EECS 504: Foundations of Computer Vision

Andrew Owens

Course staff

Haozhu Wang
Graduate student (GSI)

Anthony Liang
Instructional aide (IA)

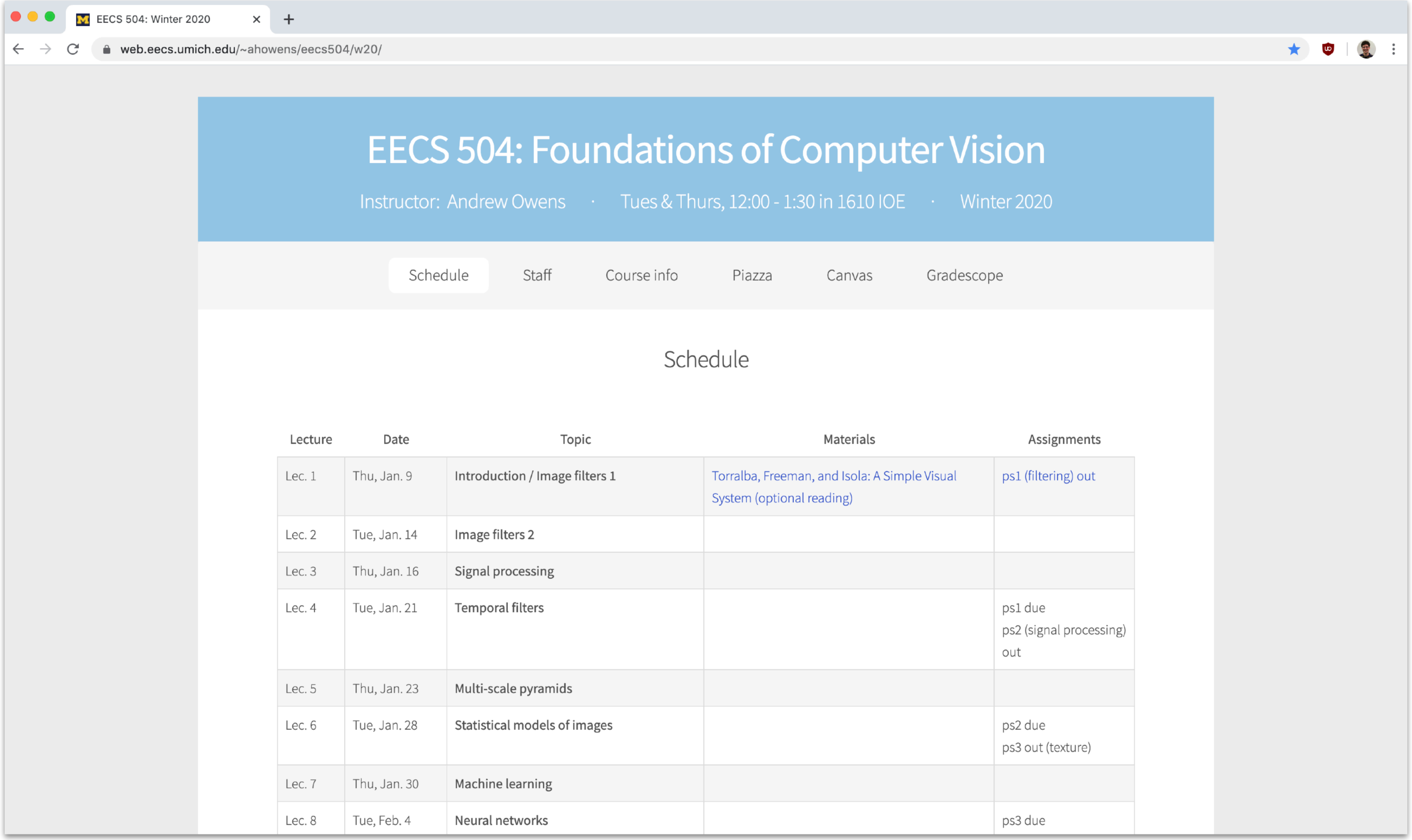
Bingqi Sun
Instructional aide (IA)

Interacting with us

- Ask questions on Piazza.
  - Sign up: https://bit.ly/36mfYeP
- Submit written work to Gradescope
- Office hours on website:

Name	Office hour times	Location
Andrew Owens	Fri. 3:00pm - 4:00pm	EECS 4231
Haozhu Wang	Thu. 7:30pm - 8:30pm	EECS 3312
Bingqi Sun	Mon. 12:00 - 1:00	EECS 3312
Anthony Liang	Mon. 1:00 - 2:00pm	EECS 3312

Course website


EECS 504: Winter 2020

web.eecs.umich.edu/~ahowens/eecs504/w20/

EECS 504: Foundations of Computer Vision

Instructor: Andrew Owens · Tues & Thurs, 12:00 - 1:30 in 1610 IOE · Winter 2020

Schedule Staff Course info Piazza Canvas Gradescope

Schedule

Lecture	Date	Topic	Materials	Assignments
Lec. 1	Thu, Jan. 9	Introduction / Image filters 1	Torralba, Freeman, and Isola: A Simple Visual System (optional reading)	ps1 (filtering) out
Lec. 2	Tue, Jan. 14	Image filters 2		
Lec. 3	Thu, Jan. 16	Signal processing		
Lec. 4	Tue, Jan. 21	Temporal filters		ps1 due ps2 (signal processing) out
Lec. 5	Thu, Jan. 23	Multi-scale pyramids		
Lec. 6	Tue, Jan. 28	Statistical models of images		ps2 due ps3 out (texture)
Lec. 7	Thu, Jan. 30	Machine learning		
Lec. 8	Tue, Feb. 4	Neural networks		ps3 due

https://web.eecs.umich.edu/~ahowens/eecs504/w20/

Grading

- Assignments (70%)
- Final project (30%)

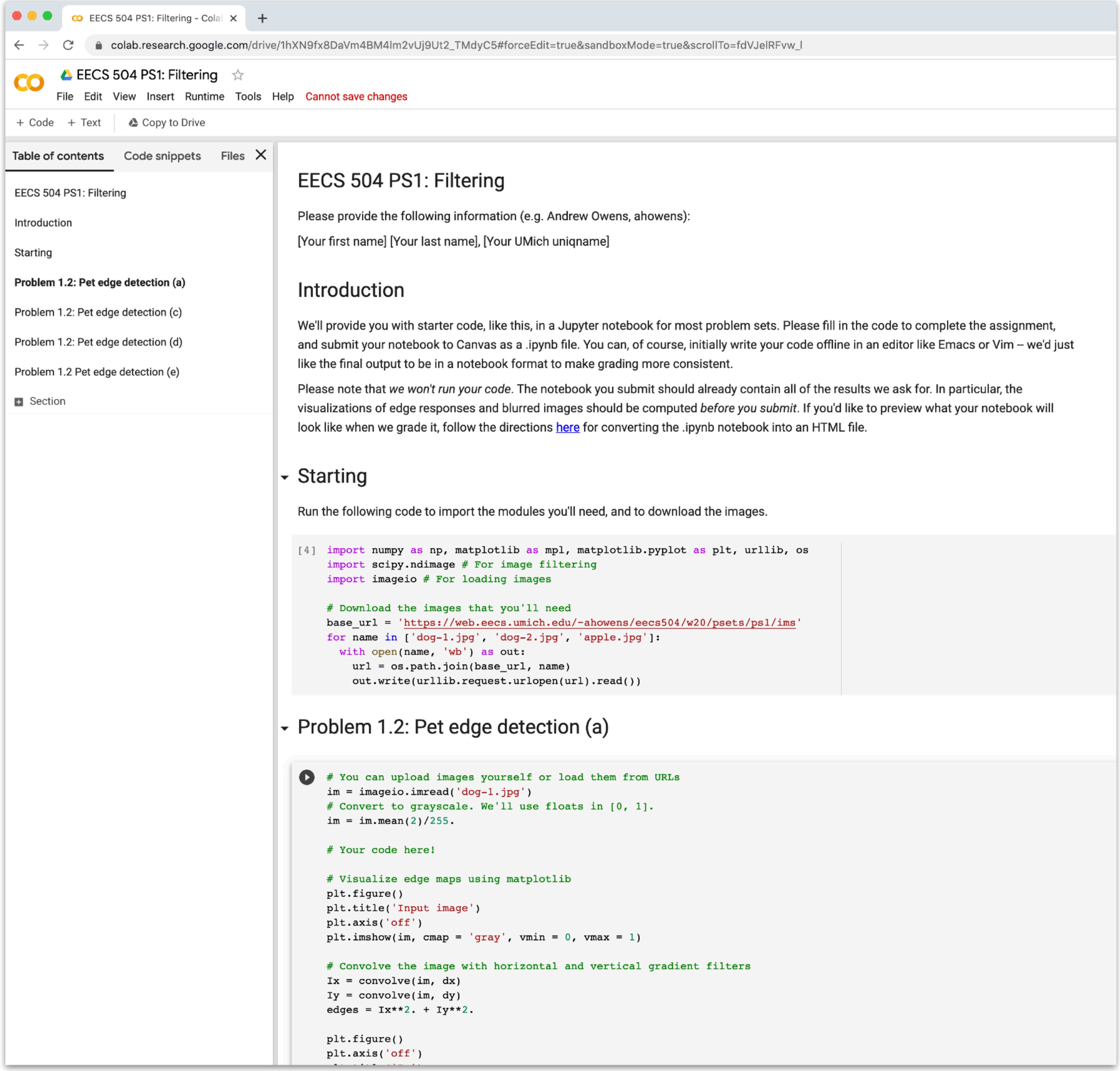
Assignments

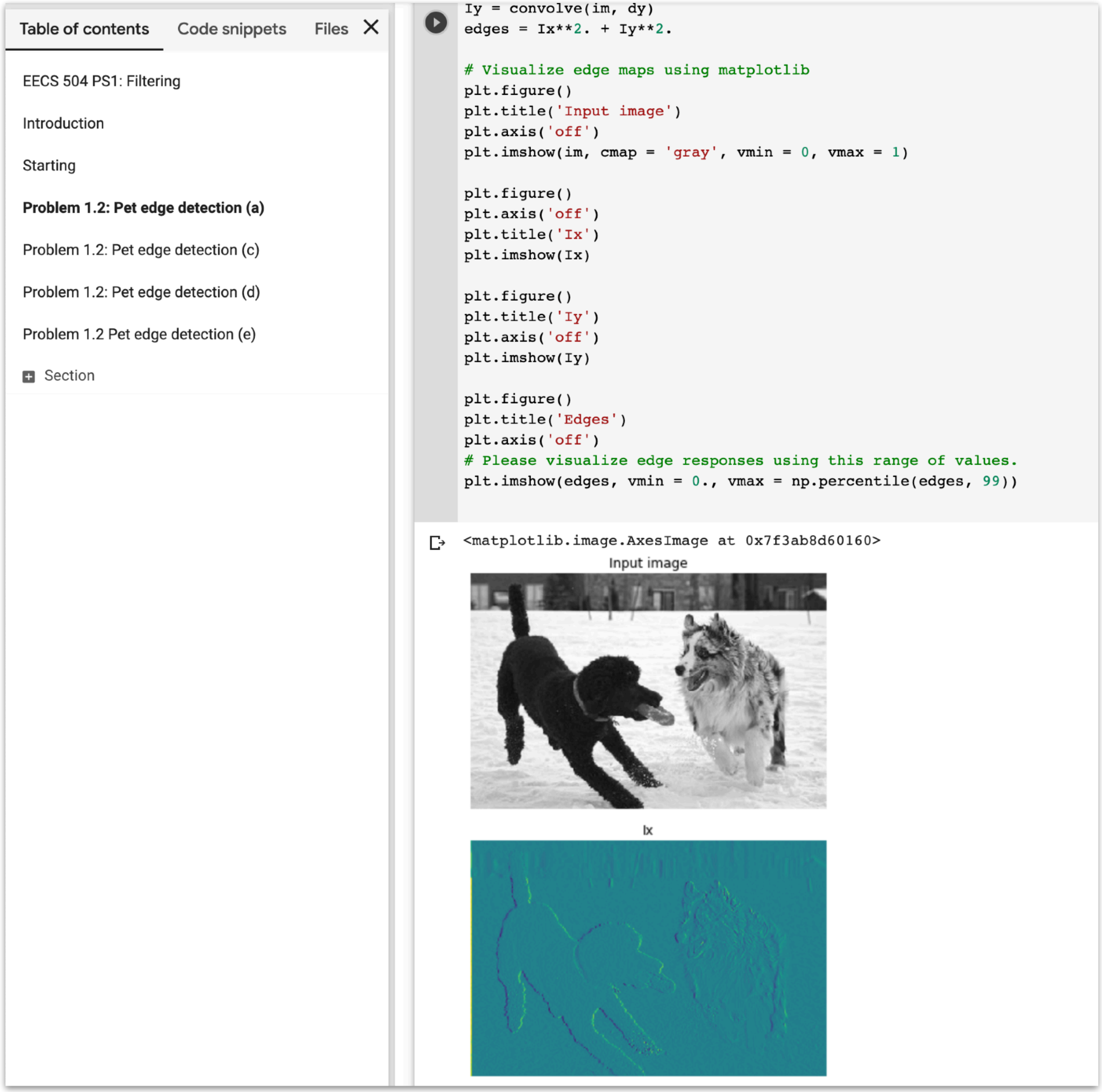
- Weekly homework assignments (≈10 total)
- Due each Tuesday at midnight
- Late submissions penalized 30% per day
  - You have 5 "late days”
- Assignments should be done independently.
  - Encouraged to discuss them
  - Programming and writing should all be yours

Assignments

- Mix of programming and written problems
- Python + numerical computing libraries (numpy, scipy, etc.)
- PyTorch for deep learning
- Linear algebra and multivariable calculus
- Jupyter notebooks and Google Colab for problem sets

