickr	YouTube
2.9 Million / hour	**0.2 Million** / hour	**18K hours** / hour

*based on article by Kimberlee Morrison in Social Times (2015)

Computer **Vision Problems**

1. Computing properties of the 3D world from visual data (***measurement***)

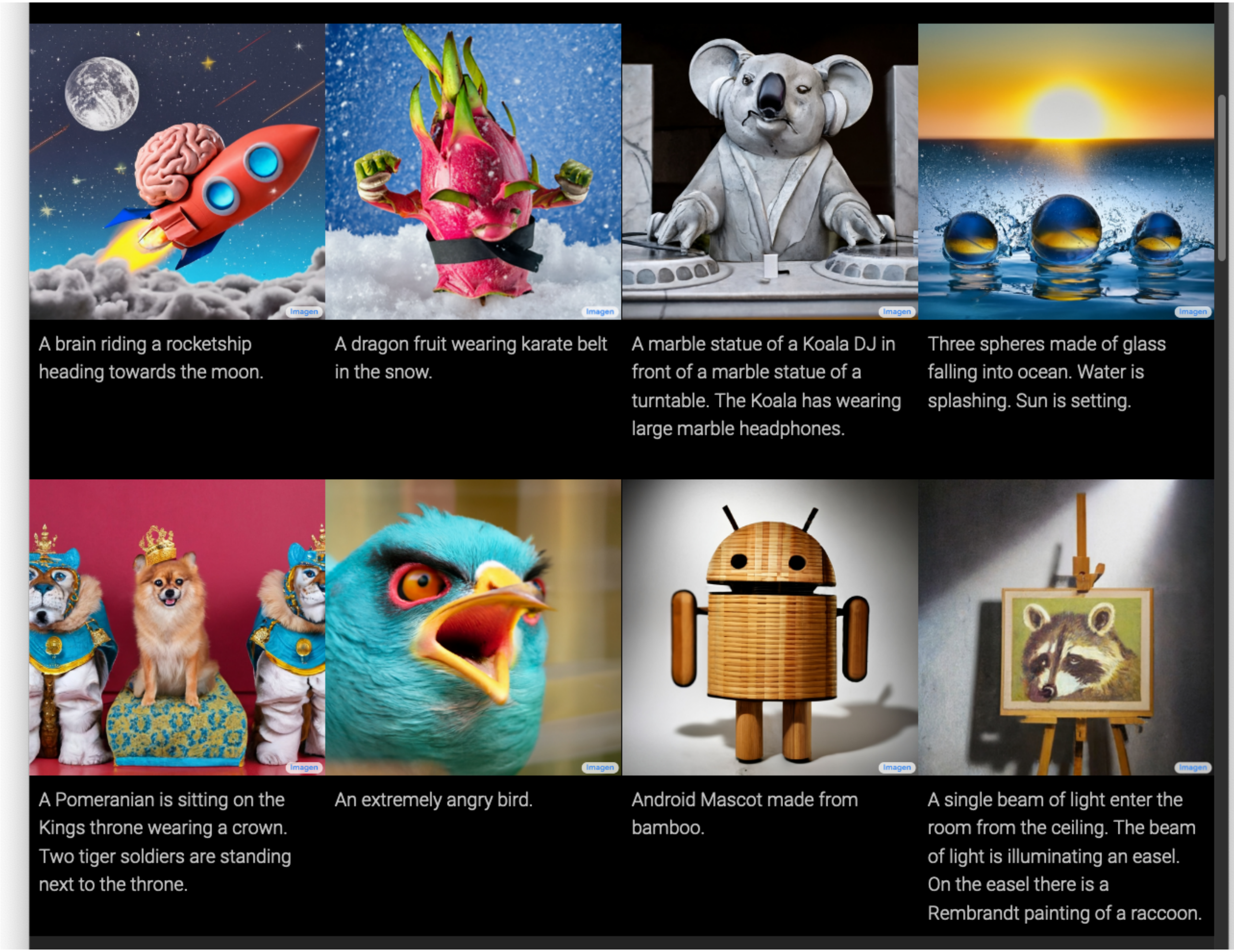
2. Algorithms and representations to allow a machine to recognize objects, people, scenes, and activities (***perception and interpretation***)

3. Algorithms to mine, search, and interact with visual data (***search and organization***)

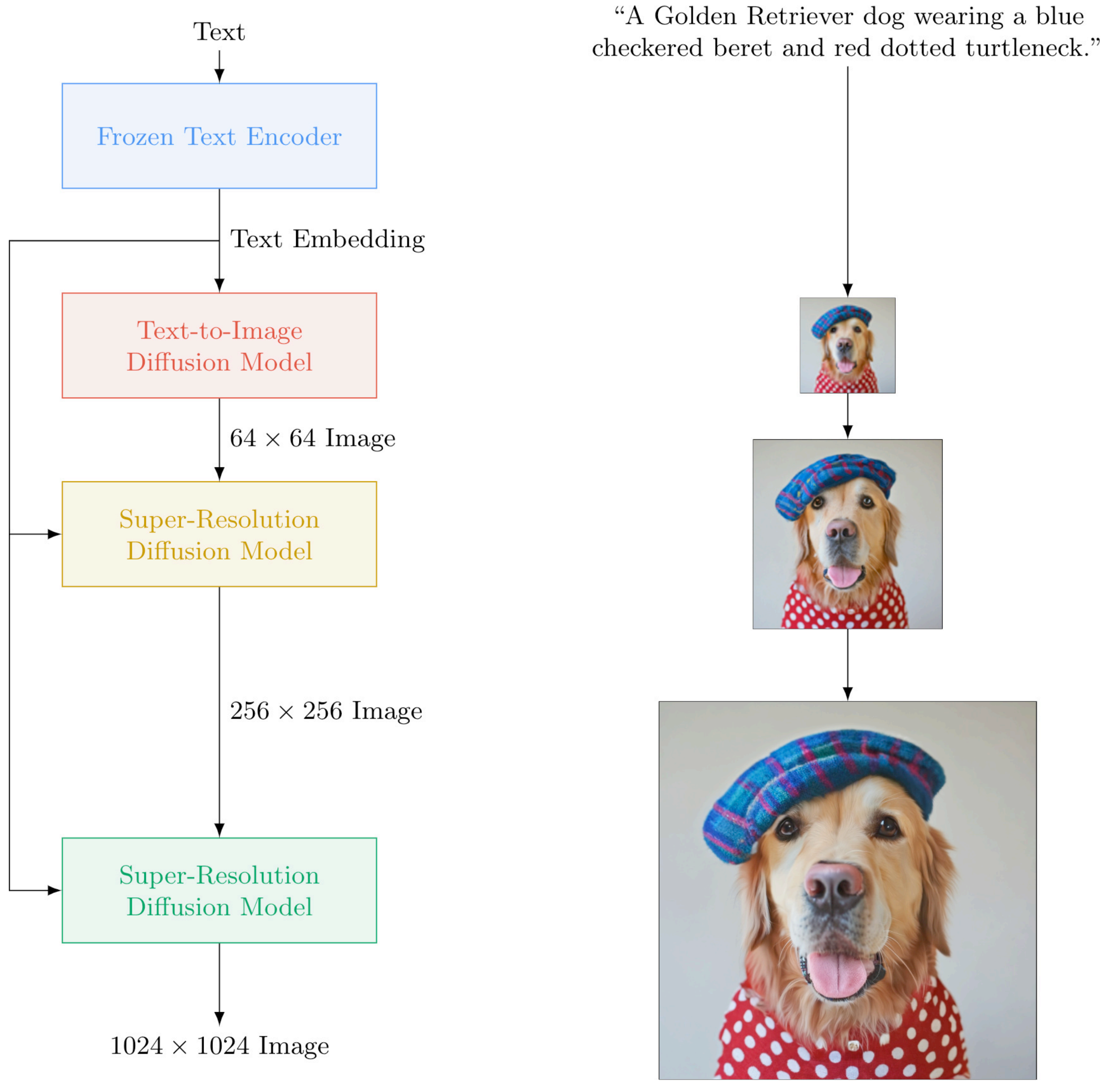
4. Algorithms for manipulation or creation of image or video content (***visual imagination***)

Slide Credit: Kristen Grauman (UT Austin)

4. Visual **Imagination**



- imagen.research.google
- Text to image generation
- Uses diffusion process, training using large dataset of text (web scale) and image-text (400M) pairs



Computer **Vision Problems**

1. Computing properties of the 3D world from visual data (***measurement***)

2. Algorithms and representations to allow a machine to recognize objects, people, scenes, and activities (***perception and interpretation***)

