THE UNIVERSITY OF BRITISH COLUMBIA

CPSC 425: Computer Vision

Image Credit: Devi Parikh

Lecture 1: Introduction and Course Logistics

Course **logistics**

Times: Tues, Thurs 5:00-6:30pm

Locations (101): Earth Sciences (ESB), Room 1012

Instructor: Leonid Sigal

E-mail: lsigal@cs.ubc.ca

Office: ICICS 119

Times: Mon, Wed 3:30-5:00pm

Locations (102): MacMillan (MCML), Room 166

Course **logistics**

Times: Tues, Thurs 5:00-6:30pm **Locations (101):** Earth Sciences (ESB), Room 1012

Instructor: Leonid Sigal

Leon

E-mail: lsigal@cs.ubc.ca

Office: ICICS 119

Times: Mon, Wed 3:30-5:00pm **Locations (102):** MacMillan (MCML), Room 166

Course **logistics**

Times: Tues, Thurs 5:00-6:30pm

Locations (101): Earth Sciences (ESB), Room 1012

Instructor: Leonid Sigal

Leon

E-mail: lsigal@cs.ubc.ca

Office: ICICS 119

PLEAST contact me through Piazza rather than e-mail

Times: Mon, Wed 3:30-5:00pm

Locations (102): MacMillan (MCML), Room 166

About **me** …

About **me …**

Software Engineer
1999 - 2001

COGNEX

About **me …**

Software Engineer
1999 - 2001

COGNEX

About **me …**

Software Engineer
1999 - 2001

COGNEX

About **me …**

PhD, MSc
2001 - 2008

BROWN

BOSTON UNIVERSITY

Software Engineer
1999 - 2001

COGNEX

About **me …**

Postdoctoral Researcher
2007 - 2009

UNIVERSITY OF TORONTO

PhD, MSc
2001 - 2008

BROWN

BOSTON UNIVERSITY

Software Engineer
1999 - 2001

COGNEX

About **me …**

Senior Research Scientist
2009 - 2017
Disney Research

Postdoctoral Researcher
2007 - 2009
UNIVERSITY OF TORONTO

PhD, MSc
2001 - 2008
BROWN
BOSTON UNIVERSITY

Software Engineer
1999 - 2001
COGNEX

About **me …**

Professor
2017 -

THE UNIVERSITY OF BRITISH COLUMBIA

Senior Research Scientist
2009 - 2017

Disney Research

Postdoctoral Researcher
2007 - 2009

UNIVERSITY OF TORONTO

PhD, MSc
2001 - 2008

BROWN

BOSTON UNIVERSITY

Software Engineer
1999 - 2001

COGNEX

About **me ...**

I have been working in **Computer Vision** for the last ~25 years

Professor
2017 -
THE UNIVERSITY OF BRITISH COLUMBIA

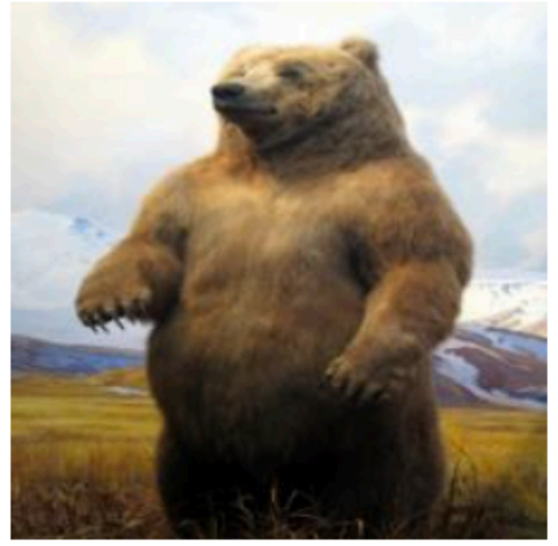
Senior Research Scientist
2009 - 2017
Disney Research

Postdoctoral Researcher
2007 - 2009
UNIVERSITY OF TORONTO

PhD, MSc
2001 - 2008
BROWN
BOSTON UNIVERSITY

Software Engineer
1999 - 2001
COGNEX

Multi-modal Learning (now known as Vision-Language Models)

Dialog Information	Input image	Attended image
Current question: *What color is it's fur ?* Predicted answer: *Brown*		
Current question: *What color is the train ?* Predicted answer: *It is white and red with some blue on it*		
Current question: *Is it a sunny day ?* Predicted answer: *Yes*		

[*Seo et al., NeurIPS 2017*]

Video Understanding

[*Xu et al., WACV 2019*]

Video Understanding

[*Xu et al., WACV 2019*]

Course **logistics**

Times: Tues, Thurs 5:00-6:30pm **Locations (101):** Earth Sciences (ESB), Room 1012

Instructor: Leonid Sigal

E-mail: lsigal@cs.ubc.ca

Office: ICICS 119

Times: Mon, Wed 3:30-5:00pm **Locations (102):** MacMillan (MCML), Room 166

Course **logistics**

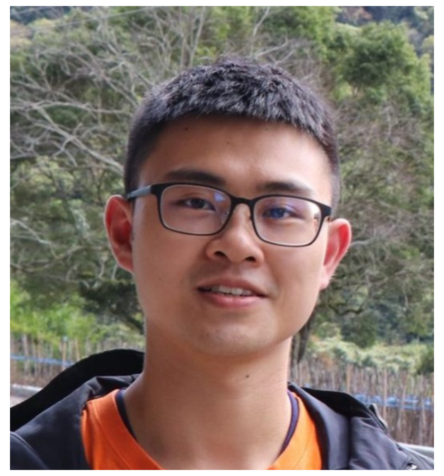
Times: Tues, Thurs 5:00-6:30pm **Locations (101):** Earth Sciences (ESB), Room 1012

Instructor: Leonid Sigal

E-mail: lsigal@cs.ubc.ca

Office: ICICS 119

TAs: Wan-Cyuan (Chris) Fan

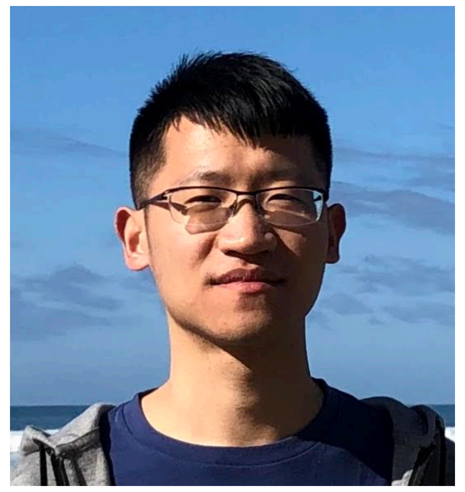


wancyuan@cs.ubc.ca

Jiayun Luo

letitial@student.ubc.ca

Bicheng Xu



bichengx@cs.ubc.ca

Shih-Han Chou

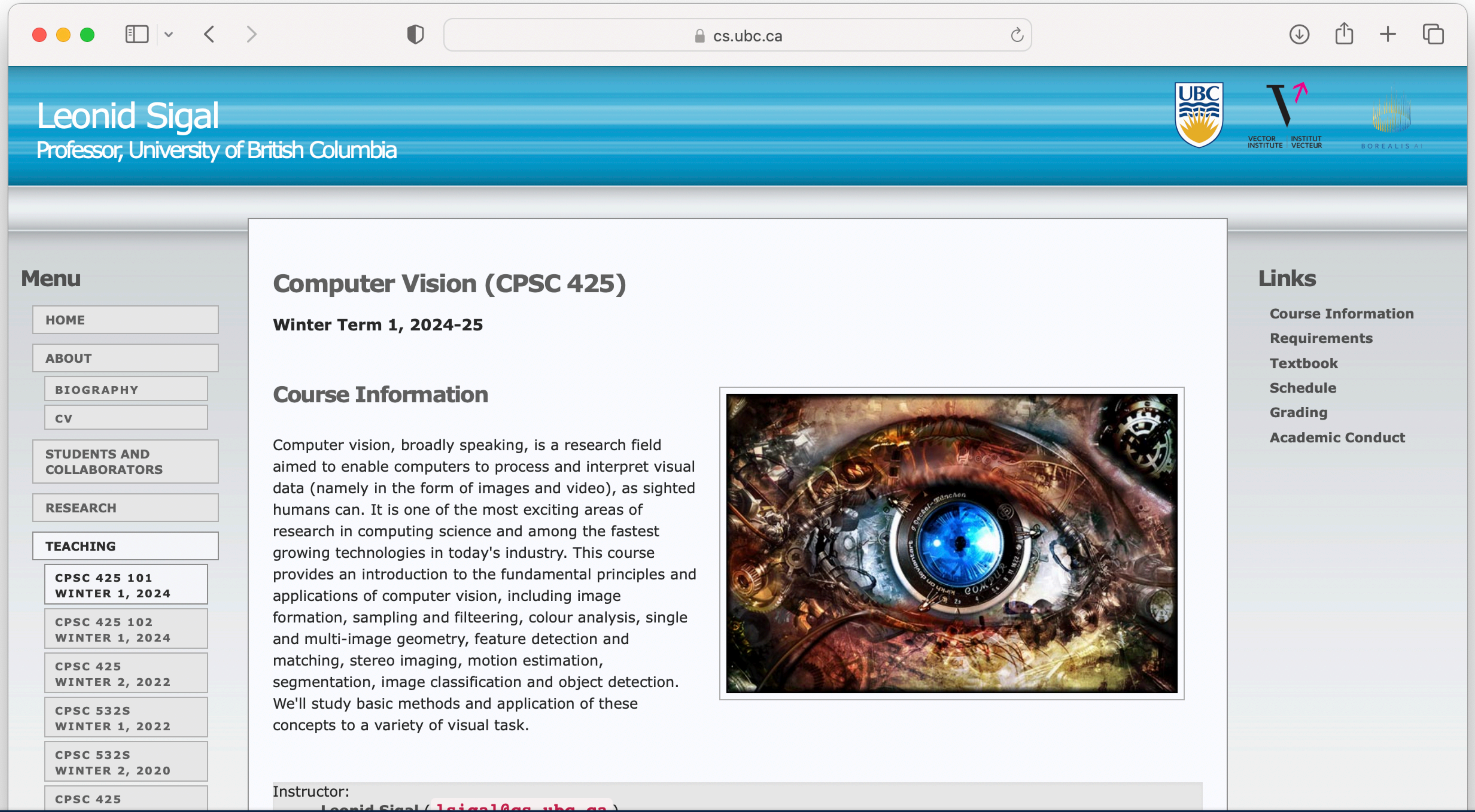
shchou75@cs.ubc.ca

Rayat Hossain

rayat137@cs.ubc.ca

Times: Mon, Wed 3:30-5:00pm **Locations (102):** MacMillan (MCML), Room 166

Course **Webpage**

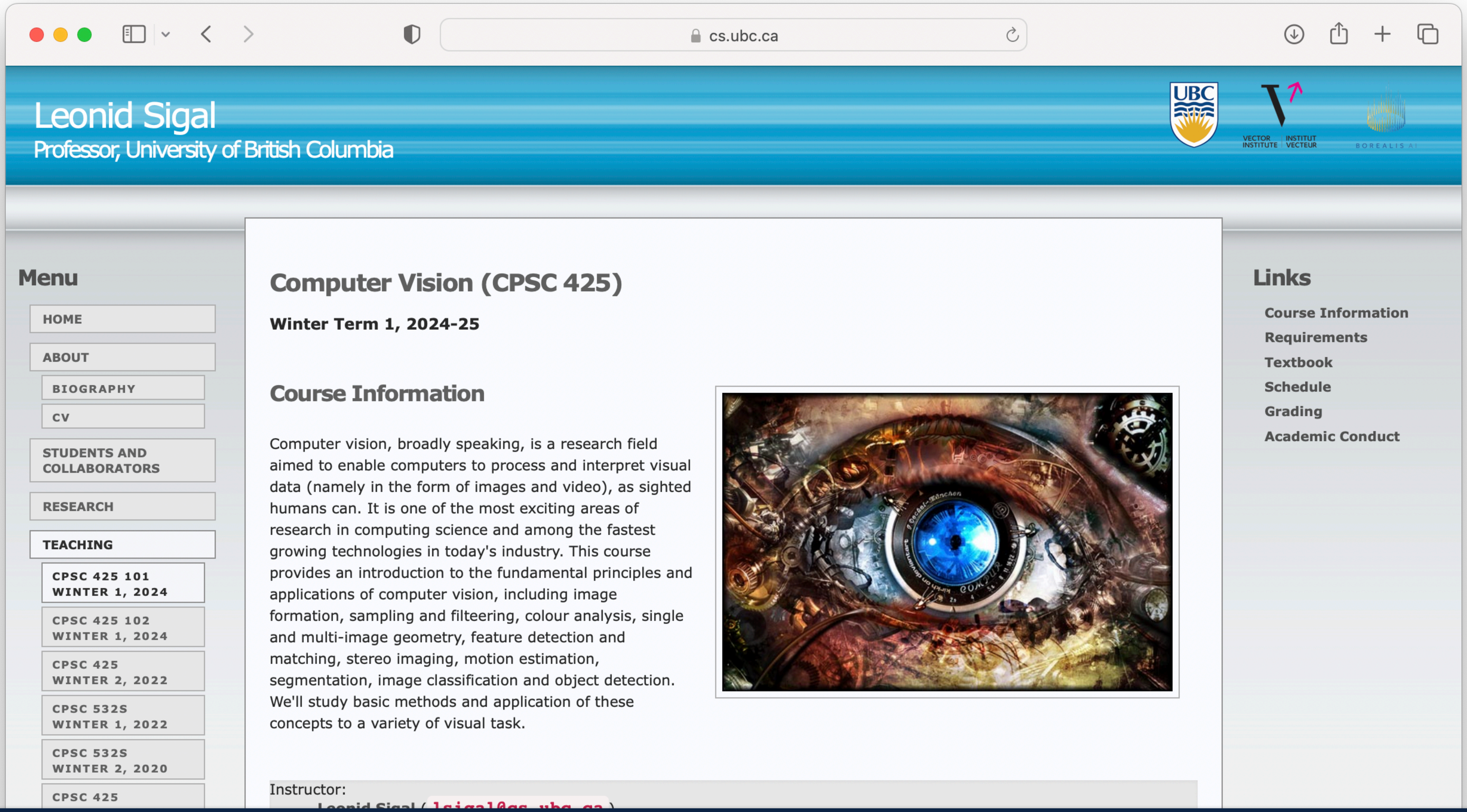


- Schedule, Assignments
- Lecture Slides and Notes
- Readings
- Course Information (public)

Section 101: https://www.cs.ubc.ca/~lsigal/teaching24_Term1a.html

Section 102: https://www.cs.ubc.ca/~lsigal/teaching24_Term1b.html

Course **Webpage**: Section 101 vs. 102

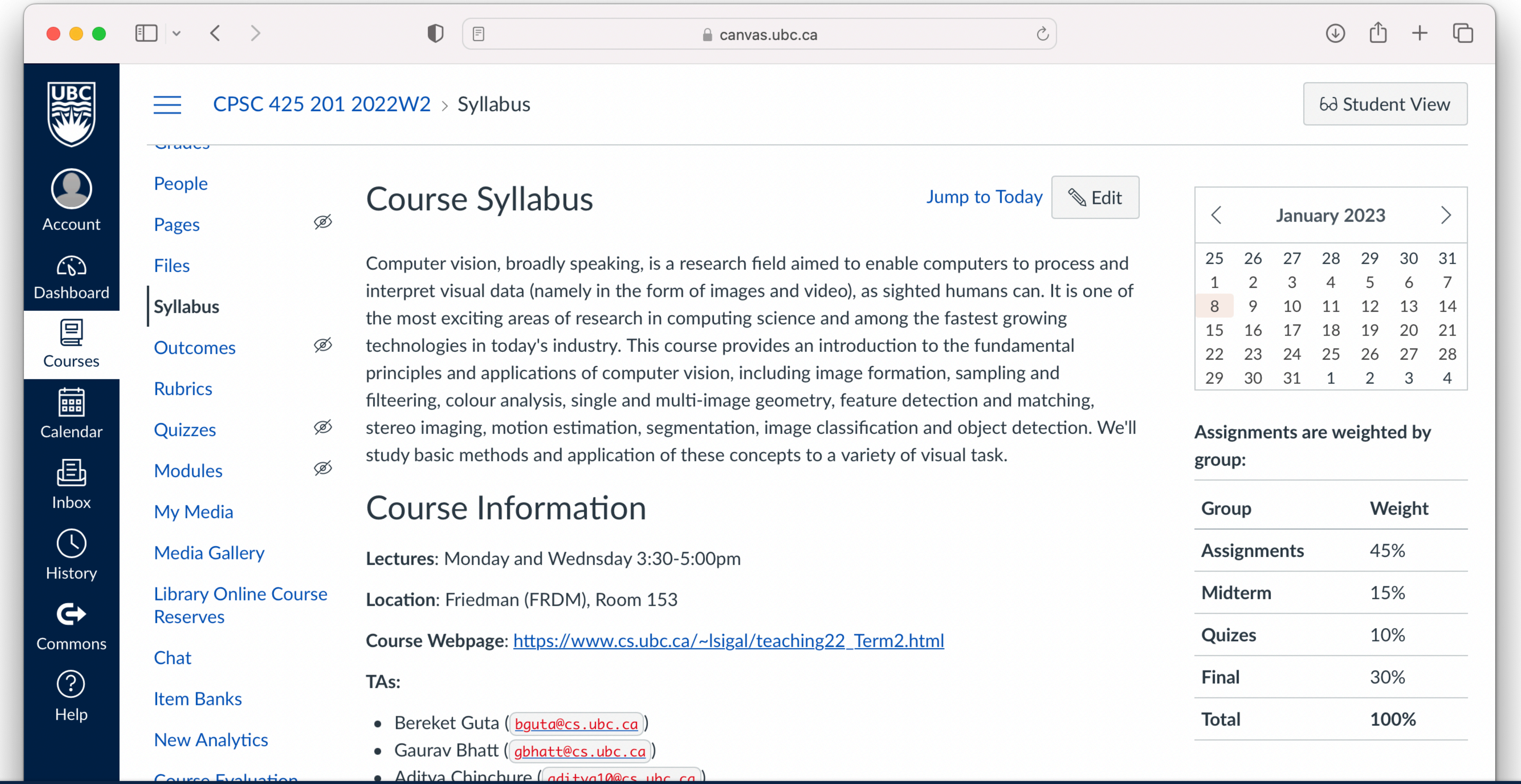


- Section 101 has three extra lectures (compared to 102)

Section 101: https://www.cs.ubc.ca/~lsigal/teaching24_Term1a.html

Section 102: https://www.cs.ubc.ca/~lsigal/teaching24_Term1b.html

Canvas



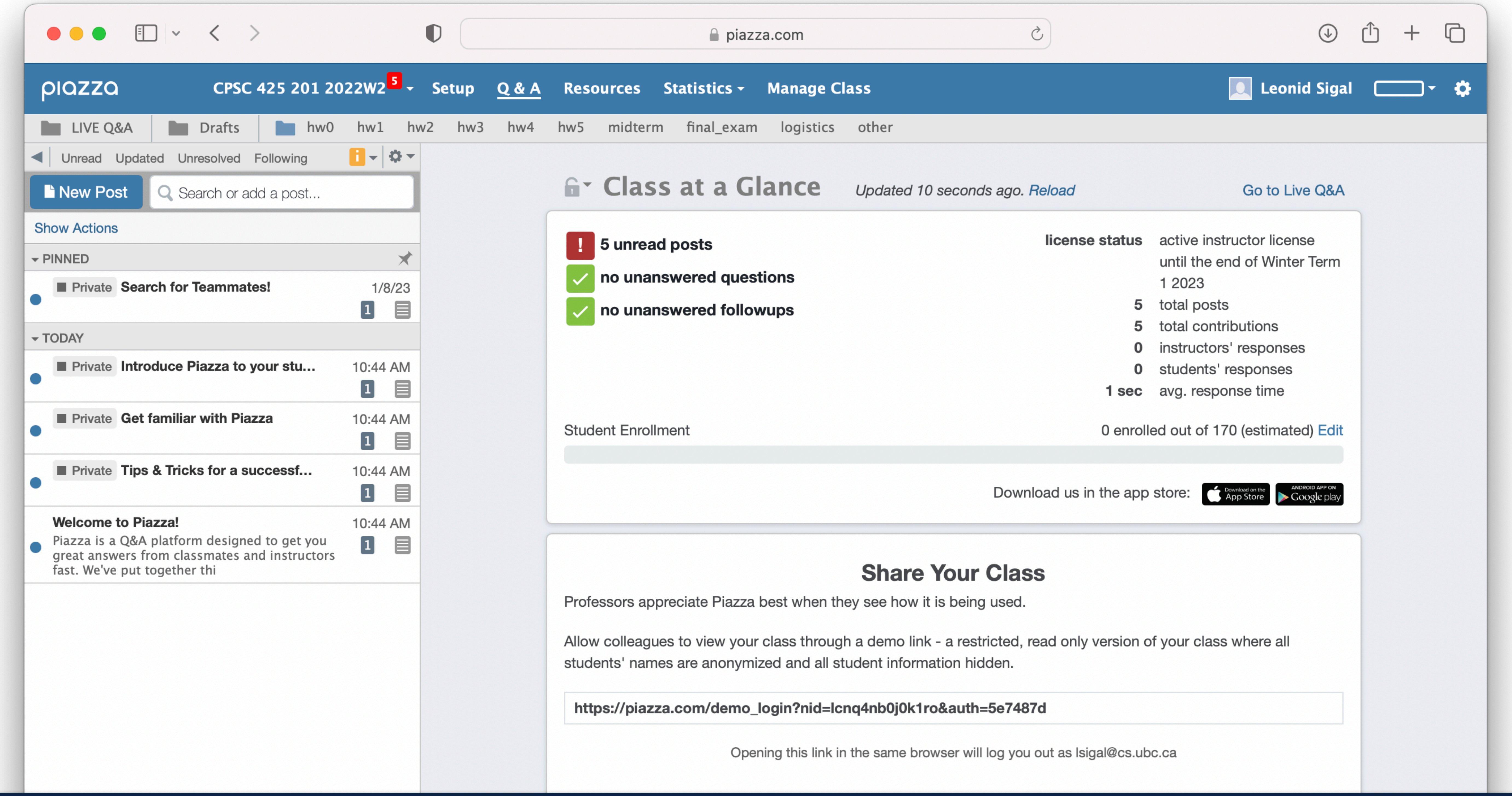
- Assignment hand-in
- Course Information (private)
- Piazza link

https://canvas.ubc.ca/courses/146440

Piazza

Discussion: https://piazza.com/ubc.ca/winterterm12024/cpsc_v4251011022024w1/home

65 students were enrolled as of AM



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

Sign up code in e-mail

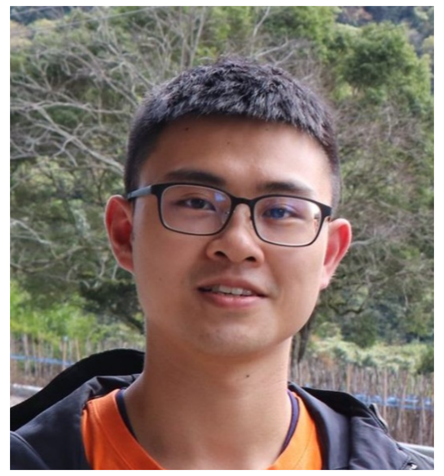
Office **Hours**

Will start **next** week

Instructor: Leonid Sigal

Friday 1-2pm

TAs: Wan-Cyuan (Chris) Fan



Friday 4-5pm

Jiayun Luo

Thursday 1-2pm

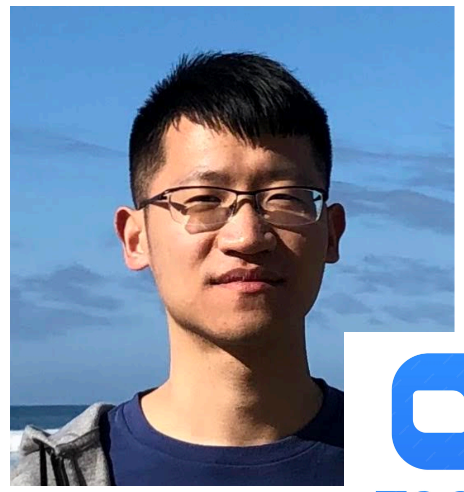
Shih-Han Chou

Wednsday 7-8pm

Rayat Hossain

Tuesday noon-1pm

Bicheng Xu



Monday 7-8pm

See Piazza for Links and Locations (mix of in-person and Zoom)

Special **Circumstances**

I will be **away** September 12th — 17th (inclusively)

Kwang Moo Yi

September 16th, 17th

I will also make recordings of these lectures from past year available

How important is **Vision**?

How important is **Vision**?

To answer this questions, we need to go back to about

…. **543 million years, B.C.**

How important is **Vision**?

To answer this questions, we need to go back to about

…. **543 million years, B.C.**

Vision is really fundamental to life and evolution

What is **Computer Vision**?

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

What is **Computer Vision**?

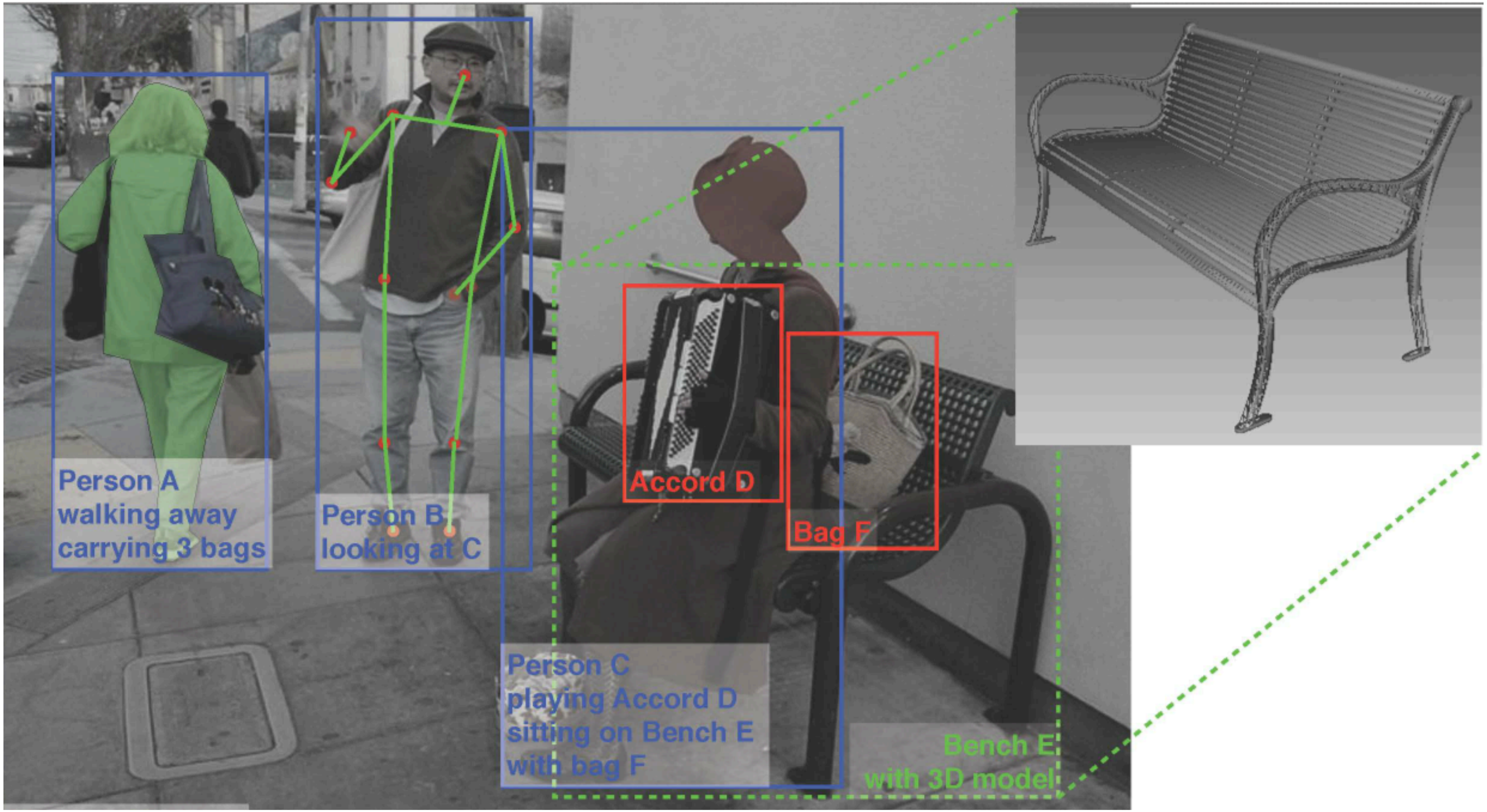
Compute vision, broadly speaking, is a research field aimed to enable computers to **process and interpret visual data**, as sighted humans can.

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

What do **you** see?