Assignments

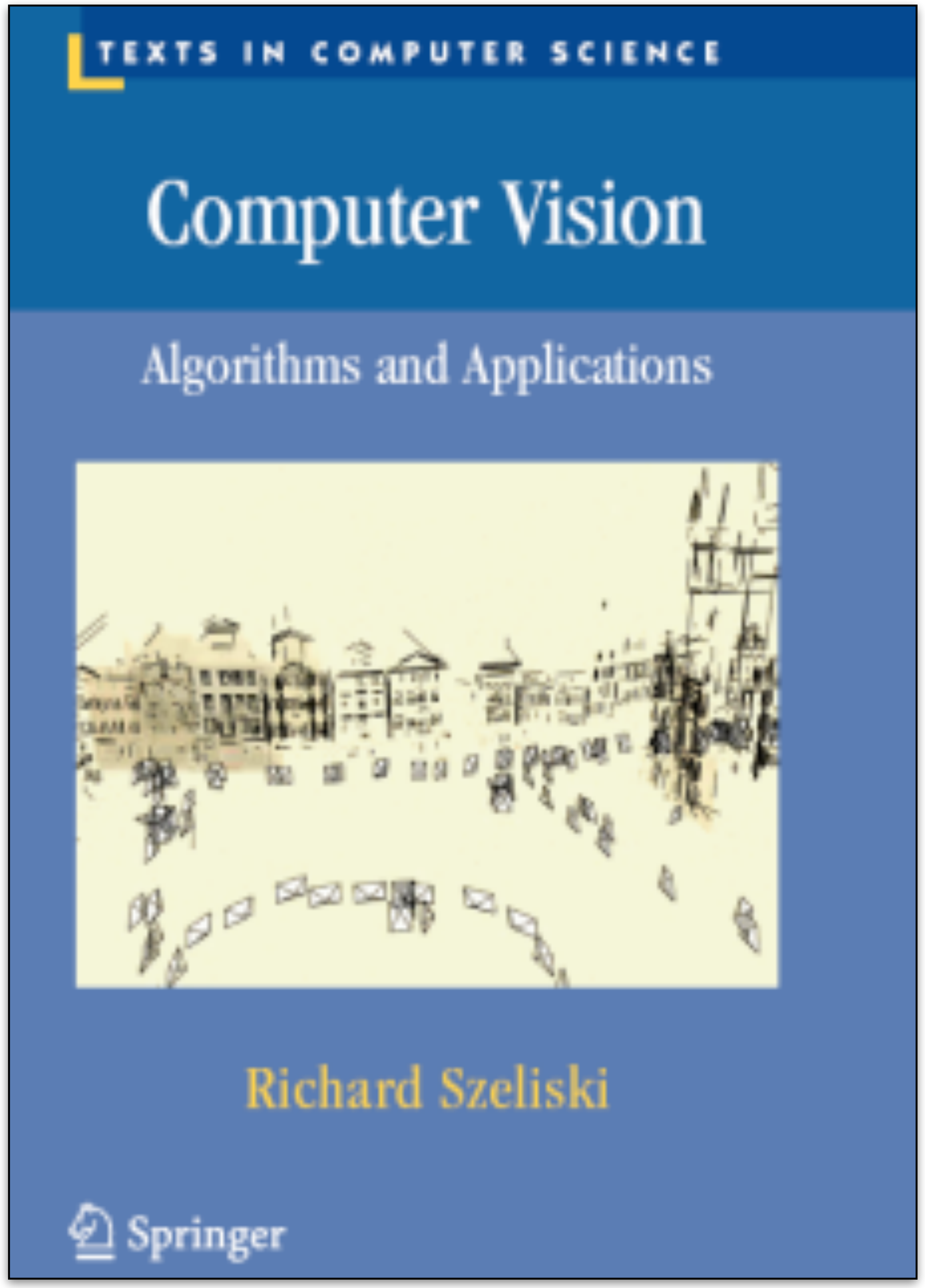




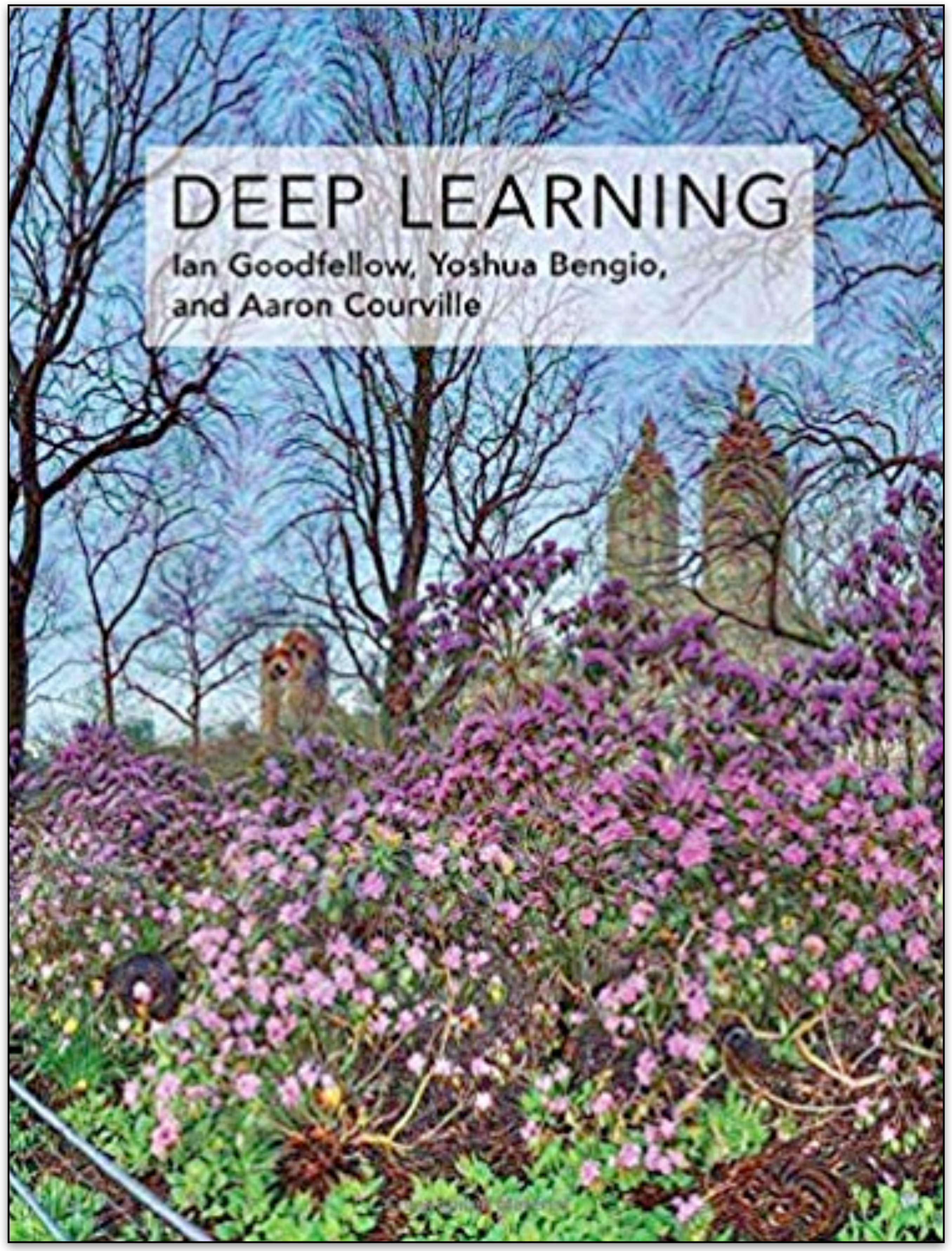
Project

- Open-ended! Example projects:
  - Implement and extend a recent computer vision paper
  - Use computer vision in your research
  - We’ll also provide a list of project ideas
- Work in small groups (up to 4 people)
- Complete in last month of class.
  - Project proposal (after spring break)
  - Short presentation (finals period)
  - Writeup (finals period)

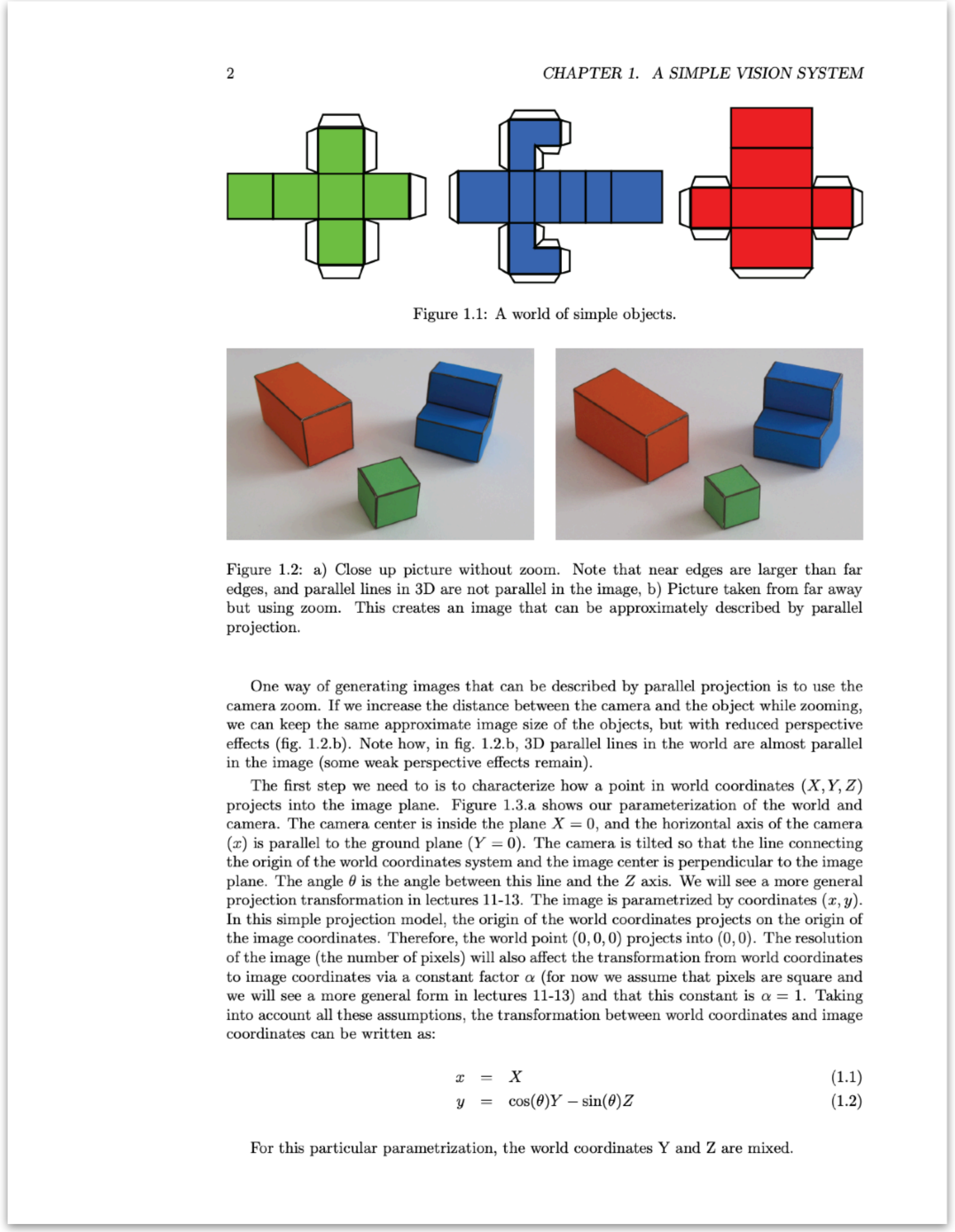
Readings



http://szeliski.org/Book



https://www.deeplearningbook.org



2 CHAPTER 1. A SIMPLE VISION SYSTEM

Figure 1.1: A world of simple objects.

Figure 1.2: a) Close up picture without zoom. Note that near edges are larger than far edges, and parallel lines in 3D are not parallel in the image, b) Picture taken from far away but using zoom. This creates an image that can be approximately described by parallel projection.

One way of generating images that can be described by parallel projection is to use the camera zoom. If we increase the distance between the camera and the object while zooming, we can keep the same approximate image size of the objects, but with reduced perspective effects (fig. 1.2.b). Note how, in fig. 1.2.b, 3D parallel lines in the world are almost parallel in the image (some weak perspective effects remain).

The first step we need to is to characterize how a point in world coordinates (X, Y, Z) projects into the image plane. Figure 1.3.a shows our parameterization of the world and camera. The camera center is inside the plane $X = 0$, and the horizontal axis of the camera (x) is parallel to the ground plane $(Y = 0)$. The camera is tilted so that the line connecting the origin of the world coordinates system and the image center is perpendicular to the image plane. The angle θ is the angle between this line and the Z axis. We will see a more general projection transformation in lectures 11-13. The image is parametrized by coordinates (x, y). In this simple projection model, the origin of the world coordinates projects on the origin of the image coordinates. Therefore, the world point $(0, 0, 0)$ projects into $(0, 0)$. The resolution of the image (the number of pixels) will also affect the transformation from world coordinates to image coordinates via a constant factor α (for now we assume that pixels are square and we will see a more general form in lectures 11-13) and that this constant is $\alpha = 1$. Taking into account all these assumptions, the transformation between world coordinates and image coordinates can be written as:

$$x = X \quad (1.1)$$
$$y = \cos(\theta)Y - \sin(\theta)Z \quad (1.2)$$

For this particular parametrization, the world coordinates Y and Z are mixed.

Manuscript chapters by Torralba, Freeman, and Isola (on course website). Class based on this coursework.

And also occasional paper readings

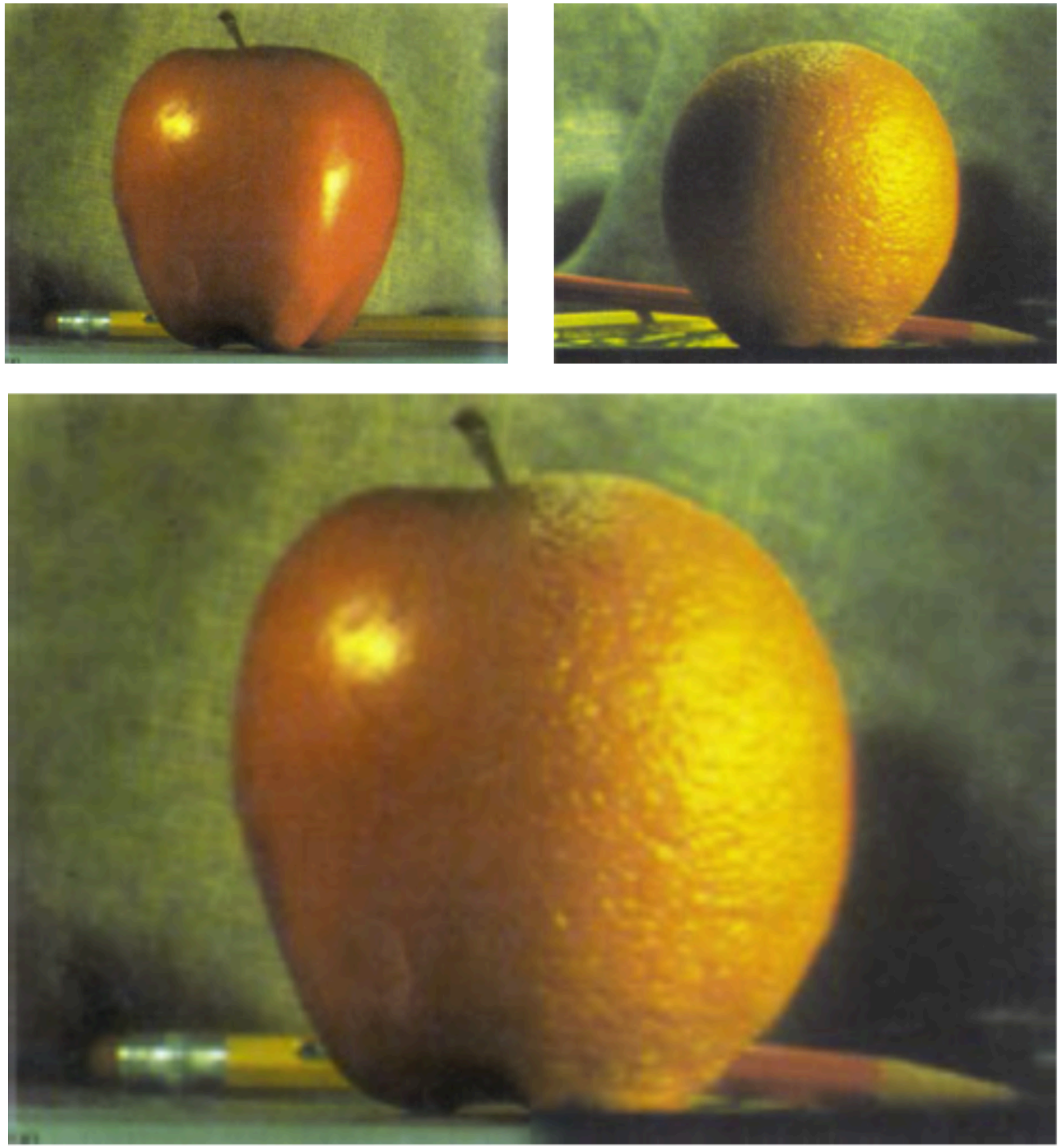
Class topics

Homework problem:

Lec. 1	Thu, Jan. 9	Introduction / Image filters 1
Lec. 2	Tue, Jan. 14	Image filters 2
Lec. 3	Thu, Jan. 16	Signal processing
Lec. 4	Tue, Jan. 21	Temporal filters
Lec. 5	Thu, Jan. 23	Multi-scale pyramids
Lec. 6	Tue, Jan. 28	Statistical models of images
Lec. 7	Thu, Jan. 30	Machine learning
Lec. 8	Tue, Feb. 4	Neural networks
Lec. 9	Thu, Feb. 6	Optimization
Lec. 10	Tue, Feb. 11	Convolutional networks
Tutorial	Tue, Feb. 11	PyTorch tutorial

...

Signa



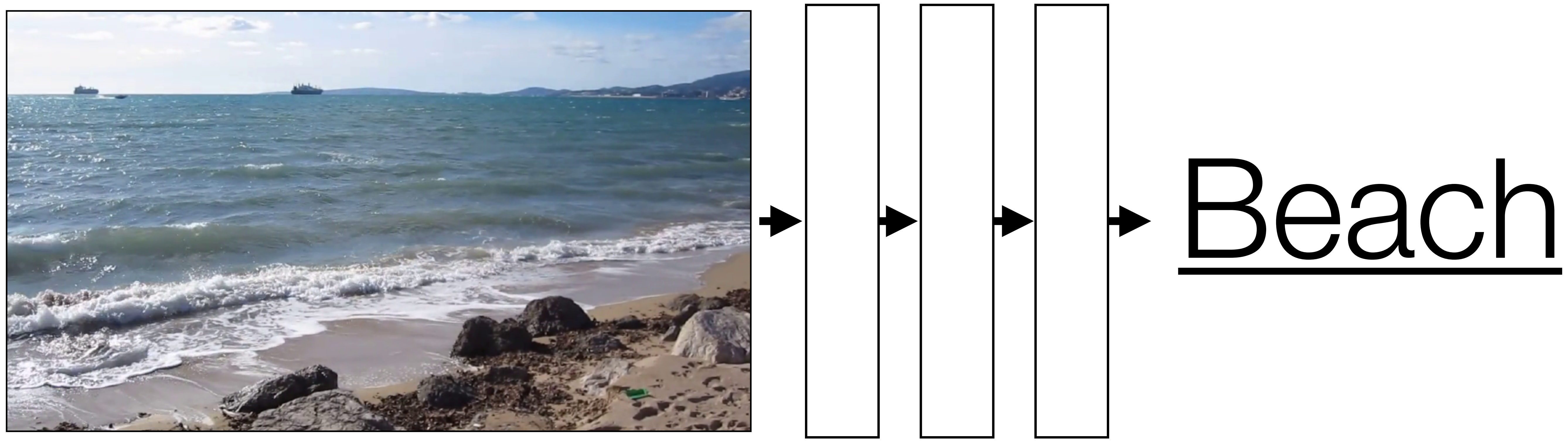
Lec. 1	Thu, Jan. 9	Introduction / Image filters 1
Lec. 2	Tue, Jan. 14	Image filters 2
Lec. 3	Thu, Jan. 16	Signal processing
Lec. 4	Tue, Jan. 21	Temporal filters
Lec. 5	Thu, Jan. 23	Multi-scale pyramids
Lec. 6	Tue, Jan. 28	Statistical models of images
Lec. 7	Thu, Jan. 30	Machine learning
Lec. 8	Tue, Feb. 4	Neural networks
Lec. 9	Thu, Feb. 6	Optimization
Lec. 10	Tue, Feb. 11	Convolutional networks
Tutorial	Tue, Feb. 11	PyTorch tutorial

...