3. Algorithms to mine, search, and interact with visual data (***search and organization***)

4. Algorithms for manipulation or creation of image or video content (***visual imagination***)

Slide Credit: Kristen Grauman (UT Austin)

Challenges: Viewpoint invariance

Michelangelo 1475-1564

*slide credit Fei-Fei, Fergus & Torralba

Challenges: Lighting

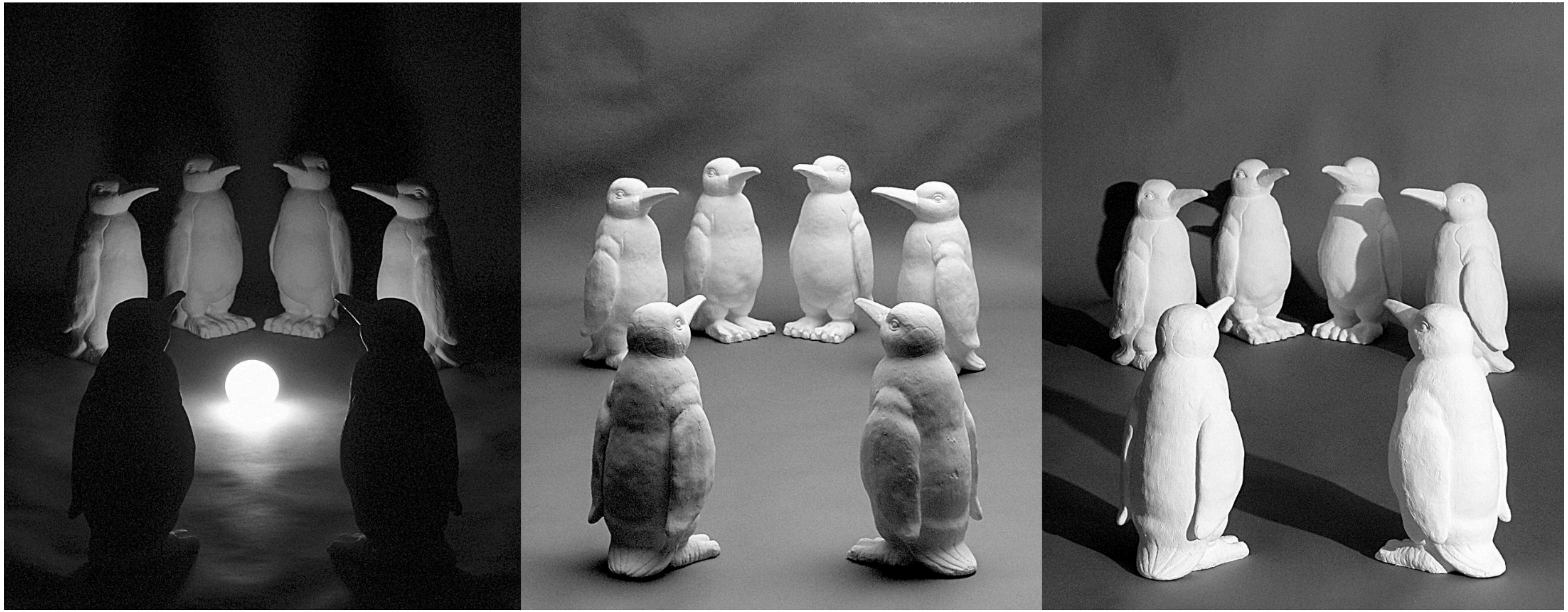


*image credit J. Koenderink

Challenges: Scale

*slide credit Fei-Fei, Fergus & Torralba

Challenges: Deformation

*image credit Peter Meer

Challenges: Occlusions

Rene Magritte 1965

Challenges: Background clutter