Slide Credit: Jitendra Malik (UC Berkeley)

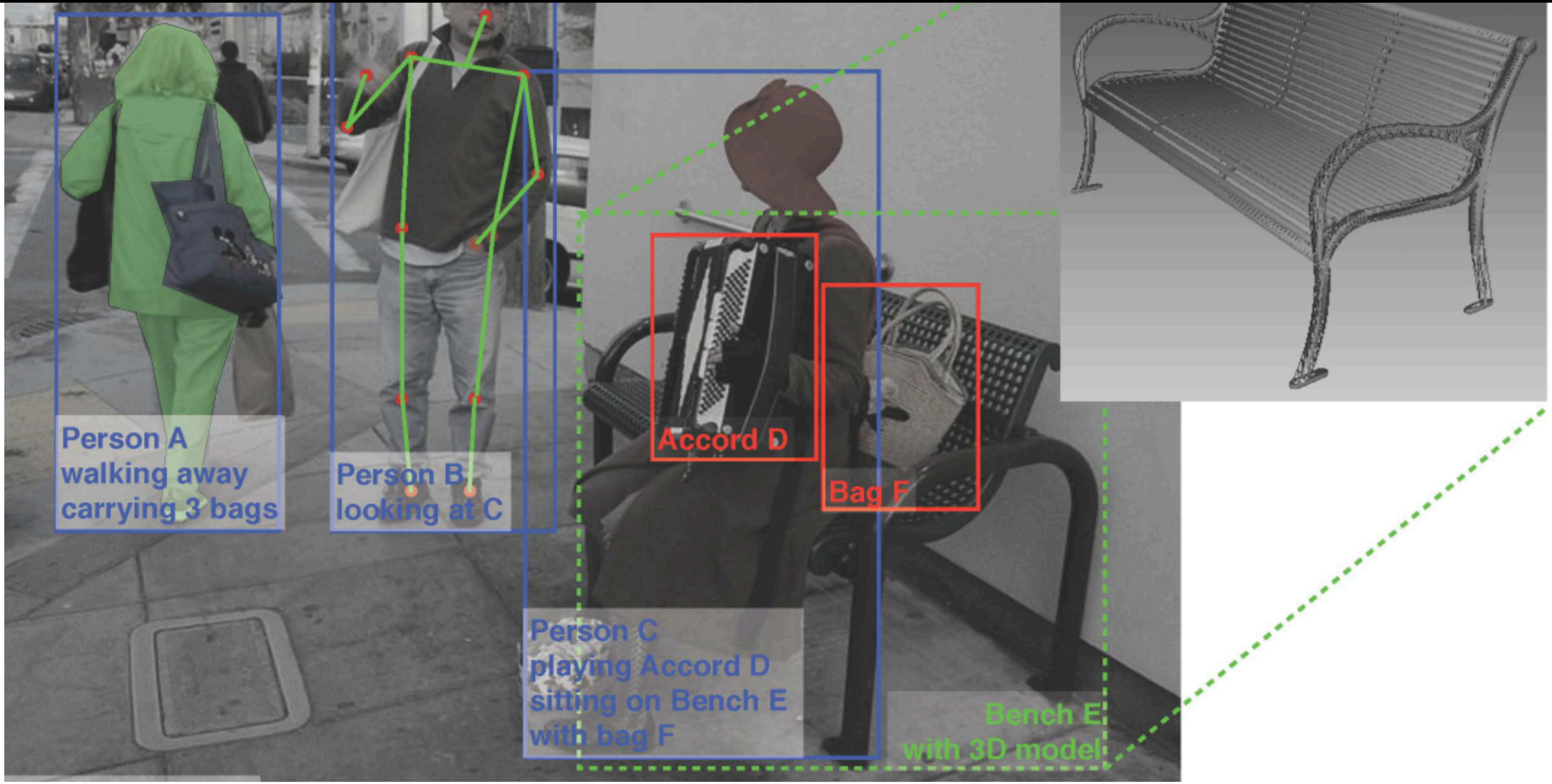
What we would like **computer to infer**?



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)

What is **Computer Vision**?

Compute vision, broadly speaking, is a research field aimed to enable computers to **process and interpret visual data**, as sighted humans can.

Sensing Device

Interpreting Device

Image (or video)

Interpretation

blue sky,
trees,
fountains,
UBC, …

What is **Computer Vision**?

Compute vision, broadly speaking, is a research field aimed to enable computers to **process and interpret visual data**, as sighted humans can.

Image (or video)

Sensing Device

Interpreting Device

Image Credit: https://www.flickr.com/photos/flamephoenix1991/8376271918

Interpretation

blue sky,
trees,
fountains,
UBC, …

What is **Computer Vision**?

Compute vision, broadly speaking, is a research field aimed to enable computers to **process and interpret visual data**, as sighted humans can.

Image (or video)

Sensing Device

Interpreting Device

Image Credit: https://www.flickr.com/photos/flamephoenix1991/8376271918

Interpretation

blue sky,
trees,
fountains,
UBC, …

What is **Computer Vision**?

Compute vision, broadly speaking, is a research field aimed to enable computers to **process and interpret visual data**, as sighted humans can.

Image (or video)

Sensing Device

Interpreting Device

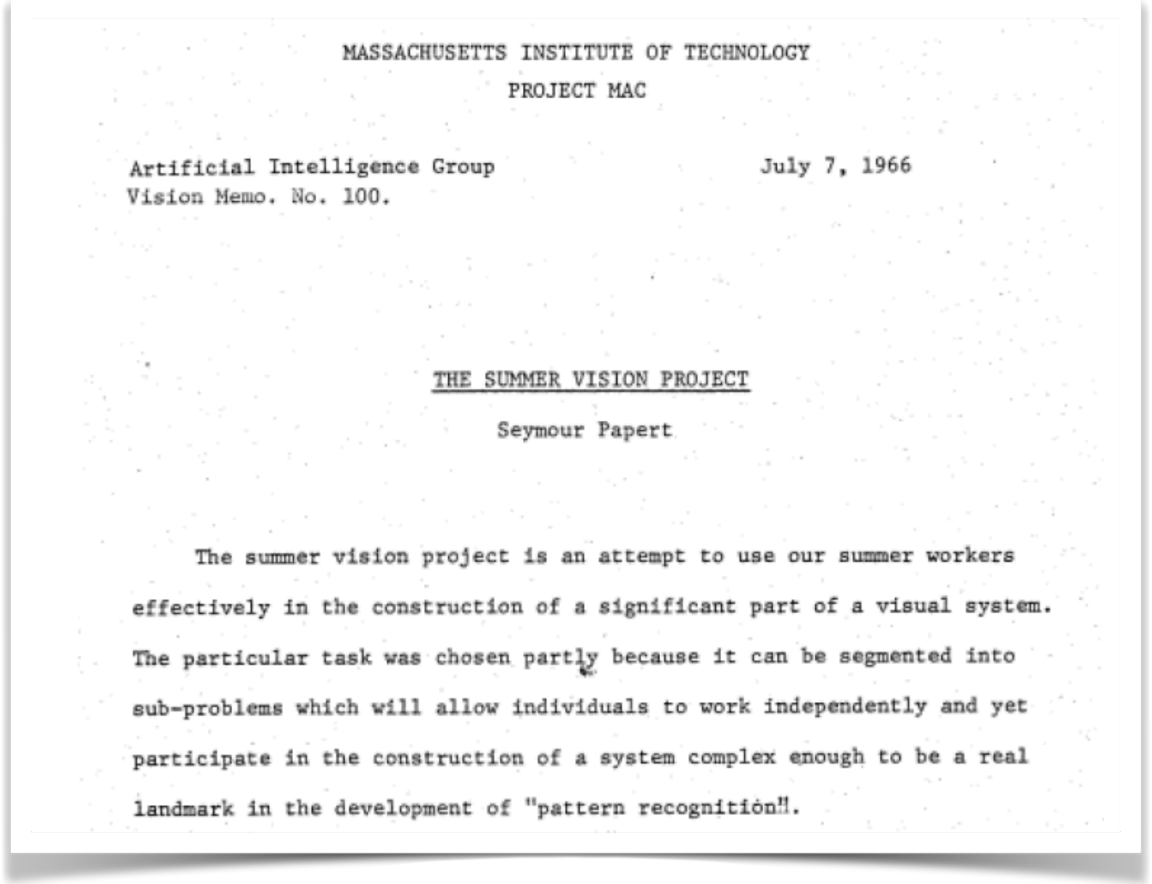
1918

Interpretation

blue sky,
trees,
fountains,
UBC, …

This implies a very strong connection between **Computer Vision** and **(Machine) Learning**

Computer vision … the beginning …



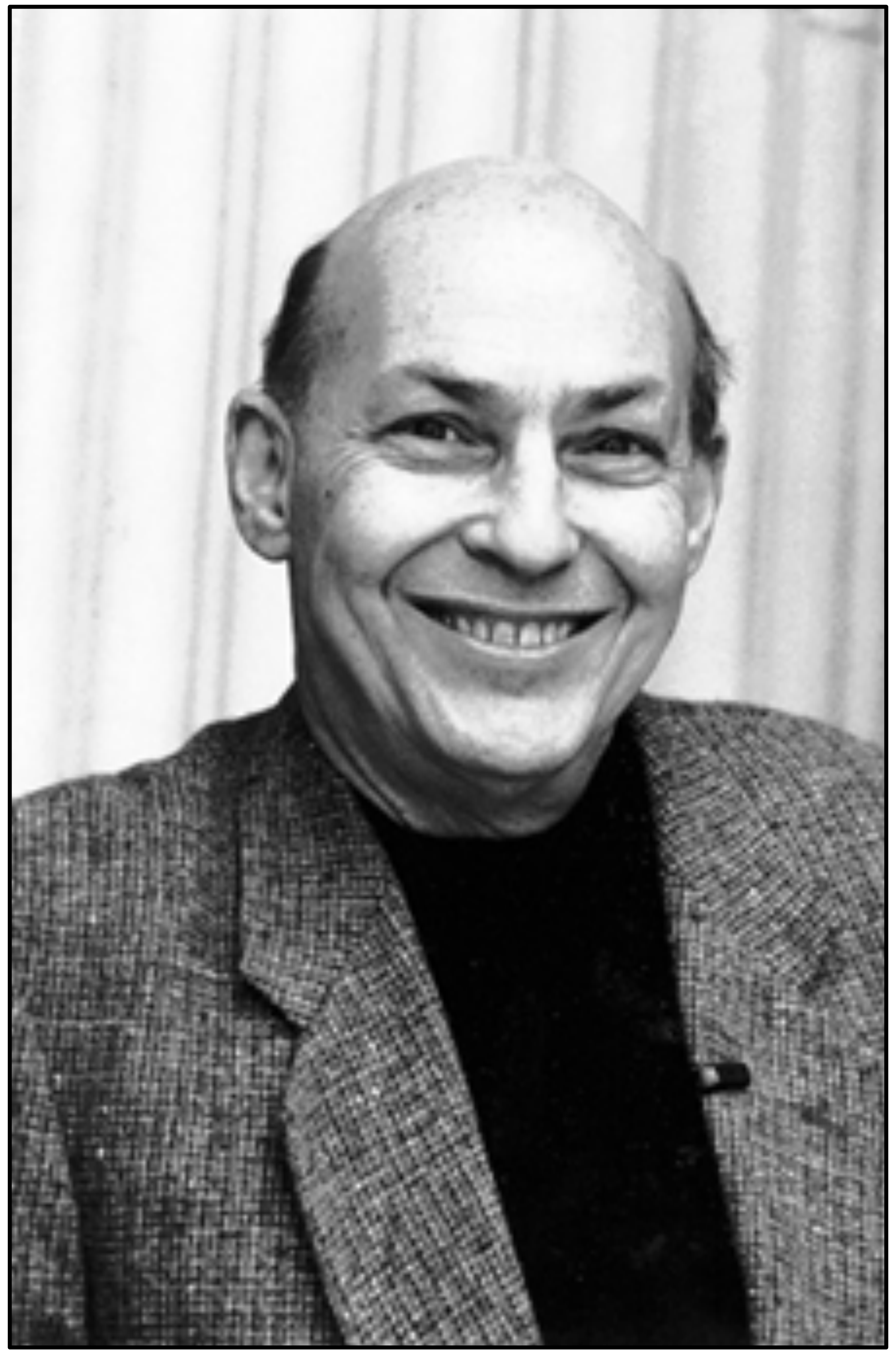
MASSACHUSETTS INSTITUTE OF TECHNOLOGY
PROJECT MAC

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)

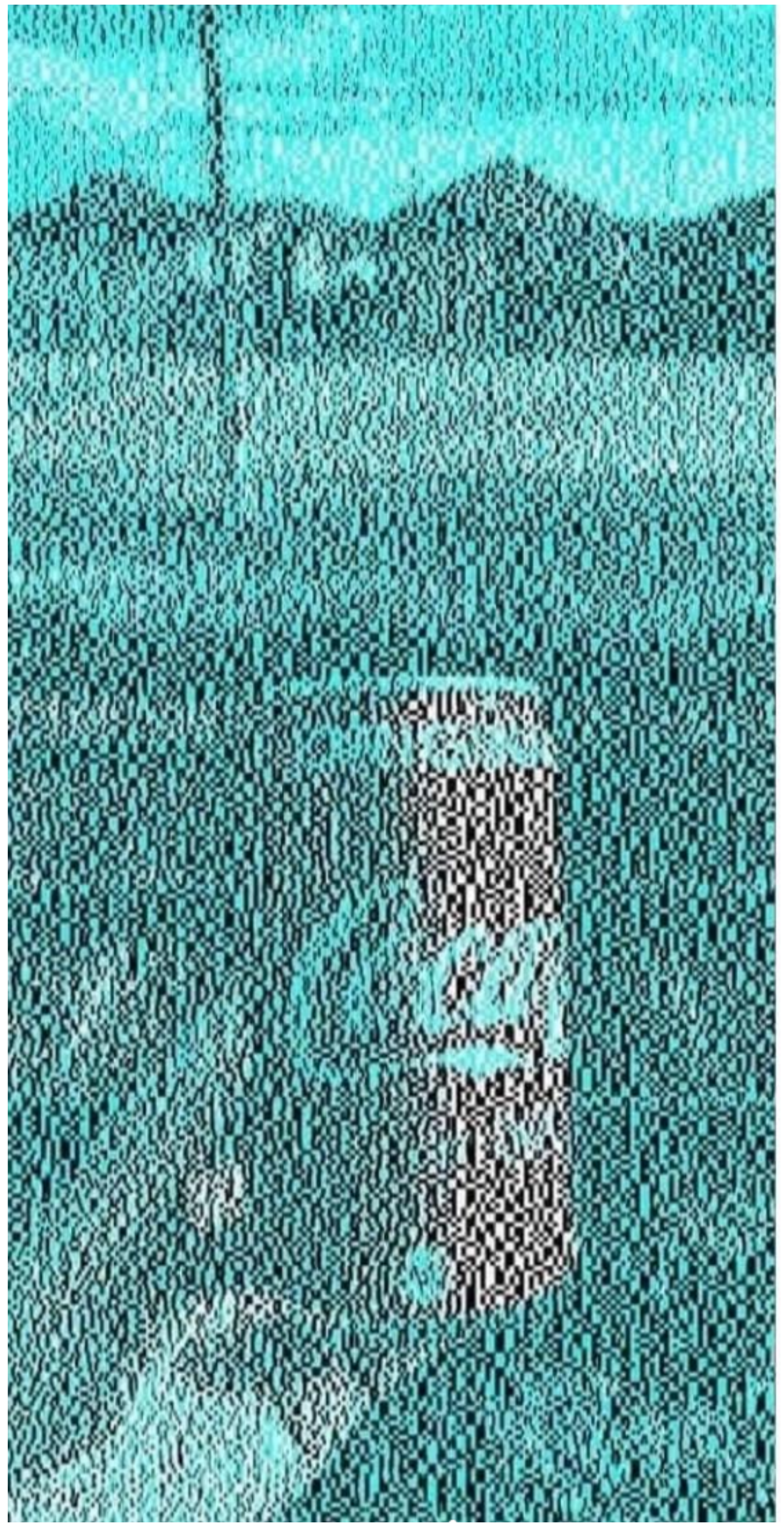
Can computers **match (or beat)** human vision?

- We've been at it for 50 years

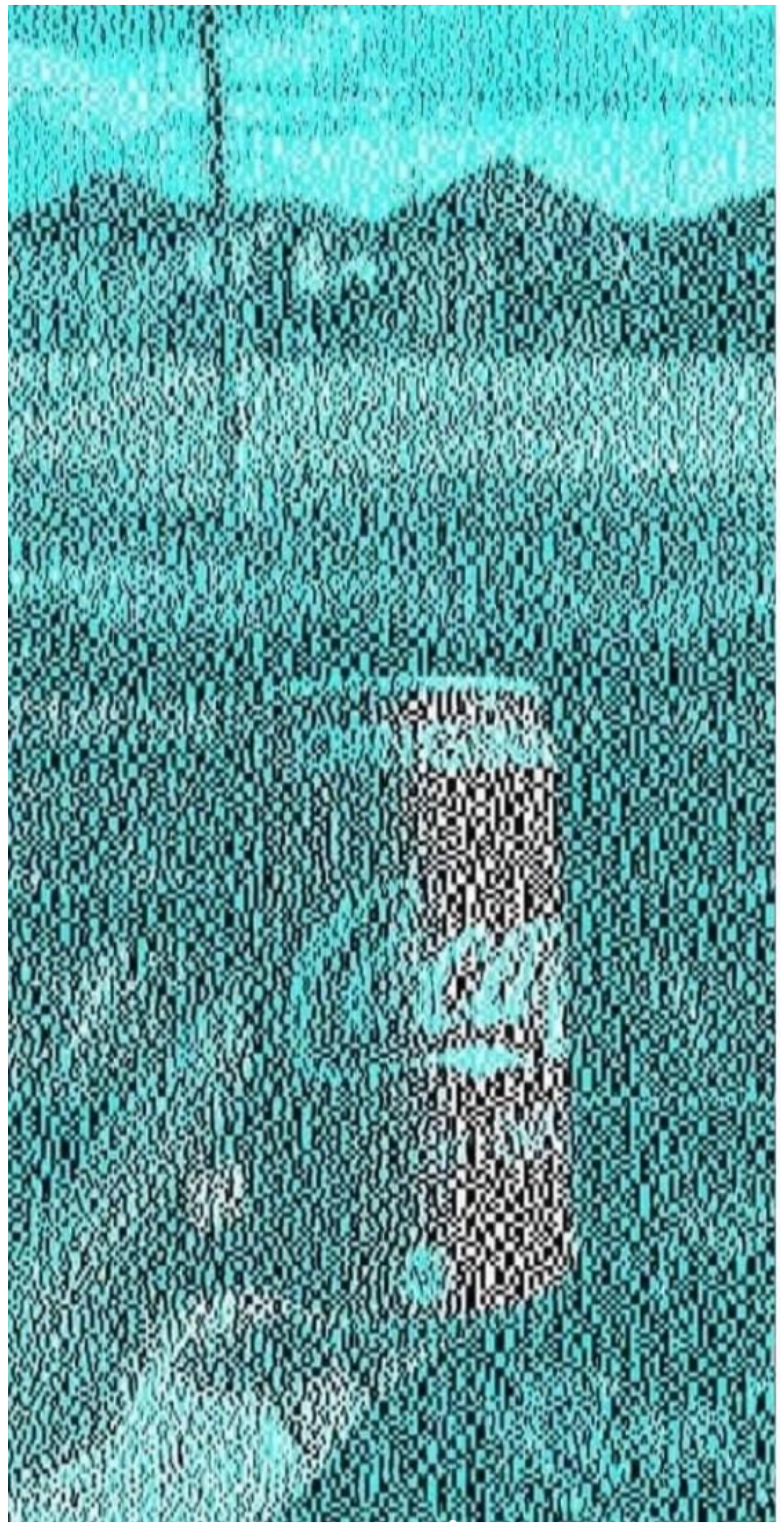
Can computers **match (or beat)** human vision?

- How good is human vision?

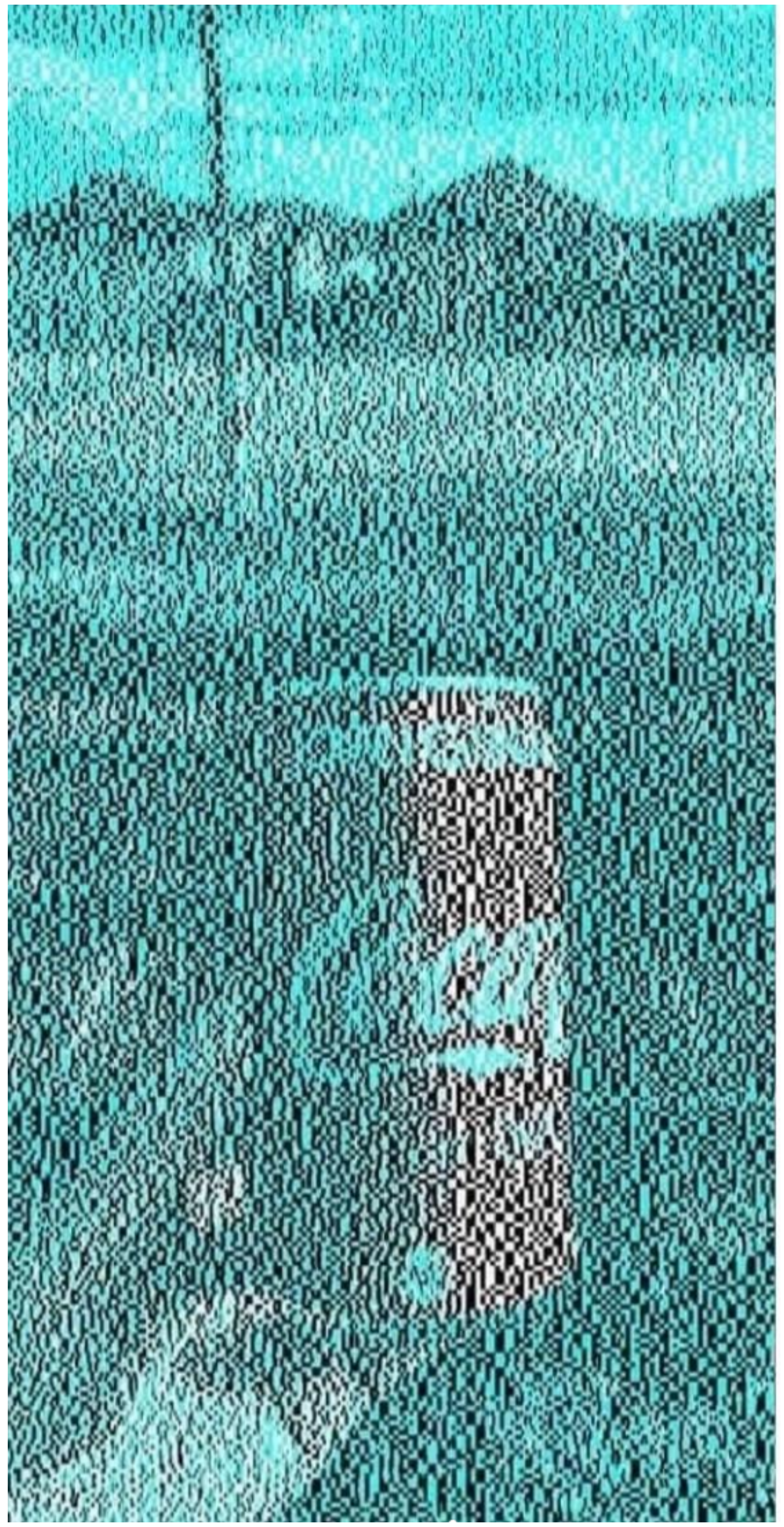
Can computers **match (or beat)** human vision?



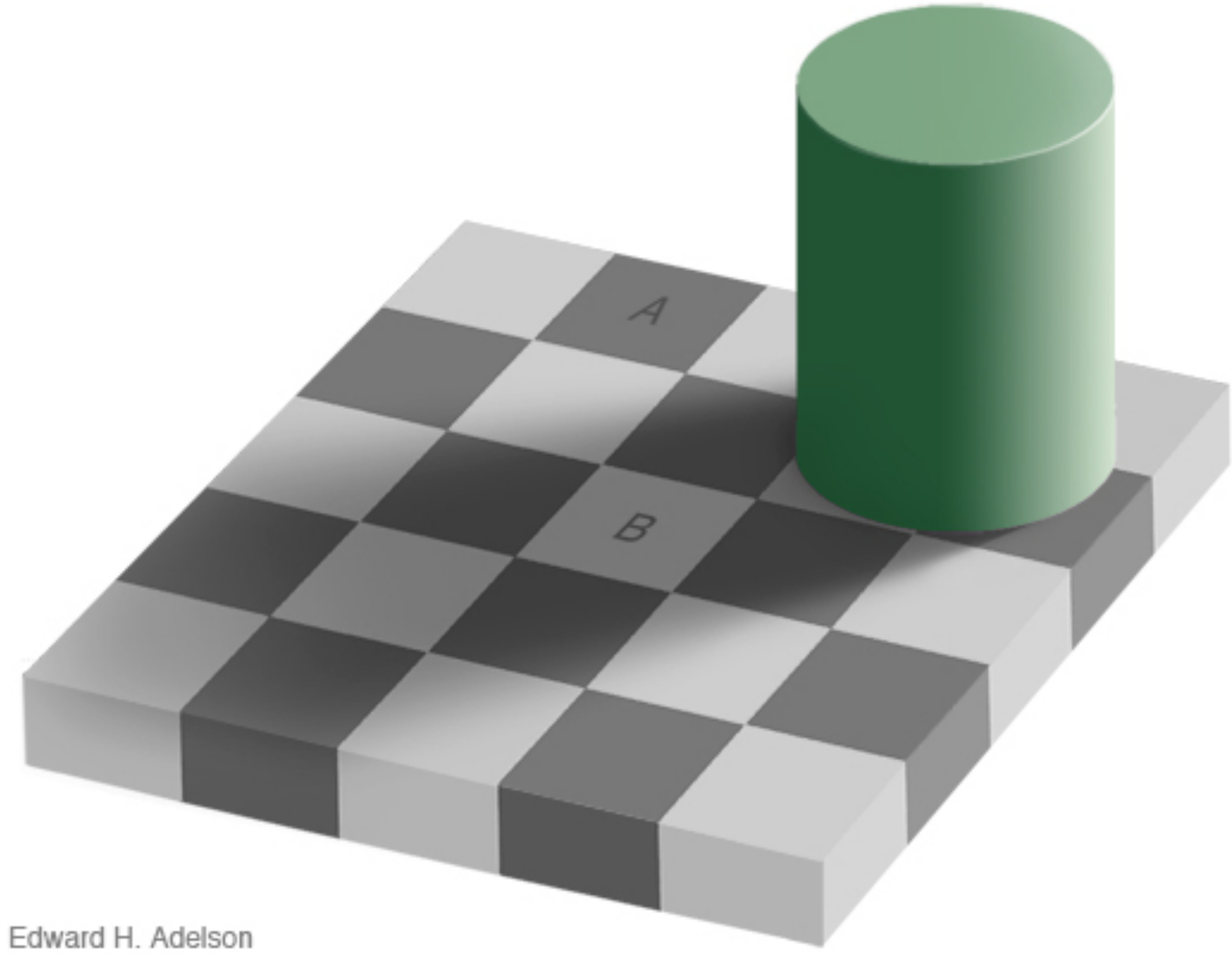
Can computers **match (or beat)** human vision?



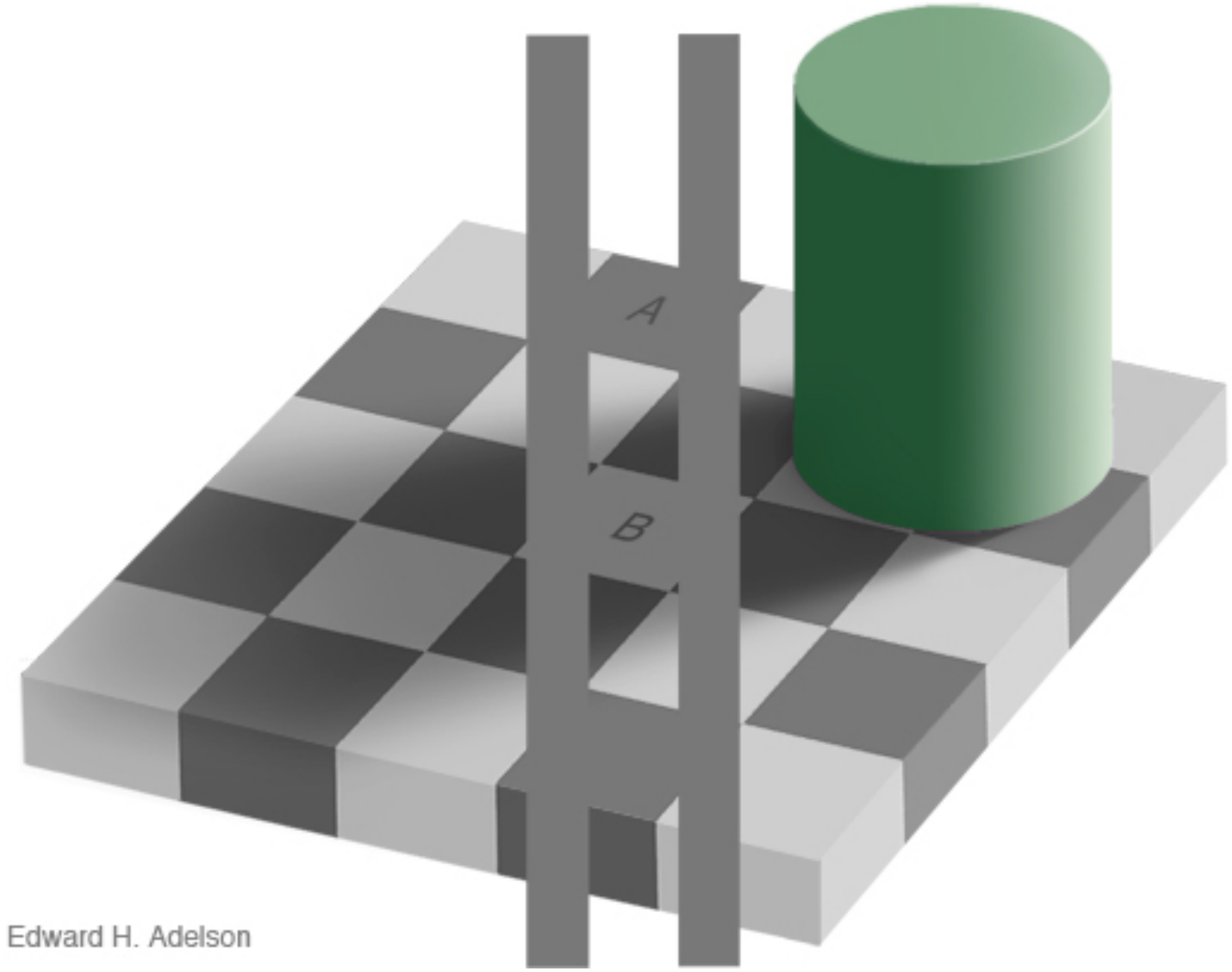
Can computers **match (or beat)** human vision?