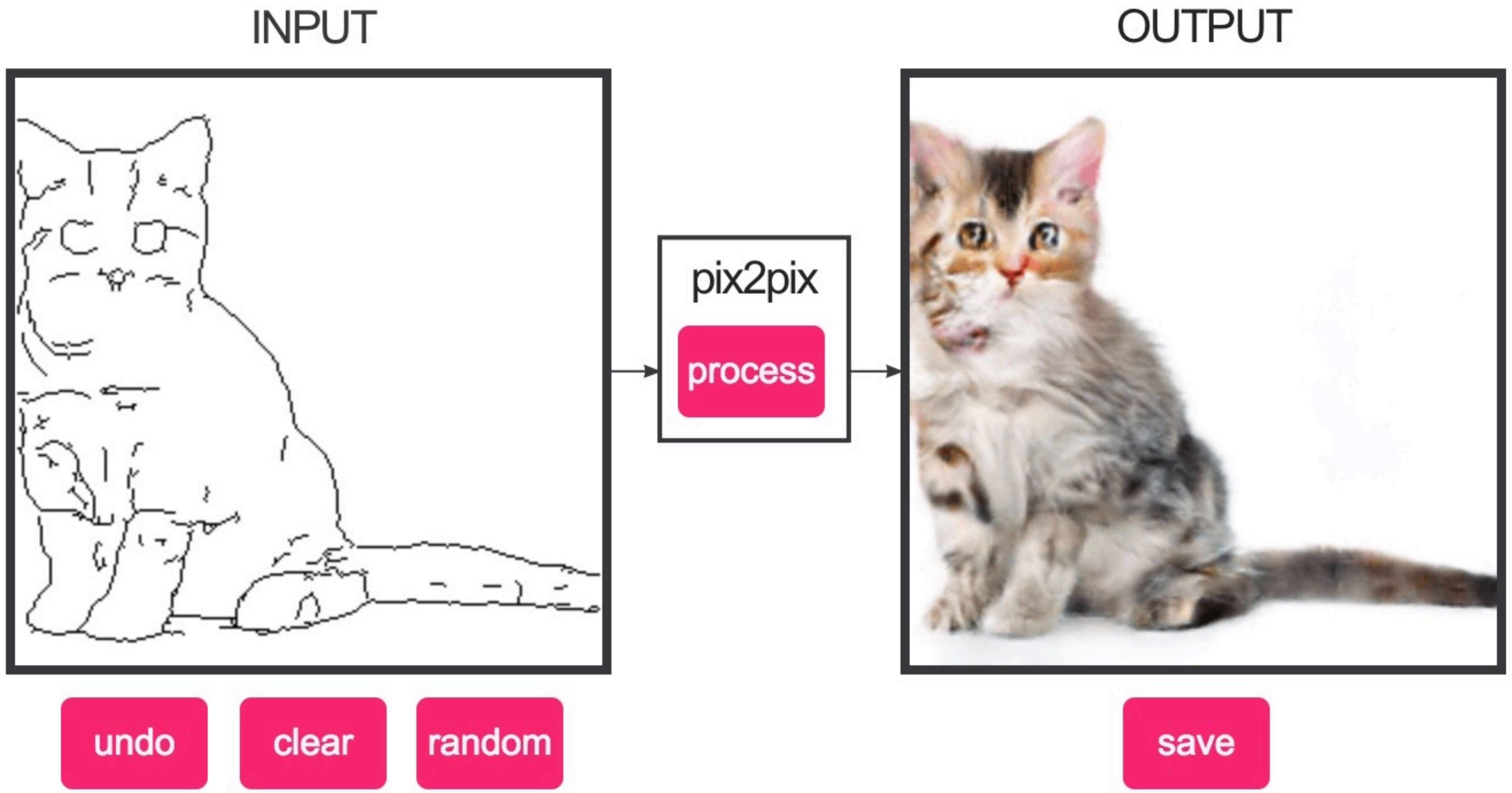
Intro to
deep learning

Tutorial	Tue, Feb. 11	PyTorch tutorial
Lec. 11	Thu, Feb. 13	Scene understanding
Lec. 12	Tue, Feb. 18	Object detection
Lec. 13	Thu, Feb. 20	Action recognition
Lec. 14	Tue, Feb. 25	GANs
Lec. 15	Thu, Feb. 27	Image synthesis
Spring break		
Lec. 16	Tue, Mar. 10	Representation learning
Lec. 17	Thu, Mar. 12	Sight, sound, and touch
Lec. 18	Tue, Mar. 17	Optical flow
Lec. 19	Thu, Mar. 19	Multi-view geometry

...



Learning for vision



Homework problem:

Lec. 17	Thu, Mar. 12	Sight, sound, and touch
Lec. 18	Tue, Mar. 17	Optical flow
Lec. 19	Thu, Mar. 19	Multi-view geometry
Lec. 20	Tue, Mar. 24	Structure from motion
Lec. 21	Thu, Mar. 26	Depth estimation
Lec. 22	Tue, Mar. 31	Graphical models
Lec. 23	Thu, Apr. 2	Color
Lec. 24	Tue, Apr. 7	Embodied vision
Lec. 25	Thu, Apr. 9	Modern CNNs
Lec. 26	Tue, Apr. 14	Image forensics
Lec. 27	Thu, Apr. 16	Language and vision

...

Ca
optic

Lec. 19	Thu, Mar. 19	Multi-view geometry
Lec. 20	Tue, Mar. 24	Structure from motion
Lec. 21	Thu, Mar. 26	Depth estimation
Lec. 22	Tue, Mar. 31	Graphical models
Lec. 23	Thu, Apr. 2	Color
Lec. 24	Tue, Apr. 7	Embodied vision
Lec. 25	Thu, Apr. 9	Modern CNNs
Lec. 26	Tue, Apr. 14	Image forensics
Lec. 27	Thu, Apr. 16	Language and vision
Lec. 28	Tue, Apr. 21	Datasets and bias

Advanced topics
and applications

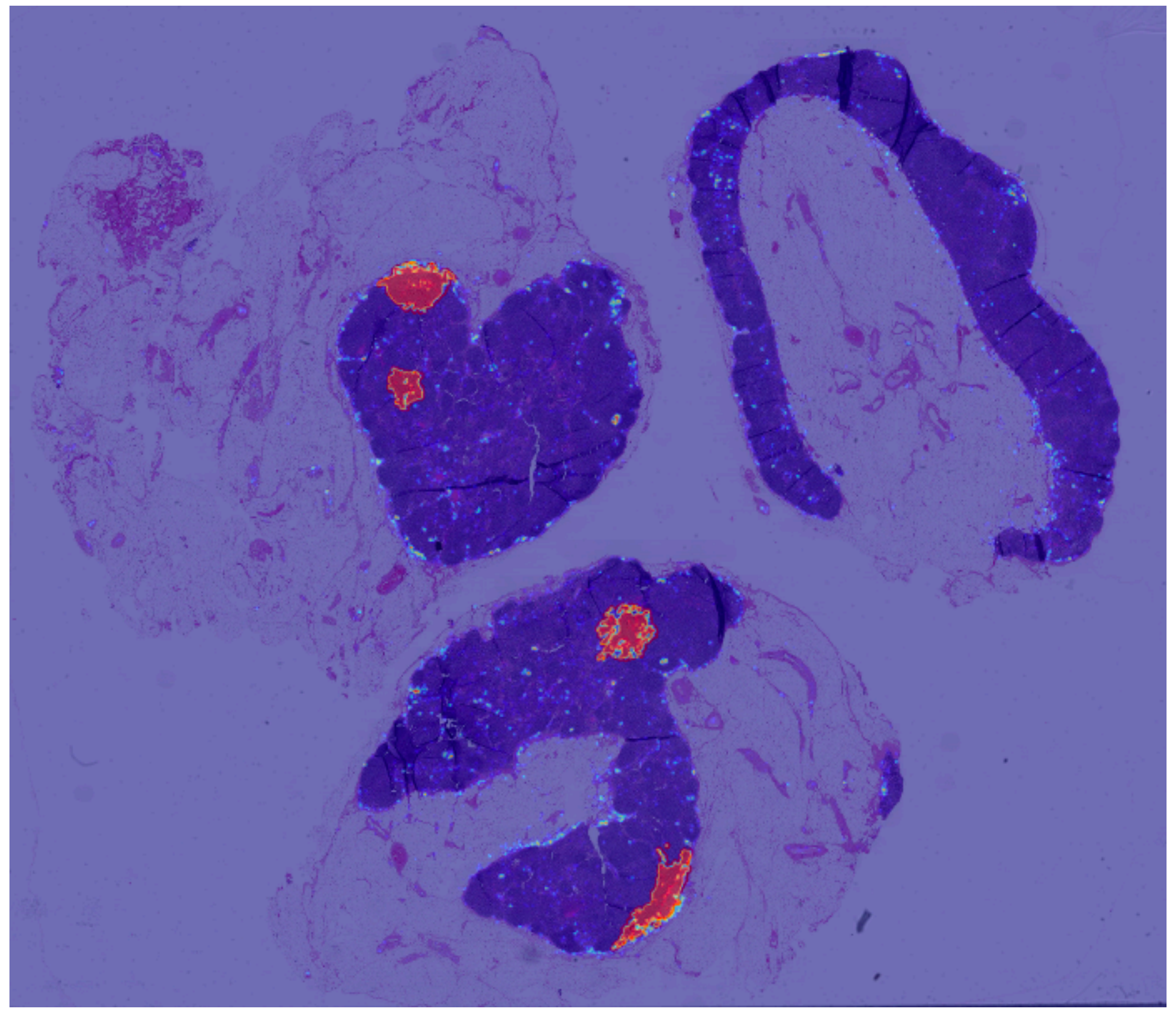
Today

1. A bit of vision history
2. Why vision is hard
3. A simple visual system

Exciting times for computer vision

Robotics

Medical applications



3D modeling

Driving

Mobile devices

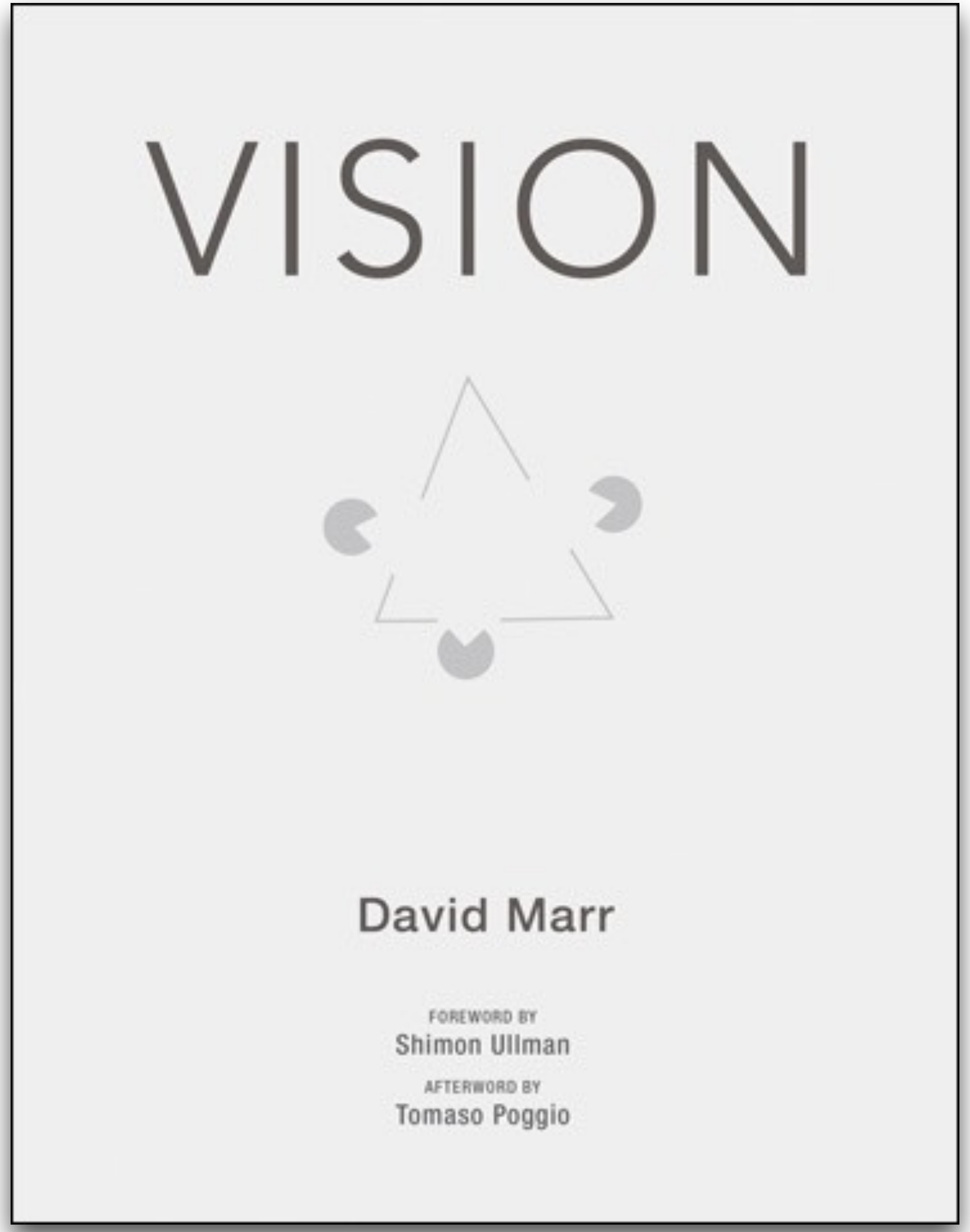
Accessibility

Slide credit: Torralba, Freeman, Isola

To see

“What does it mean, to see? The plain man's answer (and Aristotle's, too) would be, to know what is where by looking.”

To discover from images what is present in the world, where things are, what actions are taking place, to predict and anticipate events in the world.



MASSACHUSETTS INSTITUTE OF TECHNOLOGY

PROJECT MAC

Artificial Intelligence Group July 7, 1966

Vision Memo. No. 100.