Kilmeny Niland 1995

Challenges: Local ambiguity and context

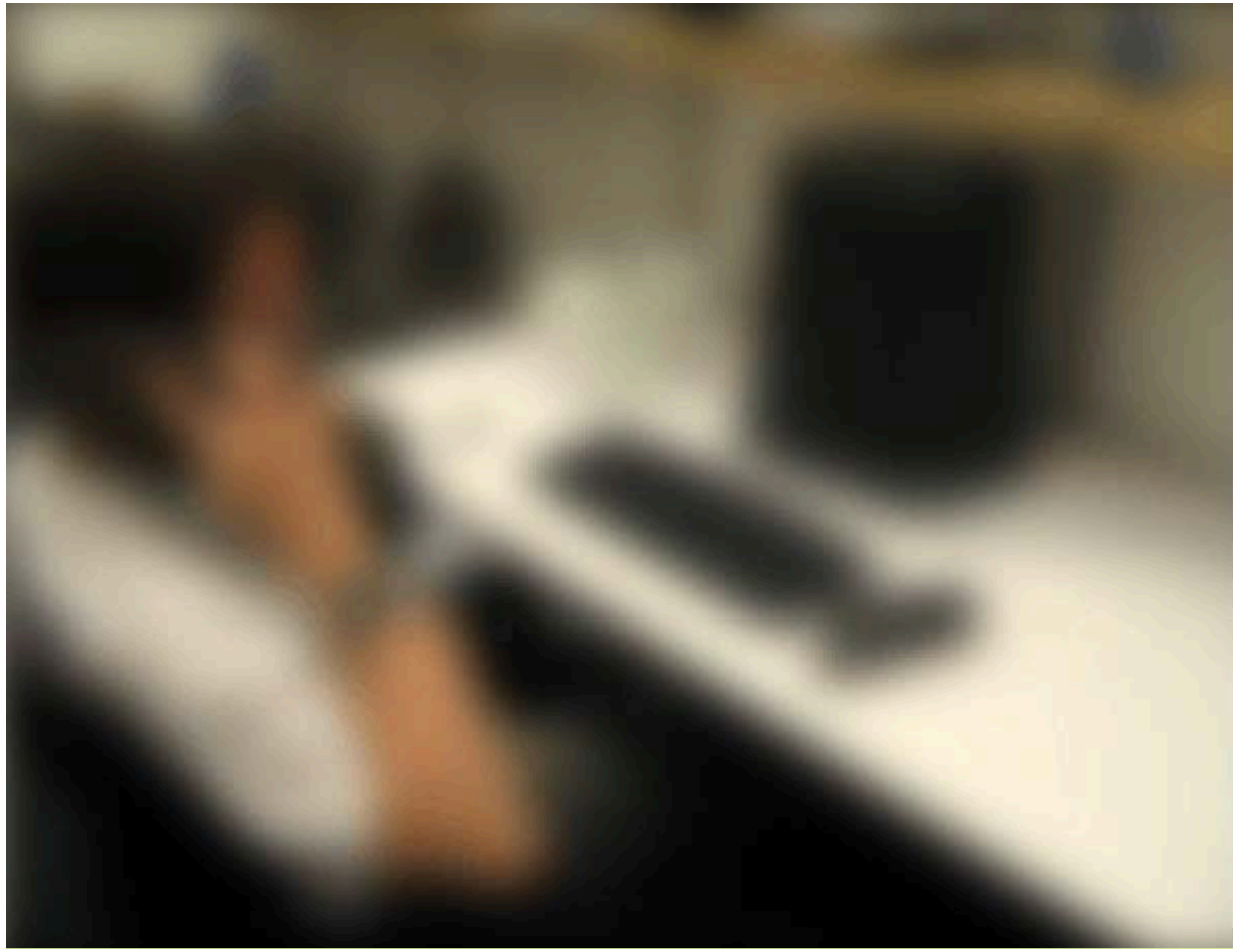


*image credit Fergus & Torralba

Challenges: Local ambiguity and context

*image credit Fergus & Torralba

Challenges: Motion

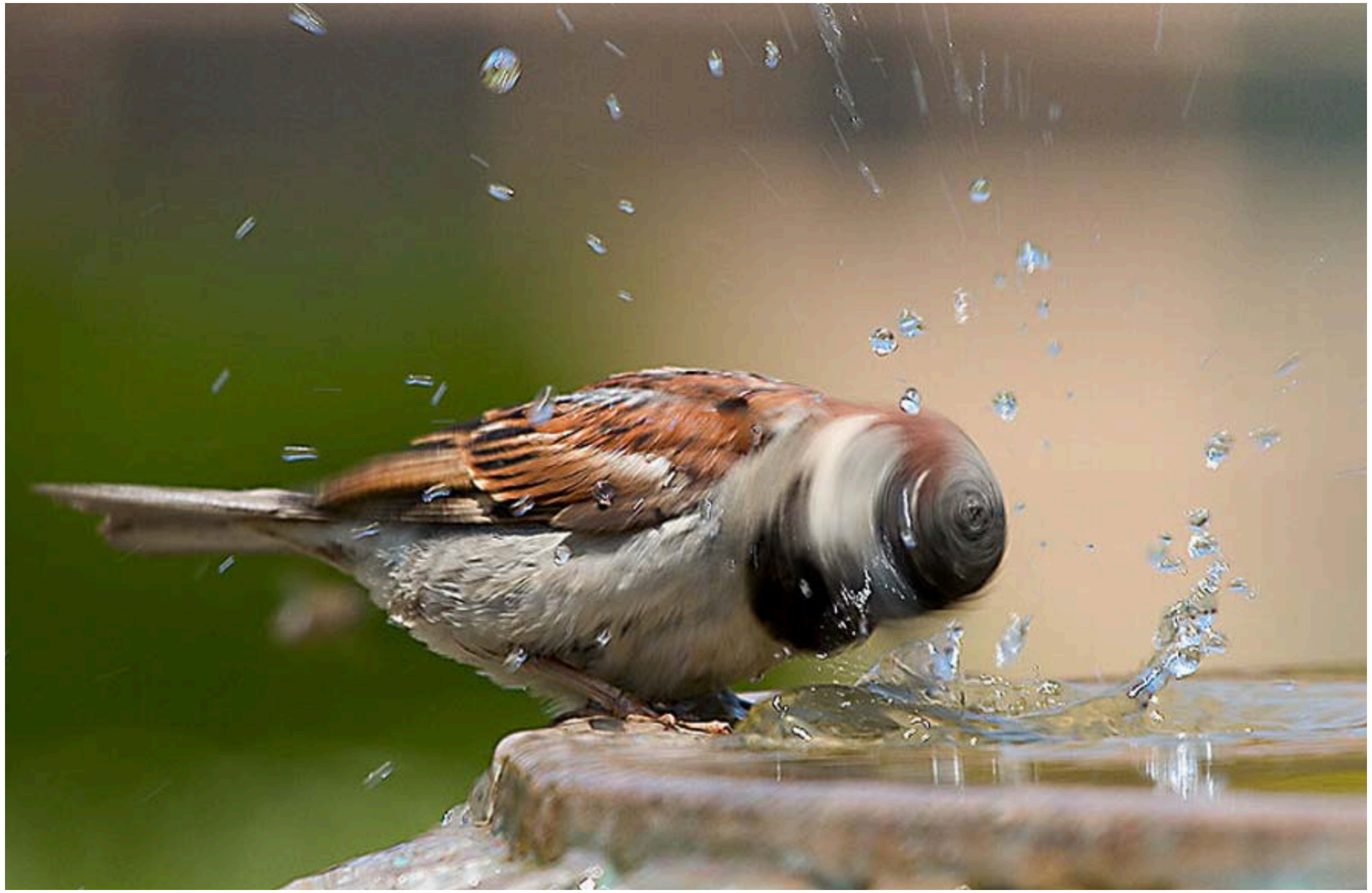


*image credit Peter Meer

Challenges: Object inter-class variation

*slide credit Fei-Fei, Fergus & Torralba

Computer Vision **Applications**

- Let's see some examples of state-of-the-art and where it is used

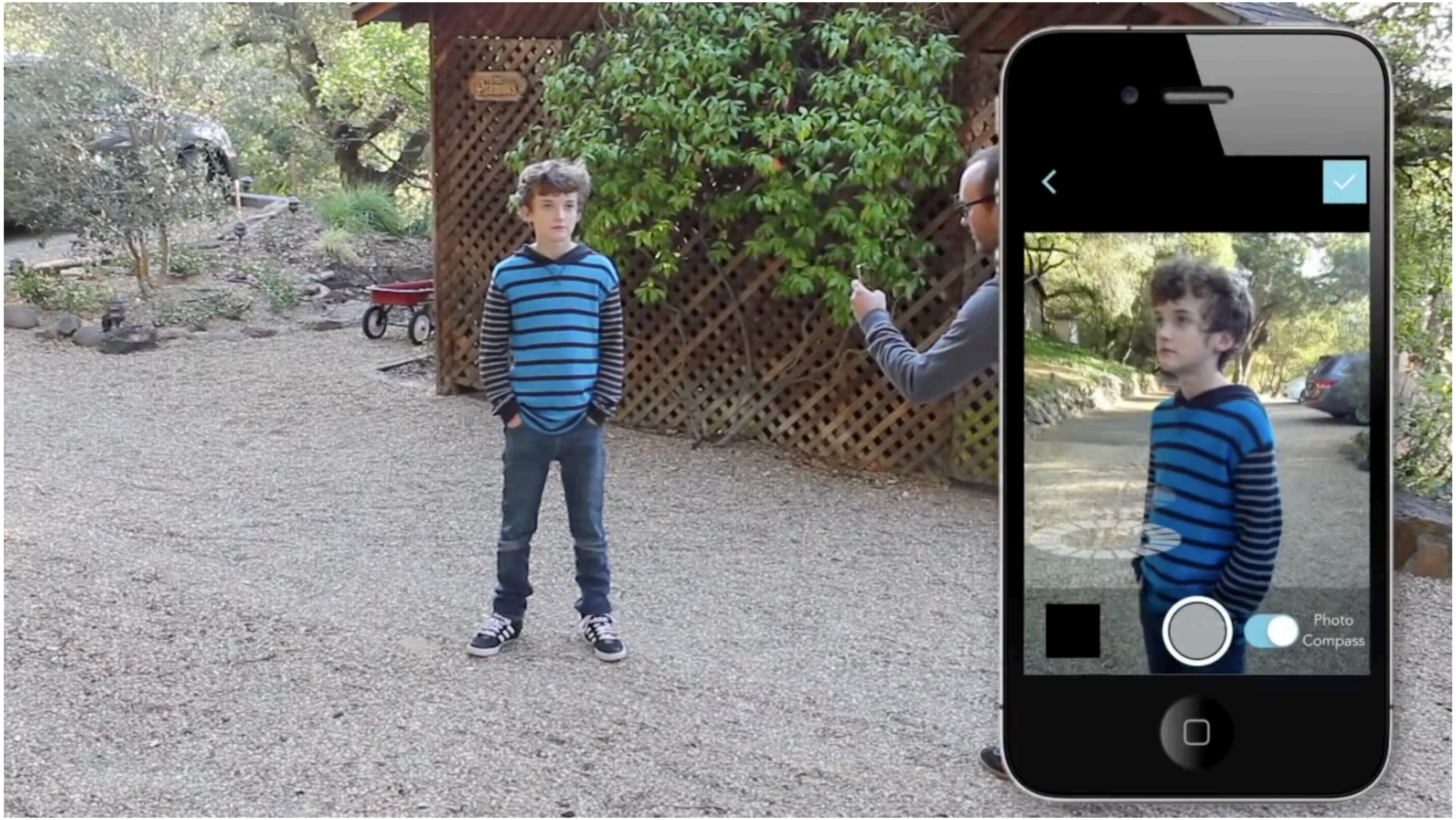
Face Detection

[Motorola]

Camera Tracking

[Boujou — Vicon / OMG]