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



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



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

- How good is human vision?

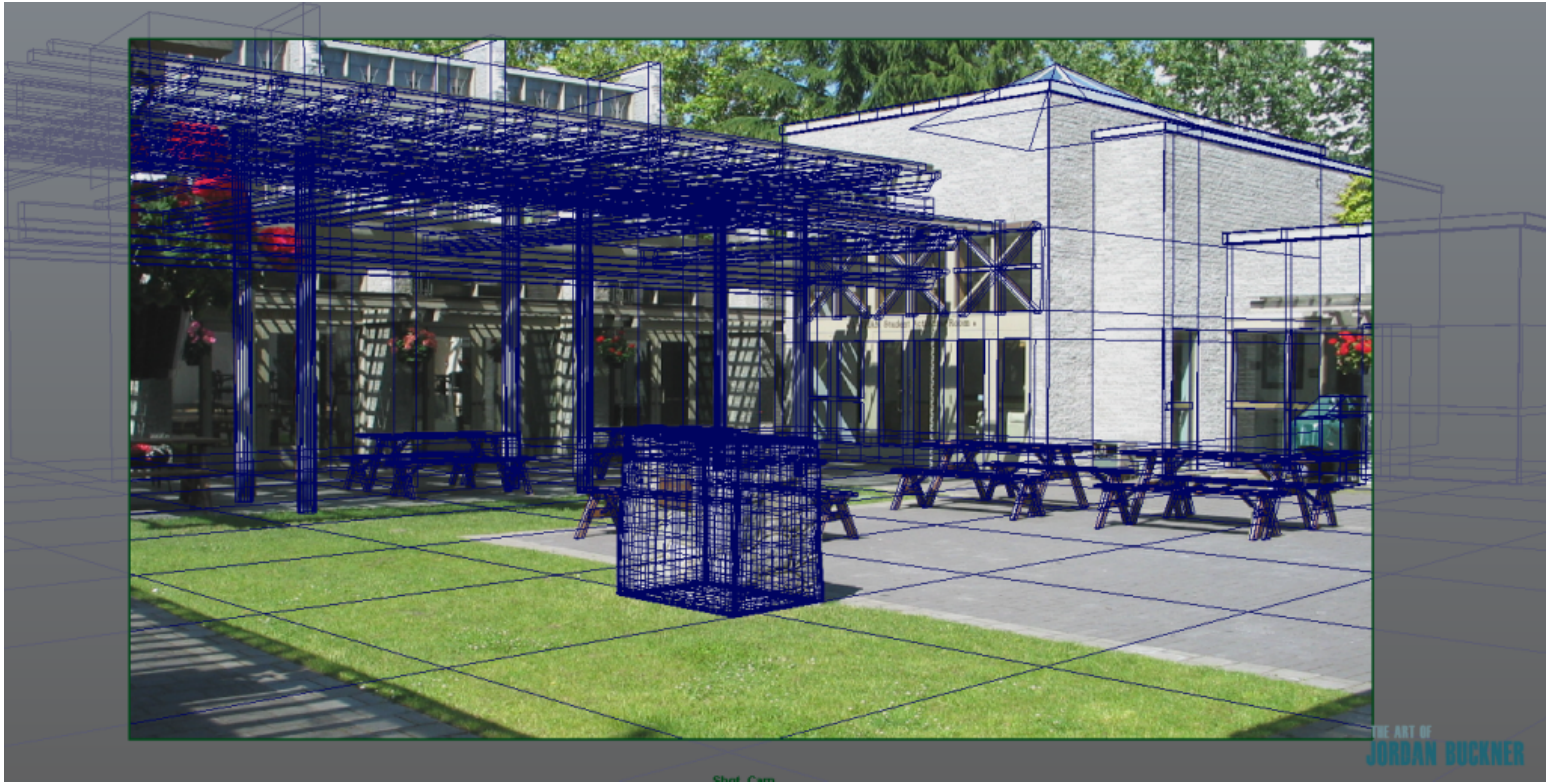
  As a measuring device not very good, as a functioning device really good

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

- **Yes and No** (mostly NO)

Alternative definition of computer vision

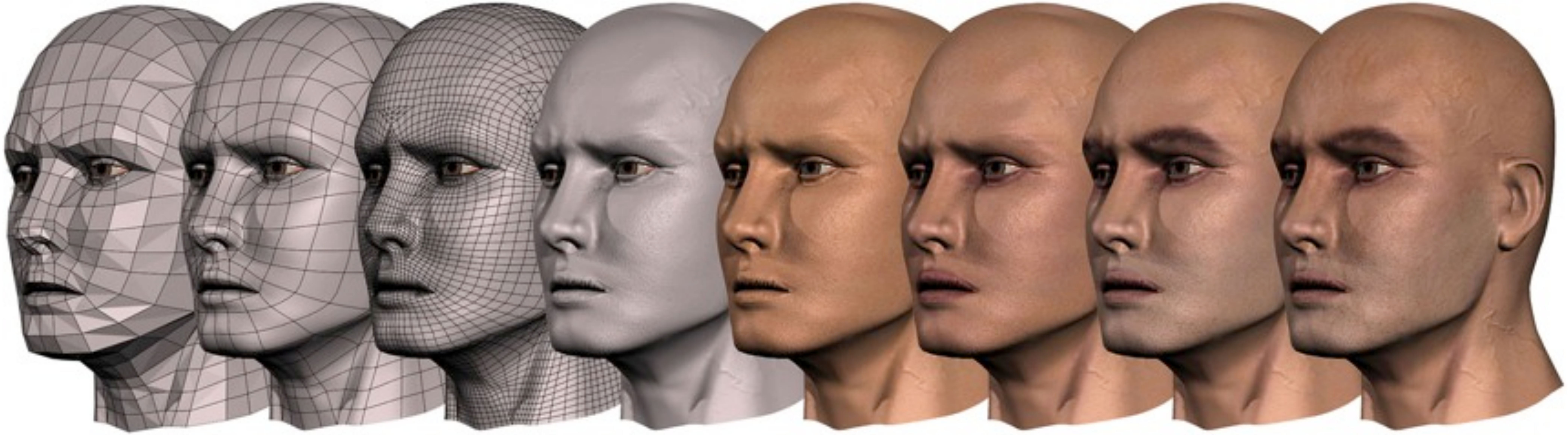
“Inverse Computer Graphics”



[J. Buckner]

Alternative definition of computer vision

"Inverse Computer Graphics"



Graphics

Vision

Computer **Vision Problems**

Slide Credit: Kristen Grauman (UT Austin)

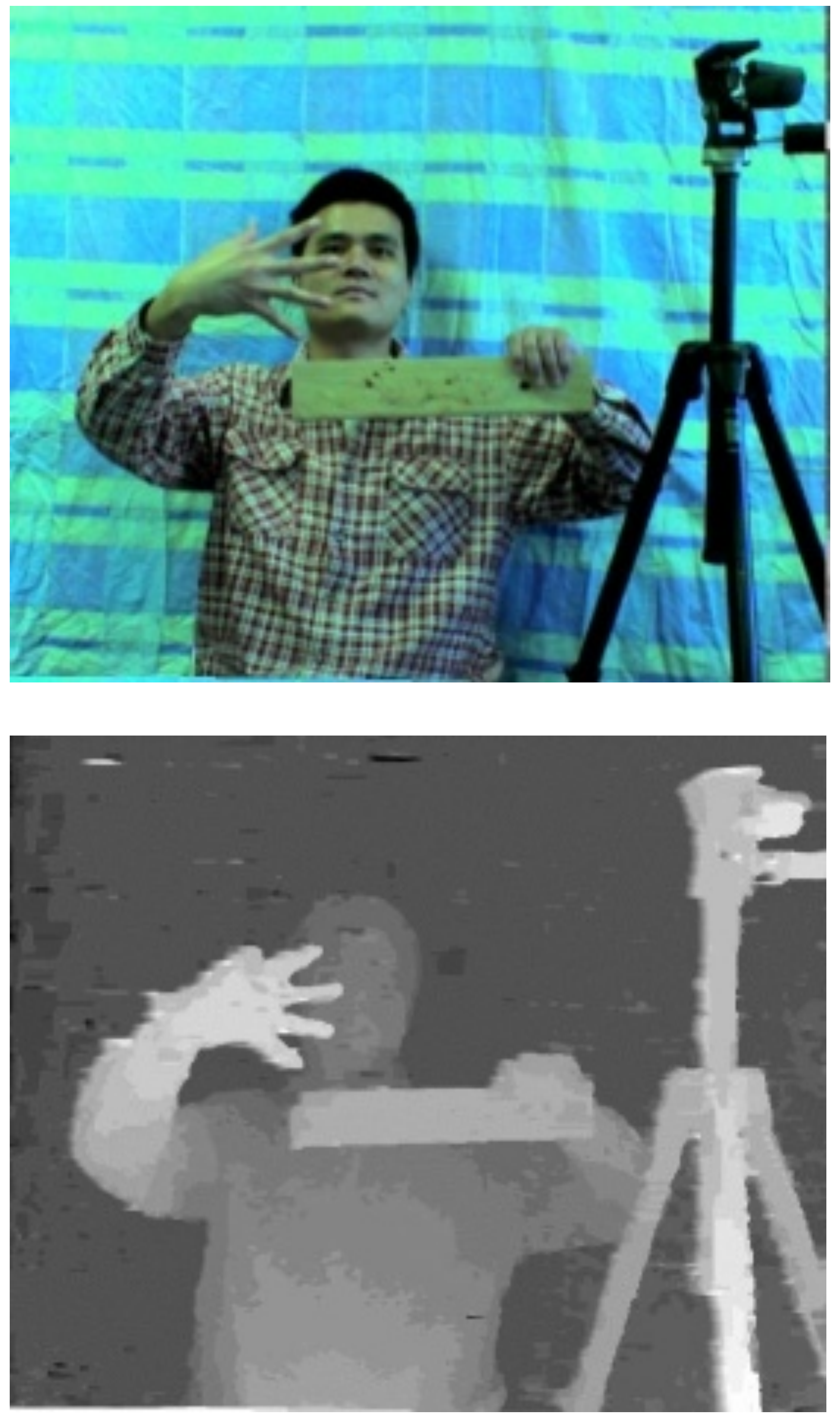
Computer **Vision Problems**

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

Slide Credit: Kristen Grauman (UT Austin)

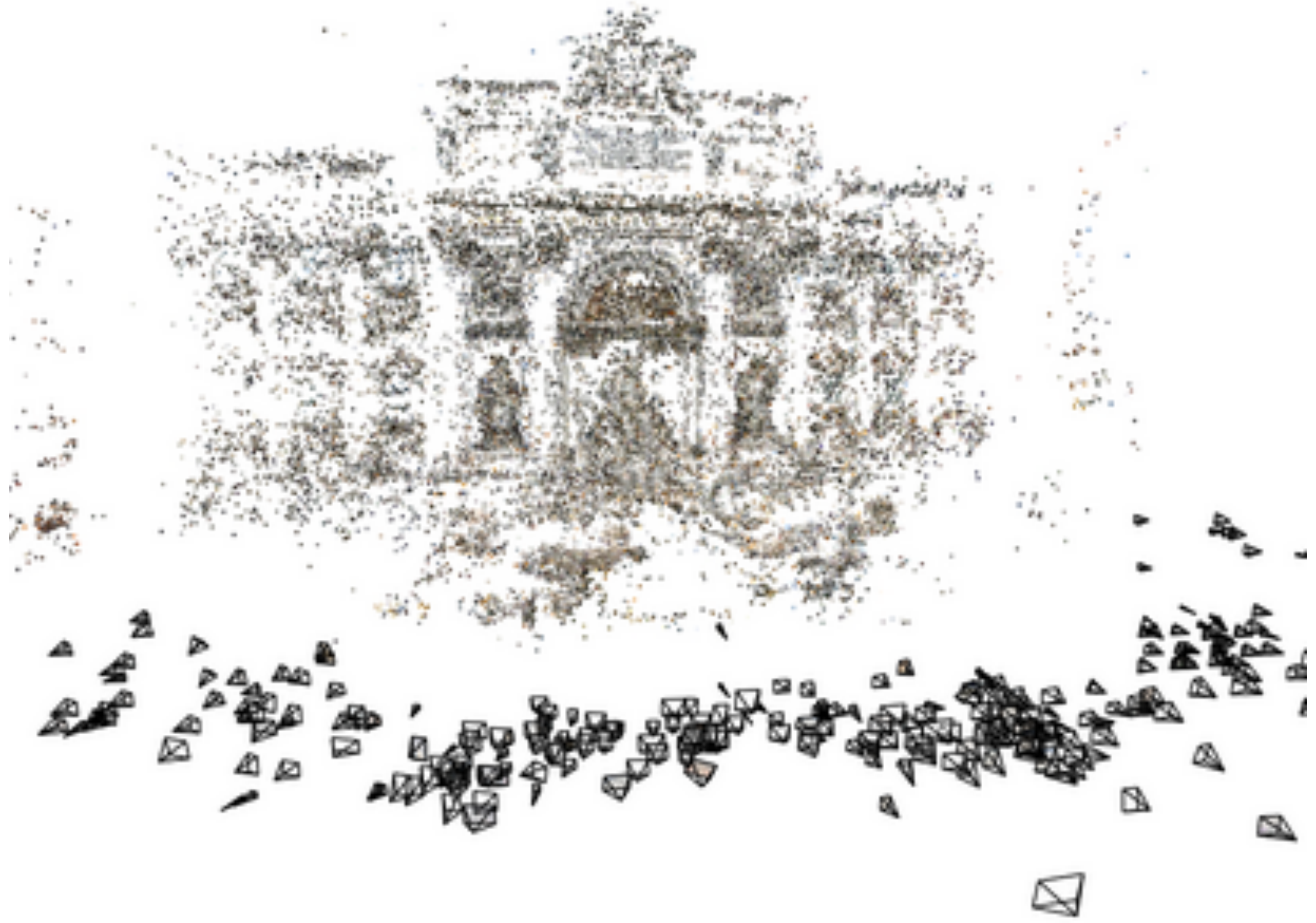
1. Vision for **Measurement**

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***)

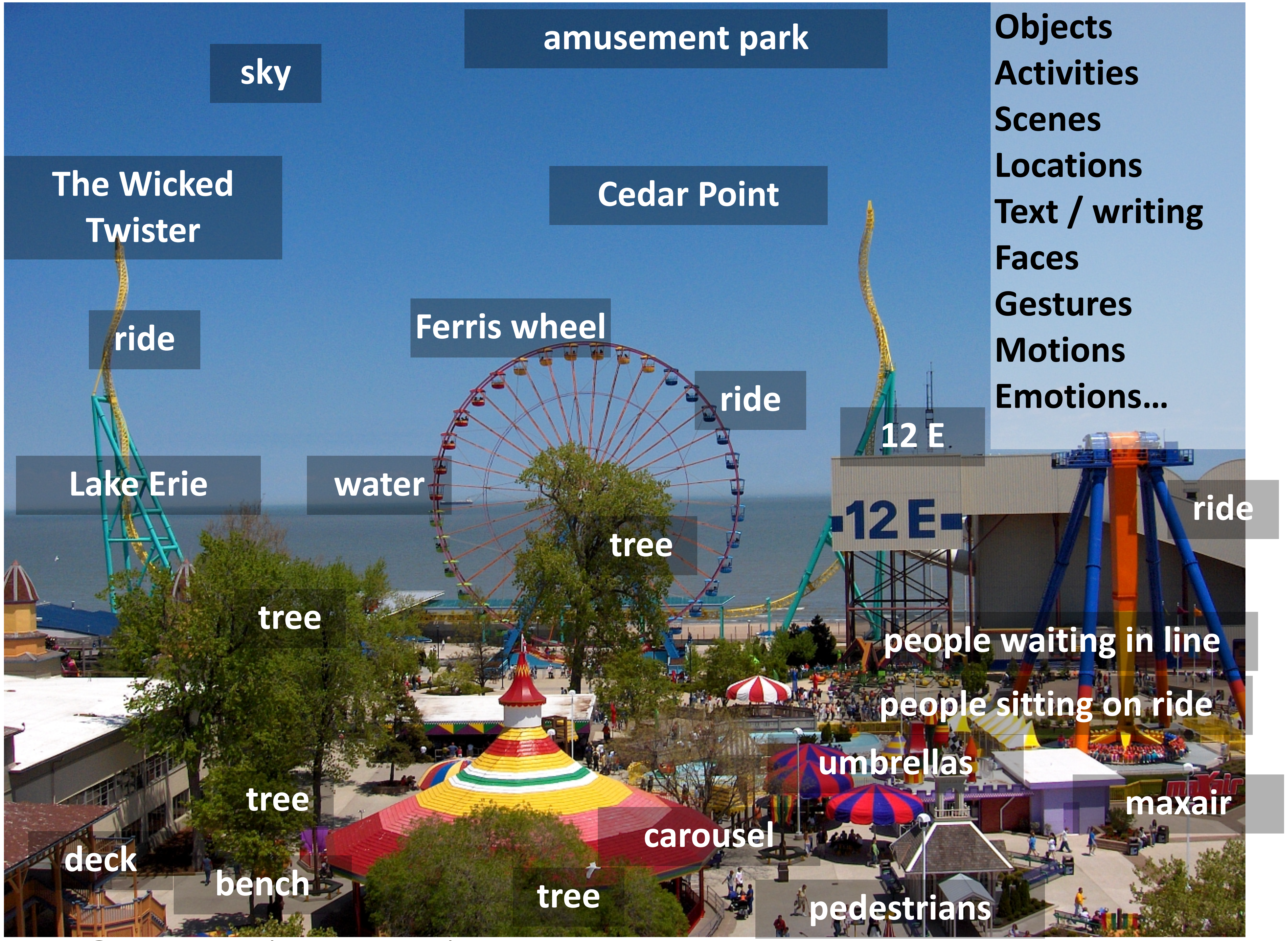
2. Algorithms and representations to allow a machine to recognize objects, people, scenes, and activities (***perception and interpretation***)

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**

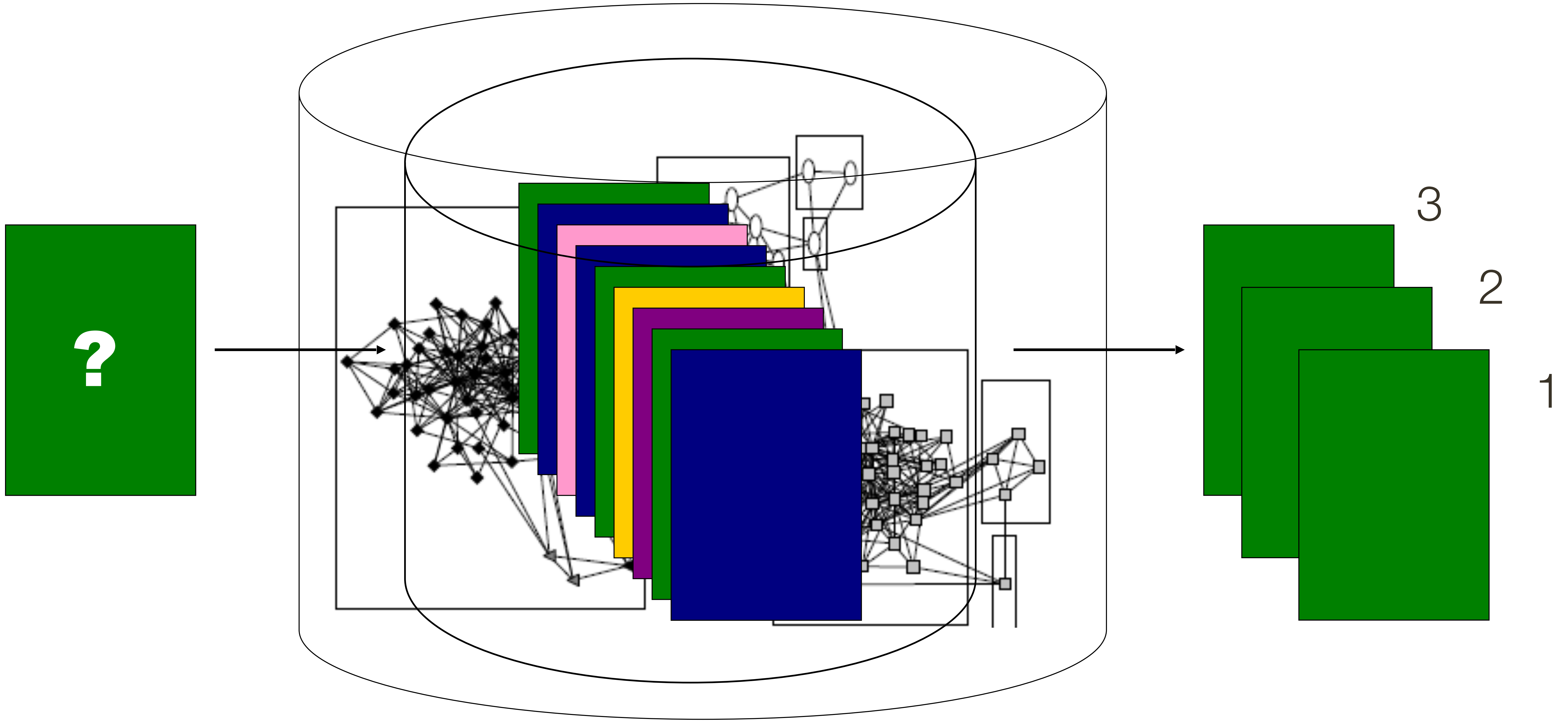
1. Computing properties of the 3D world from visual data (***measurement***)

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**



Query **Image or video archives** **Relevant content**

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

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)

3. **Search and Organization**

*from iStock by GettyImages

> 85% of all web content is multimedia content of visual form

Snapchat	WhatsApp	Facebook
31.7 Million	**29.2 Million**	**14.6 Million** our

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***)

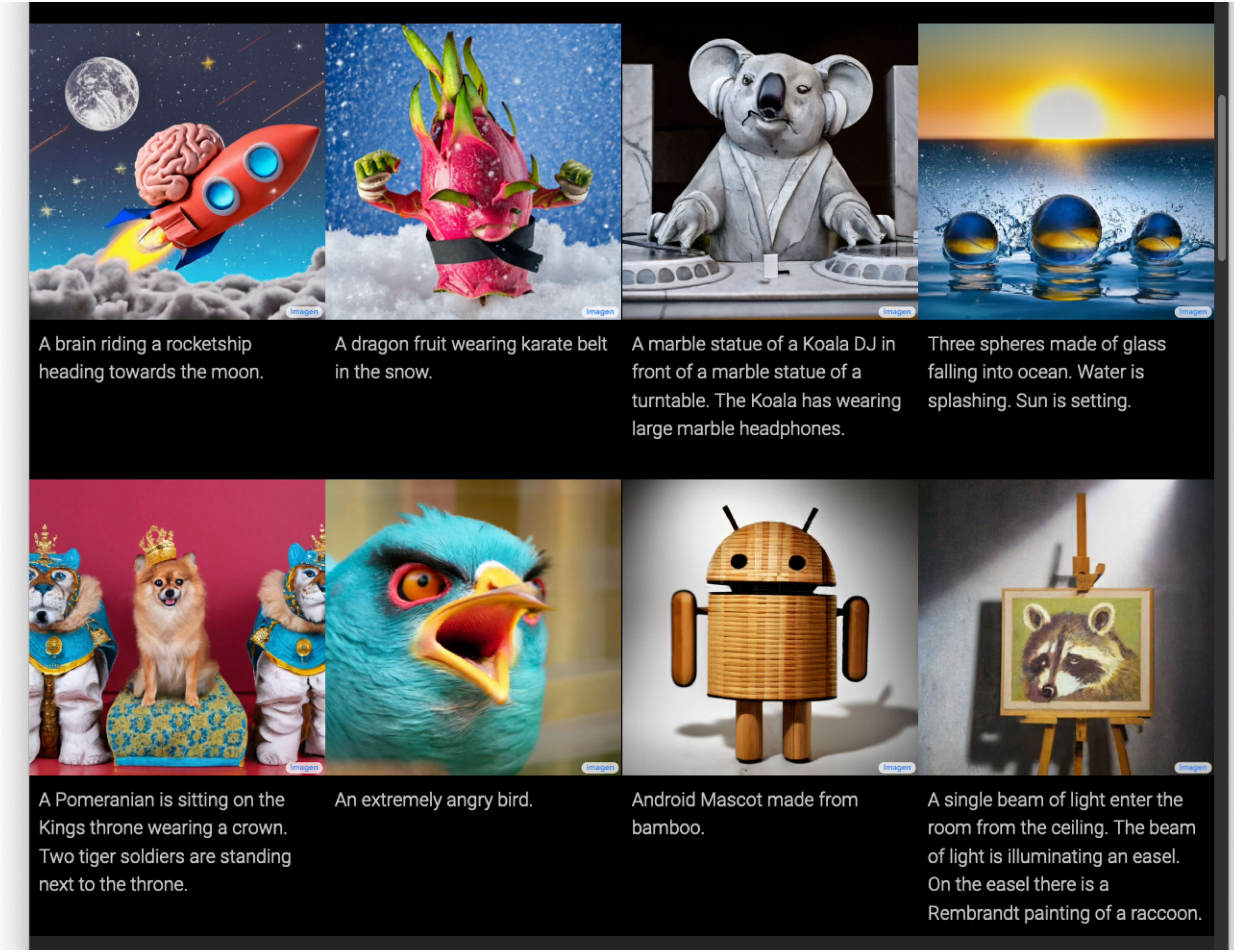
2. Algorithms and representations to allow a machine to recognize objects, people, scenes, and activities (***perception and interpretation***)

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

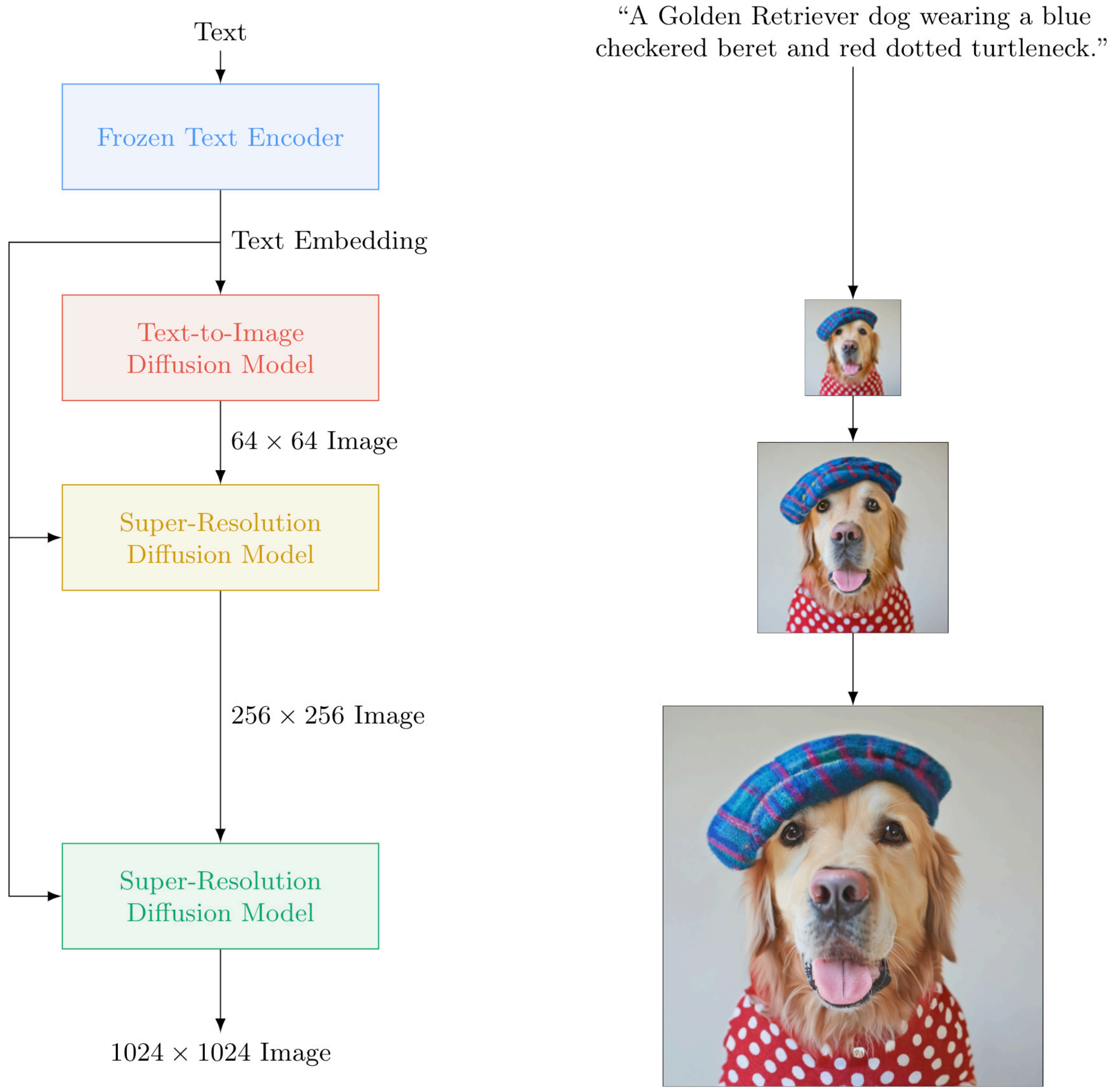
4. Algorithms for manipulation or creation of image or video content (***visual imagination***)