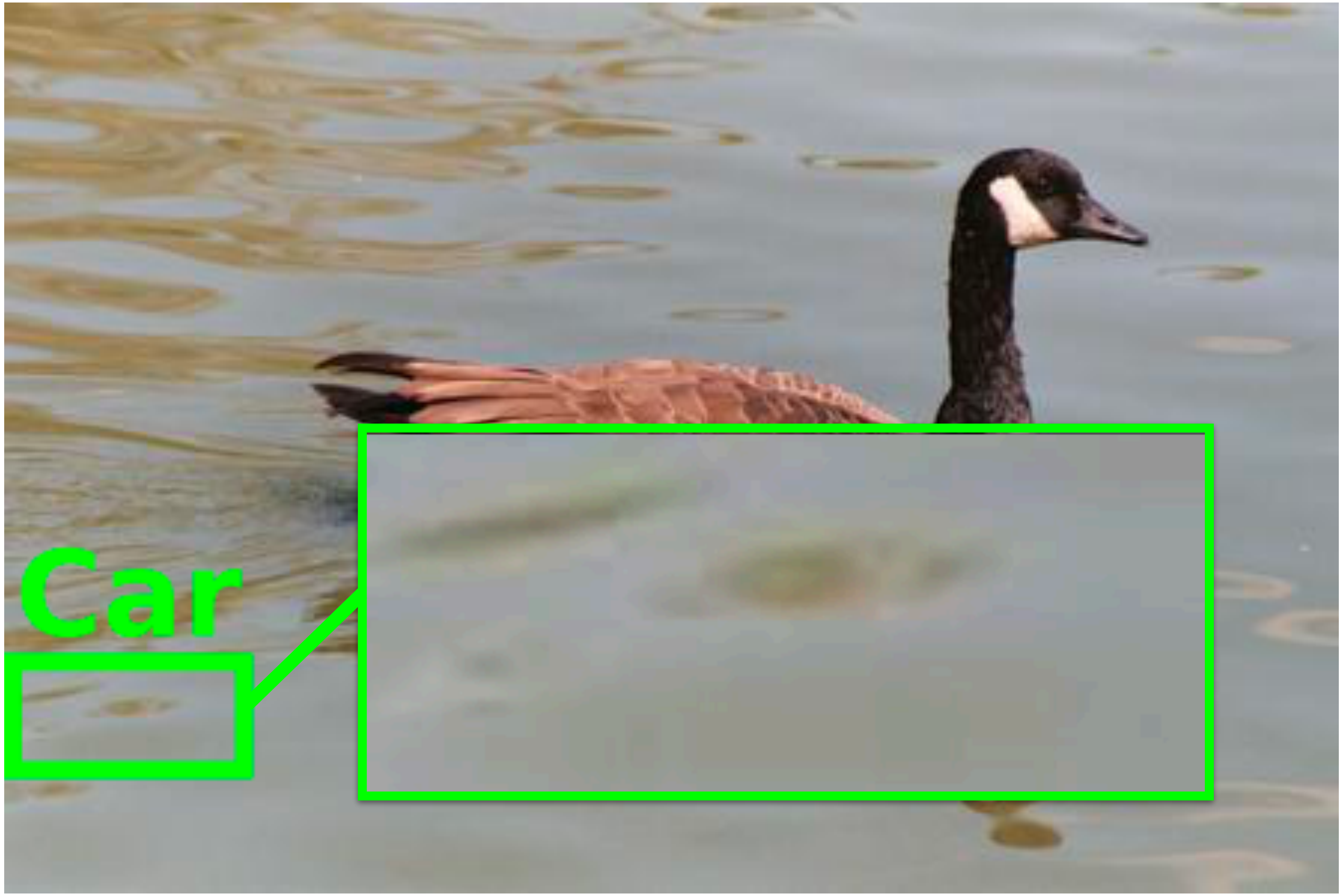
THE SUMMER VISION PROJECT

Seymour Papert

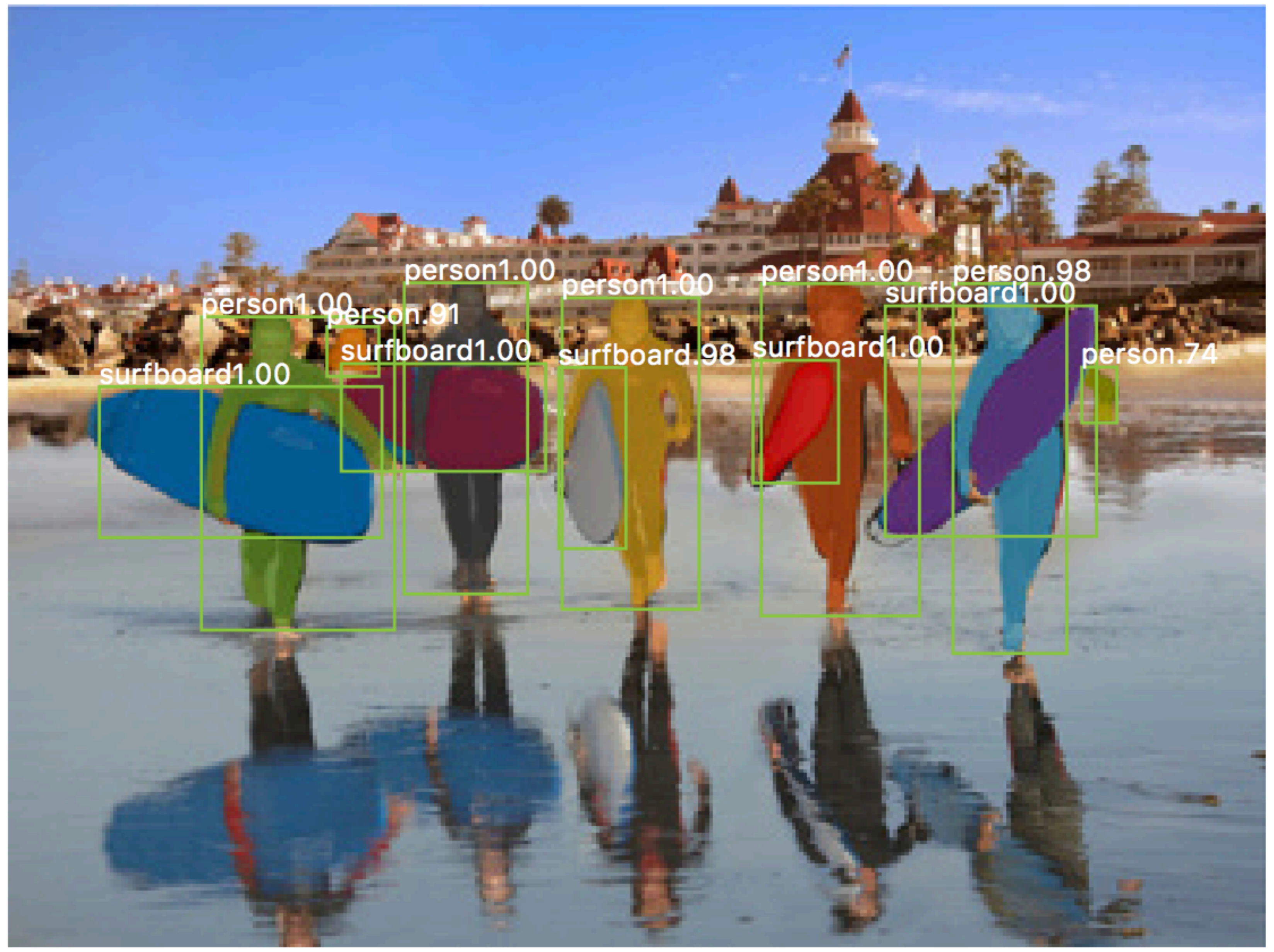
The summer vision project is an attempt to use our summer workers effectively in the construction of a significant part of a visual system. The particular task was chosen partly because it can be segmented into sub-problems which will allow individuals to work independently and yet participate in the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

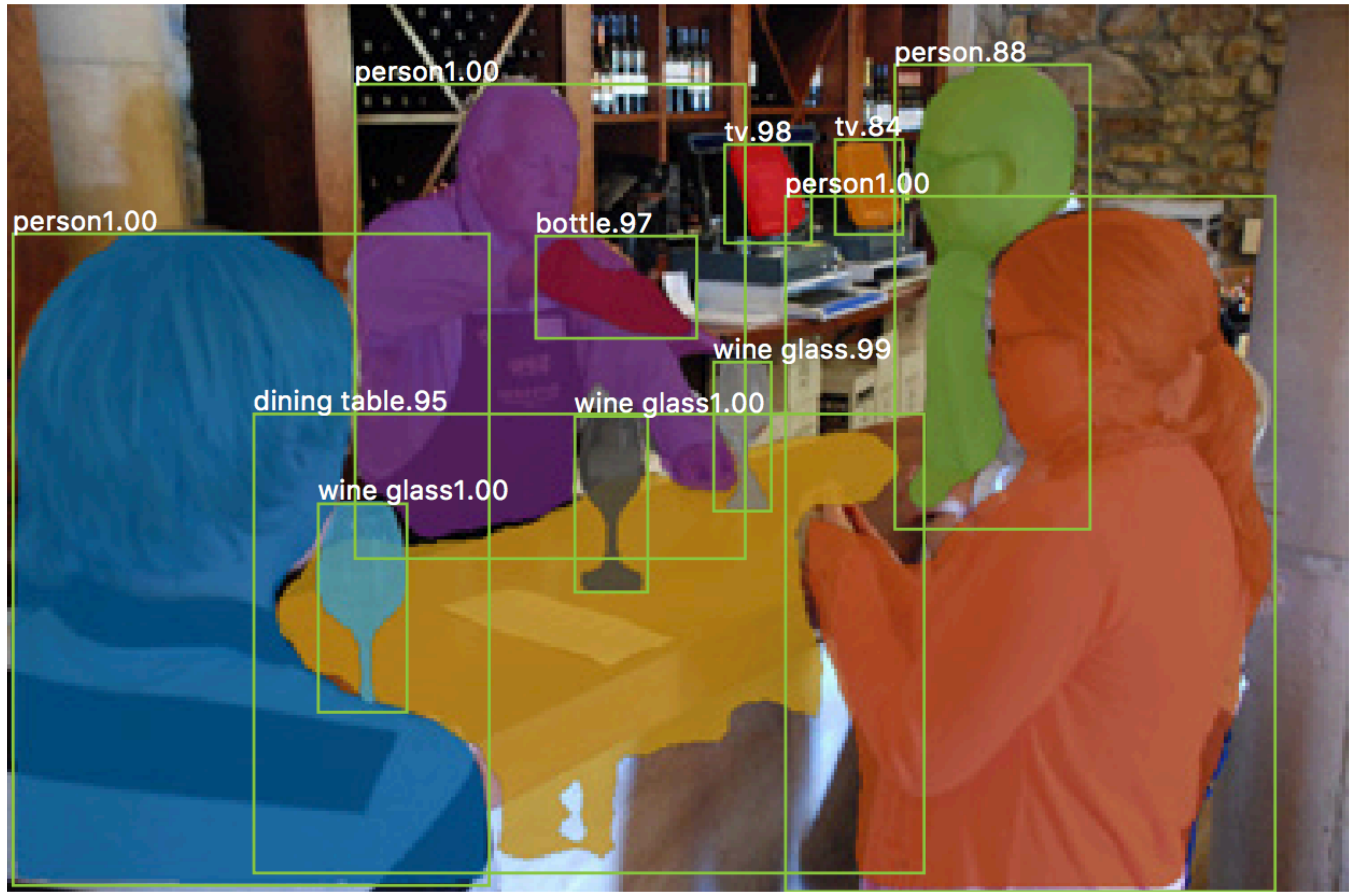
Slide credit: Torralba, Freeman, Isola

Just a few years ago…



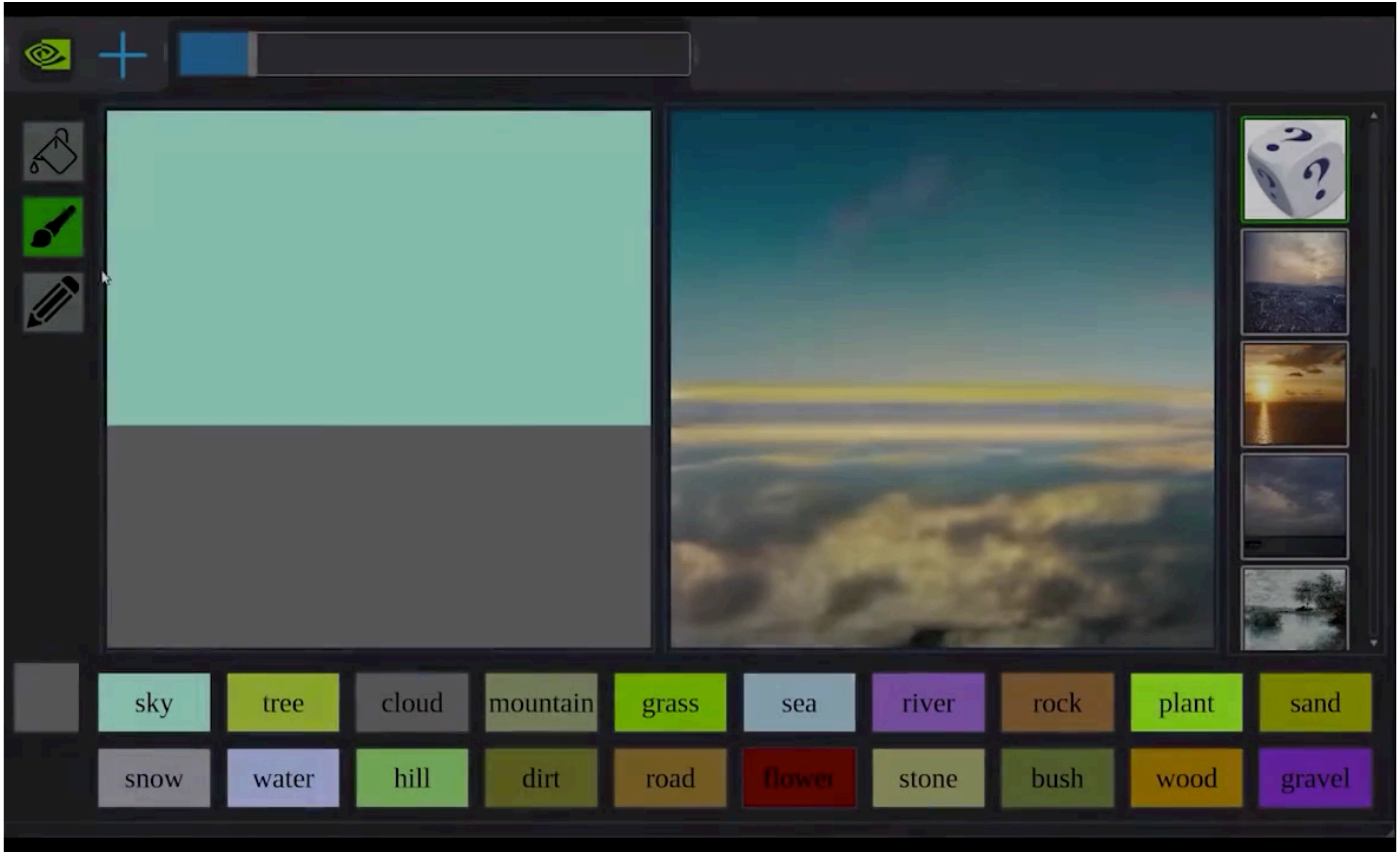
["HOGgles", Vondrick et al. , ICCV 2013]





Slide credit: Torralba, Freeman, Isola

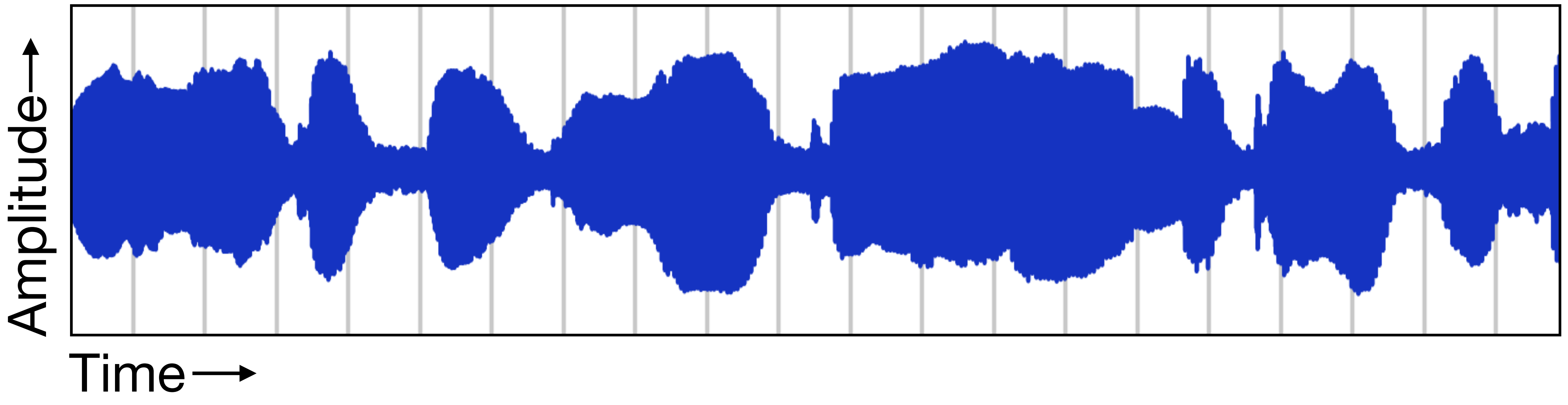
[“Mask RCNN”, He et al., ICCV 2017]



[“GauGAN”, Park et al., CVPR 2019]