3D Reconstruction

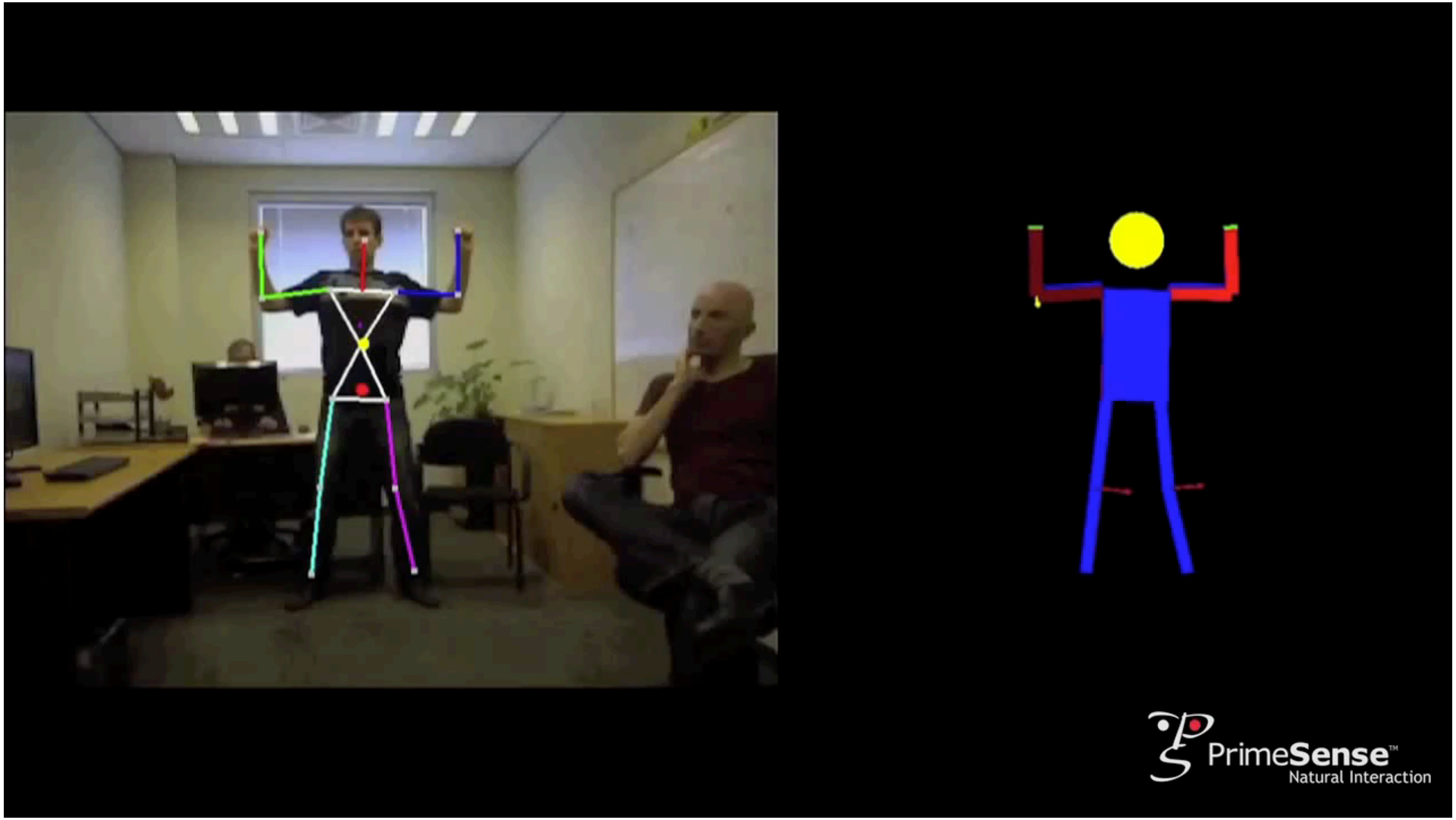


[Autodesk 123D Catch]

Body Pose Tracking

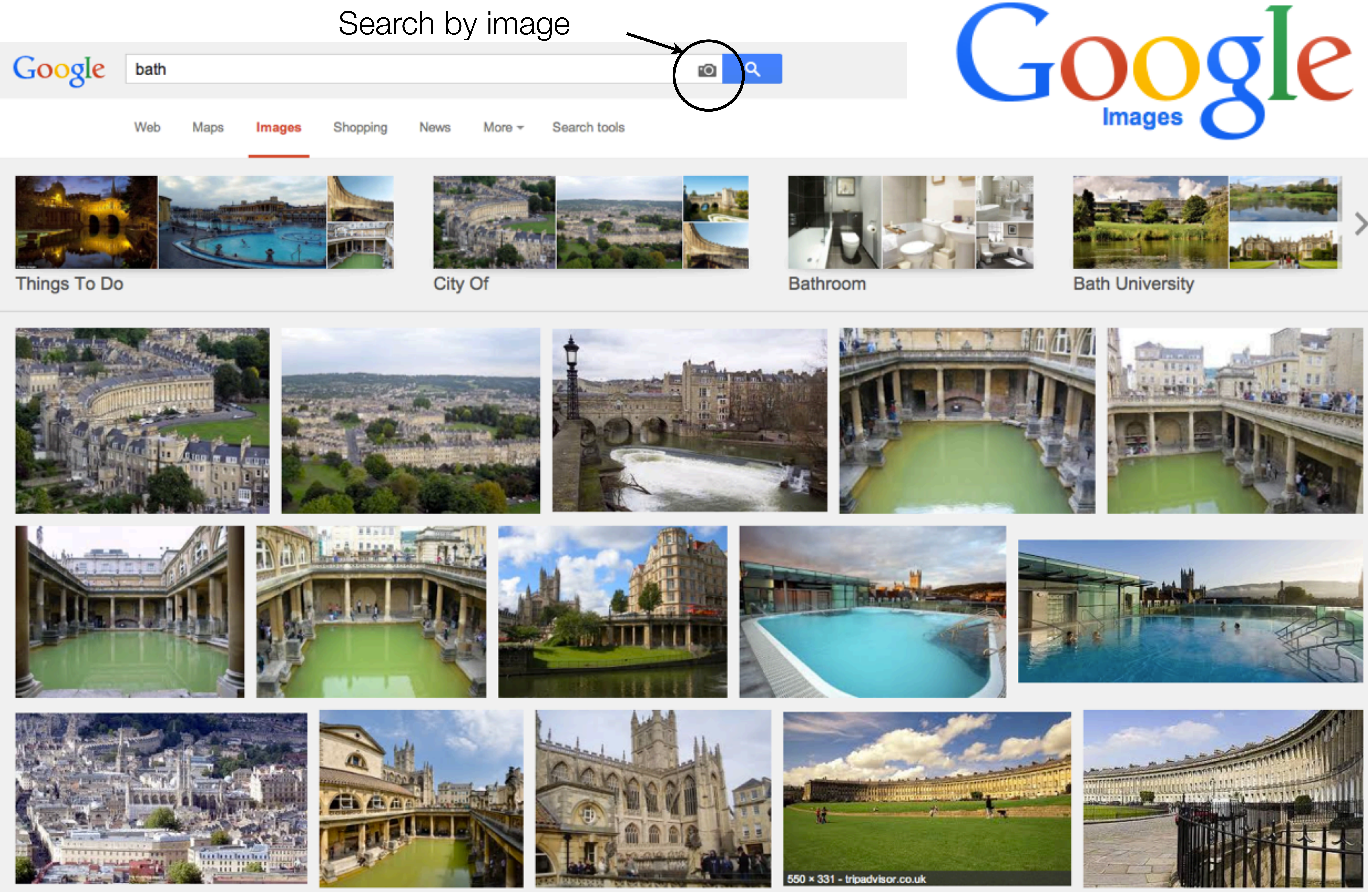
[Microsoft Xbox Kinect]

Body Pose Tracking



[PrimeSense]

Image Recognition and Search



Self-Driving Cars

[Google]