Slide Credit: Kristen Grauman (UT Austin)

4. Visual **Imagination**

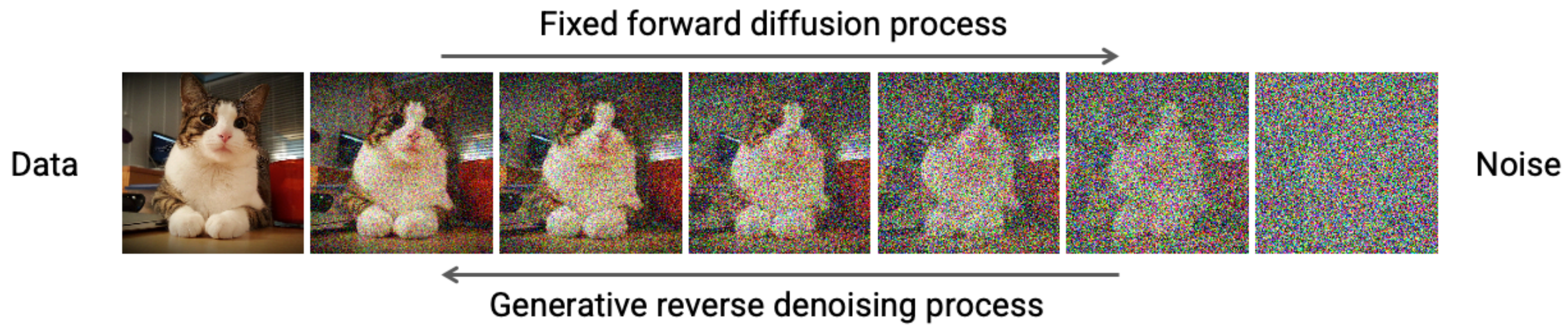


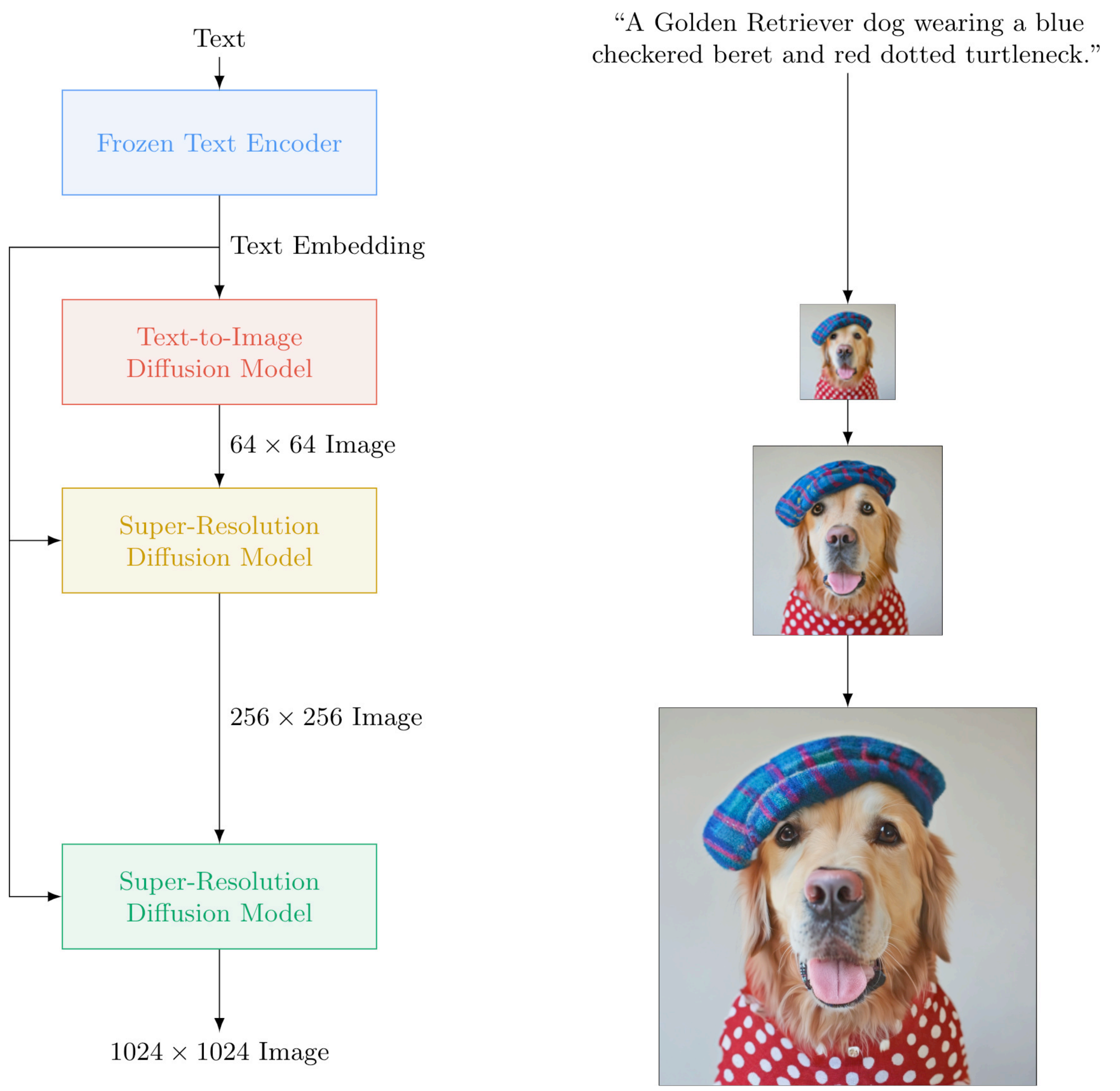
- imagen.research.google
- Text to image generation
- Uses diffusion process, training using large dataset of text (web scale) and image-text (400M) pairs



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***)

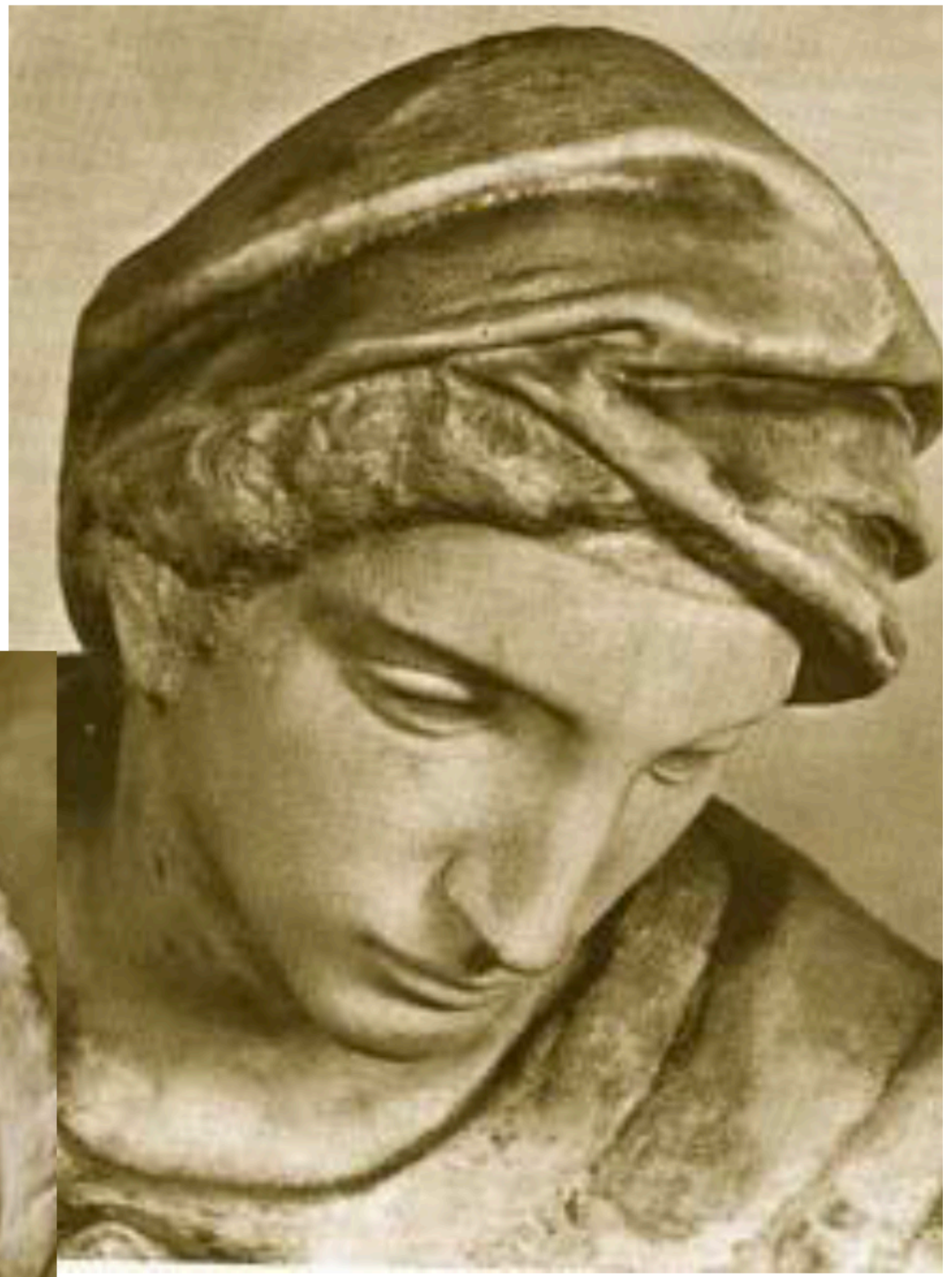
2. Algorithms and representations to allow a machine to recognize objects, people, scenes, and activities (***perception and interpretation***)

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

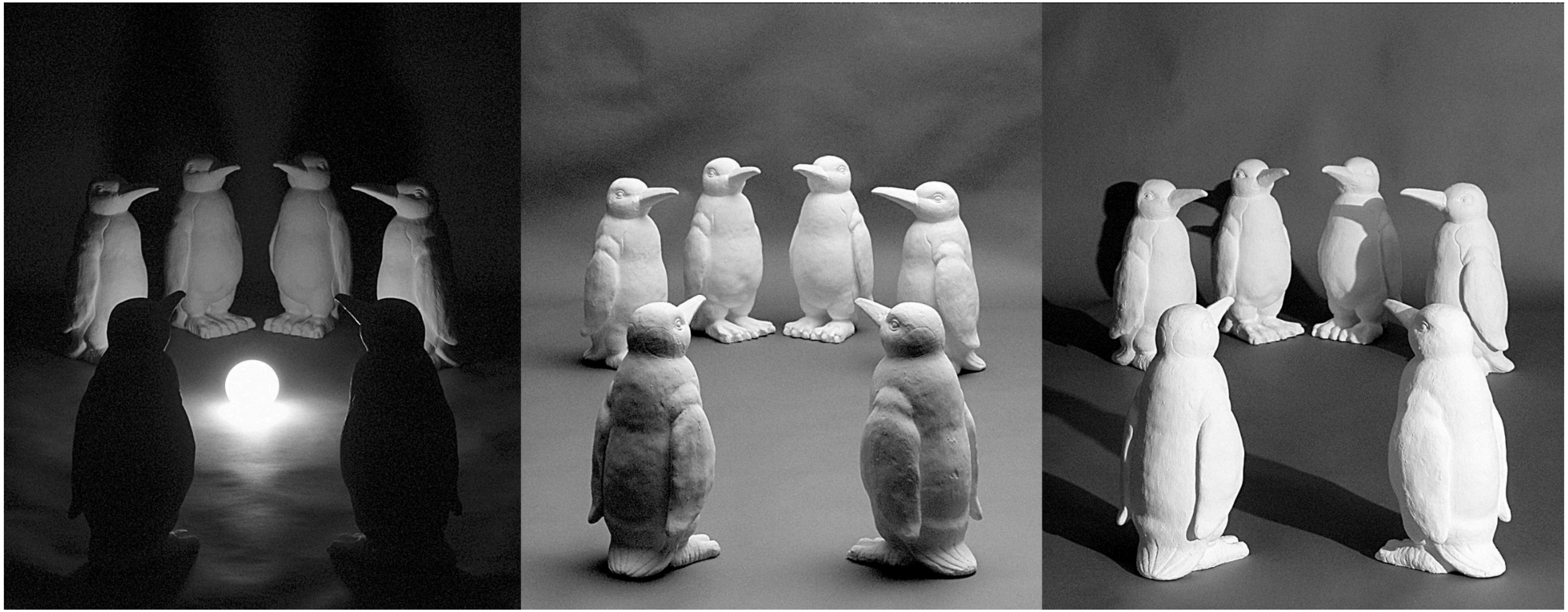


*image credit J. Koenderink

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

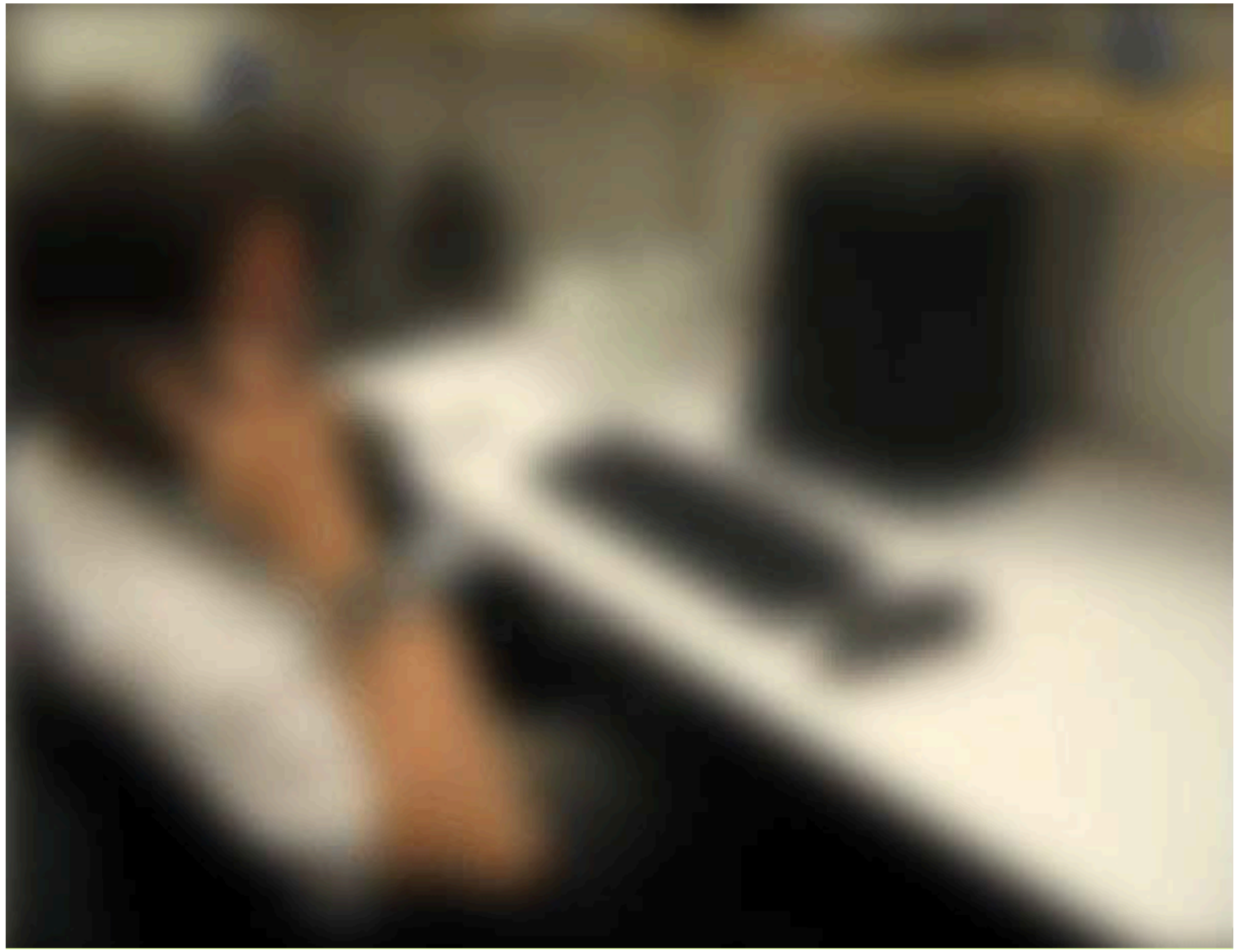


*image credit Fergus & Torralba

Challenges: Local ambiguity and context

*image credit Fergus & Torralba

Challenges: Motion

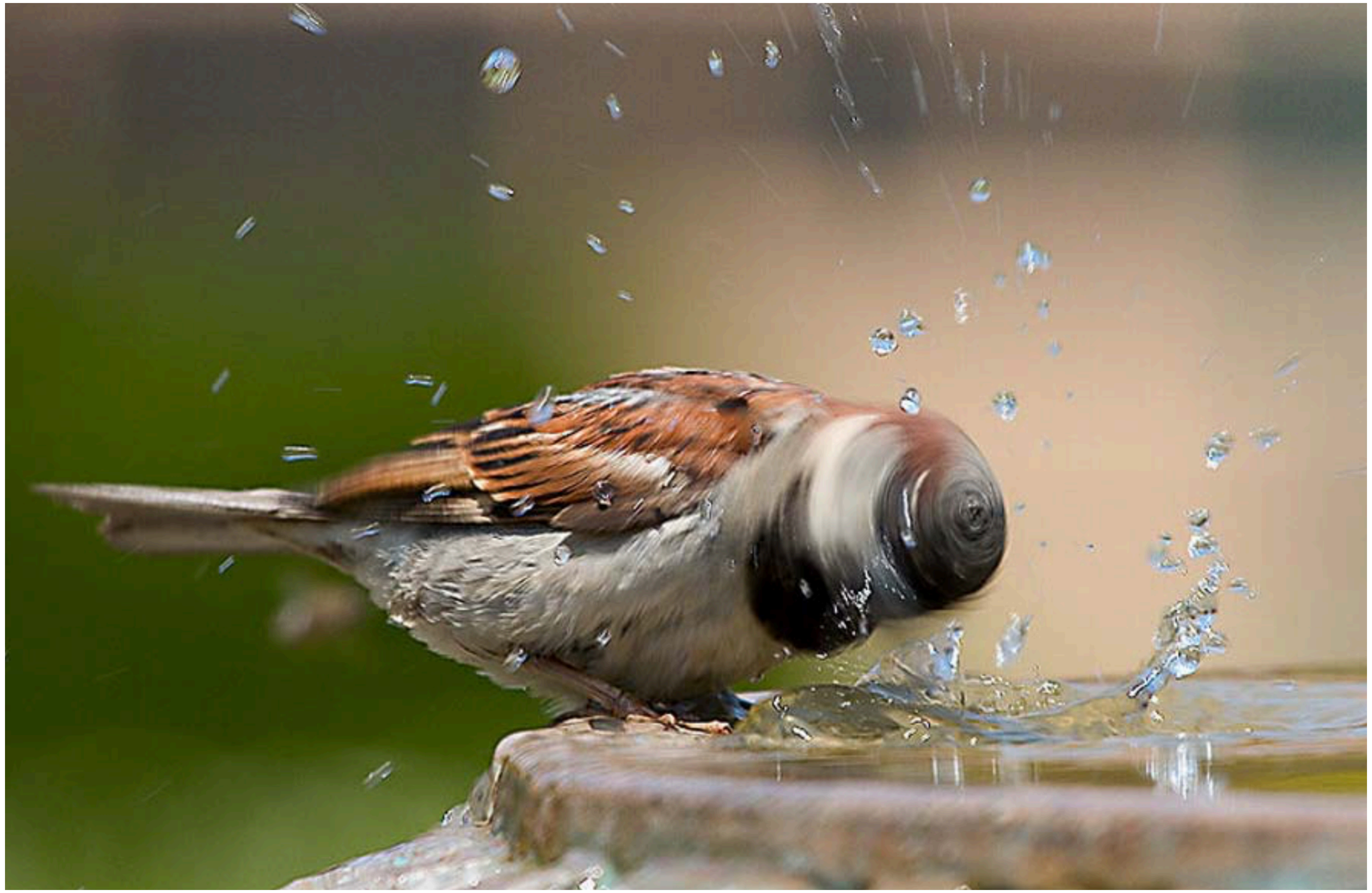


*image credit Peter Meer

Challenges: Object inter-class variation






*slide credit Fei-Fei, Fergus & Torralba

Computer Vision **Applications**

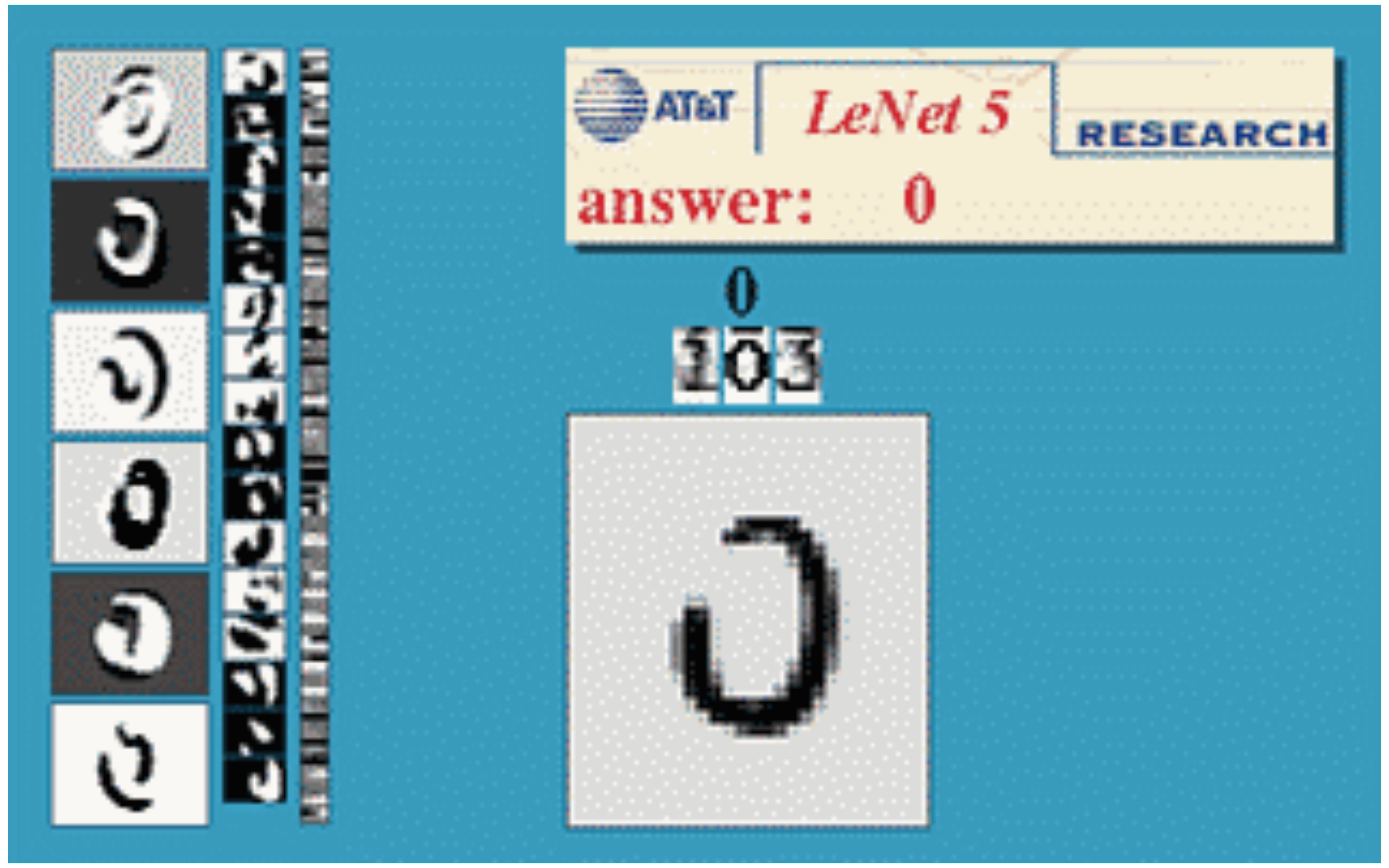
- Let’s see some examples of state-of-the-art and where it is used

Optical Character Recognition (**OCR**)

Technology to convert **scanned documents to text**

(comes with any scanner now days)

Yann LeCun



Digit recognition, AT&T labs
http://www.research.att.com/~yann/

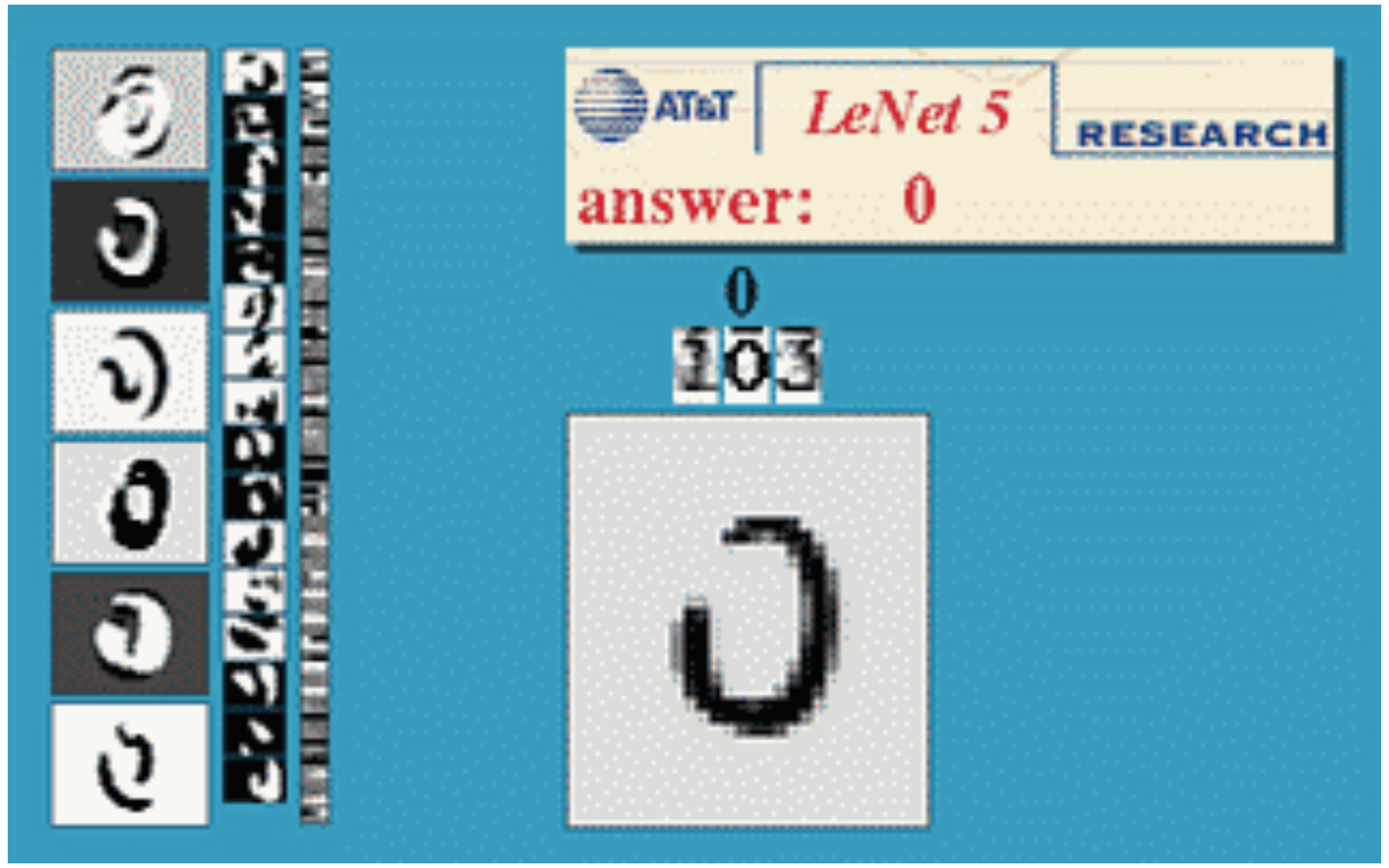
License plate readers
http://en.wikipedia.org/wiki/Automatic_number_plate_recognition

Optical Character Recognition (**OCR**)

Technology to convert **scanned documents to text**

(comes with any scanner now days)

Yann LeCun



Digit recognition, AT&T labs
http://www.research.att.com/~yann/

License plate readers
http://en.wikipedia.org/wiki/Automatic_number_plate_recognition