Different signals, same methods

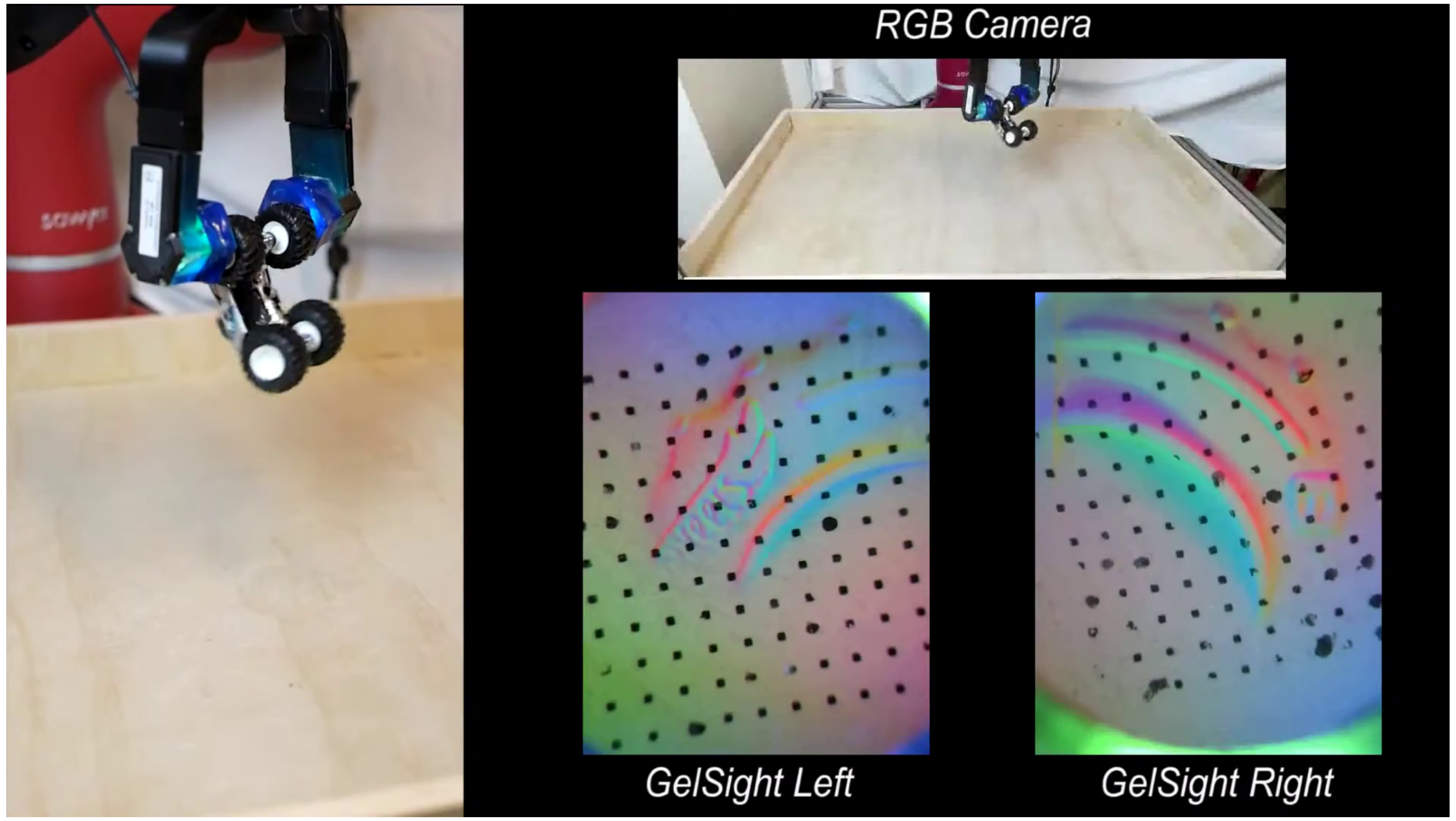
Sound



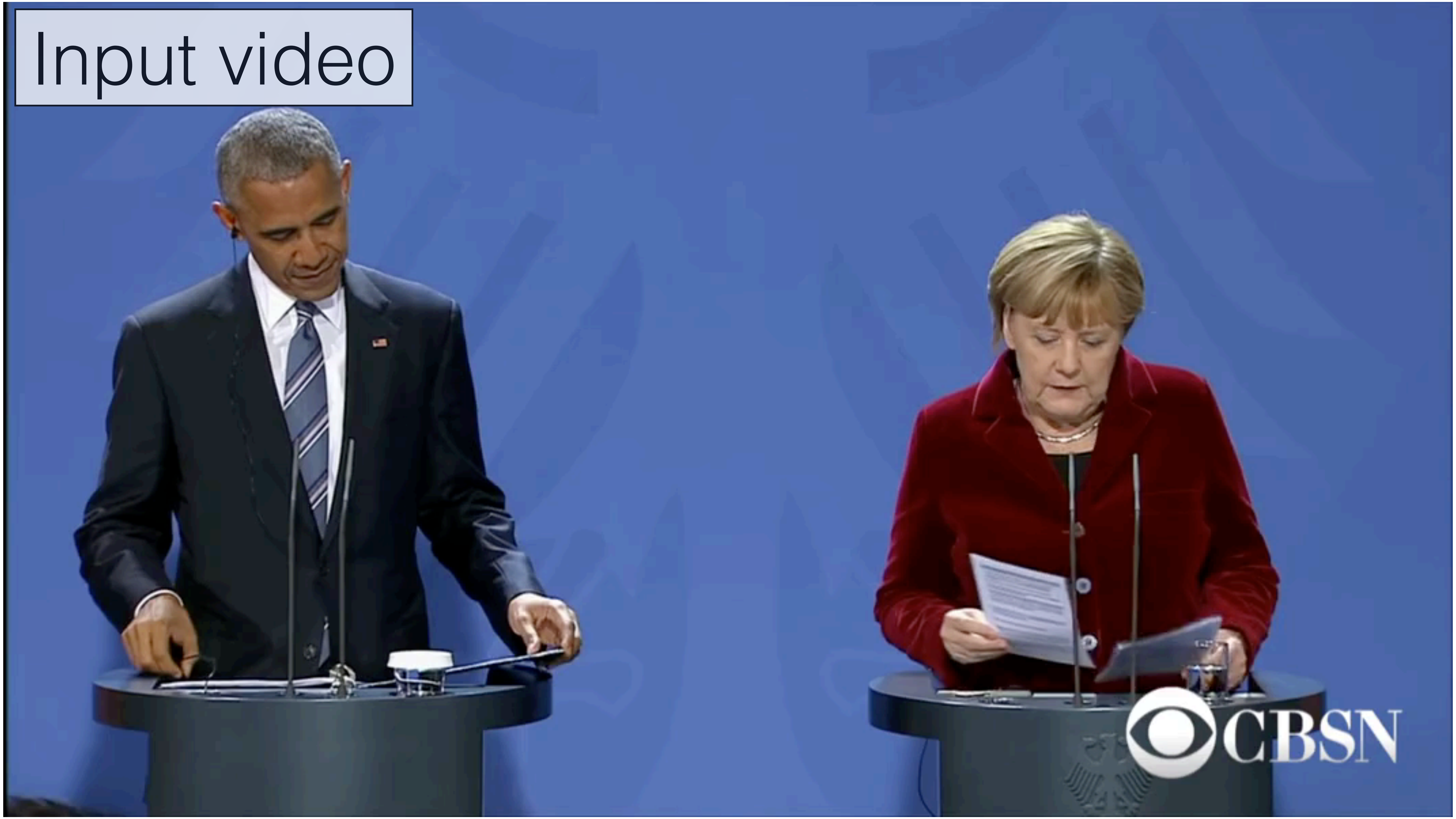
WiFi

(Zhao et al. 2019)

Touch

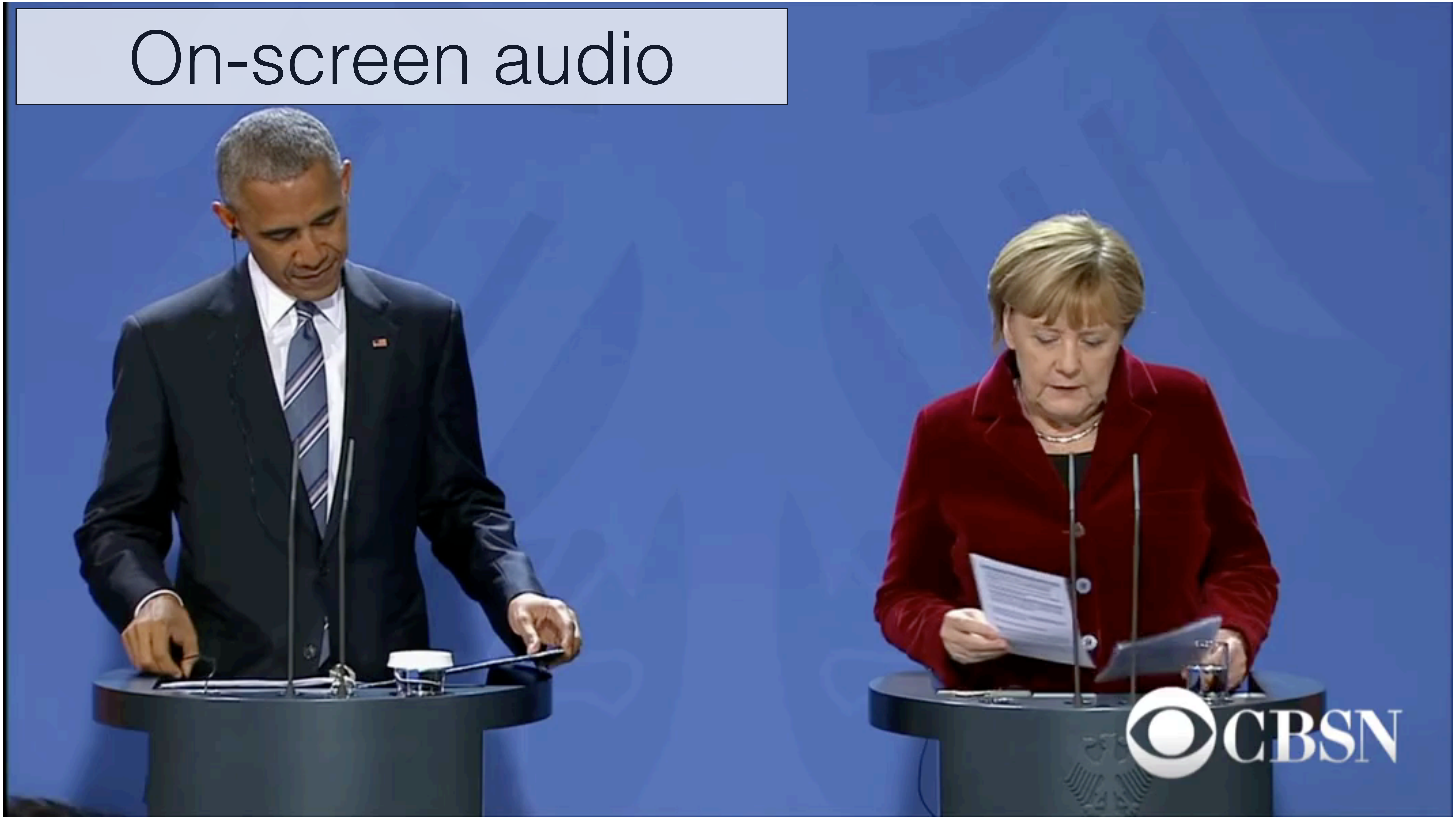


(Calandra et al. 2018)



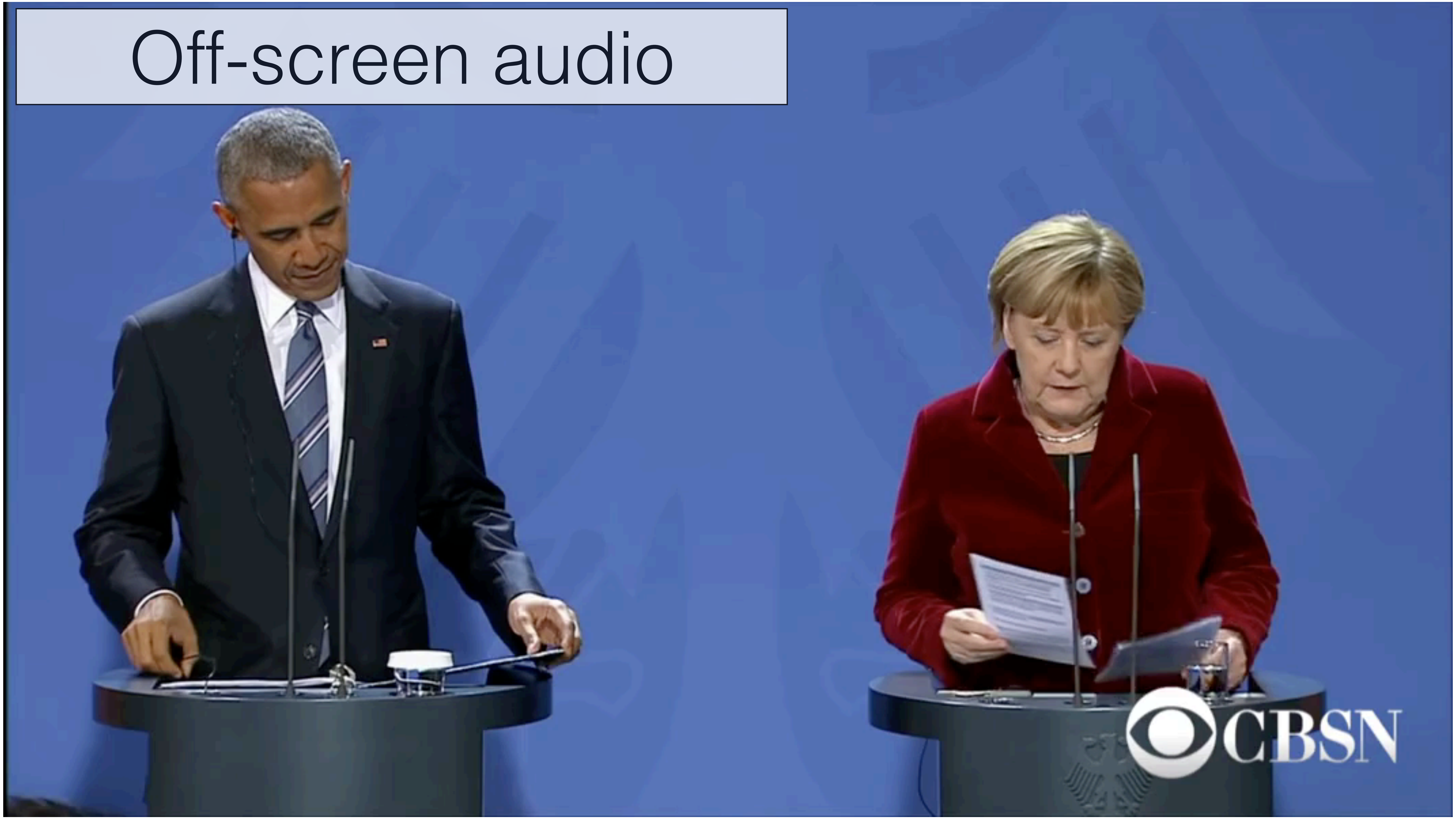
(Owens and Efros 2018)

On-screen audio



(Owens and Efros 2018)

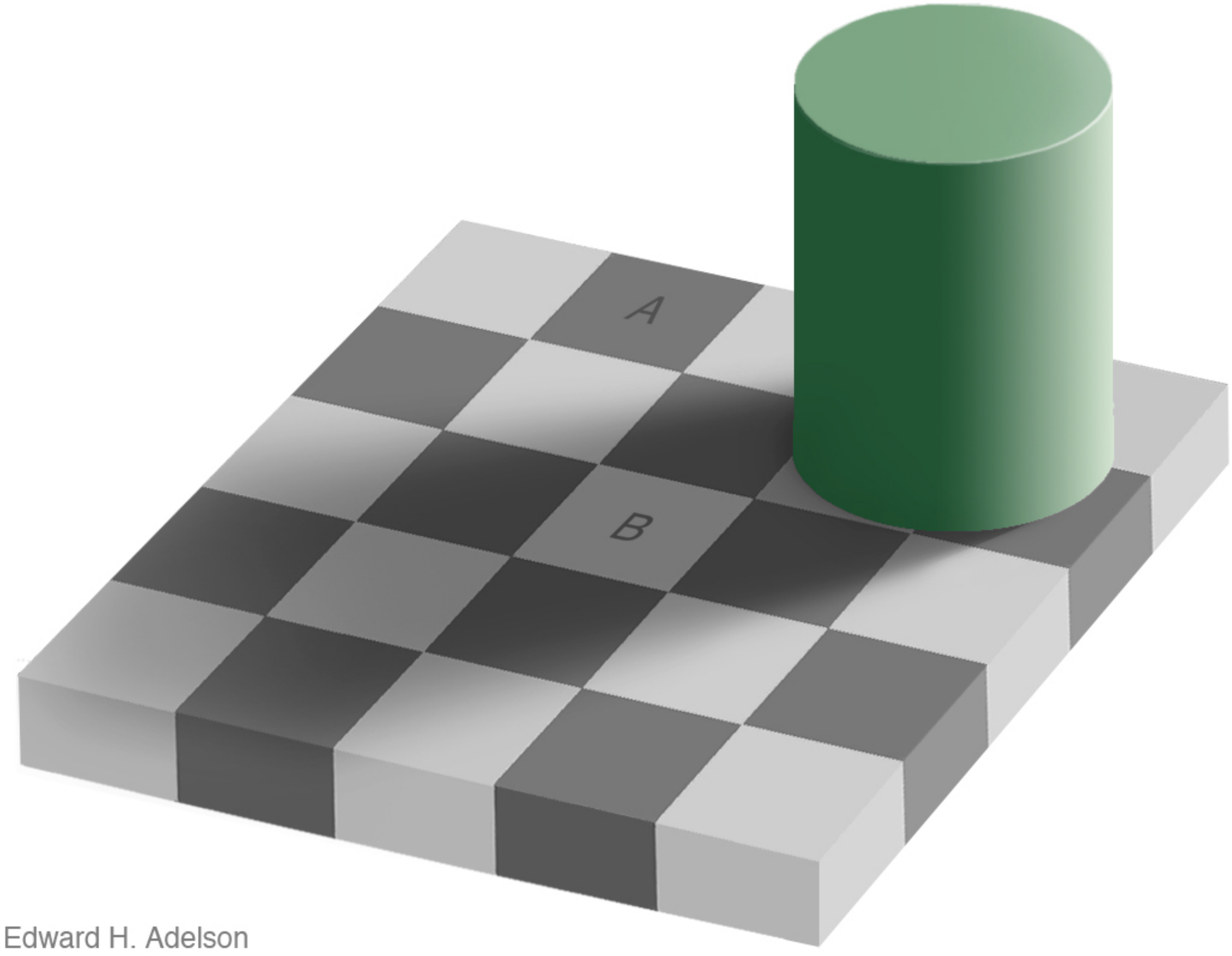
Off-screen audio



(Owens and Efros 2018)

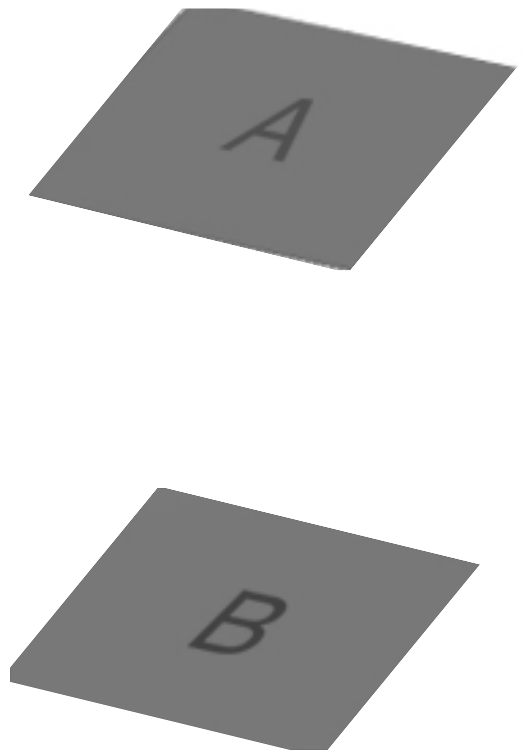
What makes vision hard?