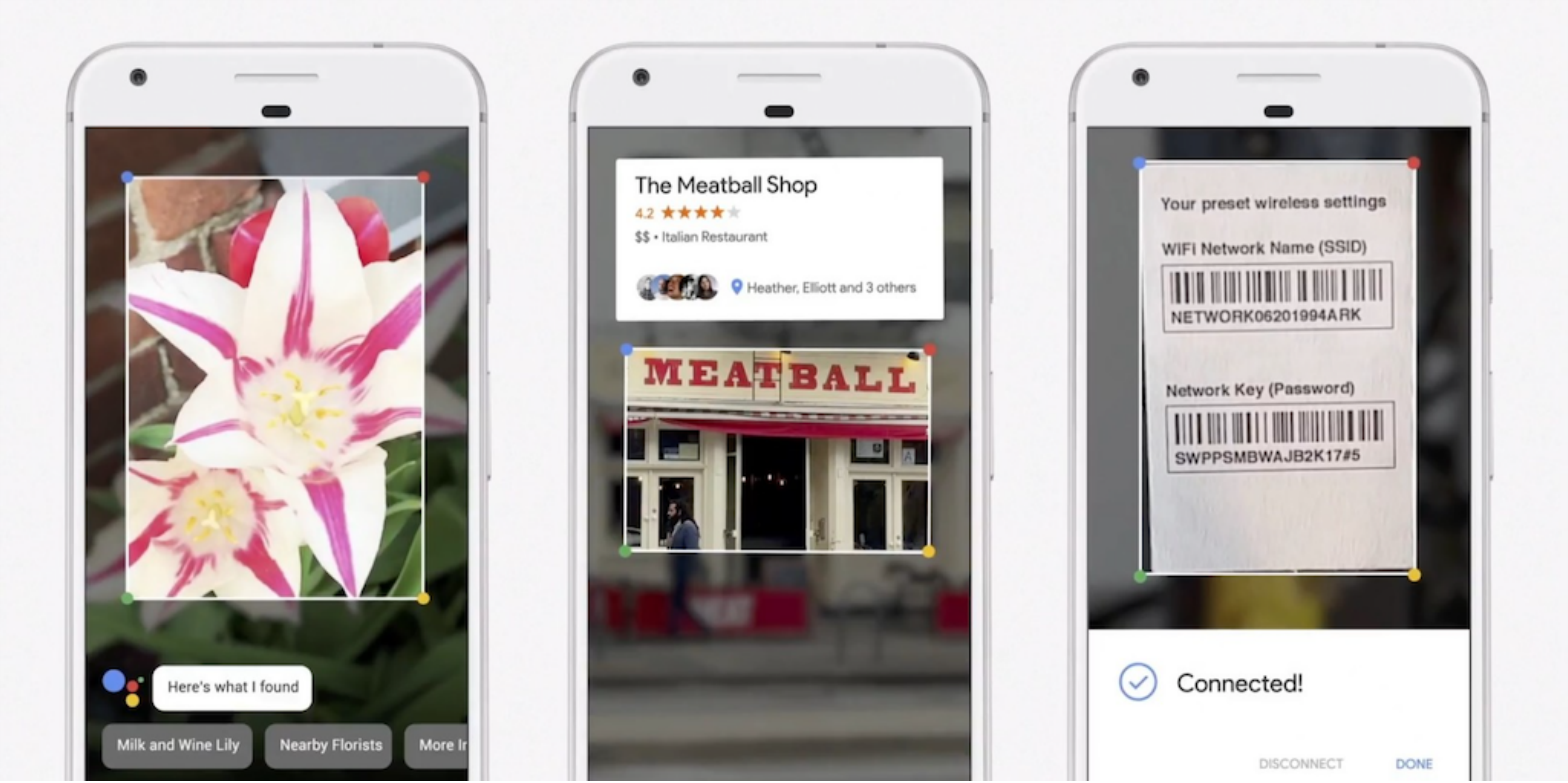
Flying Vehicles

www.skydio.com

AR / VR

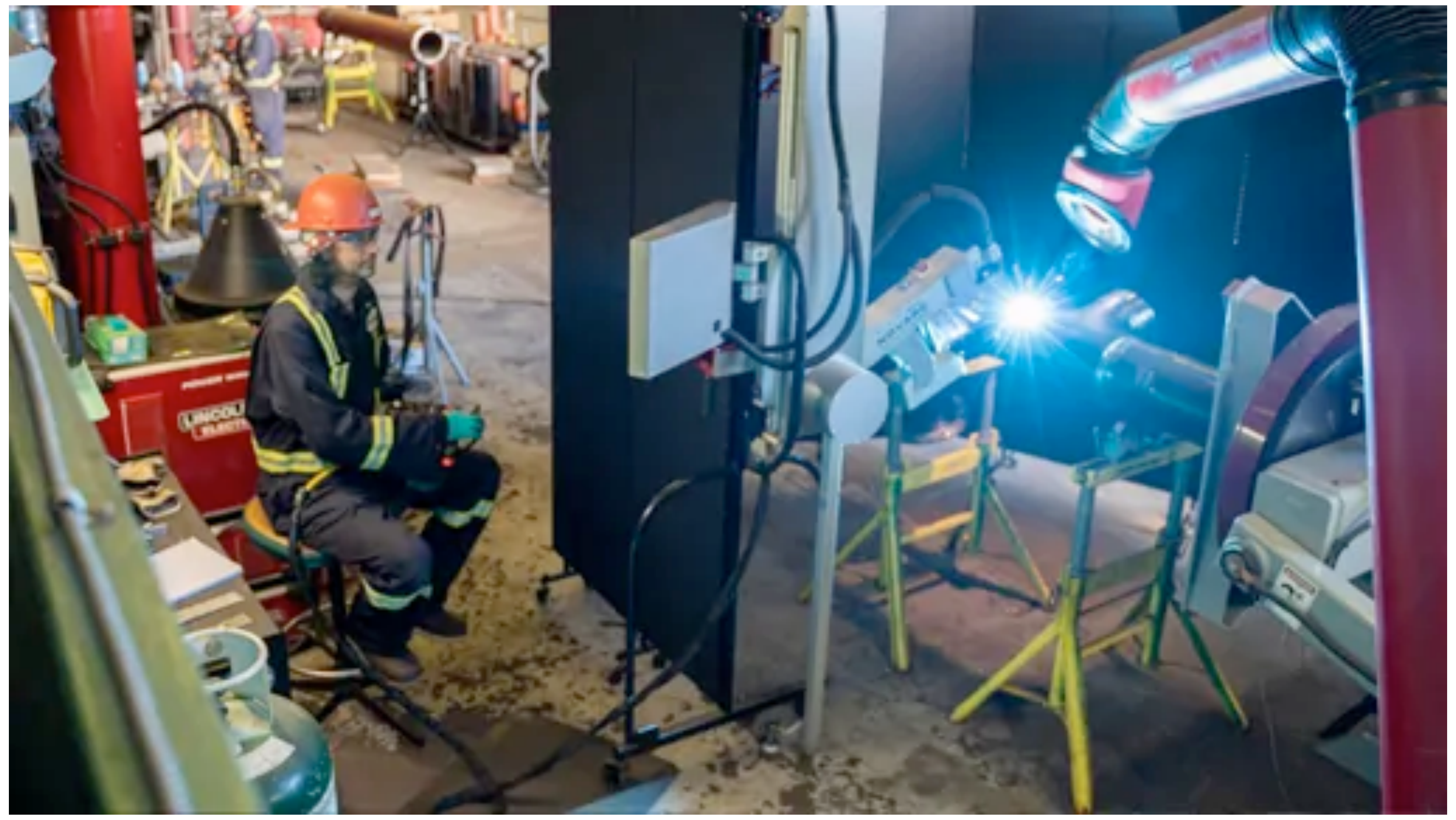
[Microsoft HoloLens]

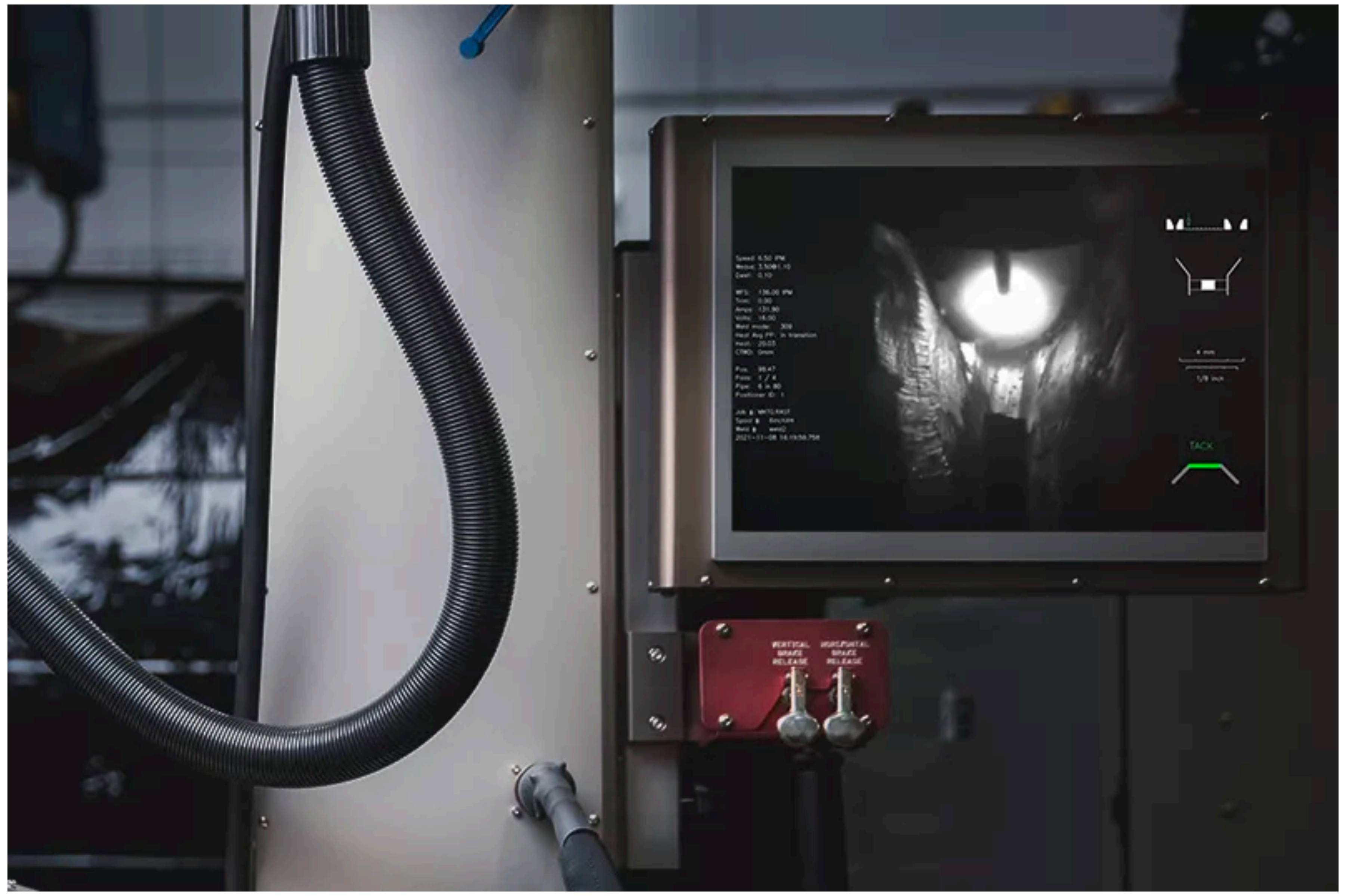
Mobile Apps



[Google Lens]