Face **Detection**

Technology available in any digital camera now

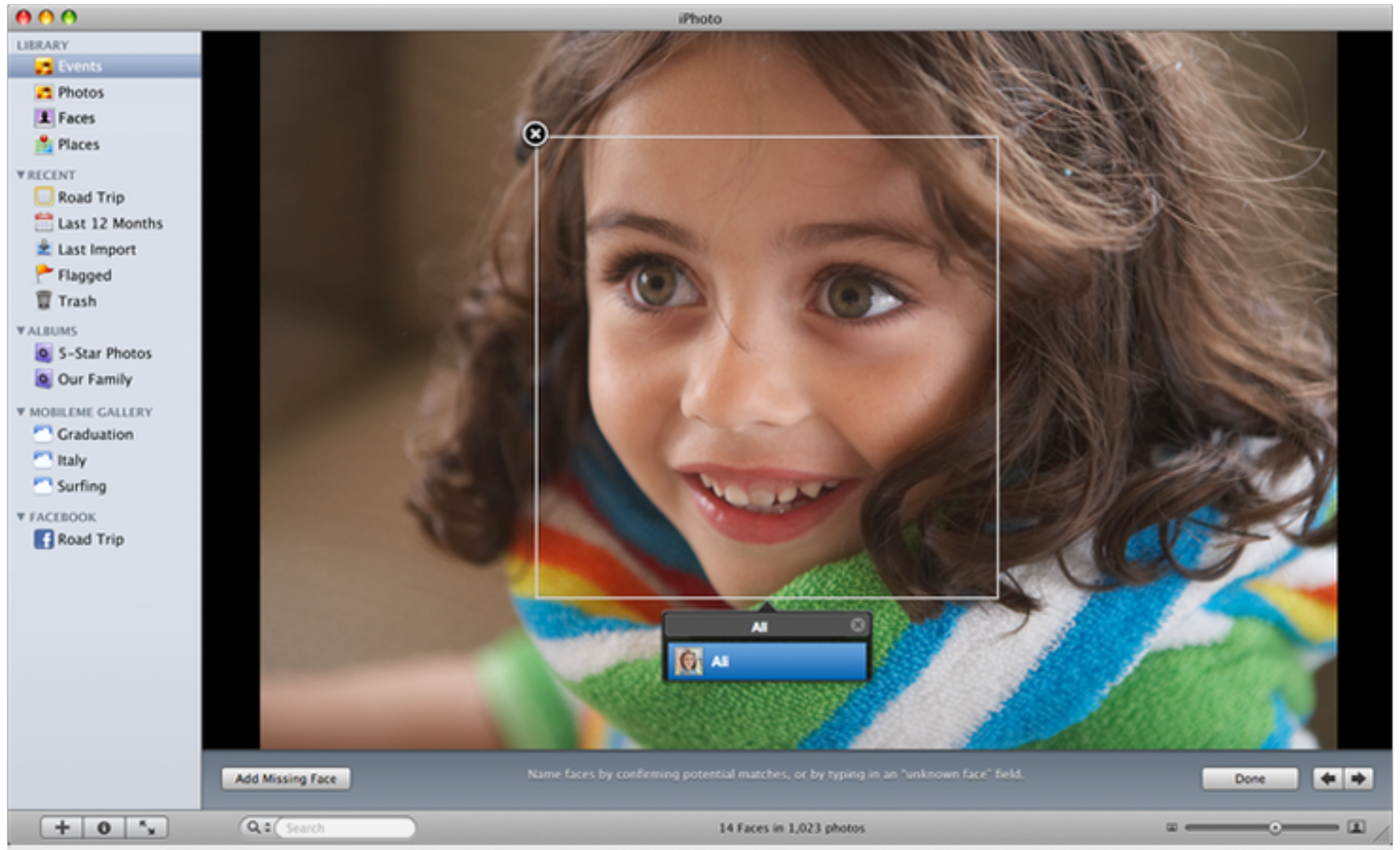
(one of the first big commercial successes of vision algorithms)

[Motorola]

Face **Recognition**

Facebook

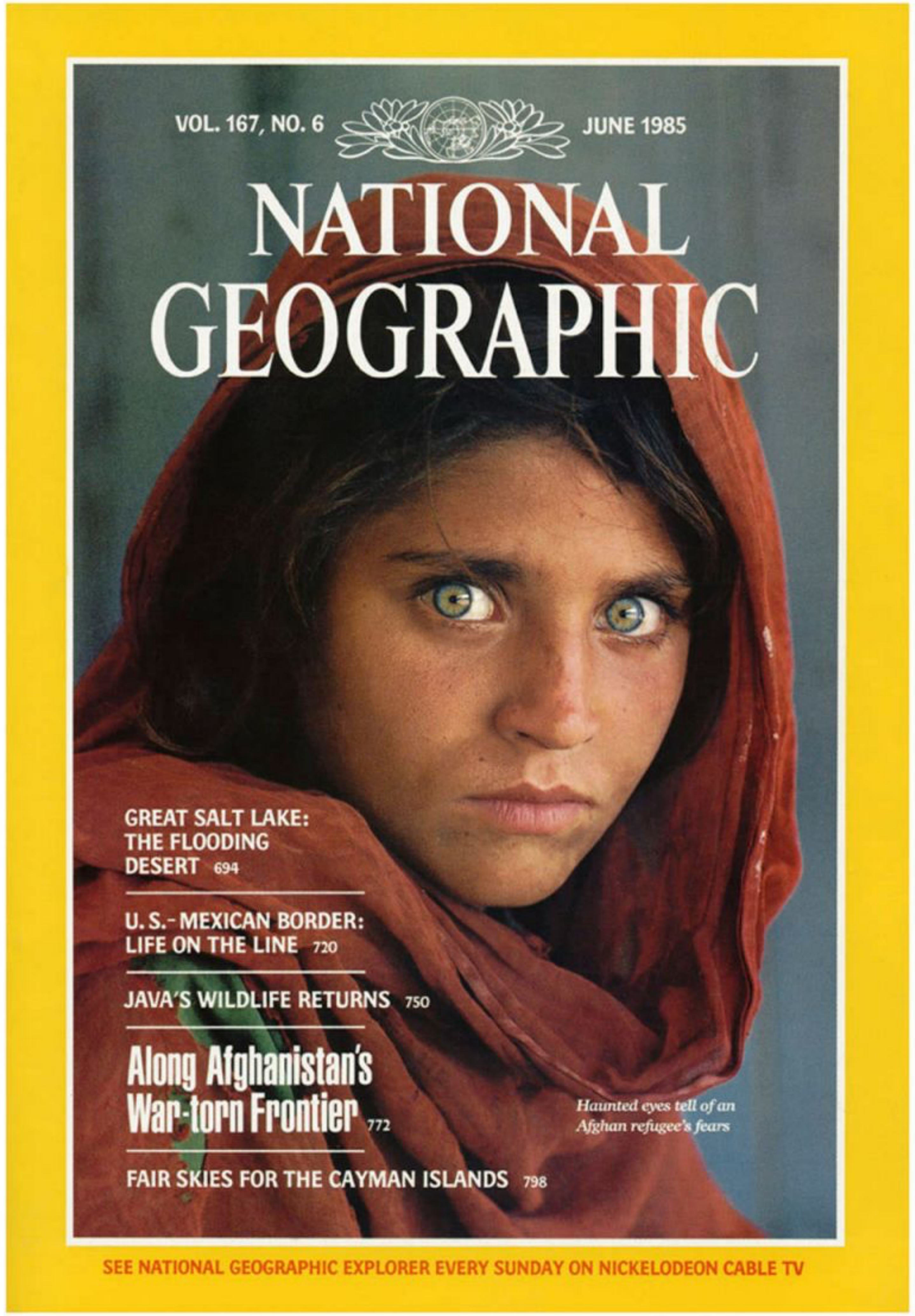
Apple's iPhoto

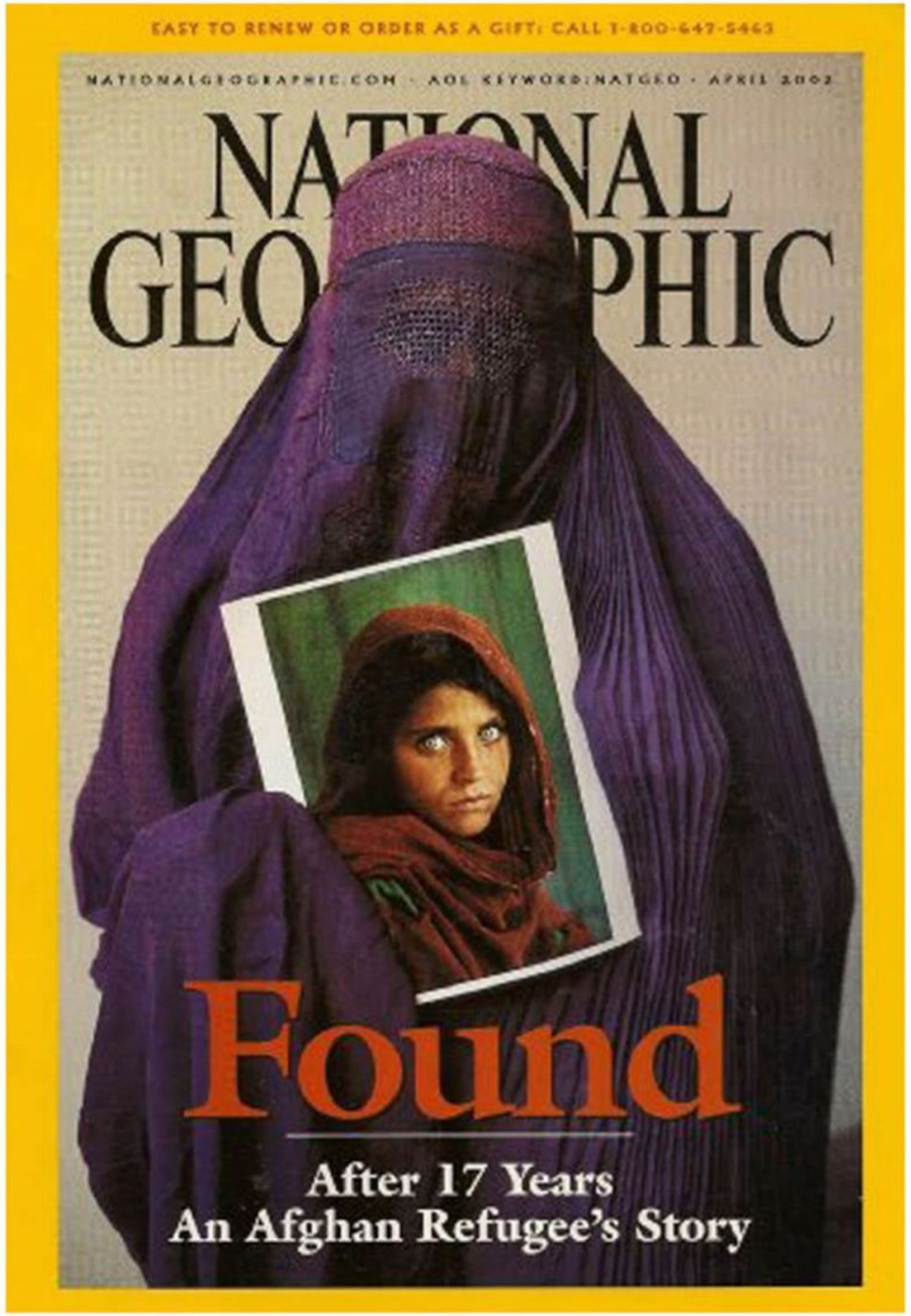


http://www.apple.com/ilife/iphoto/

Slide Credit: Devi Parikh (GA Tech) and Fei-Fei Li (Stanford)

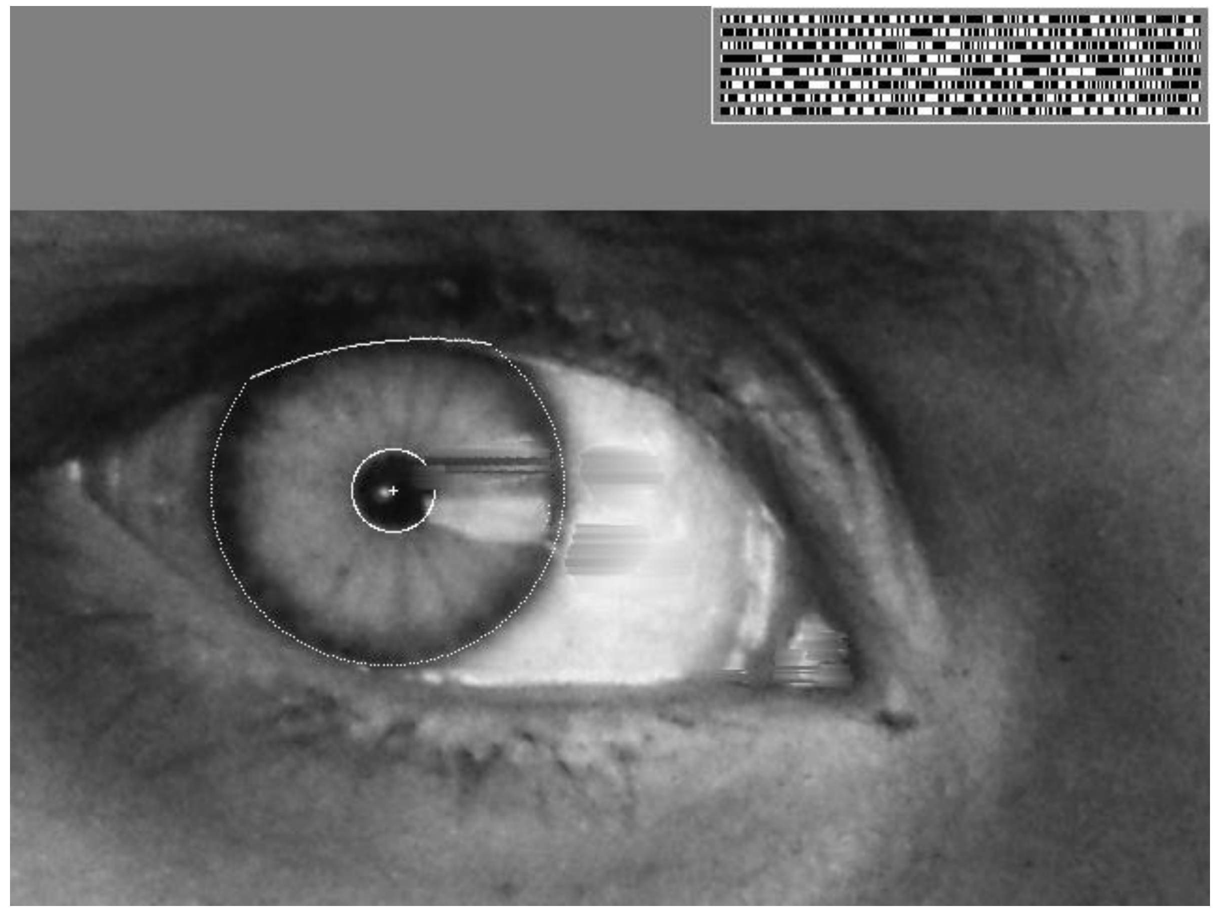
Vision for **Biometrics**

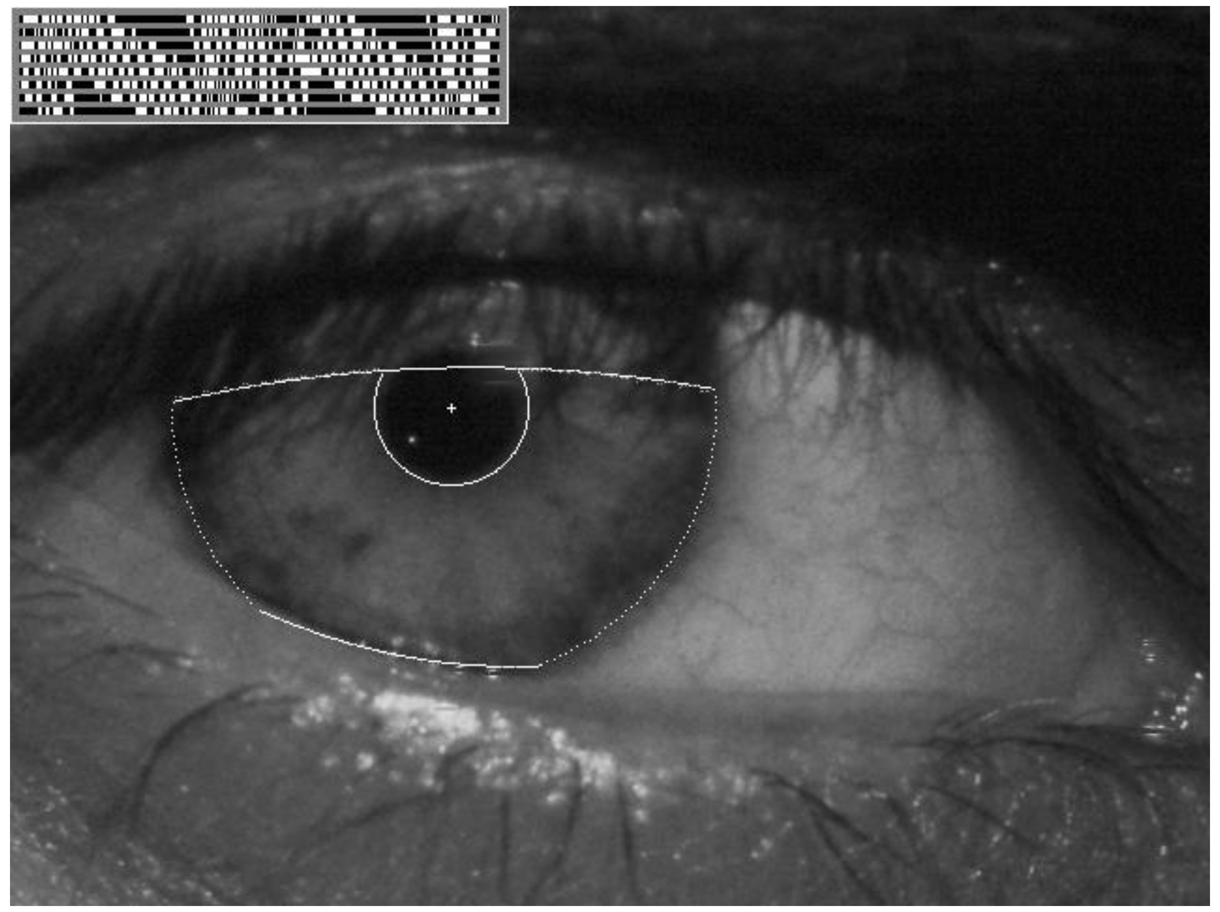




Vision for **Biometrics**

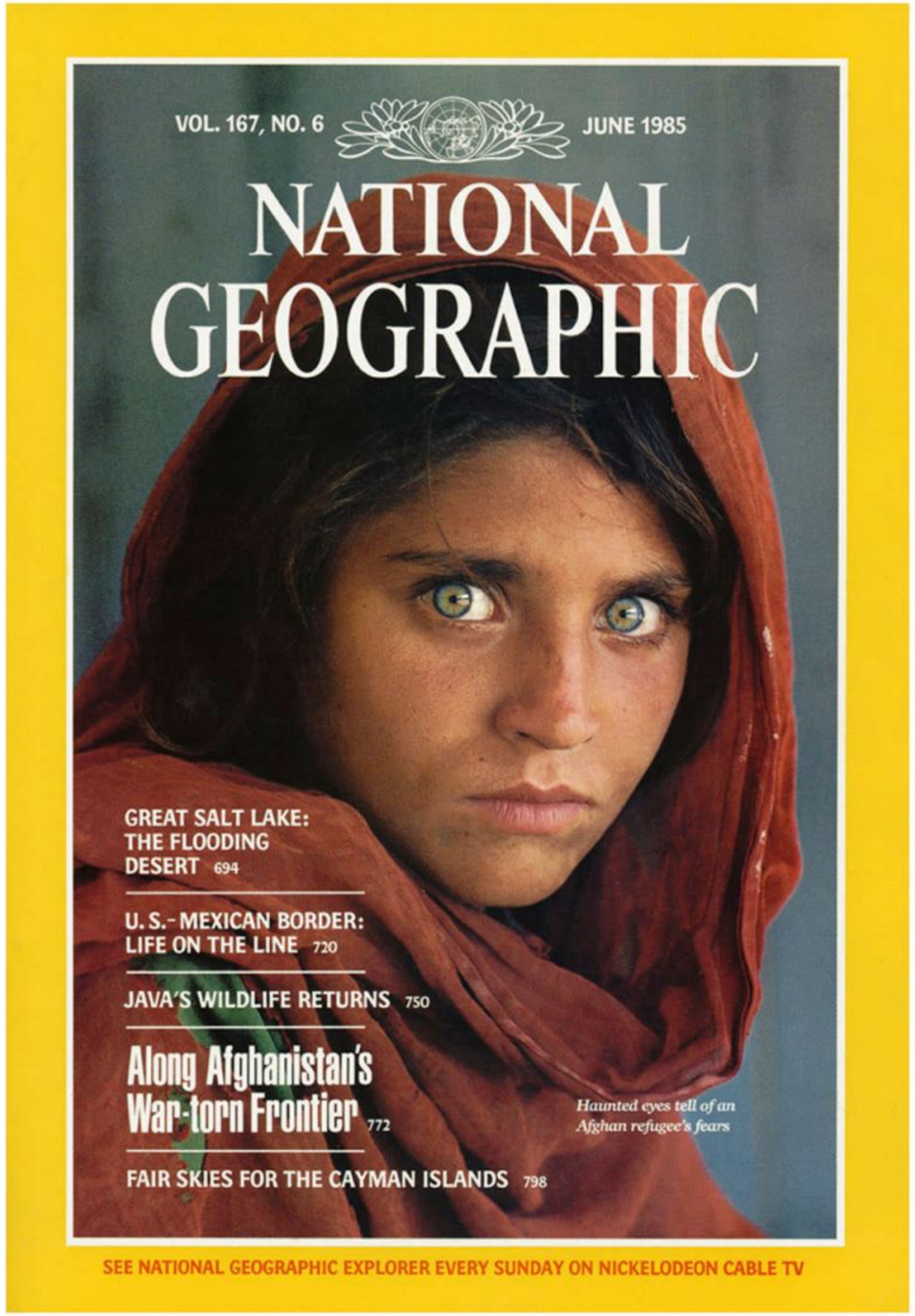
"How the Afghan Girl was Identified by Her Iris Patterns" Read the story wikipedia

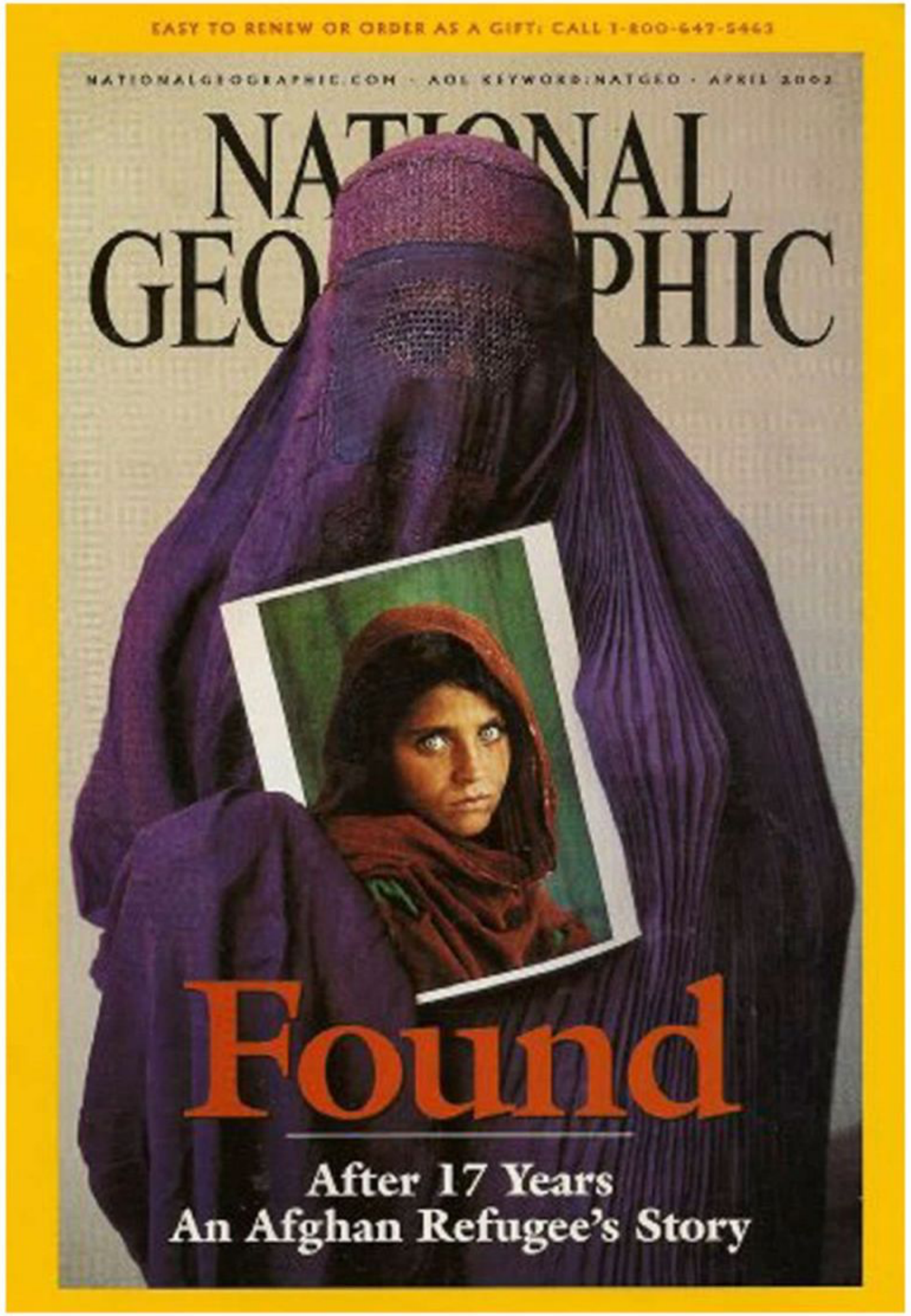




Slide Credit: James Hays (GA Tech)

Vision for **Biometrics**





Vision for **Biometrics**

Fingerprint scanners on many new laptops, other devices

iPhone X Face ID

Face recognition systems are not part of widely used technologies

How it works and how to fool it:

https://www.youtube.com/watch?v=FhbMLmsCax0

Image Credit: James Hays (GA Tech)

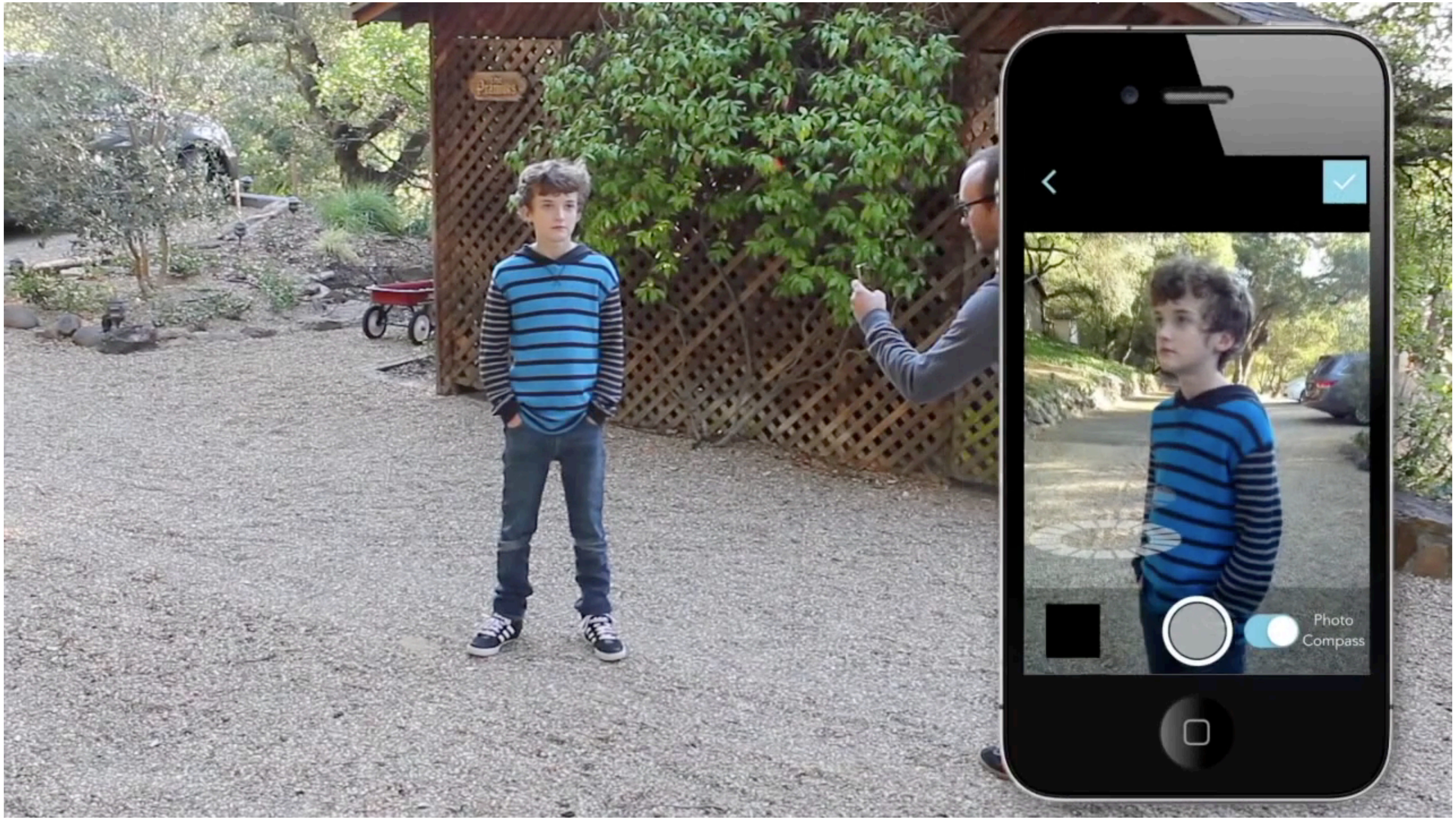
Camera **Tracking**

[Boujou — Vicon / OMG]