To see: perception vs. measurement



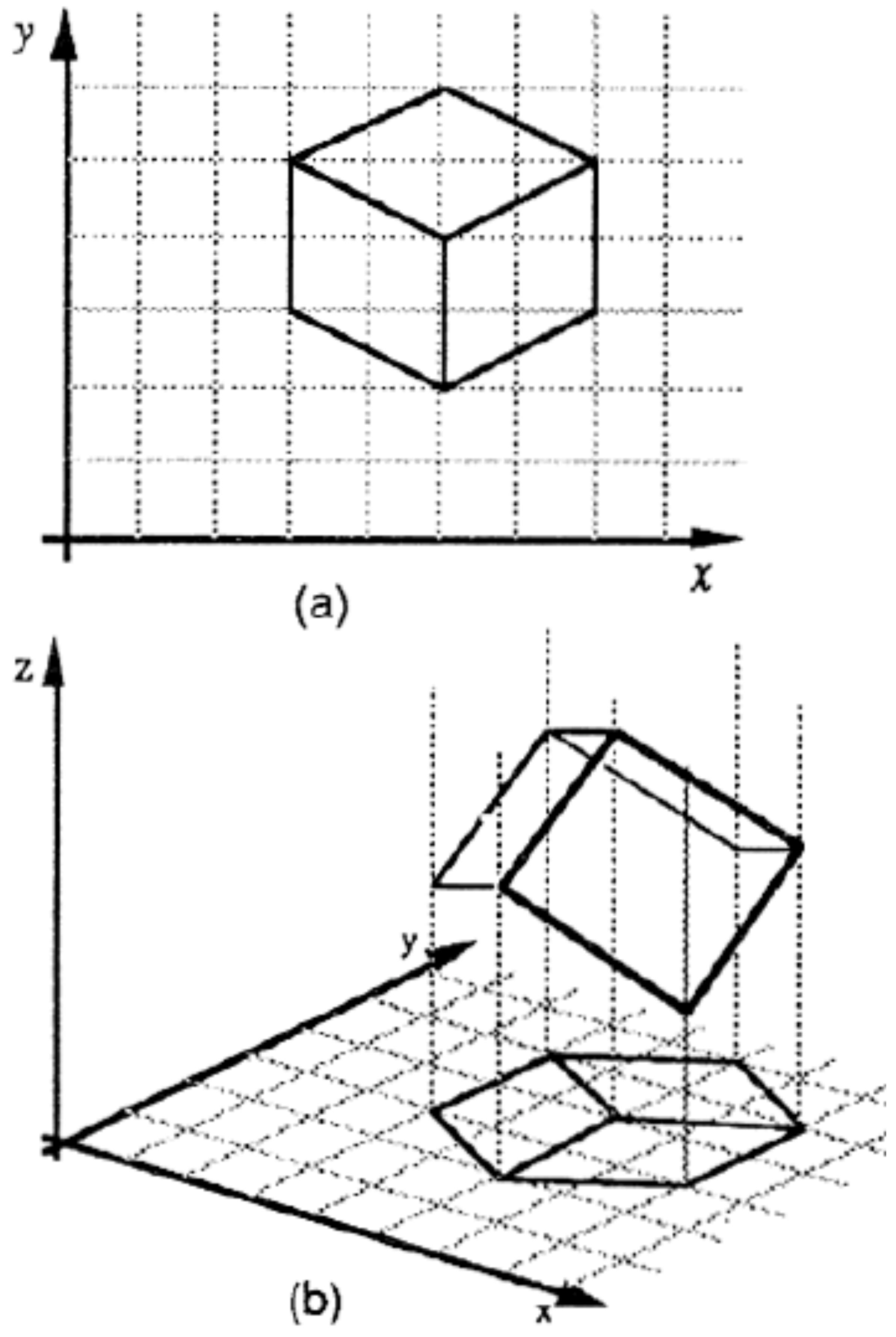
Slide credit: Torralba, Freeman, Isola

To see: perception vs. measurement



Slide credit: Torralba, Freeman, Isola

Other ambiguities



Sinha & Adelson 93

Slide credit: Antonio Torralba

Other ambiguities

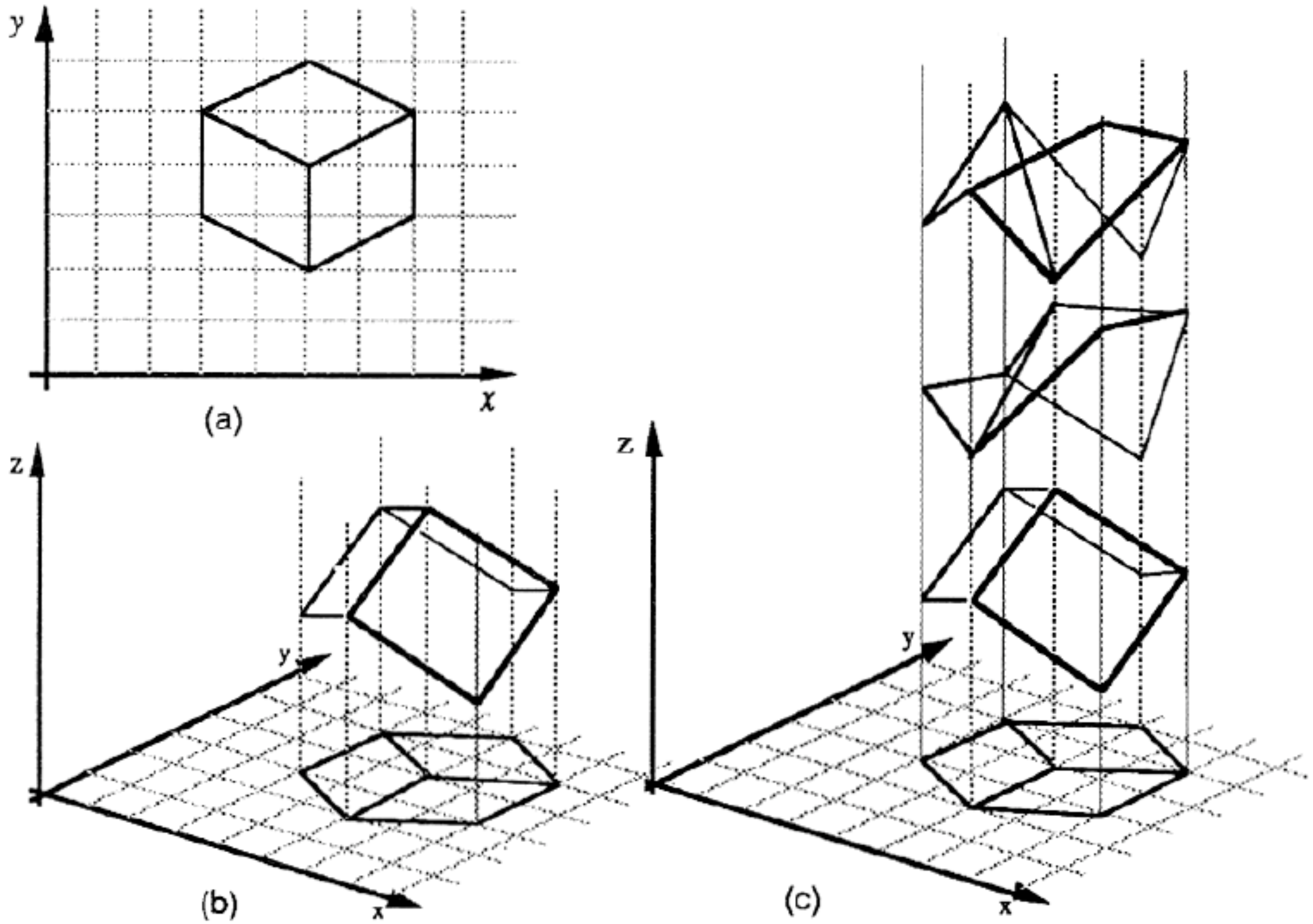


Figure 1. (a) A line drawing provides information only about the x, y coordinates of points lying along the object contours. (b) The human visual system is usually able to reconstruct an object in three dimensions given only a single 2D projection (c) Any planar line-drawing is geometrically consistent with infinitely many 3D structures.

Sinha & Adelson 93

Slide credit: Antonio Torralba

A simple visual system

- A simple world
- A simple image formation model
- A simple goal

A simple world

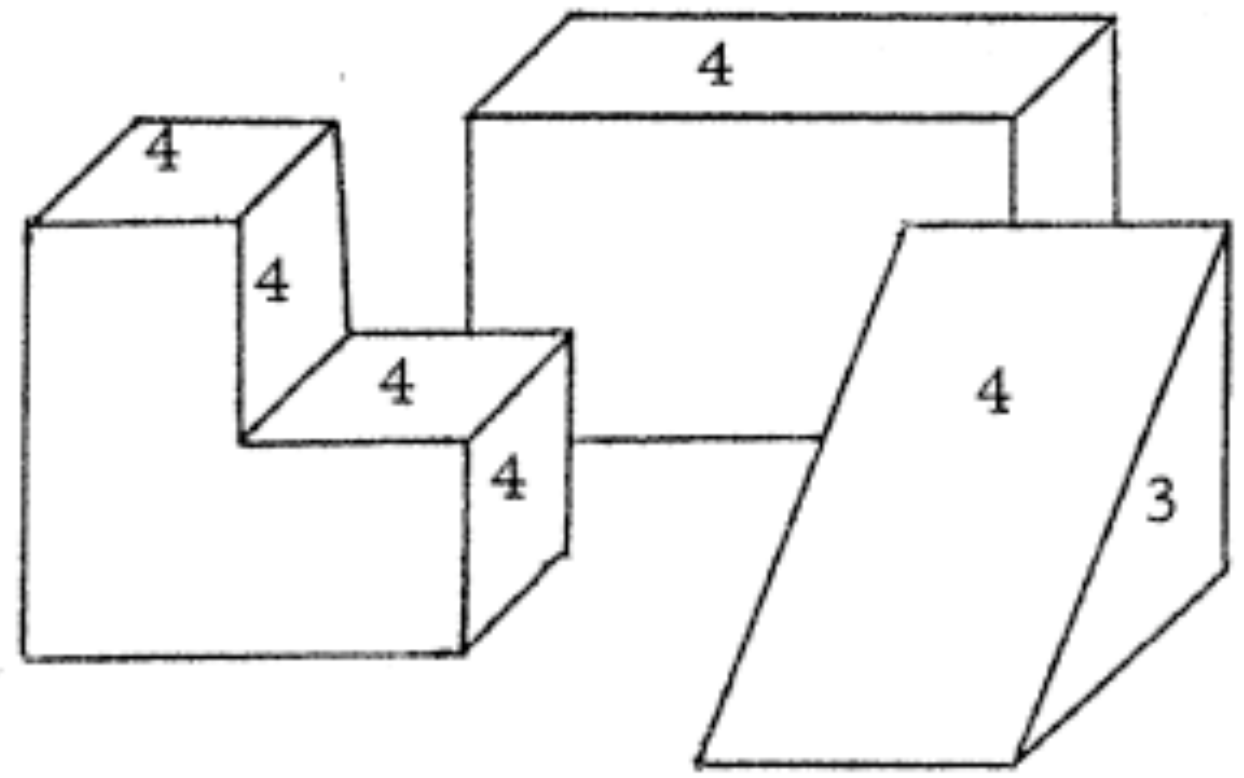
MACHINE PERCEPTION OF THREE-DIMENSIONAL SOLIDS

by

LAWRENCE GILMAN ROBERTS

Submitted to the Department of Electrical Engineering on May 10, 1963, in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

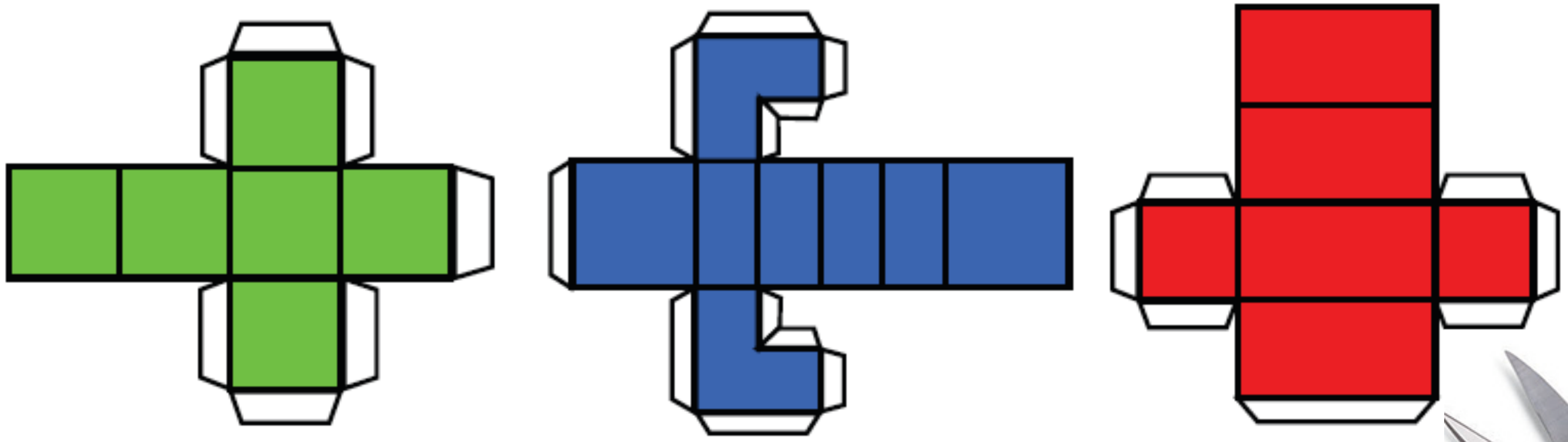
The problem of machine recognition of pictorial data has long been a challenging goal, but has seldom been attempted with anything more complex than alphabetic characters. Many people have felt that research on character recognition would be a first step, leading the way to a more general pattern recognition system. However, the multitudinous attempts at character recognition, including my own, have not led very far. The reason, I feel, is that the study of abstract, two-dimensional forms leads us away from, not toward, the techniques necessary for the recognition of three-dimensional objects. The per-