Industrial

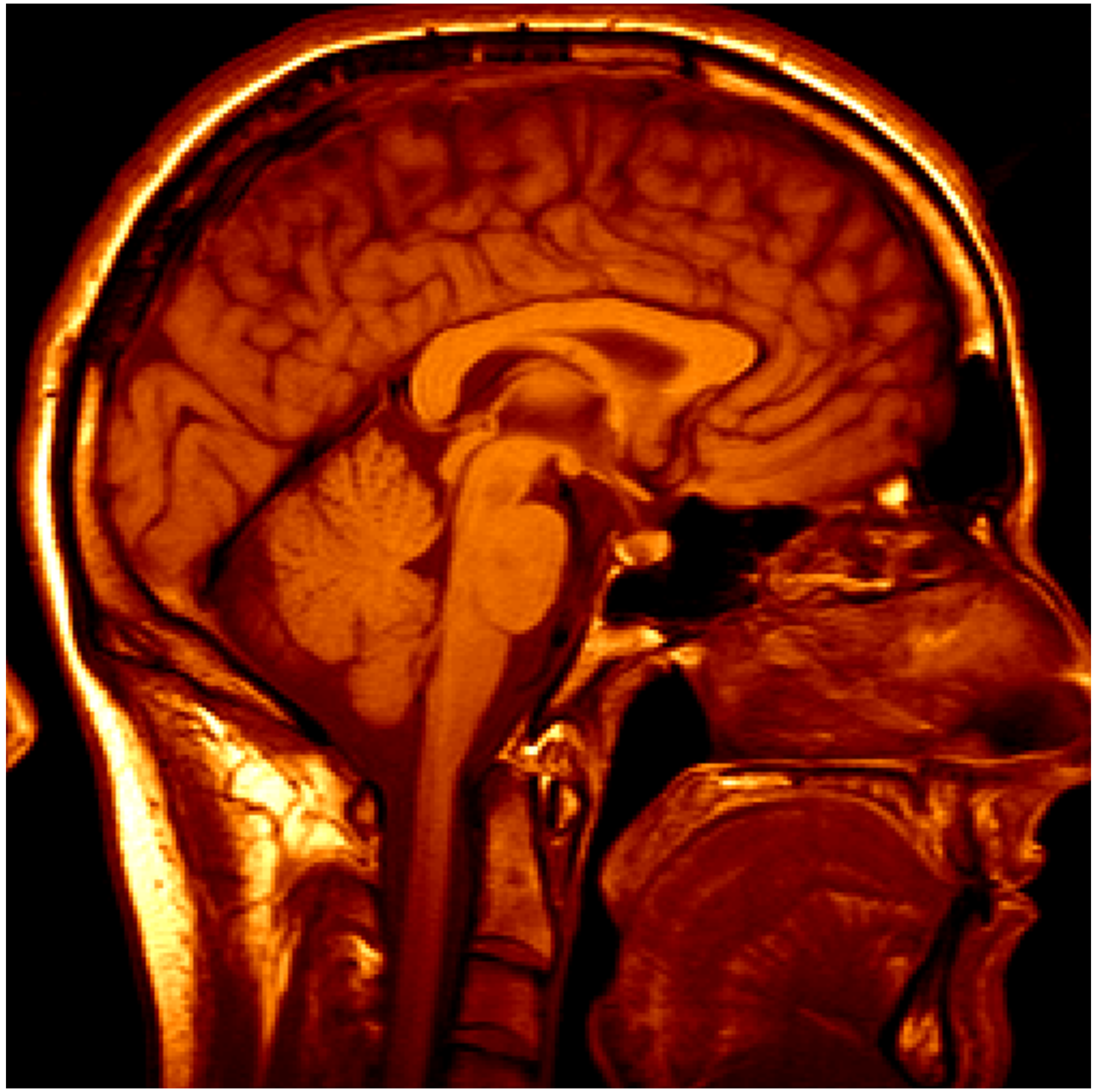




Machine Vision controlled welding robotics

NOVARC TECHNOLOGIES

Medicine

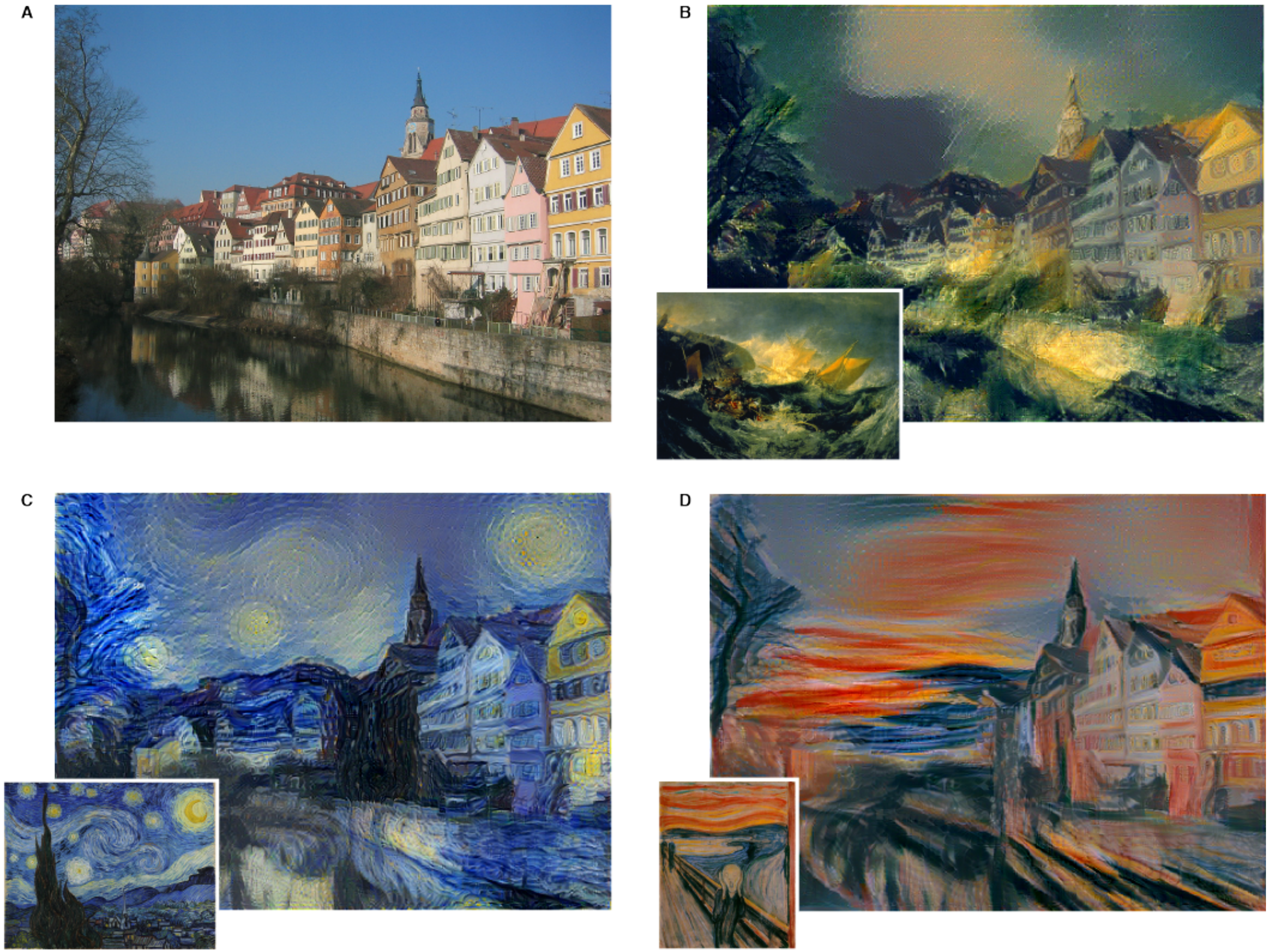


3D imaging
MRI, CT

Image guided surgery
Grimson et al., MIT

Slide Credit: James Hays (GA Tech)