Camera **Tracking**

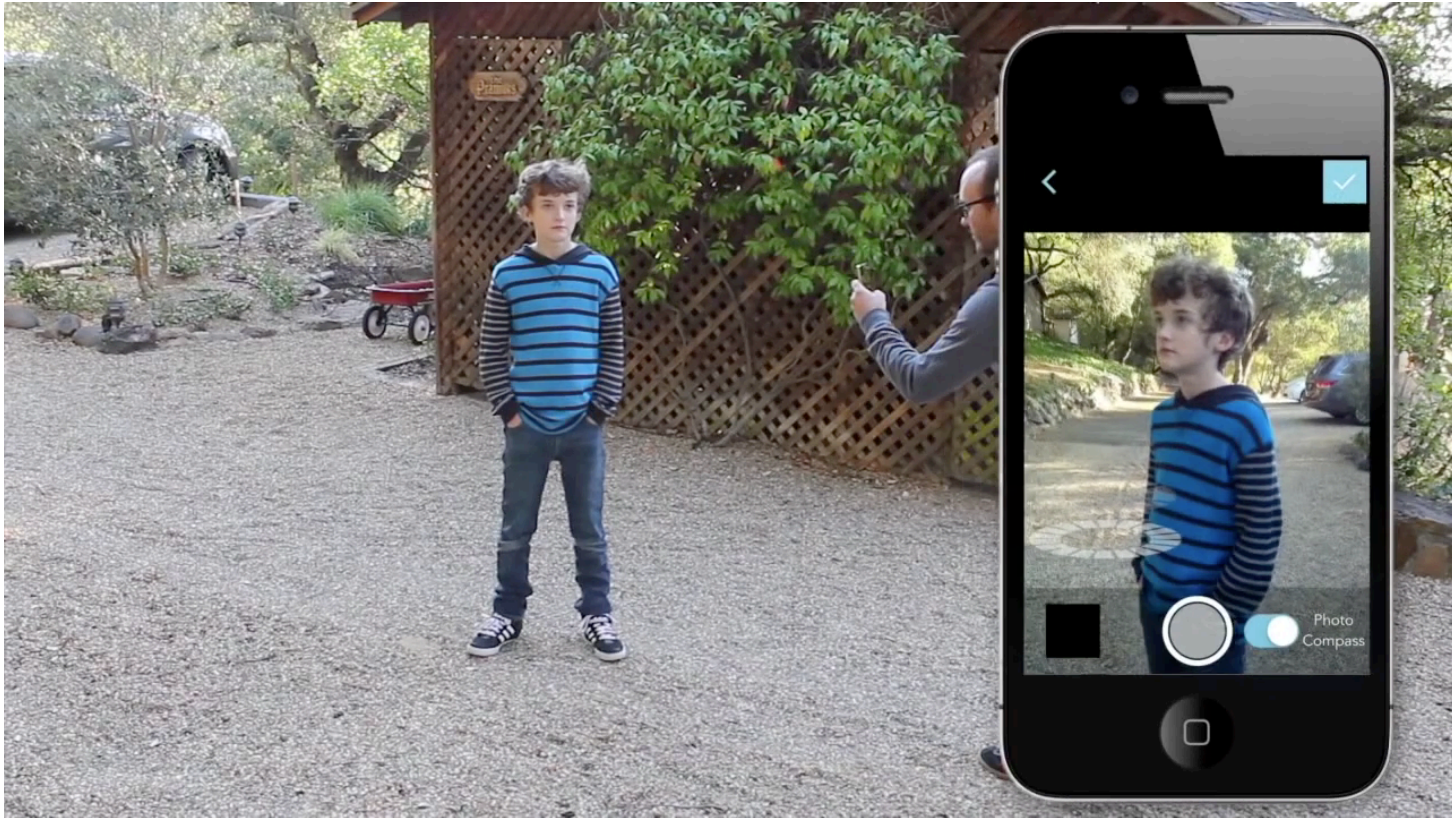
[Boujou — Vicon / OMG]

3D **Reconstruction**



[Autodesk 123D Catch]

3D **Reconstruction**

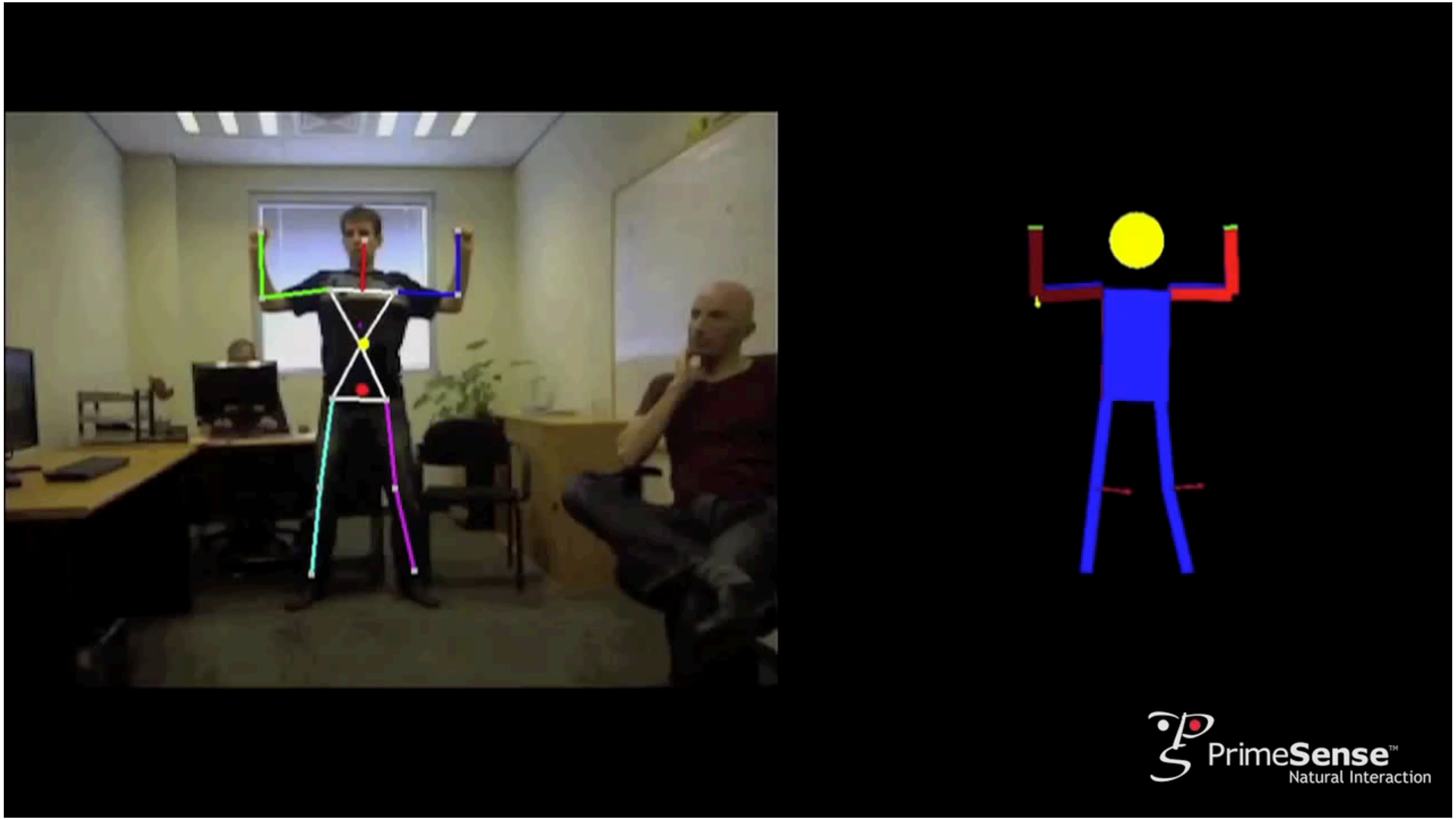


[Autodesk 123D Catch]

Body Pose Tracking

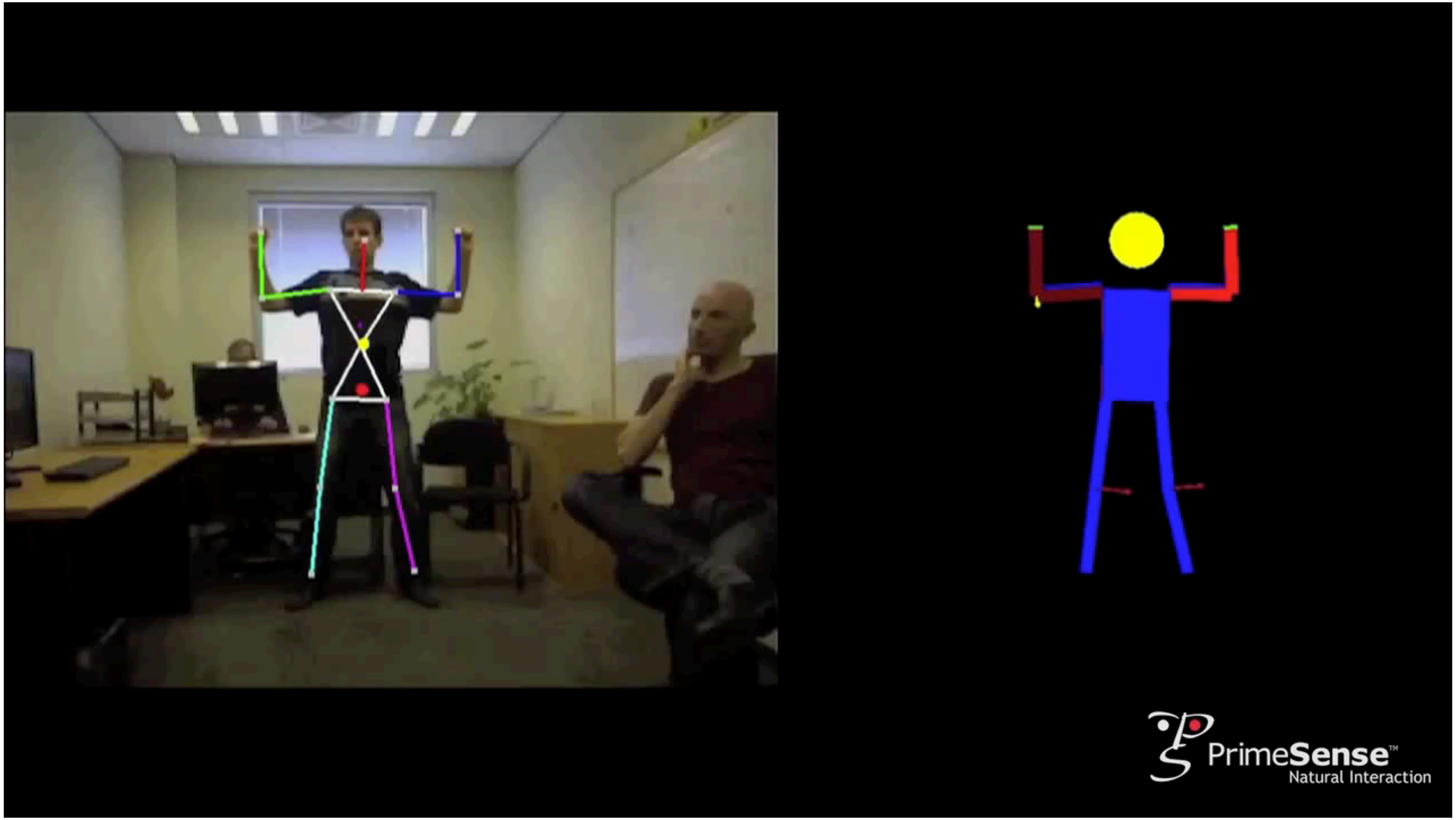
[Microsoft Xbox Kinect]

Body Pose Tracking



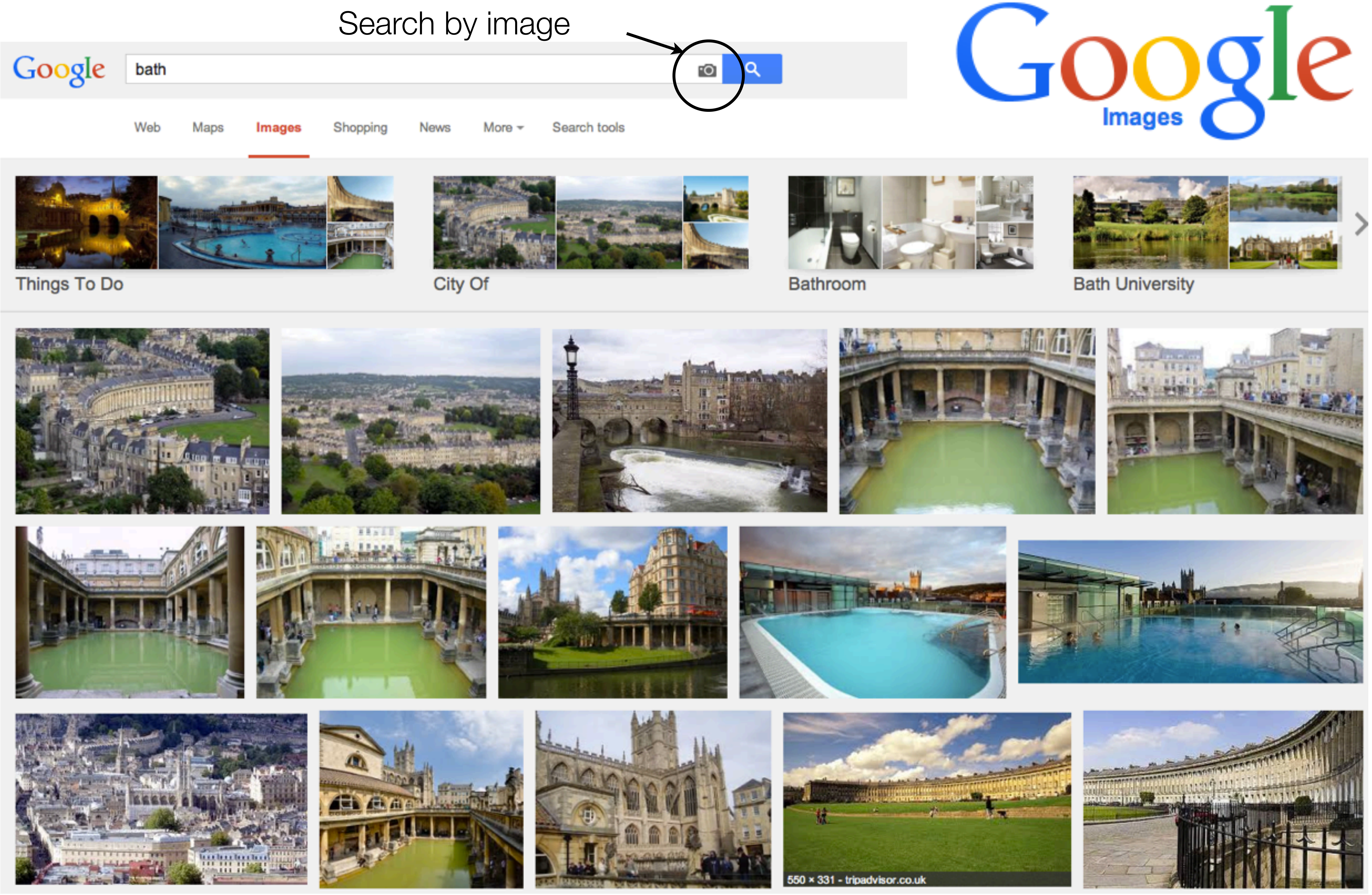
[PrimeSense]

Body Pose Tracking



[PrimeSense]

Image **Recognition** and **Search**



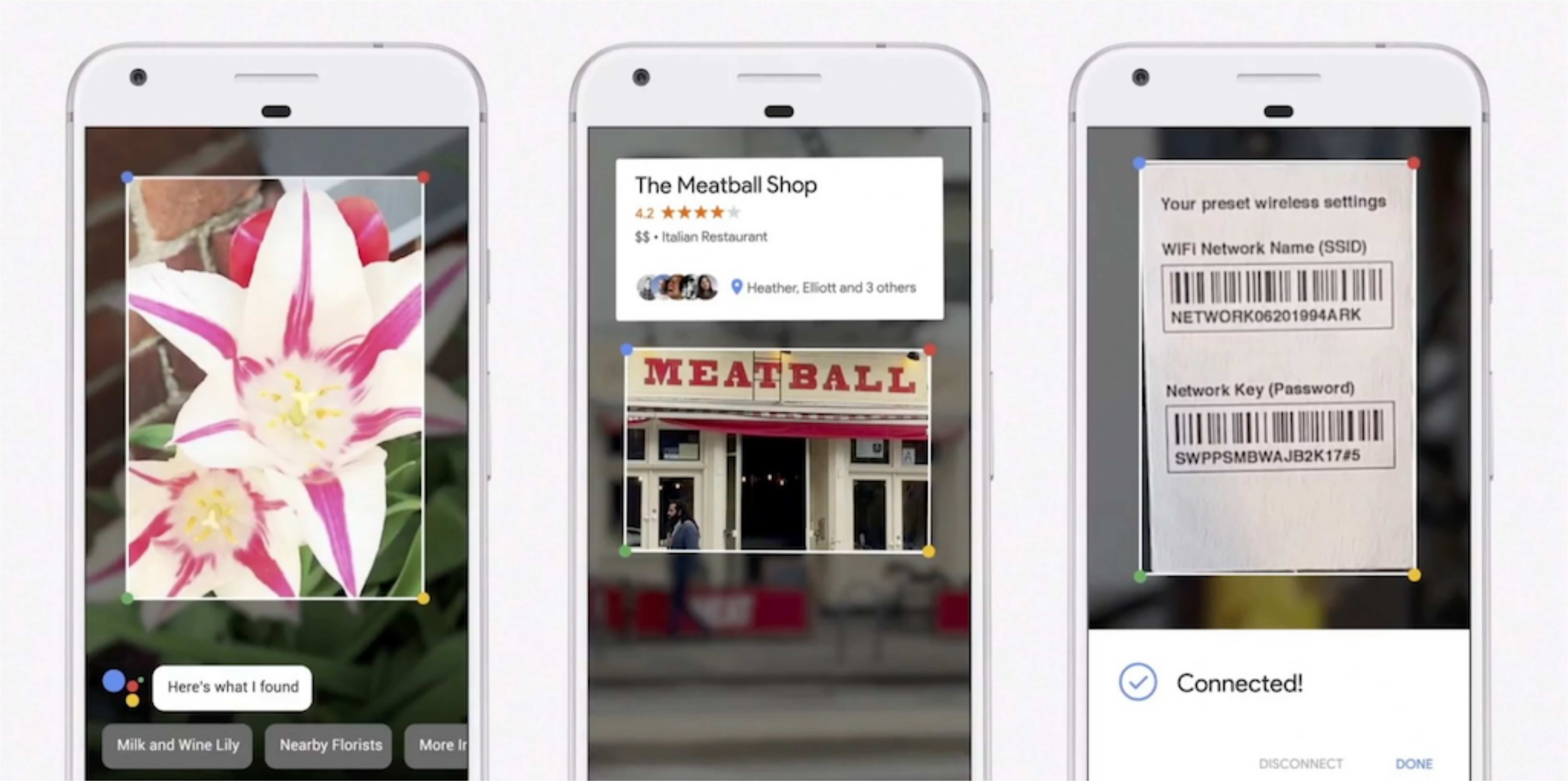
Self-Driving Cars

[Google]

AR / VR

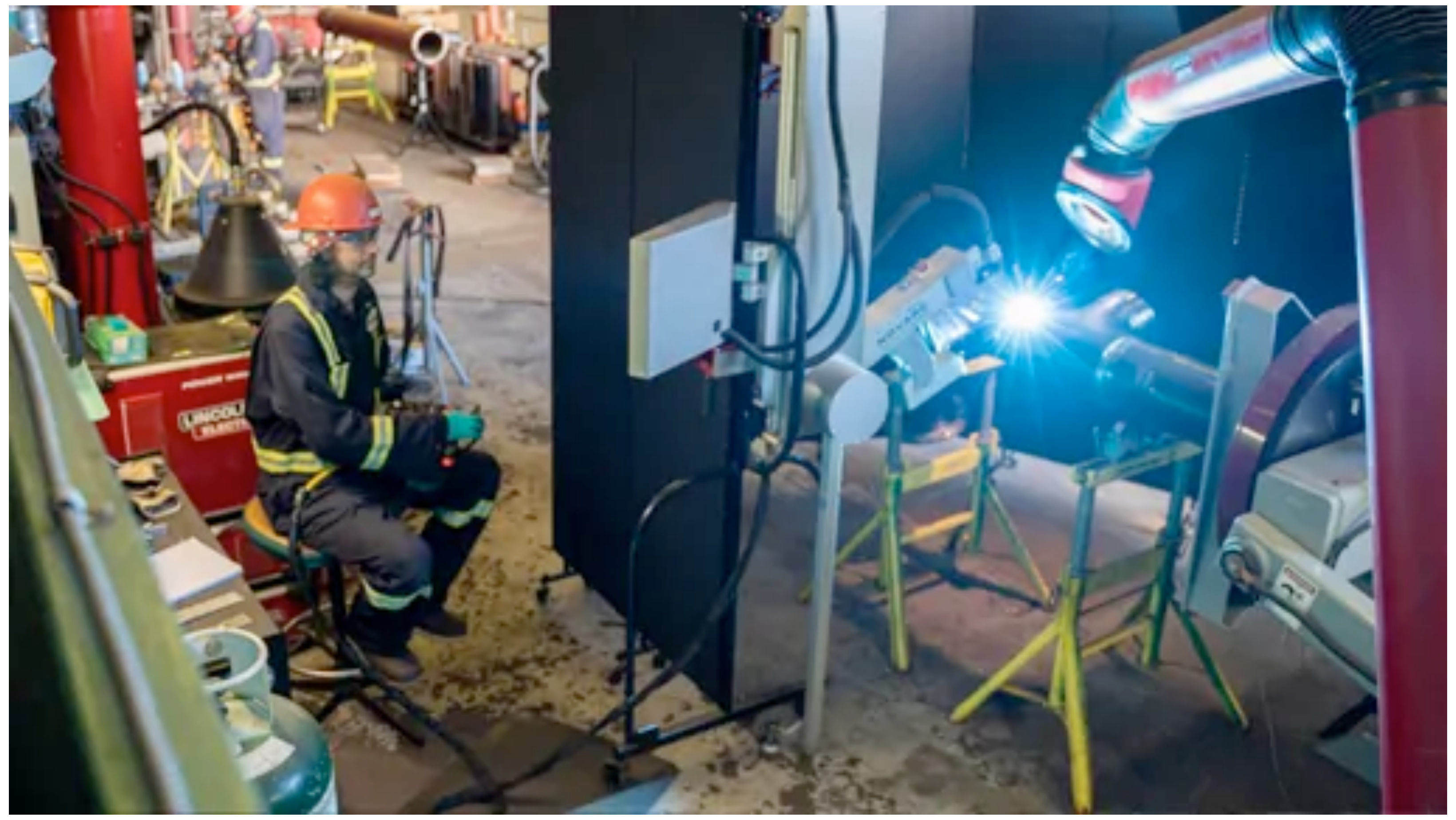
[Microsoft HoloLens]

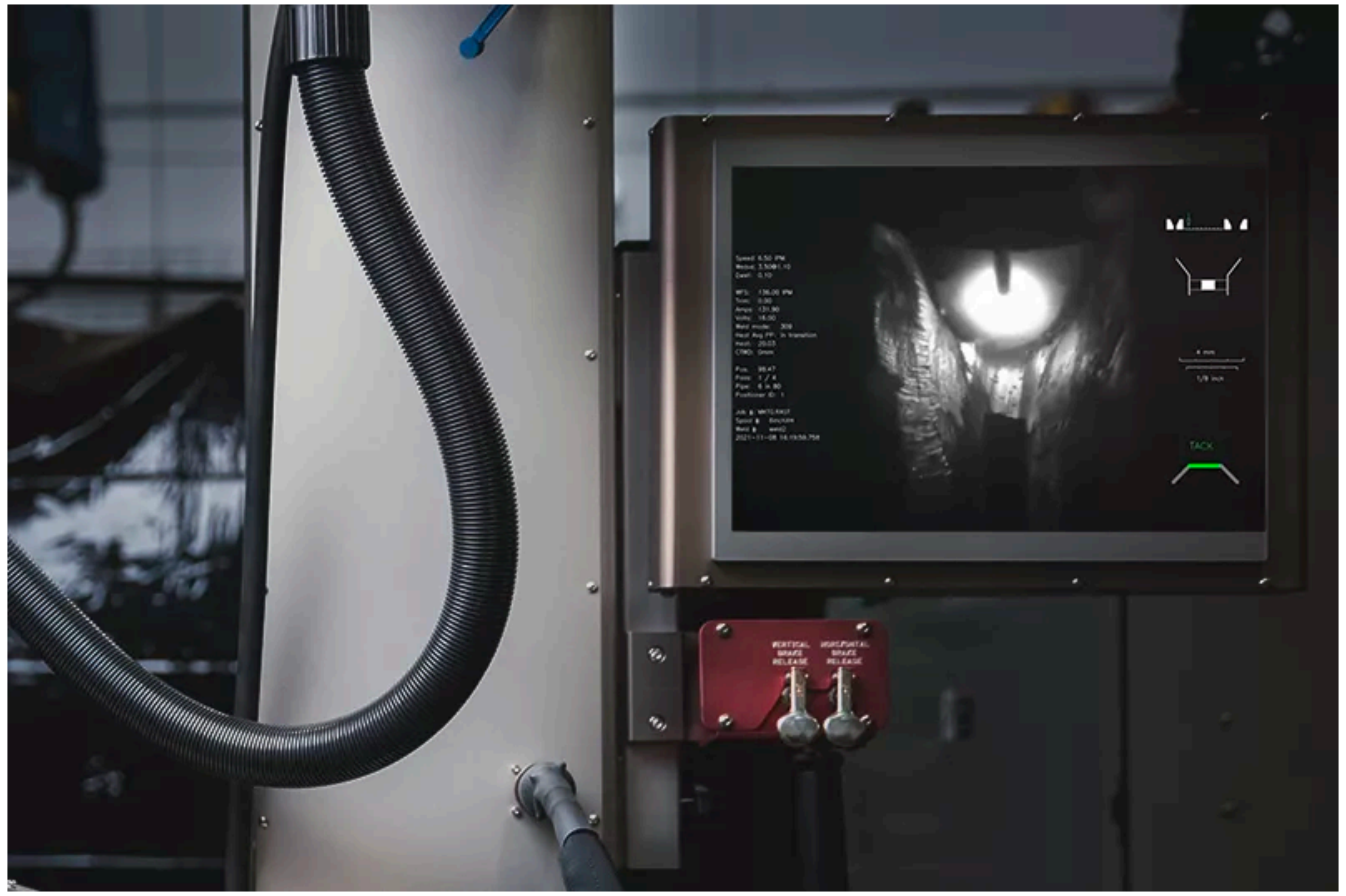
Mobile Apps

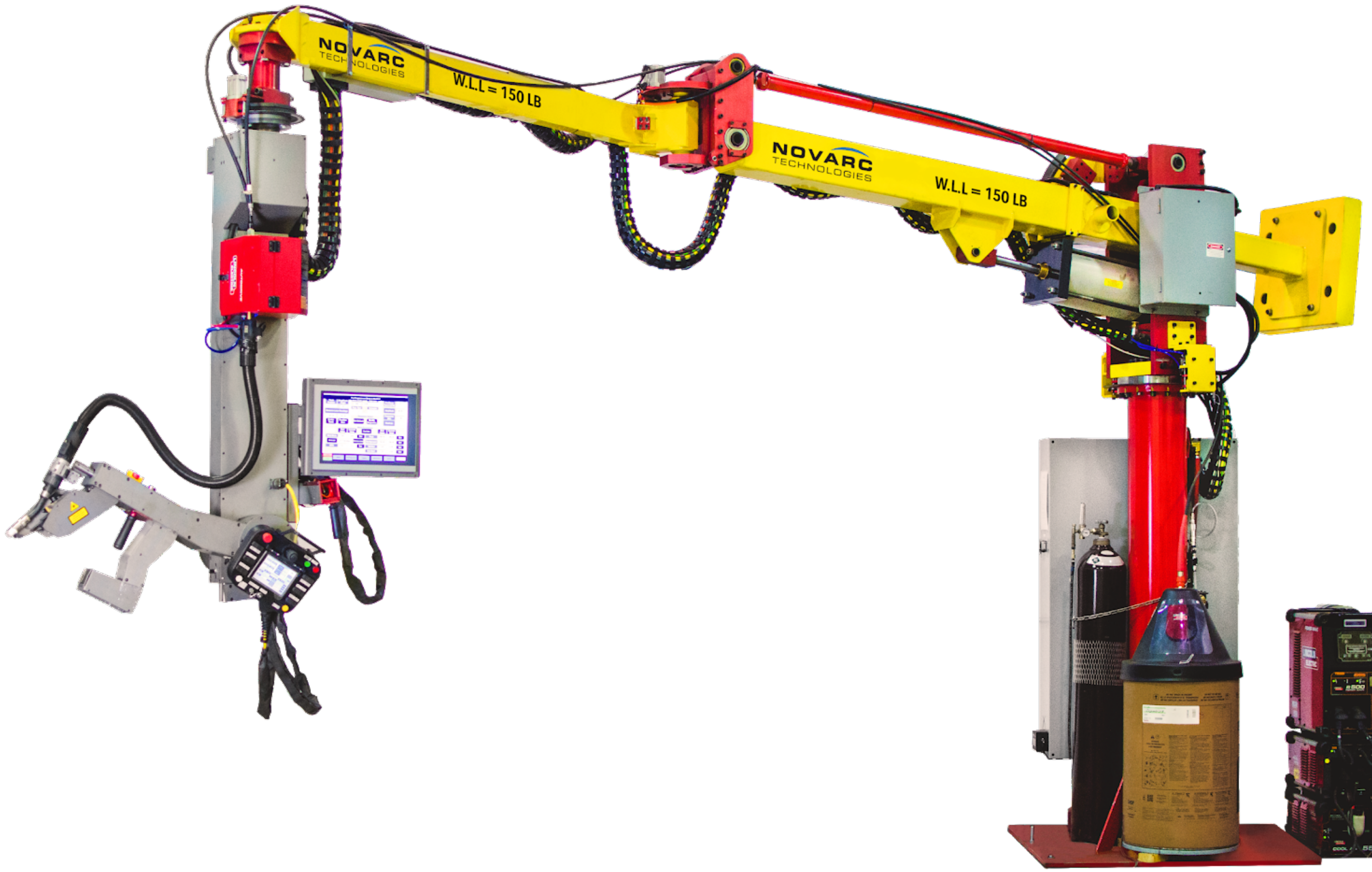


[Google Lens]