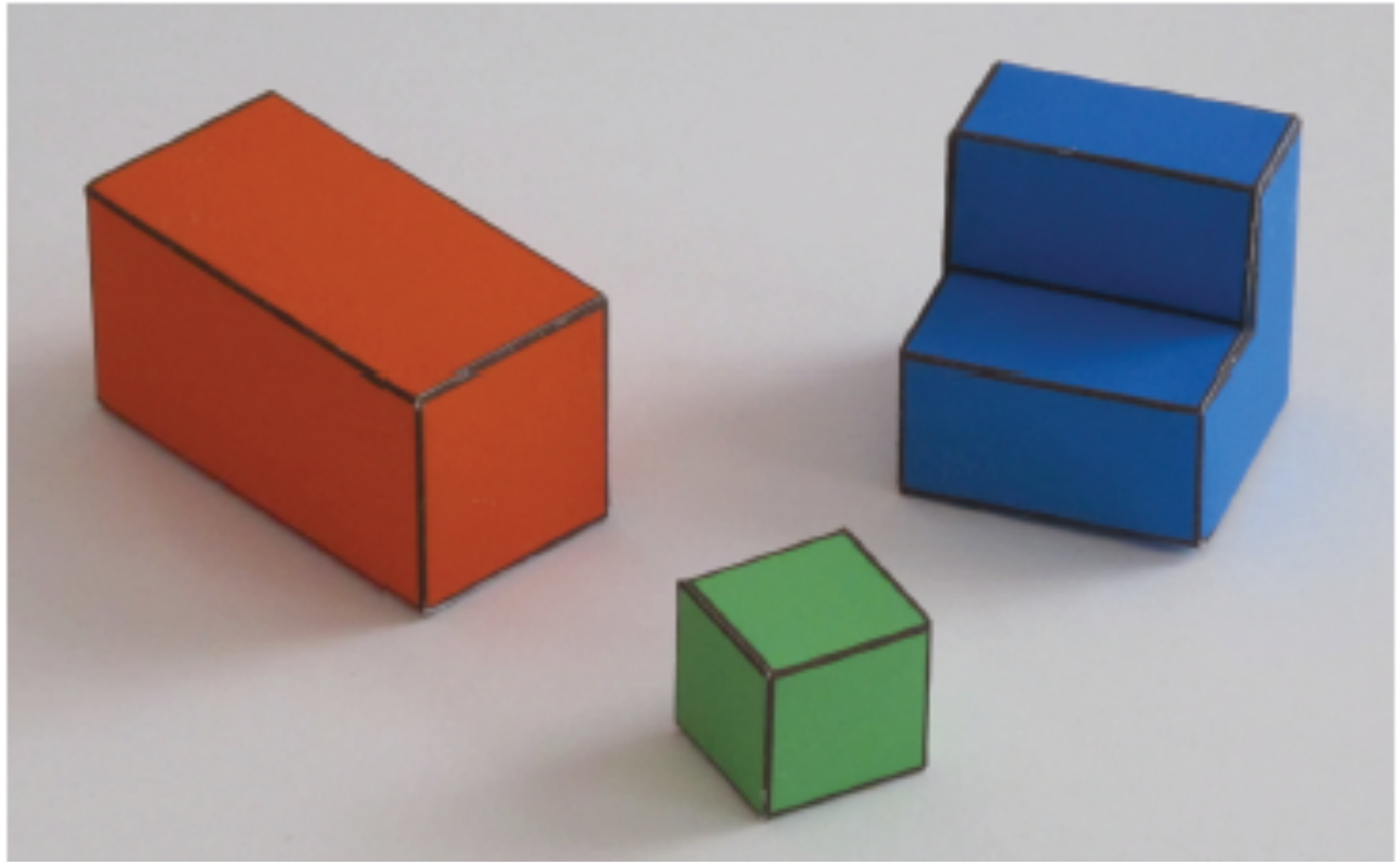


Complete Convex Polygons. The polygon selection procedure would select the numbered polygons as complete and convex. The number indicates the probable number of sides. A polygon is incomplete if one of its points is a collinear joint of another polygon.

Slide credit: Antonio Torralba

A simple world

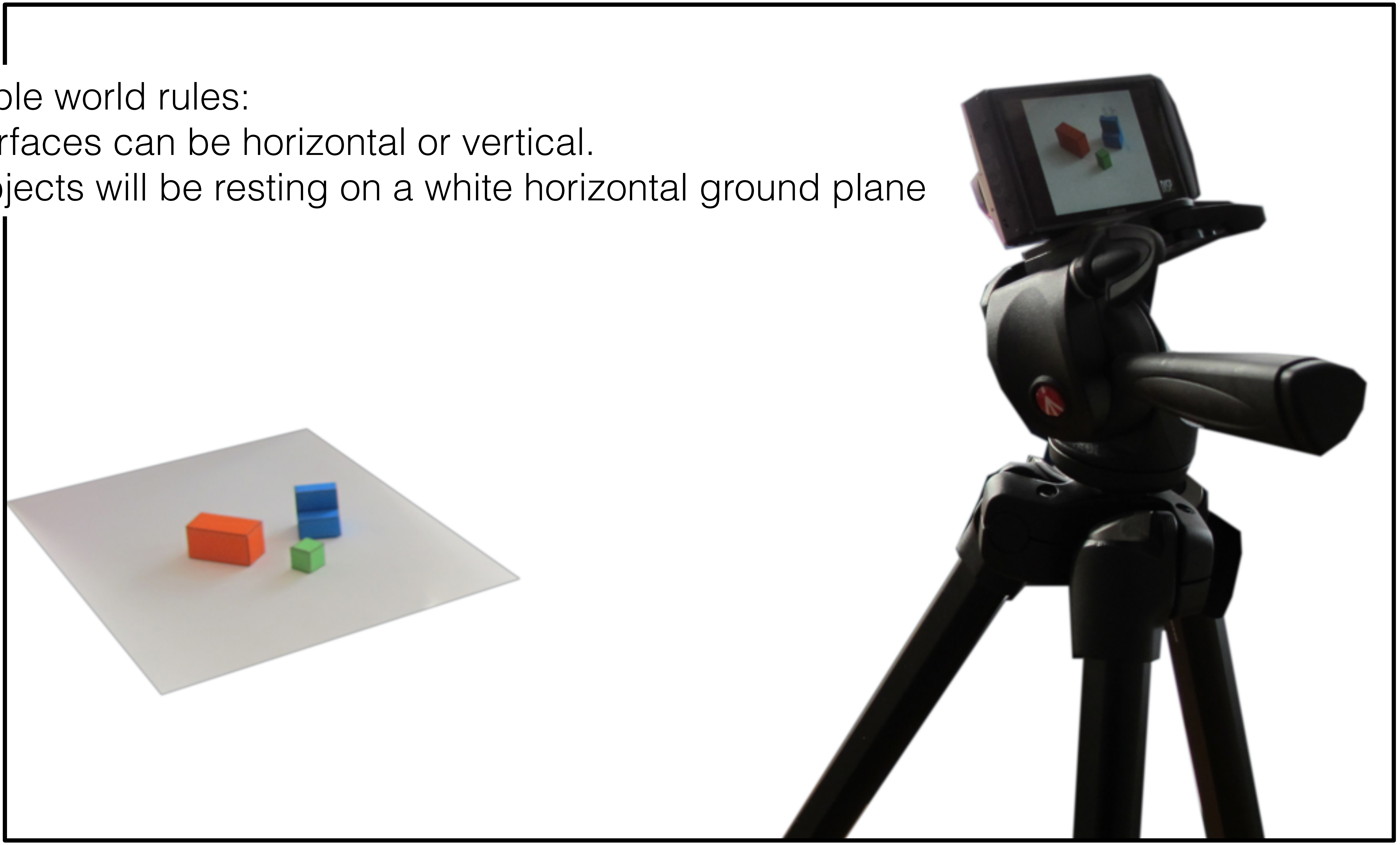




Slide credit: Antonio Torralba

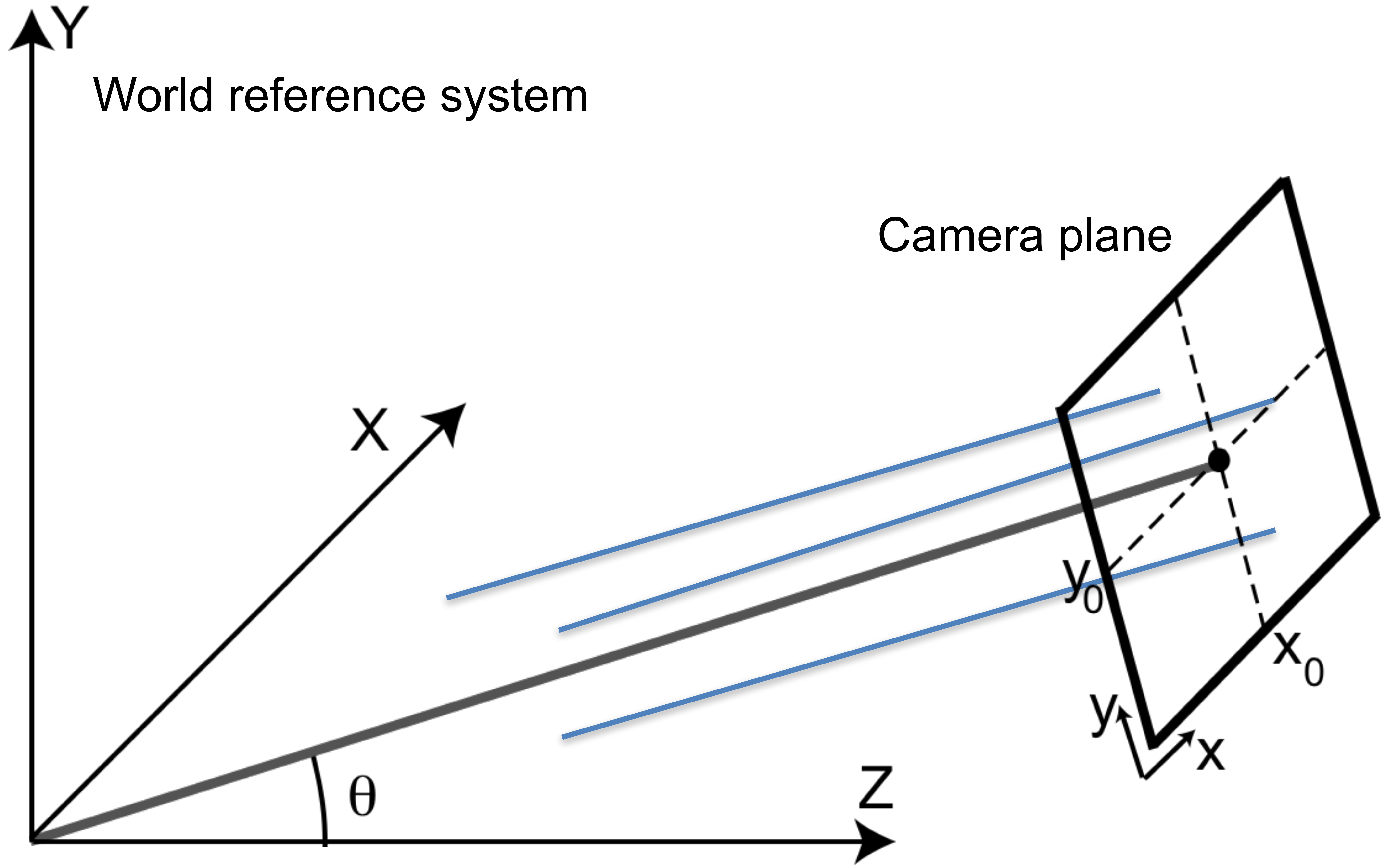
A simple image formation model

Simple world rules:
• Surfaces can be horizontal or vertical.
• Objects will be resting on a white horizontal ground plane



Slide credit: Antonio Torralba

A simple image formation model



Slide credit: Antonio Torralba

A simple image formation model

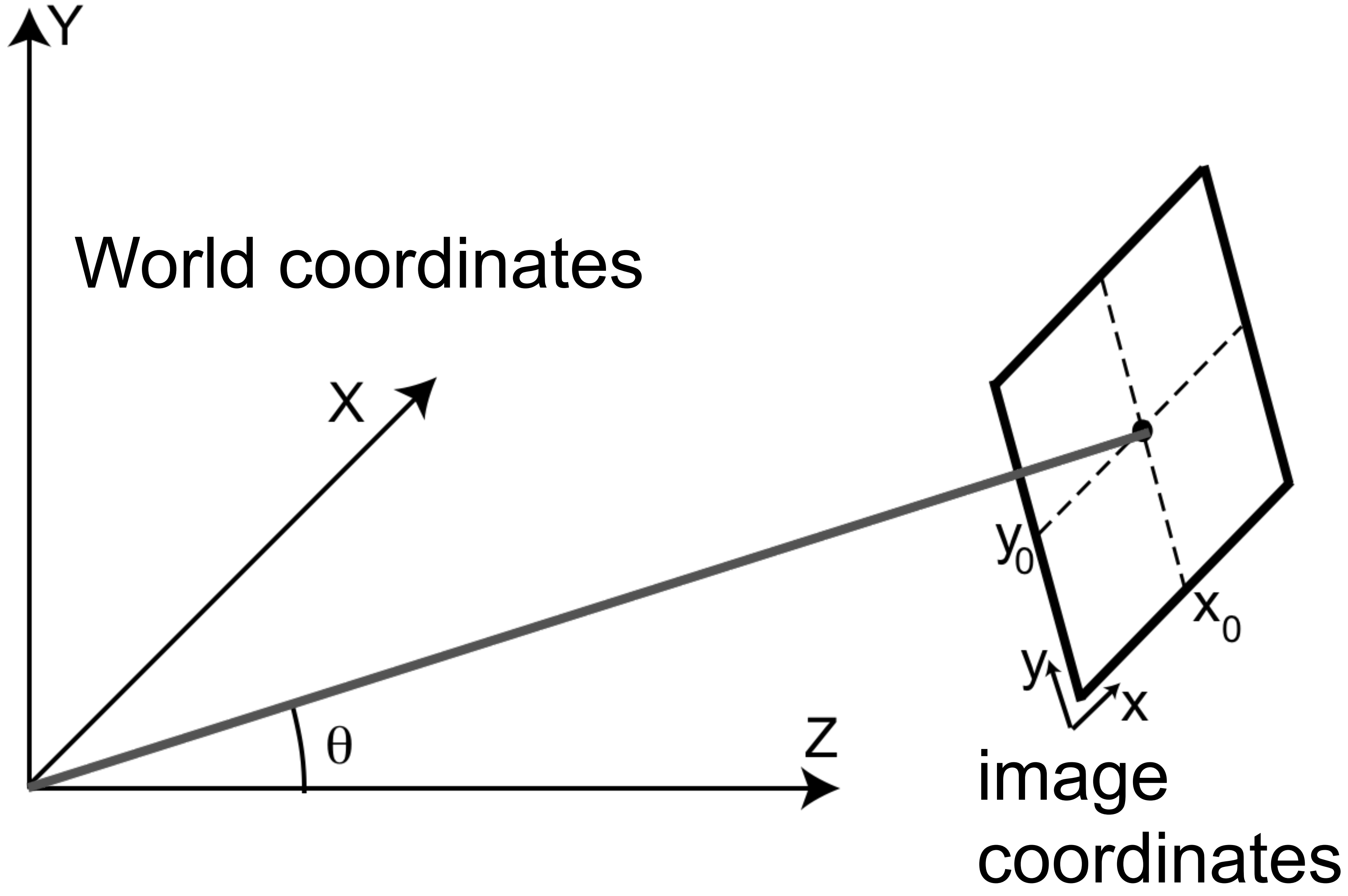
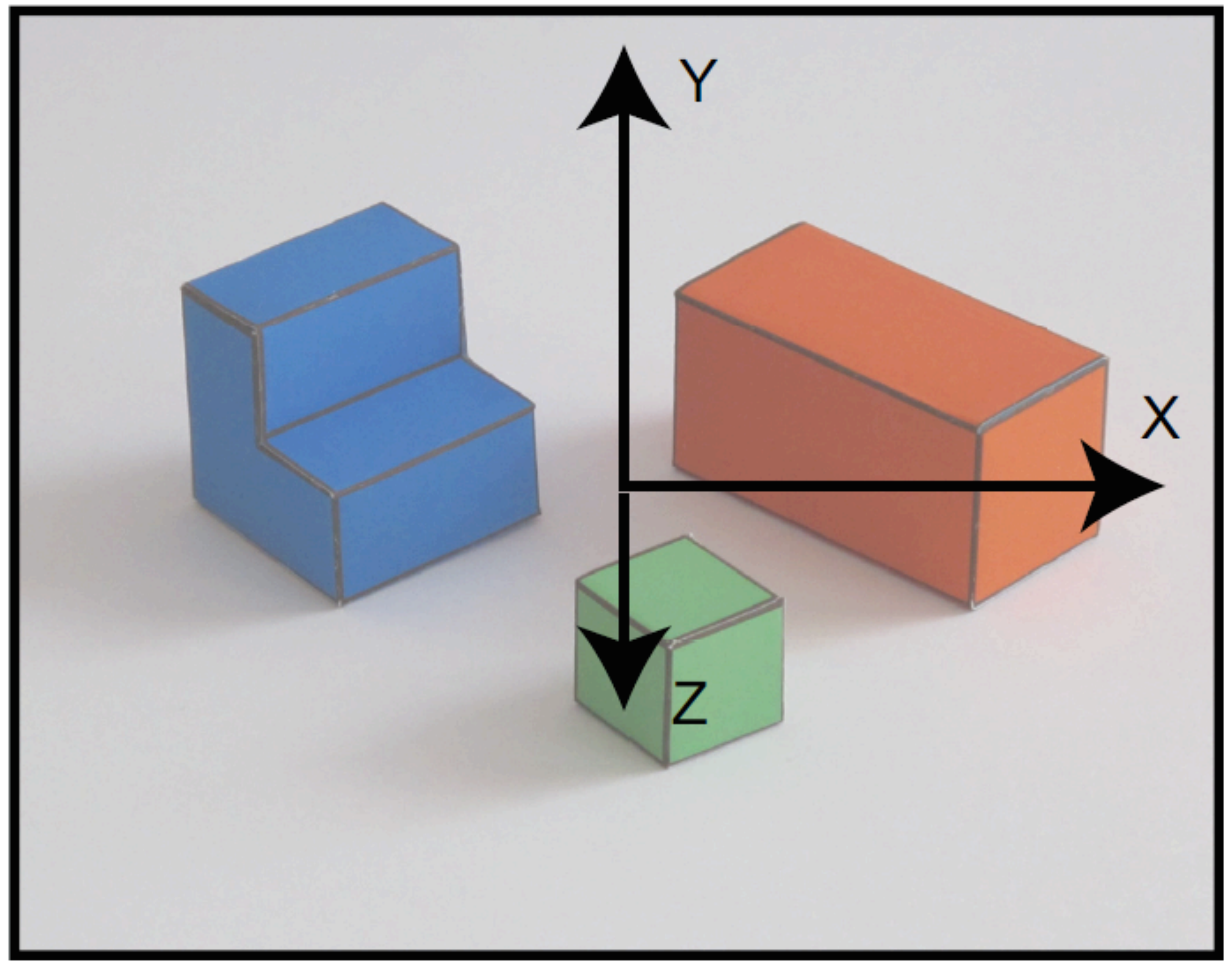