Art



[Gatys, Ecker, Bethge 2015]

Computer Vision Applications

- Digital Entertainment + Consumer
  - Camera tracking, 3D reconstruction, visual effects, virtual reality, augmented reality, product recognition
- Science and Medicine
  - Visual data analytics, anatomical measurement/analysis, tumour detection
- Engineering and Industry
  - Robotics, self driving cars, reverse engineering, visual servoing, industrial part inspection, OCR, precision agriculture
- Photography/Videography and Editing
  - Face detection, scene recognition, video stabilisation, drone camera, gap filling, image blending, panorama stitching, high dynamic range
- Mapping and Environmental
  - Image registration, 3D building modelling, streetview, numberplate recognition, landmark recognition, species identification

Course Schedule

Course Overview

Date	Lecture	Description	Notes and Resources
Sep 7	Introduction	Intro to computer vision, Course logistics	Week 1 Notes
	Image Formation	Pinhole, Perspective, Weak perspective and Orthographic projection, Lenses	
Sep 14	Image Formation	Lenses, Human Eye, Image Filtering: Image as a function, Image transformations	
	Image Filtering	Linear filters, Correlation and Convolution	
	Image Filtering	Gaussian filter, Separability, Pillbox filter, Speeding up Convolution, Fourier representation	
Sep 21	Image Filtering	Non-linear Filters, Bilateral Filter	
	Sampling	Sampling Theory, Bandlimited Signal, Nyquist Rate, Aliasing	
	Sampling	Color Filter Arrays, Demosaicing, Template Matching: Introduction	
Sep 28	Scaled Representations	Template Matching, Image Pyramids, Gaussian Pyramid	
	Scaled Representations	Gaussian Pyramid, Image Gradients	
	Local Image Features	Image Gradients, Sobel, Marr / Hildreth Edge Detection	
Oct 5	Local Image Features	Canny Edge Detection, Image Boundaries	

- Schedule, Assignments
- Lecture Slides and Notes
- Course Information (public)

https://mattabrown.github.io/425

Topics Covered

- Image Processing (Linear Filtering, Convolution)
- Filters as Templates
- Image Feature Detection (Edges & Corners)
- Texture & Colour
- Image Feature Description (SIFT)
- Model Fitting (RANSAC, The Hough Transform)
- Camera Models, Stereo Geometry
- Motion and Optical Flow
- Clustering and Image Segmentation
- Learning and Image Classification
- Deep Learning Introduction

Course Origins

CPSC 425 was originally developed by **Bob Woodham** and has evolved over the years. Much of the material this year is adapted from material prepared by Bob, as well extensions developed by others who taught this course

Previously taught by:

- 2020-2021 Term 1 by **Leonid Sigal**
- 2019-2020 Term 2 by **Leonid Sigal**
- 2019-2020 Term 1 by **Jim Little**
- 2018-2019 Term 1 & 2 by **Leonid Sigal**
- 2016-2017 Term 2 by **Jim Little**
- 2015-2016 Term 2 by **Fred Tung**
- 2015-2015 Term 2 by **Jim Little**

How to **Learn** from the **Course**?

— The course is very **broad**, but relatively **shallow** introduction to a very diverse and complex field that draws material from geometry, statistics, AI, machine learning, computer graphics, psychology and many others.

— It is easy to think that material is easy and course requires no studying

— Part of your job should be going over the slides and carefully analyzing not just what is on them, but the underlying assumptions, algorithmic steps and so on

— Don't strive for "**template matching**" strive for true "**understanding**"

Grading Criteria

In-class **clicker questions**: 10%

Programming Assignments: 45%

6 graded and 1 ungraded (optional) assignment

Midterm Exam (October 19th): 15%

Final Exam (TBD): 30%

iClicker Setup

Quizzes will be run via **iClicker**

Please make sure you have an iClicker account with your student ID:

https://lthub.ubc.ca/guides/iclicker-cloud-student-guide/

You should set **UBC** as the **institution**, use the **same email** as for your **canvas** account, and enter your **student number** in the student ID field.

You should be automatically added and the course:

CPSC 425 101 2023W1 Computer Vision

iClicker Quizzes

Setup before class! We'll do a test next week

Join the class at student.iclicker.com

There will be around 6 multiple choice questions per quiz
— 1/2 point for participation
— 1/2 point for correct answer
*not all clicker quizzes are worth the same # of points, depends on # of questions.

The clicker questions contribute 10% to your total grade

Missing Quiz Policy: If you miss a quiz for a legitimate and documented reason, that quiz will be dropped (legitimate reasons: illness, conference travel, etc.) You are required to contact instructor and provide proof within 1 week of missed quiz.

Assignments

There will be **6+1 assignments** in total (6 marked)

— Approximately 1 every 2 weeks

— You will hand these in by 11:59pm on the due date (read hand in instructions and late policy on course webpage)

You will use the **Python**, with the following libraries: Python Imaging Library (PIL), NumPy, Matplotlib, SciPy, Scikit-Learn

— Assignment 0 (which is ungraded) will introduce you to this.

Assignments contribute 45% to your final score

Midterm Exam

Scheduled for **October 19th**

— Here in class during the lecture period

— Closed book, no notes allowed

Multiple choice, true / false and short answer questions

— Aimed to test your “understanding” of the content of the course

The Midterm exam will contribute <u>15%</u> to your final score

Final Exam

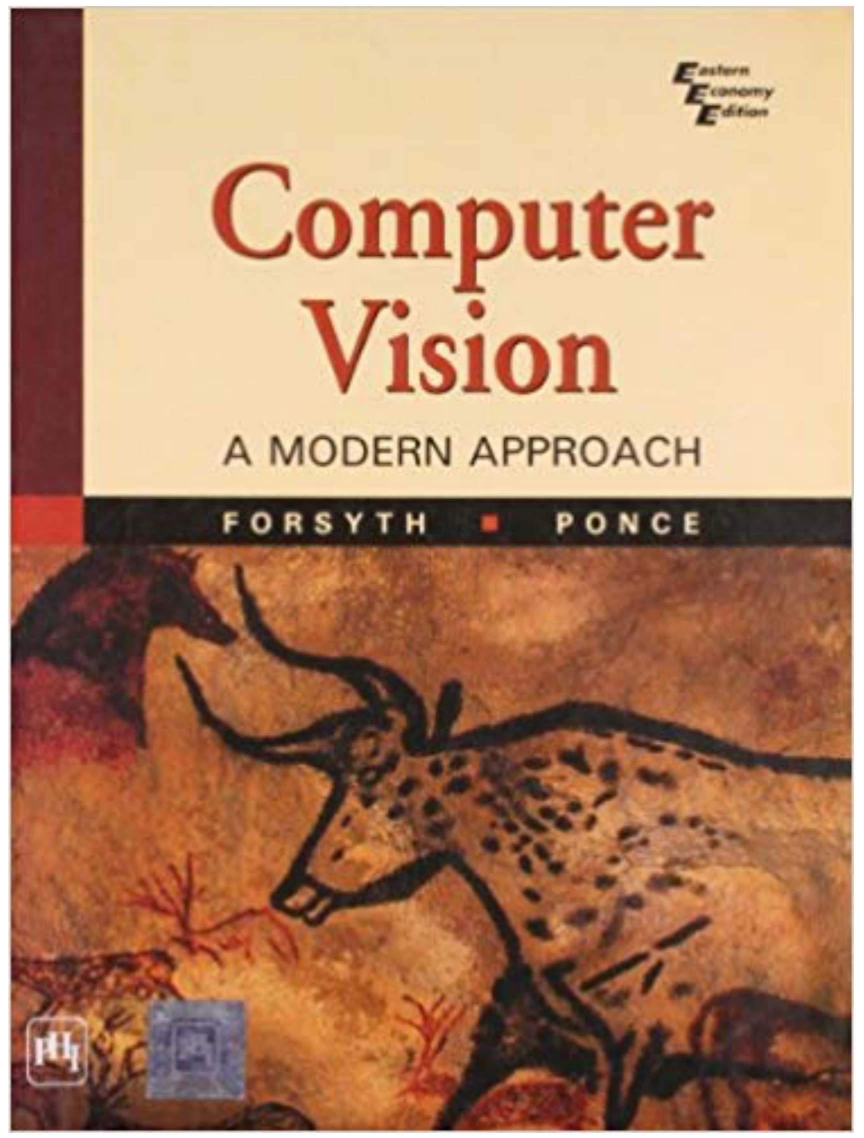
The Final exam is held during the regular examination period, **Dec 11-22**, and is scheduled by the Registrar's Office