Industrial



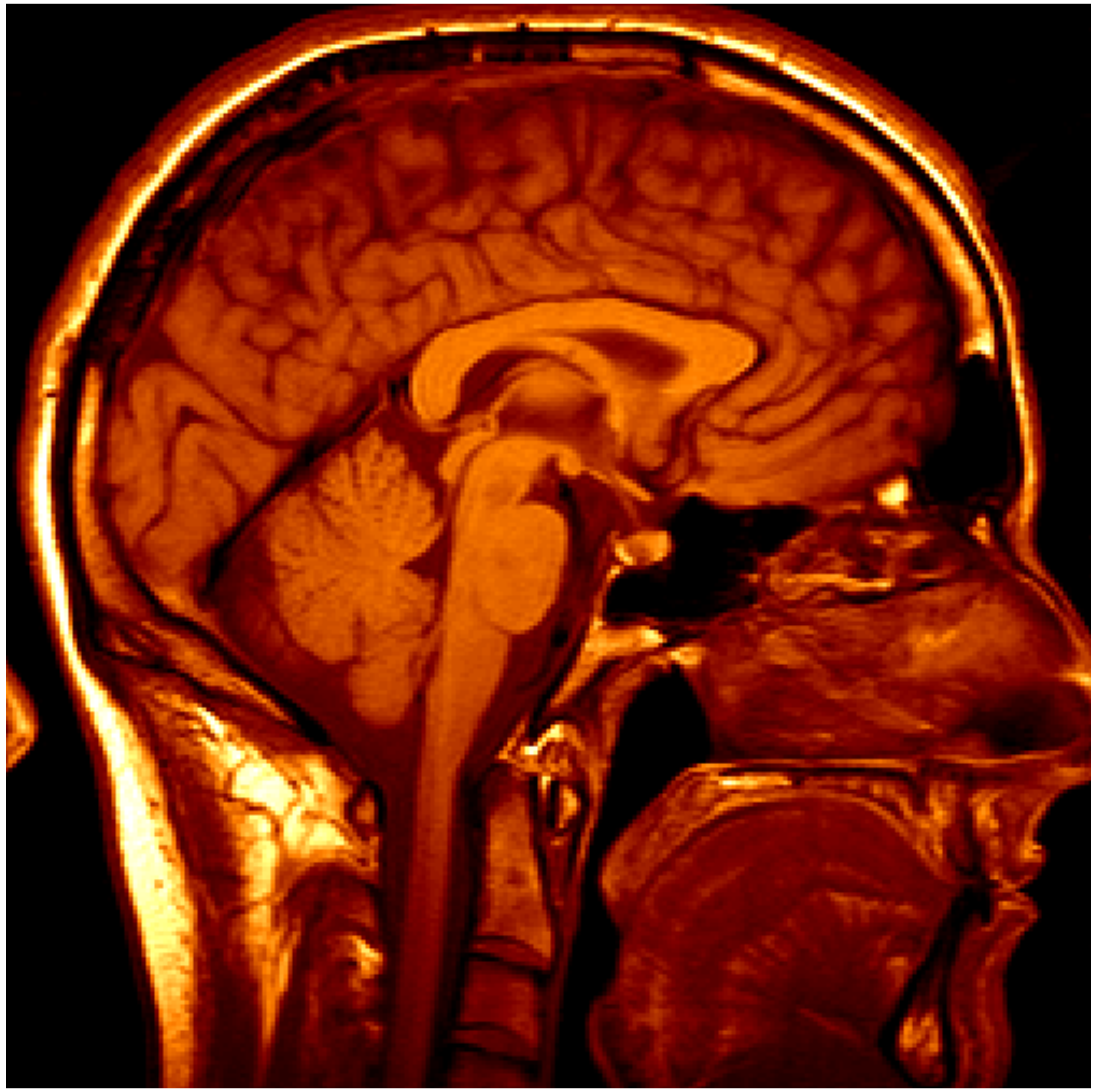




Machine Vision controlled welding robotics

NOVARC
TECHNOLOGIES

Medicine

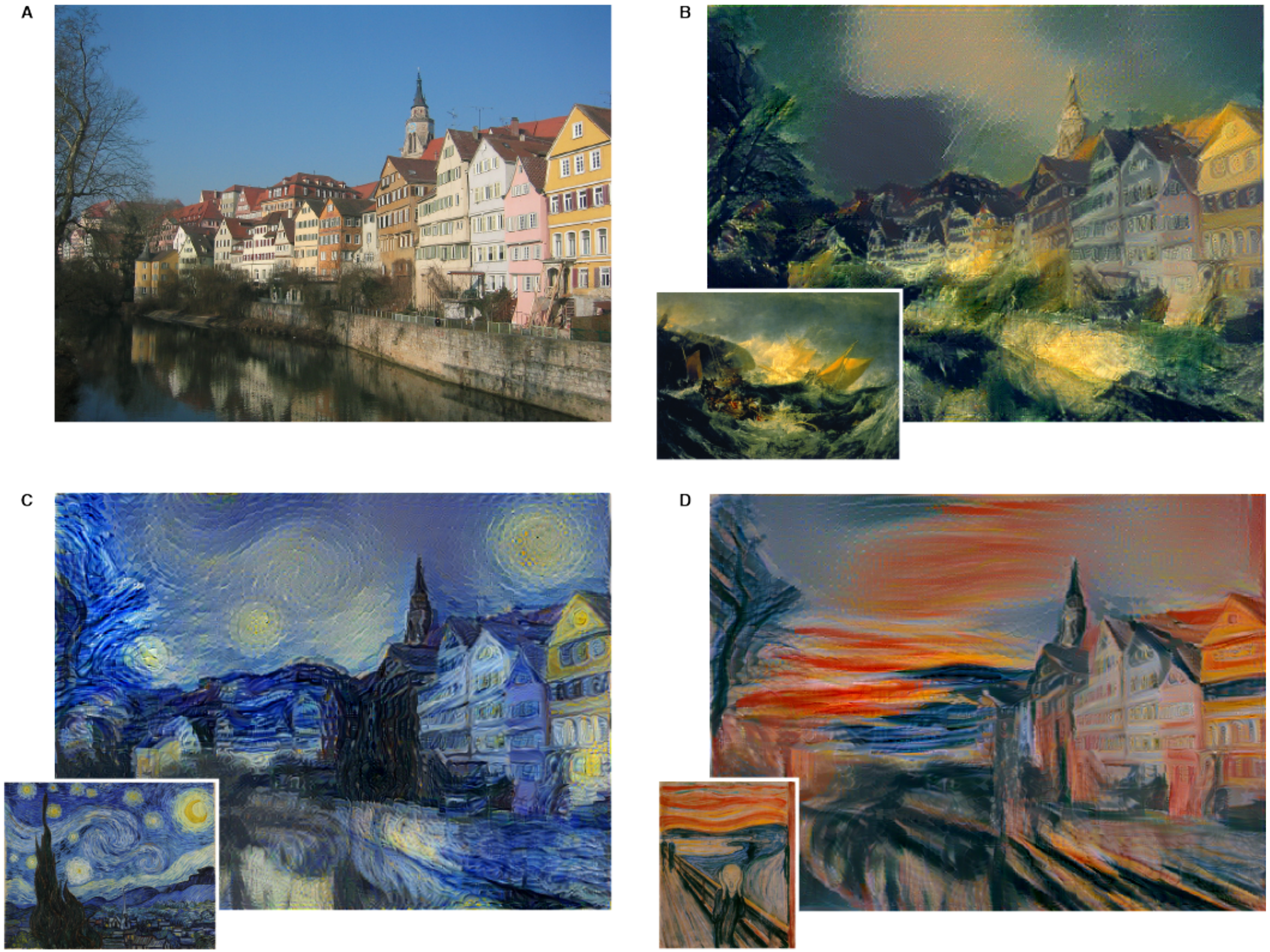


3D imaging
MRI, CT

Image guided surgery
[Grimson et al., MIT]

Slide Credit: James Hays (GA Tech)

Art



[Gatys, Ecker, Bethge 2015]

Art

TEXT DESCRIPTION

An astronaut Teddy bears A bowl of soup

riding a horse lounging in a tropical resort in space playing basketball with cats in space

in a photorealistic style in the style of Andy Warhol as a pencil drawing

→

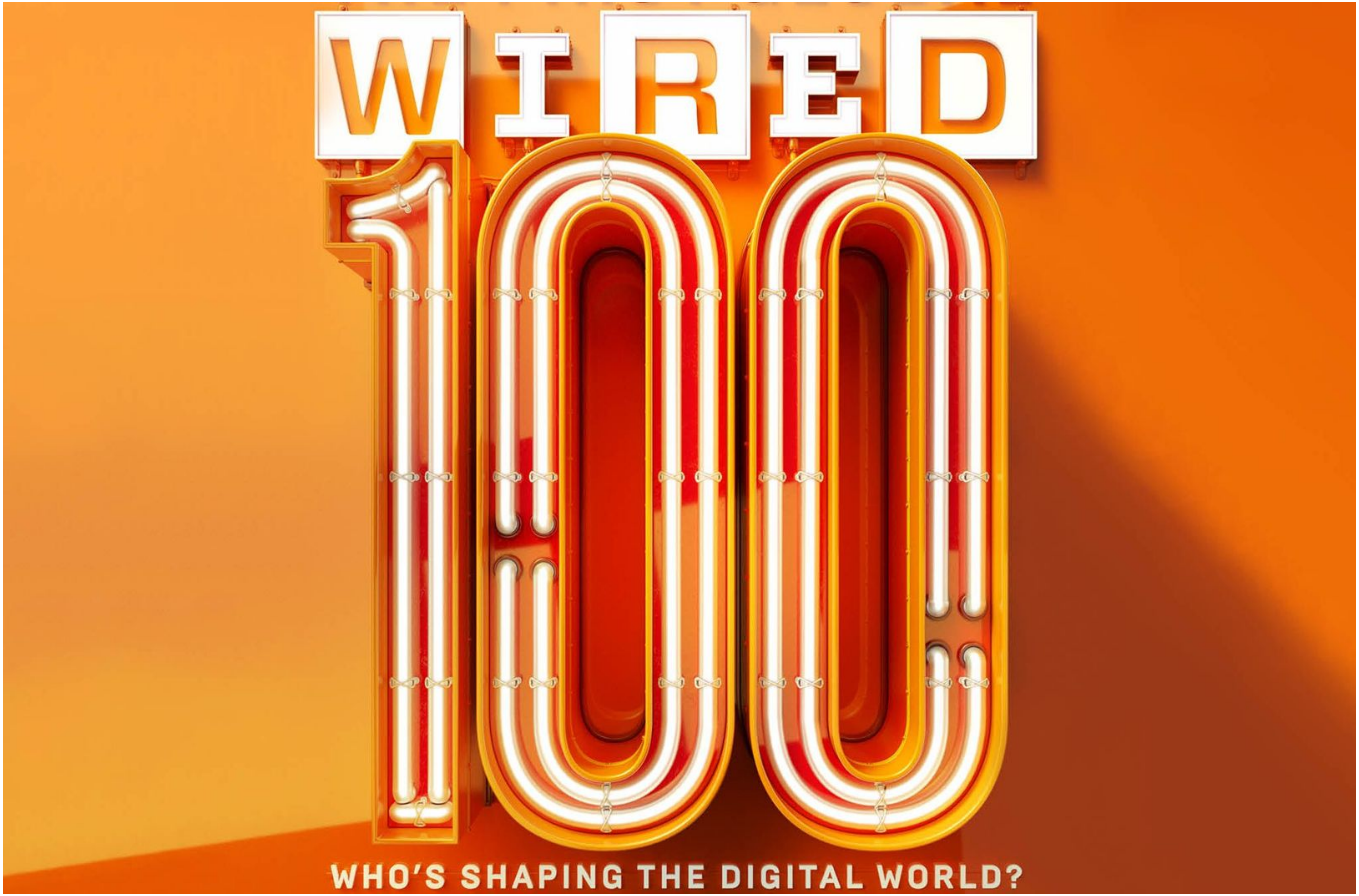
DALL·E 2

[Dall-E v2]

Why Study Computer Vision?

It is one of the **most exciting areas of research** in computer science

Among the **fastest growing technologies** in the industry today


WIRED
100
WHO'S SHAPING THE DIGITAL WORLD?

Wired's 100 **Most Influential People** in the World

63. Yann Lecun

Director of AI research, Facebook, Menlo Park

LeCun is a leading expert in deep learning and heads up what, for Facebook, could be a hugely significant source of revenue: understanding its user's intentions.

62. Richard Branson

Founder, Virgin Group, London

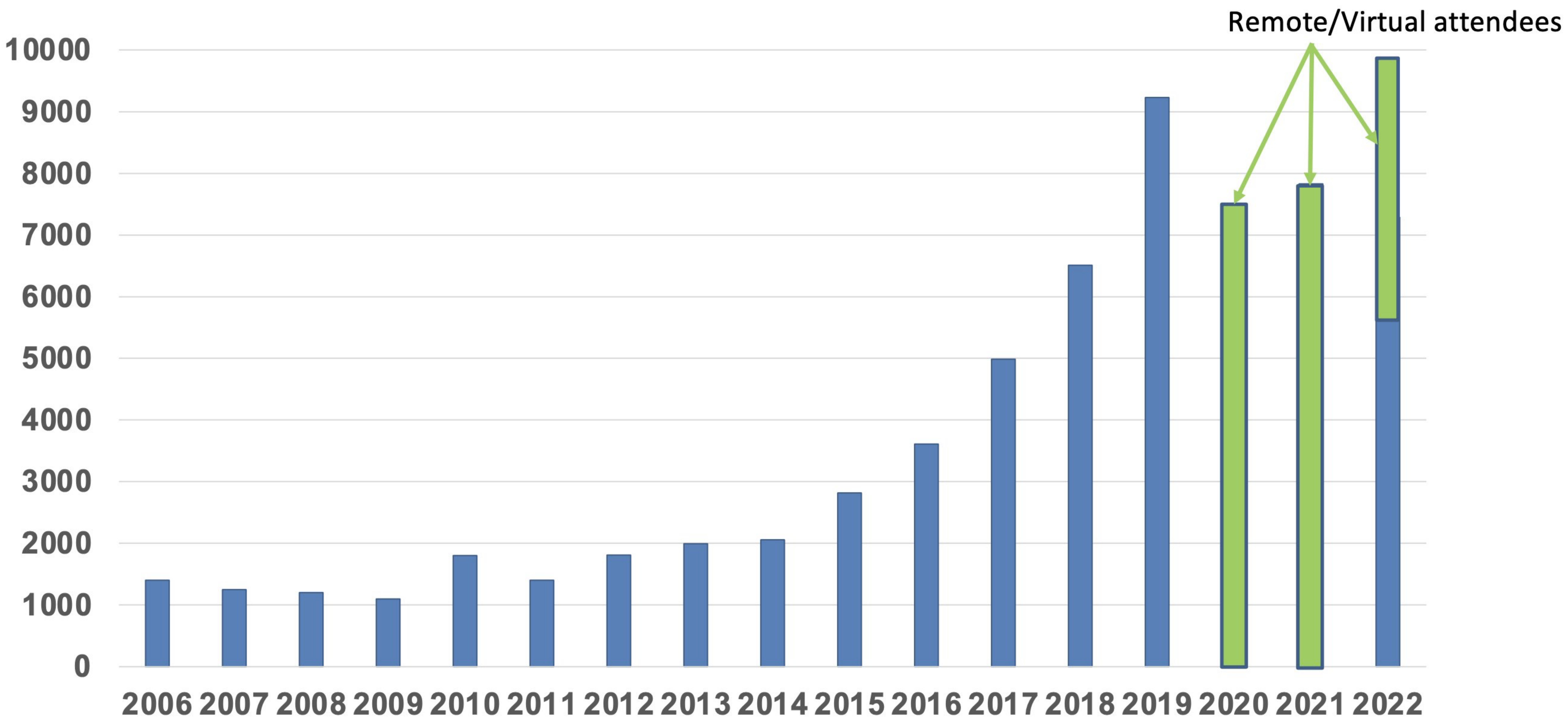
Branson saw his personal fortune grow £550 million when Alaska Air bought Virgin America for $2.6 billion in April. He is pressing on with civilian space travel with Virgin Galactic.

61. Taylor Swift

Entertainer, Los Angeles

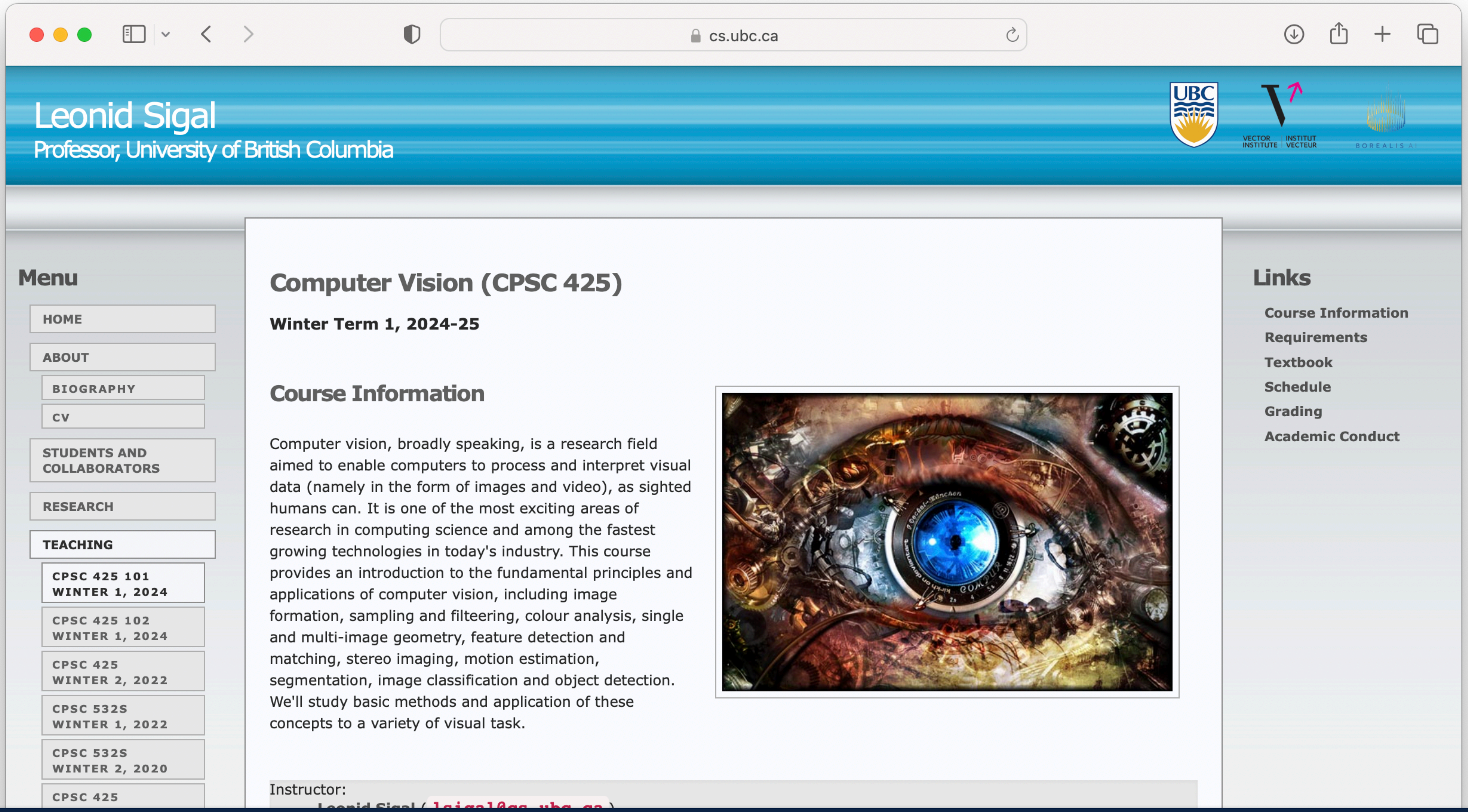

CVPR
JUNE 19-24 2022
NEW ORLEANS
LOUISIANA

CVPR **Attendance**




VANCOUVER
CVPR 2023
June 17-23, 2023

Course **Schedule**



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

Section 101: https://www.cs.ubc.ca/~lsigal/teaching24_Term1a.html

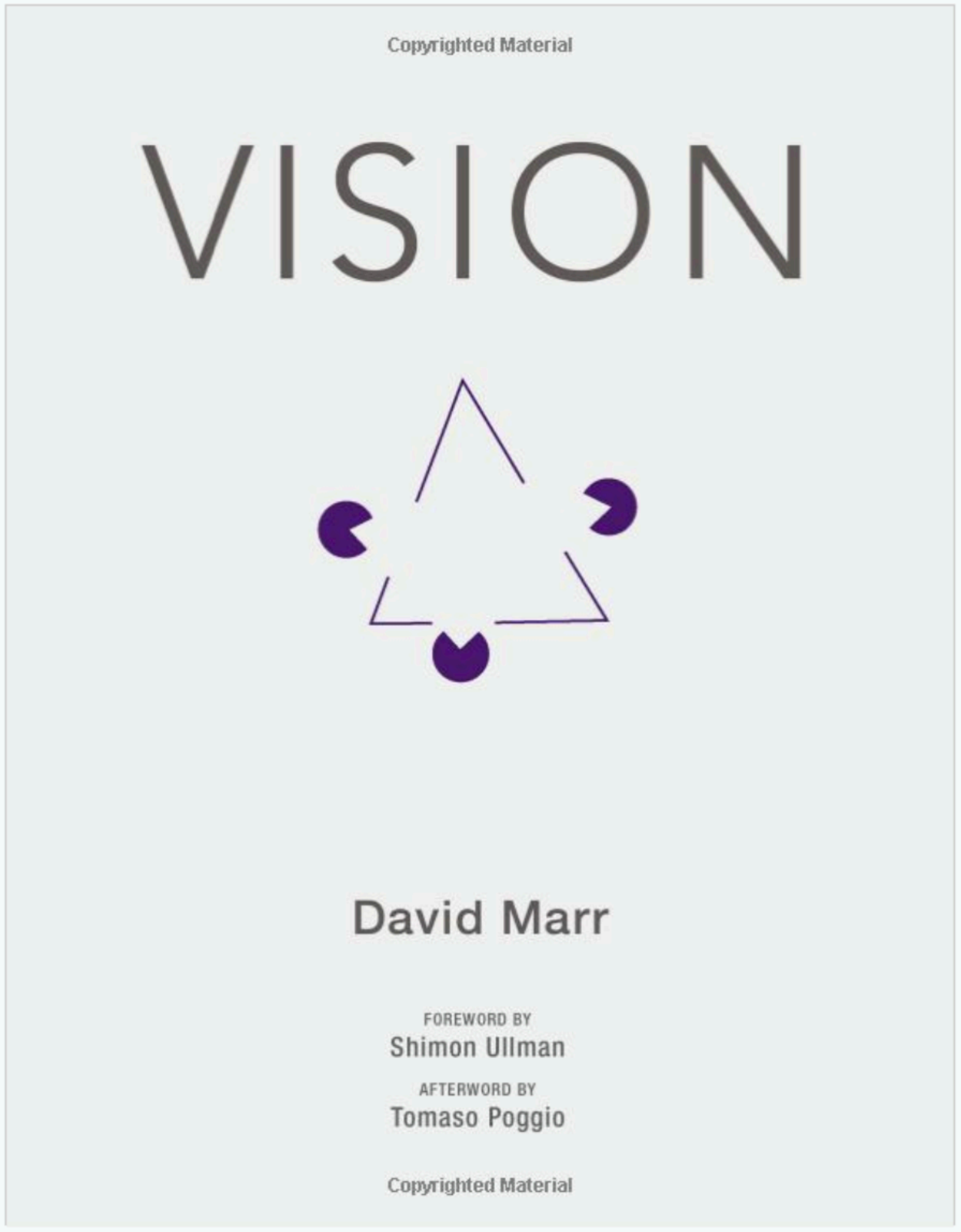
Section 102: https://www.cs.ubc.ca/~lsigal/teaching24_Term1b.html

Topics Covered

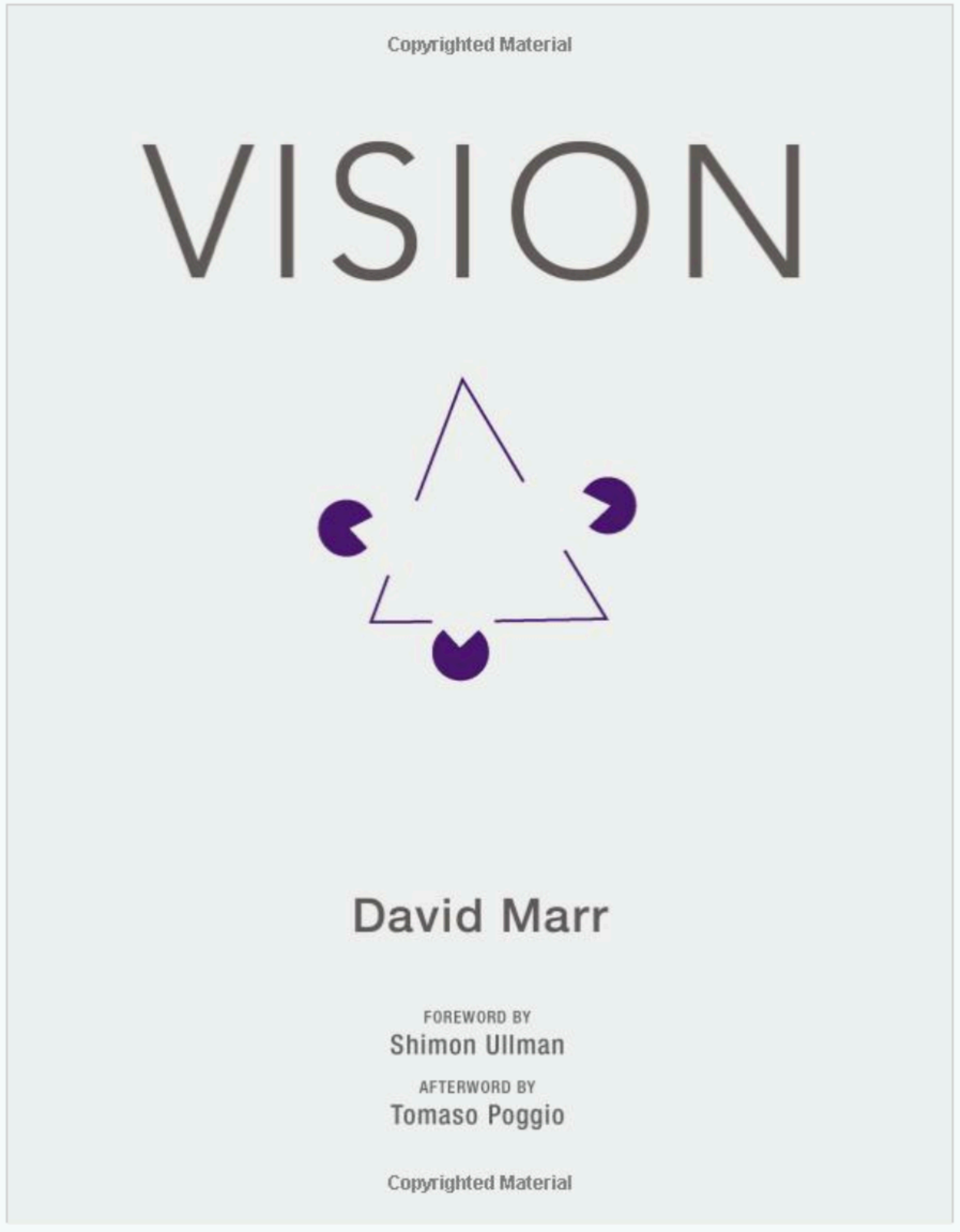
– Image Processing (Linear Filtering, Convolution)
– Filters as Templates
– Image Feature Detection (Edges & Corners)
– Texture & Colour
– Image Feature Description (SIFT)
– Model Fitting (RANSAC, The Hough Transform)
– Camera Models, Stereo Geometry
– Motion and Optical Flow
– Clustering and Image Segmentation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

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



David **Marr**, 1970s



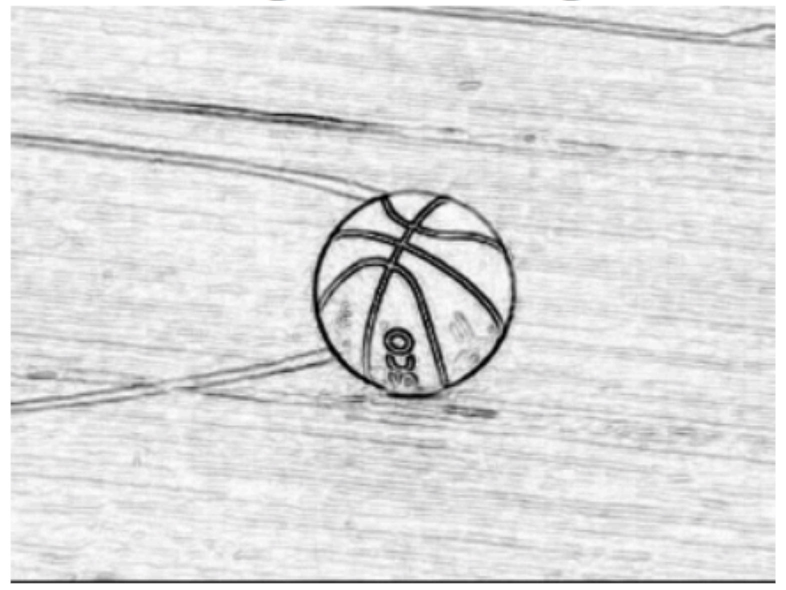
* slide from Fei-Dei Li, Justin Johnson, Serena Yeung, **cs231n Stanford**