Similar to the midterm but longer and with more extensive short/medium answer questions

The Final exam will contribute 30% to your final score

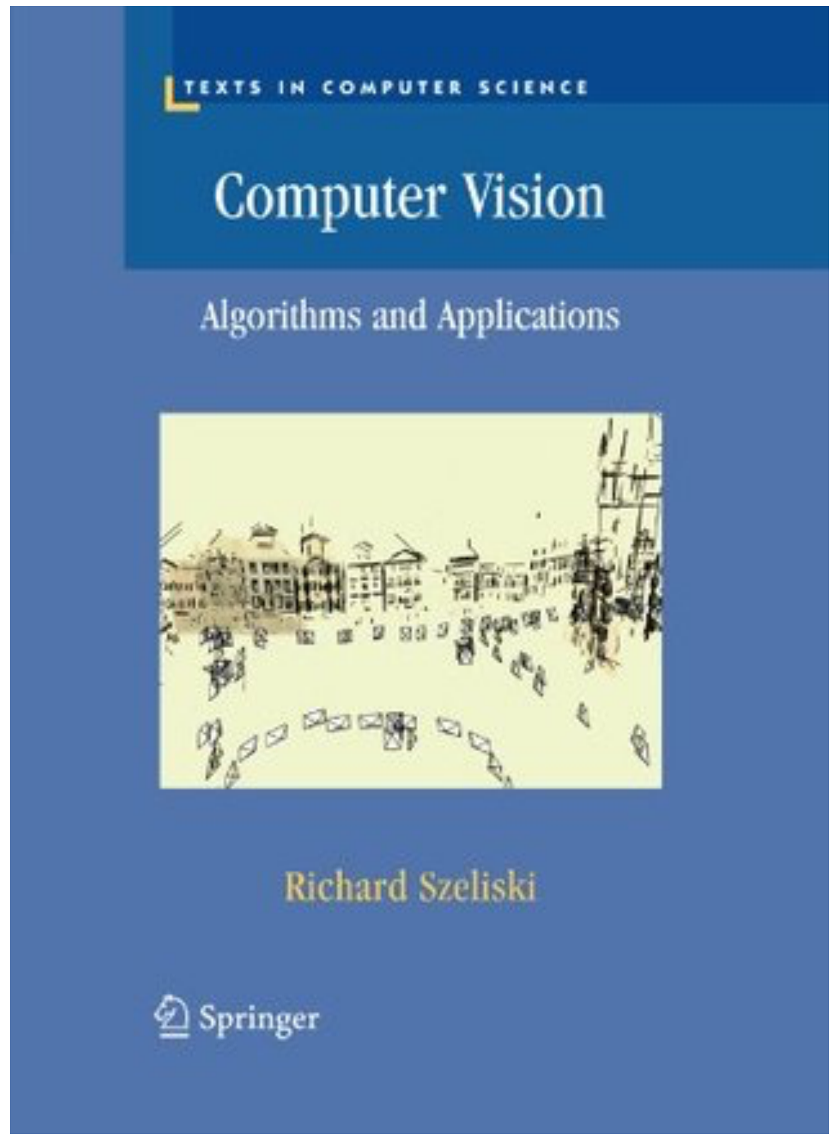
Textbooks

The course uses the following textbooks, which are recommended (but **not required**):



Computer Vision: A Modern Approach (2nd ed)

By: D. Forsyth & J. Ponce
Publisher: Pearson 2012



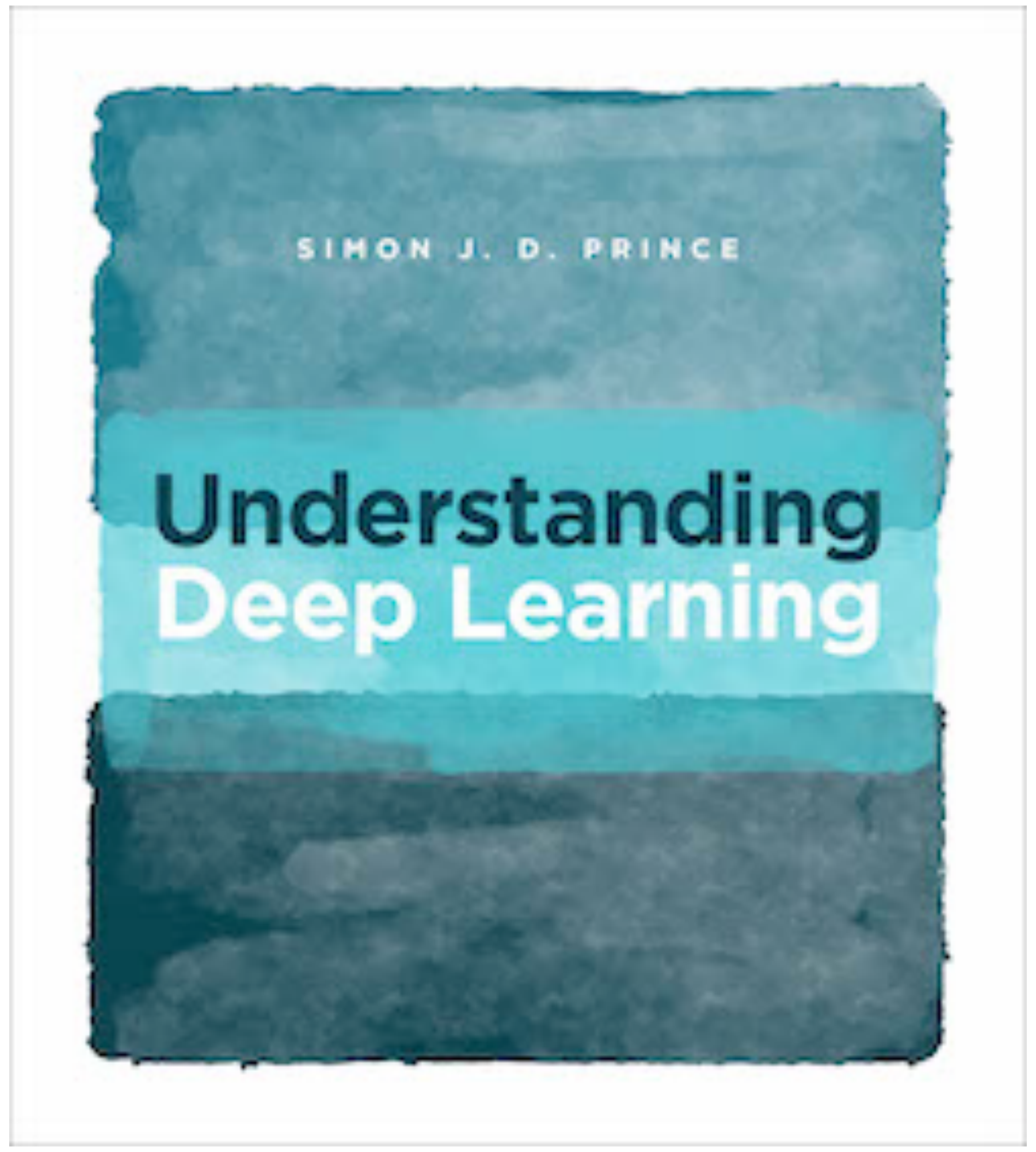
Computer Vision: Algorithms and Applications (2nd ed)

By: R. Szeliski
Publisher: Springer 2022

https://szeliski.org/Book/

Text**books**

The course uses the following textbooks, which are recommended (but **not required**):



Understanding Deep Learning

By: Simon J.D. Prince
Publisher: MIT Press 2023

https://udlbook.github.io/udlbook/

Readings

You will be assigned **readings**.

— Sometimes you will be assigned readings from other sources

Do the reading **after coming** to the lecture

— Reading assignments will be posted on course webpage

— They will also be mentioned in class

Prepare for the **Next Lecture**

Readings:

— **Next** Lecture: Szeliski Chapter 2, Forsyth & Ponce (2nd ed.) 1.1.1 — 1.1.3

Reminders:

— Start working on **Assignment 0** (ungraded) suggest complete by Sept 14

— **[optional]** Watch TED talk by Prof. Fei-Fei Li
https://www.youtube.com/watch?v=40riCqvRoMs