David **Marr**, 1970s

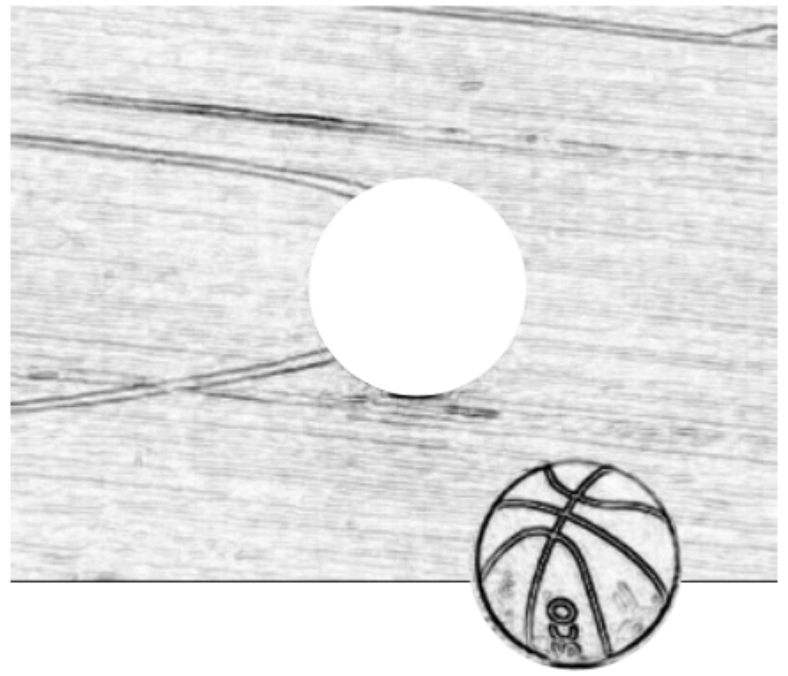
Input image

This image is CC0 1.0 public domain

Edge image



2 ½-D sketch



3-D model

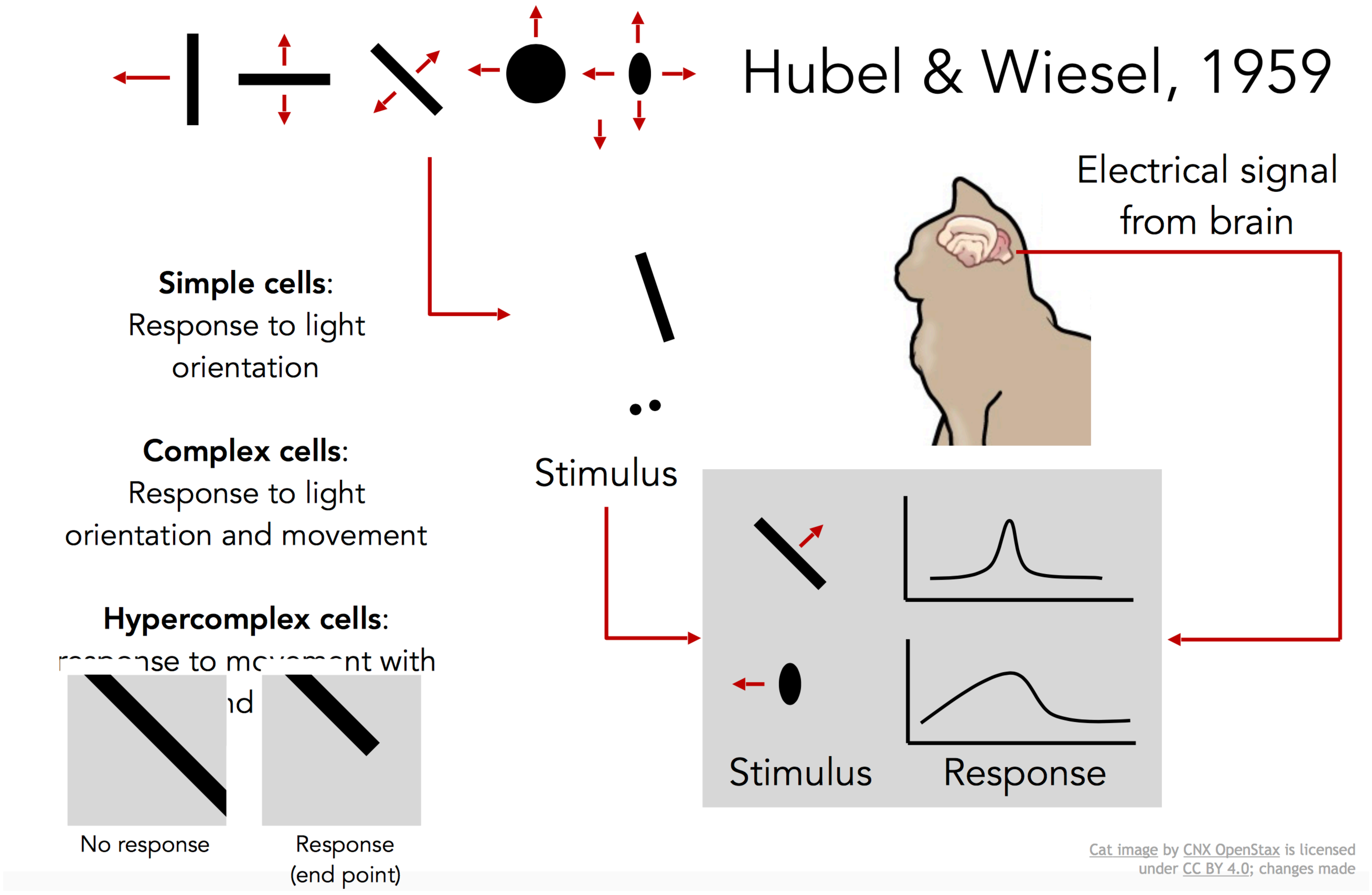
This image is CC0 1.0 public domain

Input Image	Primal Sketch	2 ½-D Sketch	3-D Model Representation
Perceived intensities	Zero crossings, blobs, edges, bars, ends, virtual lines, groups, curves boundaries	Local surface orientation and discontinuities in depth and in surface orientation	3-D models hierarchically organized in terms of surface and volumetric primitives

[Stages of Visual Representation, **David Marr**]

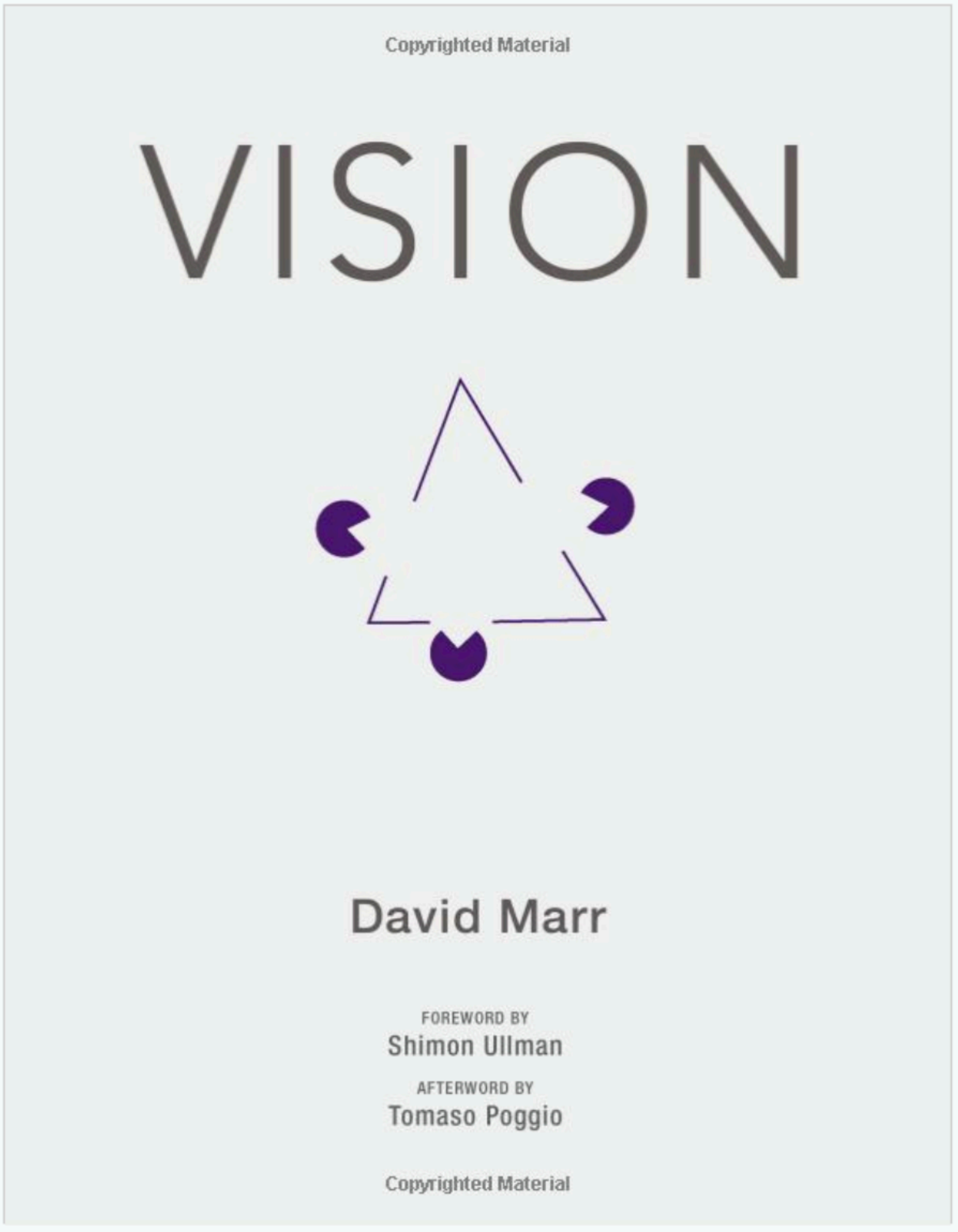
* slide from Fei-Dei Li, Justin Johnson, Serena Yeung, **cs231n Stanford**

Human vision ...



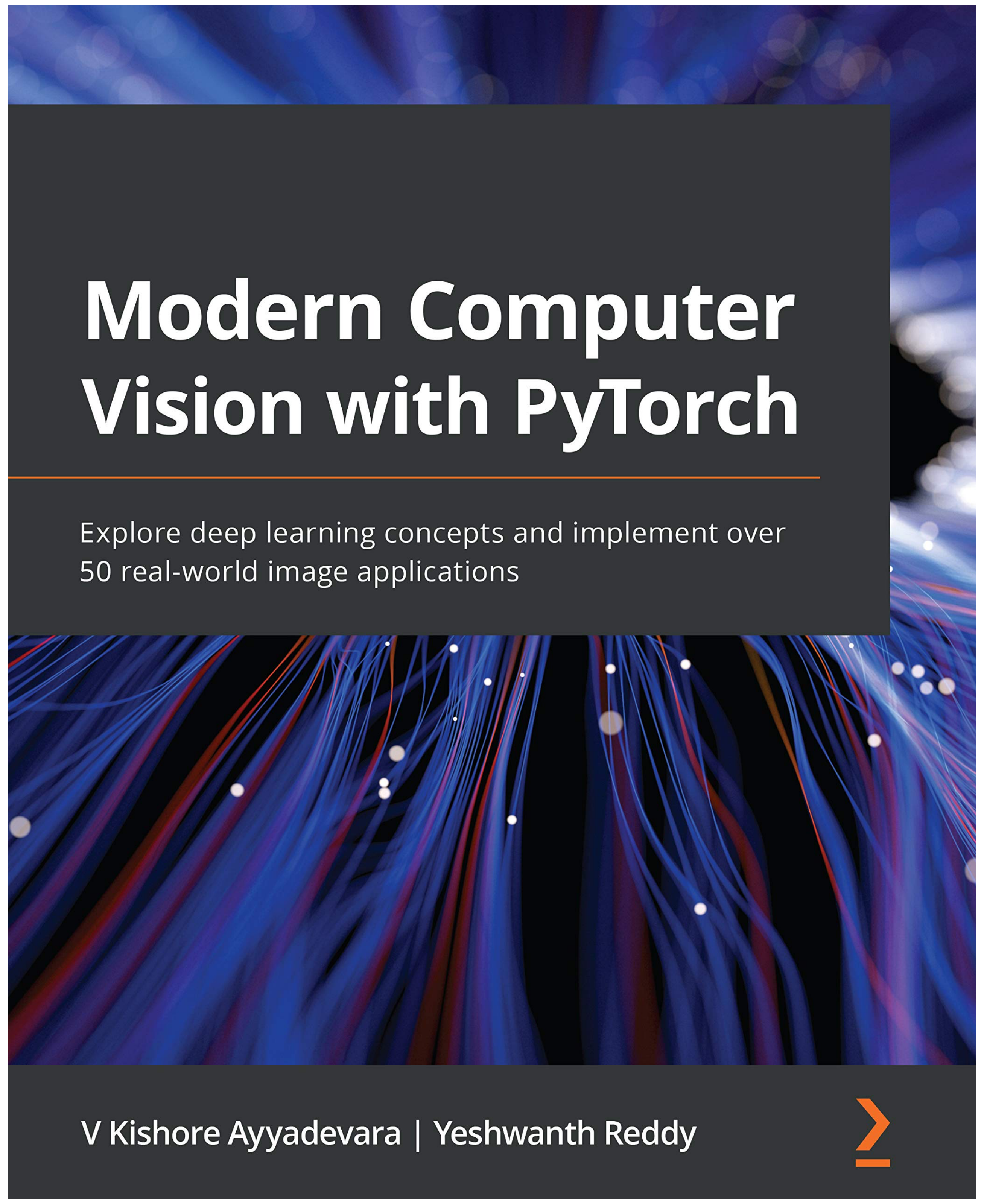
Topics Covered

– 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

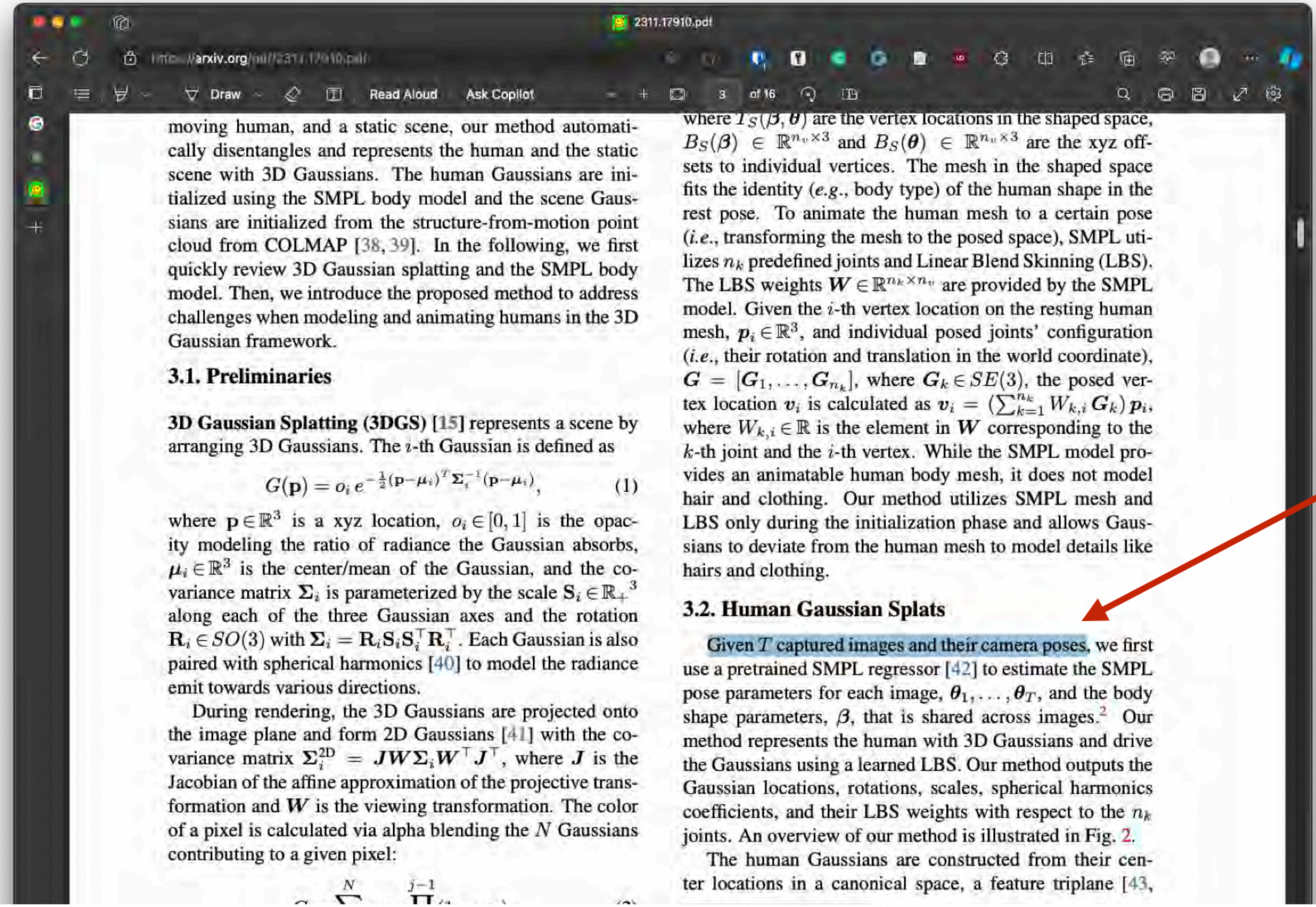


Topics Covered

- 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



Modern Approaches **Rely on a Lot of Traditional Stuff**



2311.17910.pdf

https://arxiv.org/pdf/2311.17910.pdf

Draw Read Aloud Ask Copilot 3 of 16

moving human, and a static scene, our method automatically disentangles and represents the human and the static scene with 3D Gaussians. The human Gaussians are initialized using the SMPL body model and the scene Gaussians are initialized from the structure-from-motion point cloud from COLMAP [38, 39]. In the following, we first quickly review 3D Gaussian splatting and the SMPL body model. Then, we introduce the proposed method to address challenges when modeling and animating humans in the 3D Gaussian framework.

3.1. Preliminaries

3D Gaussian Splatting (3DGS) [15] represents a scene by arranging 3D Gaussians. The i-th Gaussian is defined as

$$G(\mathbf{p}) = o_i \, e^{-\frac{1}{2}(\mathbf{p}-\boldsymbol{\mu}_i)^T \boldsymbol{\Sigma}_i^{-1}(\mathbf{p}-\boldsymbol{\mu}_i)}, \tag{1}$$

where $\mathbf{p} \in \mathbb{R}^3$ is a xyz location, $o_i \in [0,1]$ is the opacity modeling the ratio of radiance the Gaussian absorbs, $\boldsymbol{\mu}_i \in \mathbb{R}^3$ is the center/mean of the Gaussian, and the covariance matrix $\boldsymbol{\Sigma}_i$ is parameterized by the scale $\mathbf{S}_i \in \mathbb{R}_+{}^3$ along each of the three Gaussian axes and the rotation $\mathbf{R}_i \in SO(3)$ with $\boldsymbol{\Sigma}_i = \mathbf{R}_i \mathbf{S}_i \mathbf{S}_i^\top \mathbf{R}_i^\top$. Each Gaussian is also paired with spherical harmonics [40] to model the radiance emit towards various directions.

During rendering, the 3D Gaussians are projected onto the image plane and form 2D Gaussians [41] with the covariance matrix $\boldsymbol{\Sigma}_i^{2D} = \boldsymbol{J}\boldsymbol{W}\boldsymbol{\Sigma}_i\boldsymbol{W}^\top\boldsymbol{J}^\top$, where $\boldsymbol{J}$ is the Jacobian of the affine approximation of the projective transformation and $\boldsymbol{W}$ is the viewing transformation. The color of a pixel is calculated via alpha blending the N Gaussians contributing to a given pixel:

where $T_S(\boldsymbol{\beta}, \boldsymbol{\theta})$ are the vertex locations in the shaped space, $B_S(\boldsymbol{\beta}) \in \mathbb{R}^{n_v \times 3}$ and $B_S(\boldsymbol{\theta}) \in \mathbb{R}^{n_v \times 3}$ are the xyz offsets to individual vertices. The mesh in the shaped space fits the identity (*e.g.*, body type) of the human shape in the rest pose. To animate the human mesh to a certain pose (*i.e.*, transforming the mesh to the posed space), SMPL utilizes n_k predefined joints and Linear Blend Skinning (LBS). The LBS weights $\boldsymbol{W} \in \mathbb{R}^{n_k \times n_v}$ are provided by the SMPL model. Given the i-th vertex location on the resting human mesh, $\boldsymbol{p}_i \in \mathbb{R}^3$, and individual posed joints' configuration (*i.e.*, their rotation and translation in the world coordinate), $\boldsymbol{G} = [\boldsymbol{G}_1, \ldots, \boldsymbol{G}_{n_k}]$, where $\boldsymbol{G}_k \in SE(3)$, the posed vertex location $\boldsymbol{v}_i$ is calculated as $\boldsymbol{v}_i = (\sum_{k=1}^{n_k} W_{k,i} \boldsymbol{G}_k) \boldsymbol{p}_i$, where $W_{k,i} \in \mathbb{R}$ is the element in $\boldsymbol{W}$ corresponding to the k-th joint and the i-th vertex. While the SMPL model provides an animatable human body mesh, it does not model hair and clothing. Our method utilizes SMPL mesh and LBS only during the initialization phase and allows Gaussians to deviate from the human mesh to model details like hairs and clothing.

3.2. Human Gaussian Splats

Given T captured images and their camera poses, we first use a pretrained SMPL regressor [42] to estimate the SMPL pose parameters for each image, $\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_T$, and the body shape parameters, $\boldsymbol{\beta}$, that is shared across images.[2] Our method represents the human with 3D Gaussians and drive the Gaussians using a learned LBS. Our method outputs the Gaussian locations, rotations, scales, spherical harmonics coefficients, and their LBS weights with respect to the n_k joints. An overview of our method is illustrated in Fig. 2.

The human Gaussians are constructed from their center locations in a canonical space, a feature triplane [43,

- 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

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

Short **canvas quizzes**: 10%

Programming Assignments: 45%

6 graded and 1 ungraded (optional) assignment

Midterm Exam (October 21 & 22*): 15%

Final Exam (TBD): 30%

*depending on section (subject to change)

Assignments (done individually)

There will be **7 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 21 & 22** (depending on the section)

— 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 15% to your final score

Final Exam

The Final exam is held during the regular examination period, 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

Final Exam

You **don't need to pass the final** to pass the course

The grade you earn is the grade you get!

Grading issues & **Academic** Misconduct

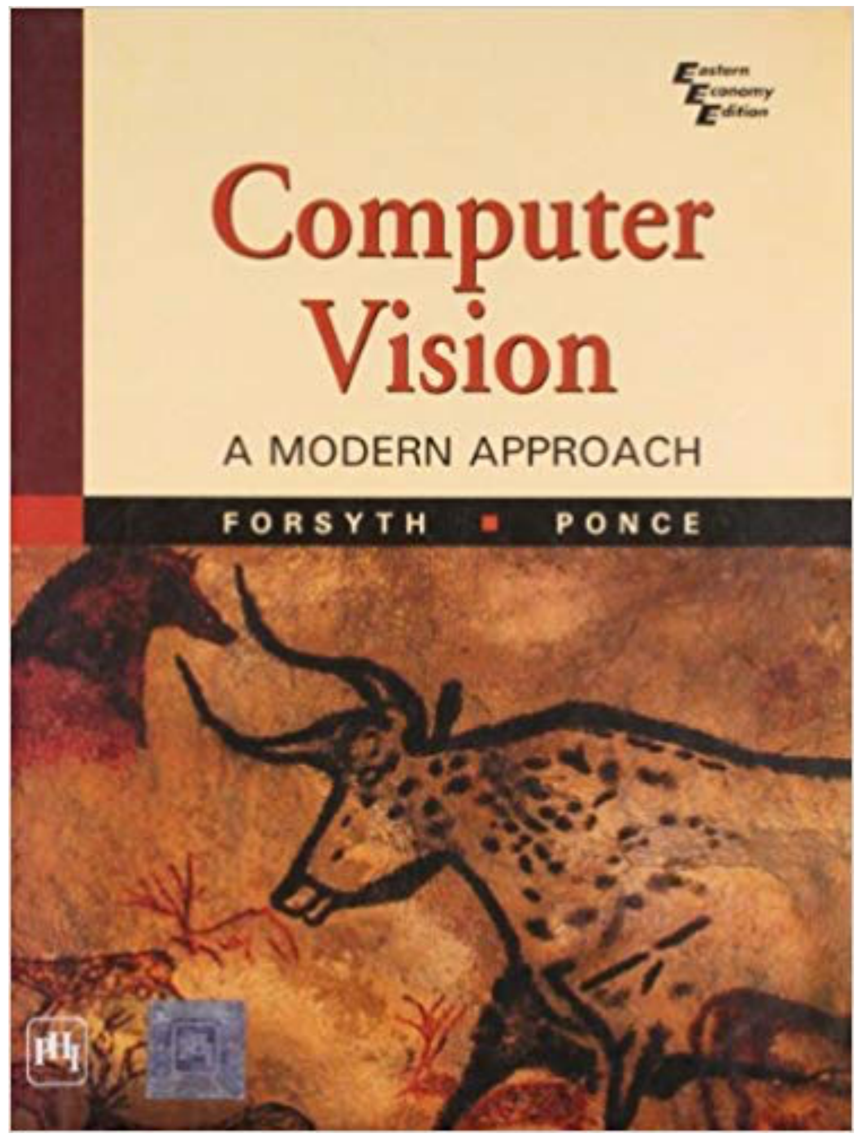
<u>Strict policy</u>: Grading mistakes happen, it’s just a nature of life. If you see an issue with your grade, you have 1 week from the release of any assignment grade to bring a <u>specific issue</u> to our attention.

(Regrade requests are handled as private messages through Piazza)

<u>Academic Misconduct</u>: Please don’t do it. Trust me it is not worth it.

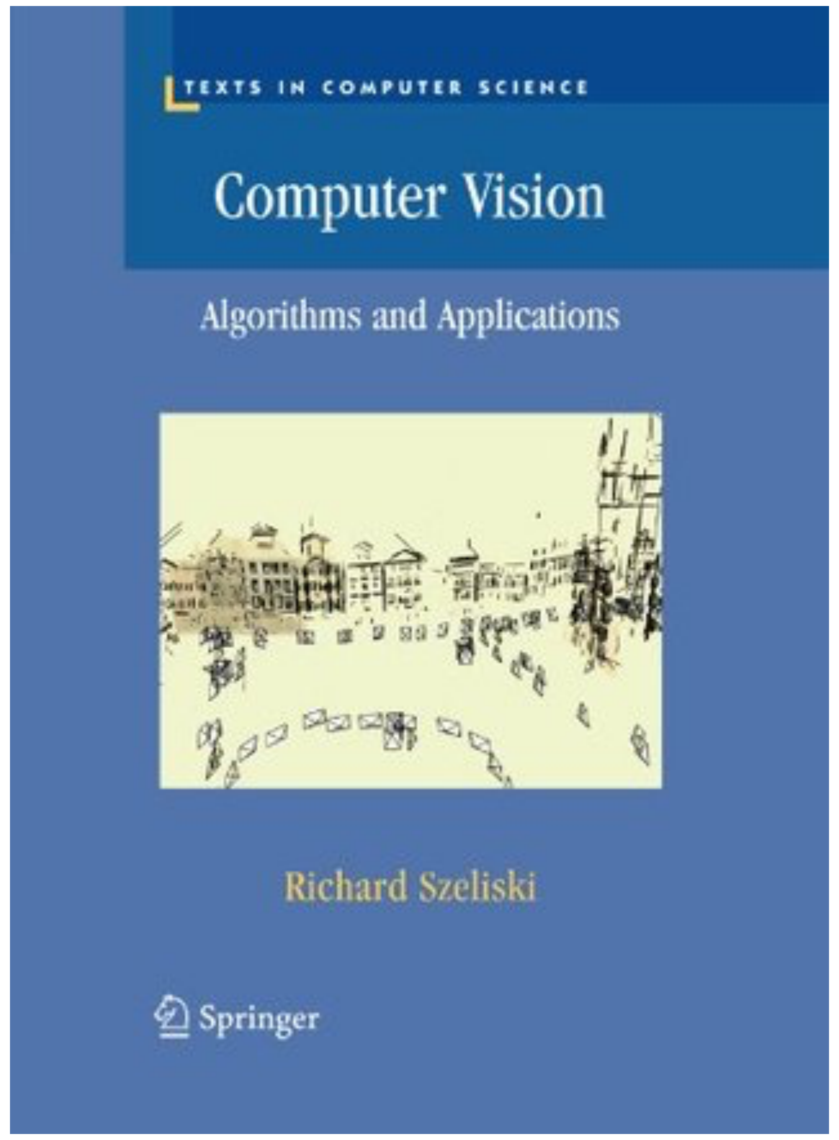
Textbooks

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



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/

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

Borealis AI’s Let’s SOLVE it **undergraduate program**

Borealis AI’s Let’s SOLVE it, AI undergrad student mentorship program, now is open for applications for our upcoming fall cycle. It would be a great help if you could spread the news among the undergrad students at your department, and/or courses! We are looking for undergrad students from diverse backgrounds to support kickstart their career in A, while helping them solve a problem from their communities. Your help with this can go a long way!

For application guidelines, please visit our Let's SOLVE it webpage.
Application deadline: **Sept. 8, 2024** (@11.59pm ET)
Program dates: October-November 2023 (2 months)
Additional info here: University Students Harness the Power of AI for Social Good

For questions, please email us at: mi.research@borealisai.com
Use subject line: Let’s SOLVE it

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 11