Image and projection of the world coordinate axes into the image plane



World coordinates

$$x = X + x_0$$

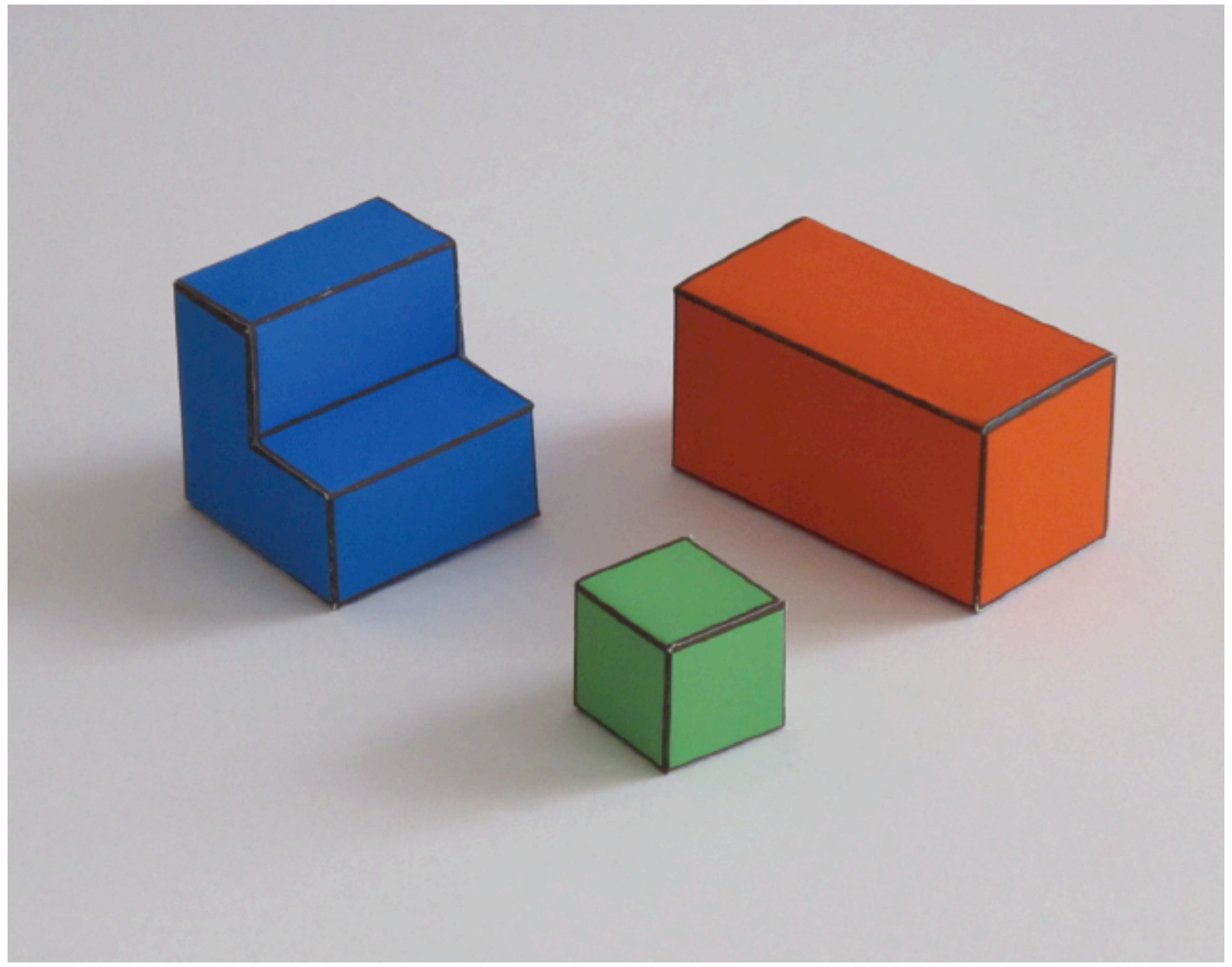
$$y = \cos(\theta)\,Y - \sin(\theta)\,Z + y_0$$

image coordinates

Slide credit: Antonio Torralba

A simple goal

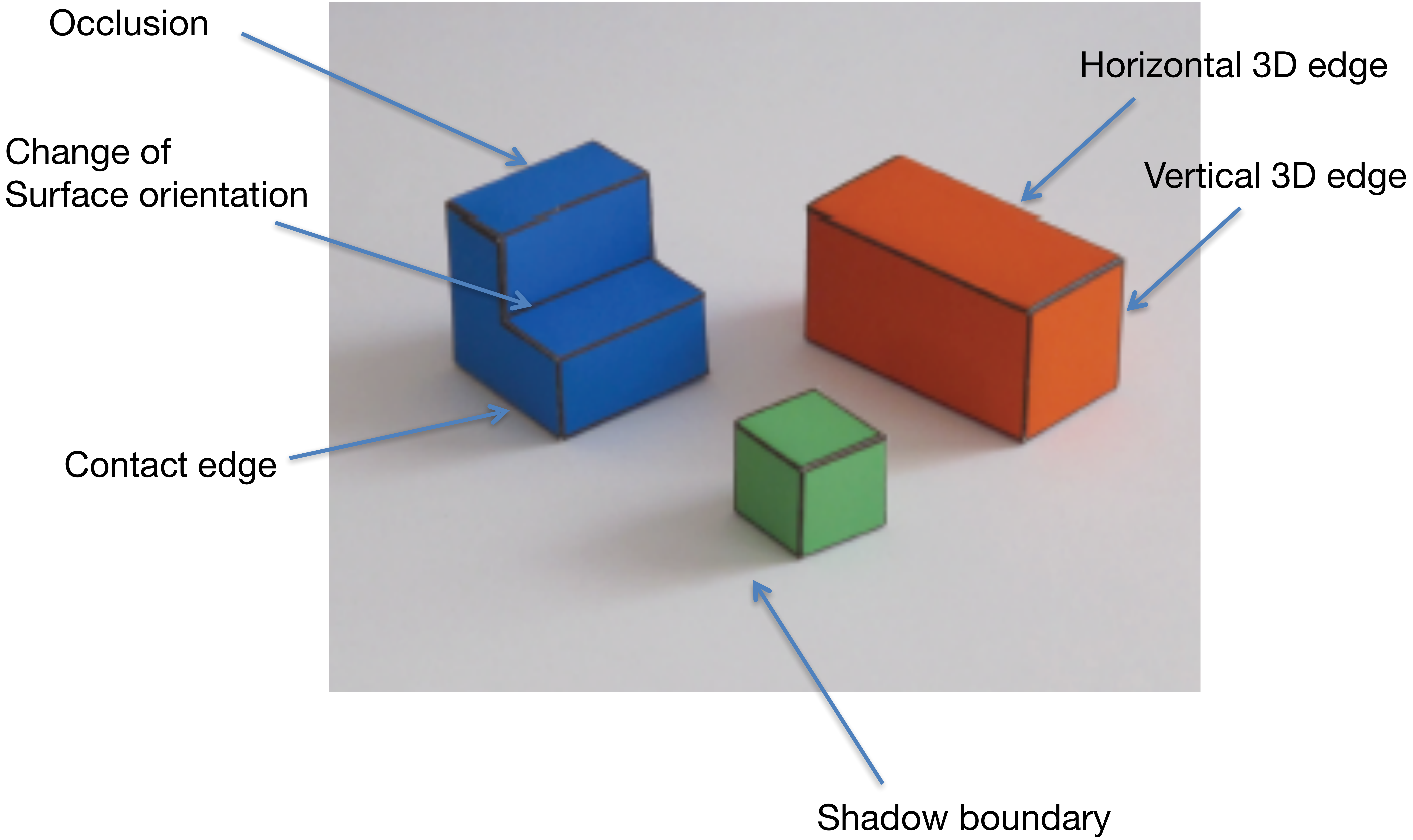
Recover the 3D structure of the world



We want to recover X(x,y), Y(x,y), Z(x,y) using as input I(x,y)

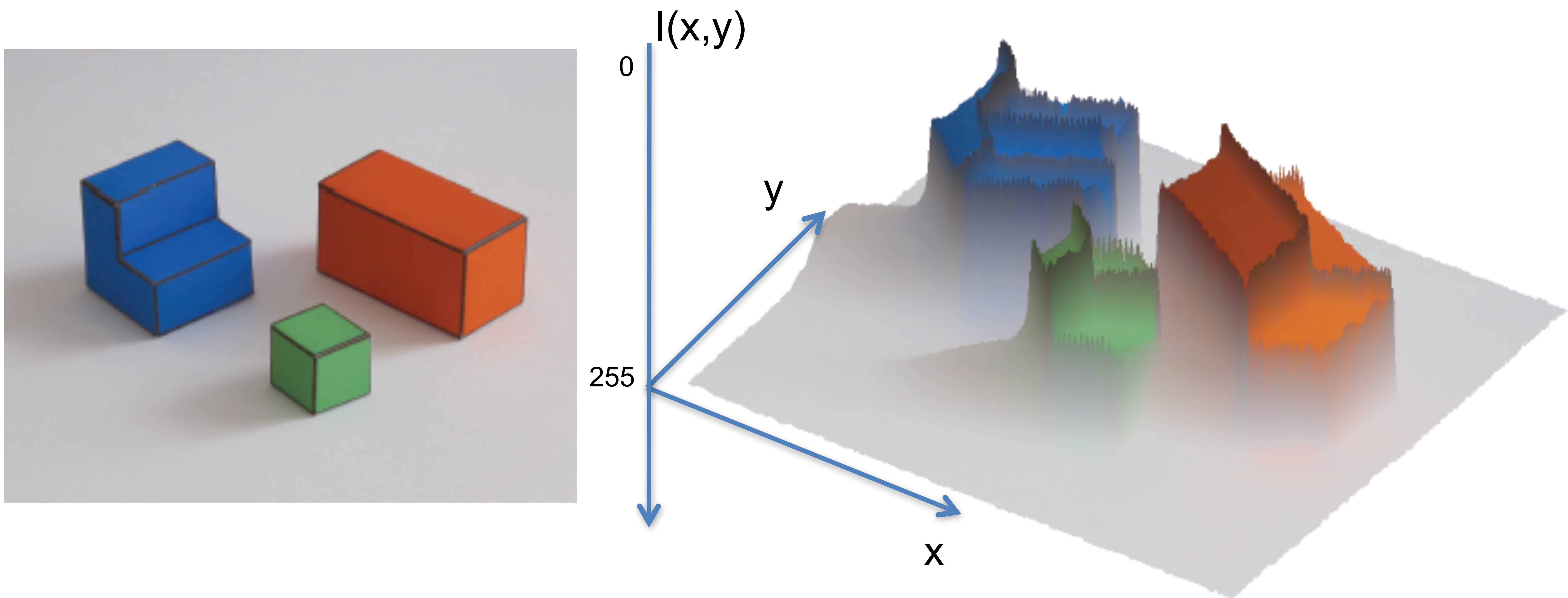
Slide credit: Antonio Torralba

Edges



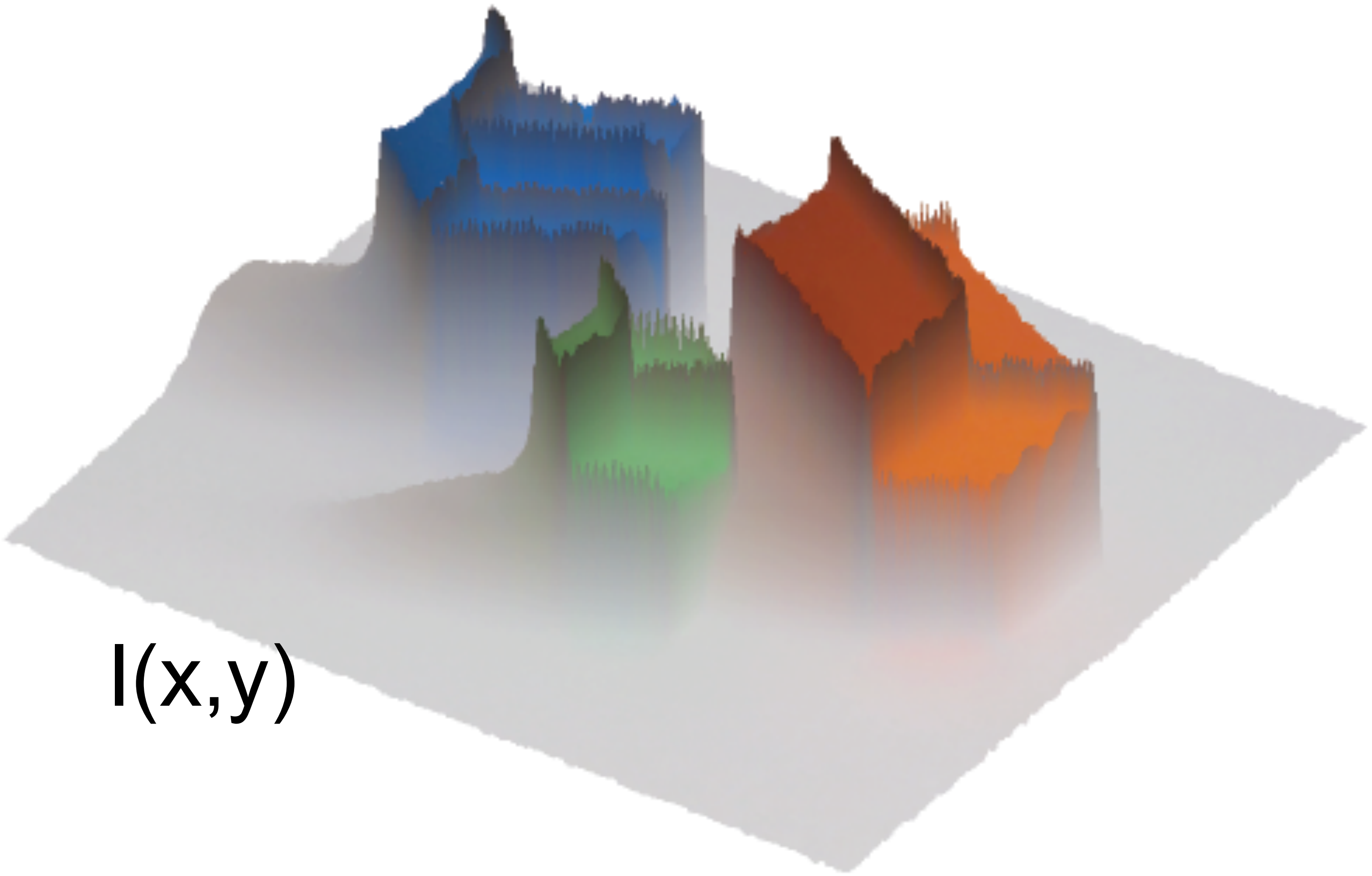
Slide credit: Antonio Torralba

Treating the image as a function



Slide credit: Antonio Torralba

Finding edges in the image



I(x,y)

Image gradient:

$$\nabla \mathbf{I} = \left(\frac{\partial \mathbf{I}}{\partial x}, \frac{\partial \mathbf{I}}{\partial y}\right)$$

Approximation image derivative:

$$\frac{\partial \mathbf{I}}{\partial x} \simeq \mathbf{I}(x,y) - \mathbf{I}(x-1,y)$$

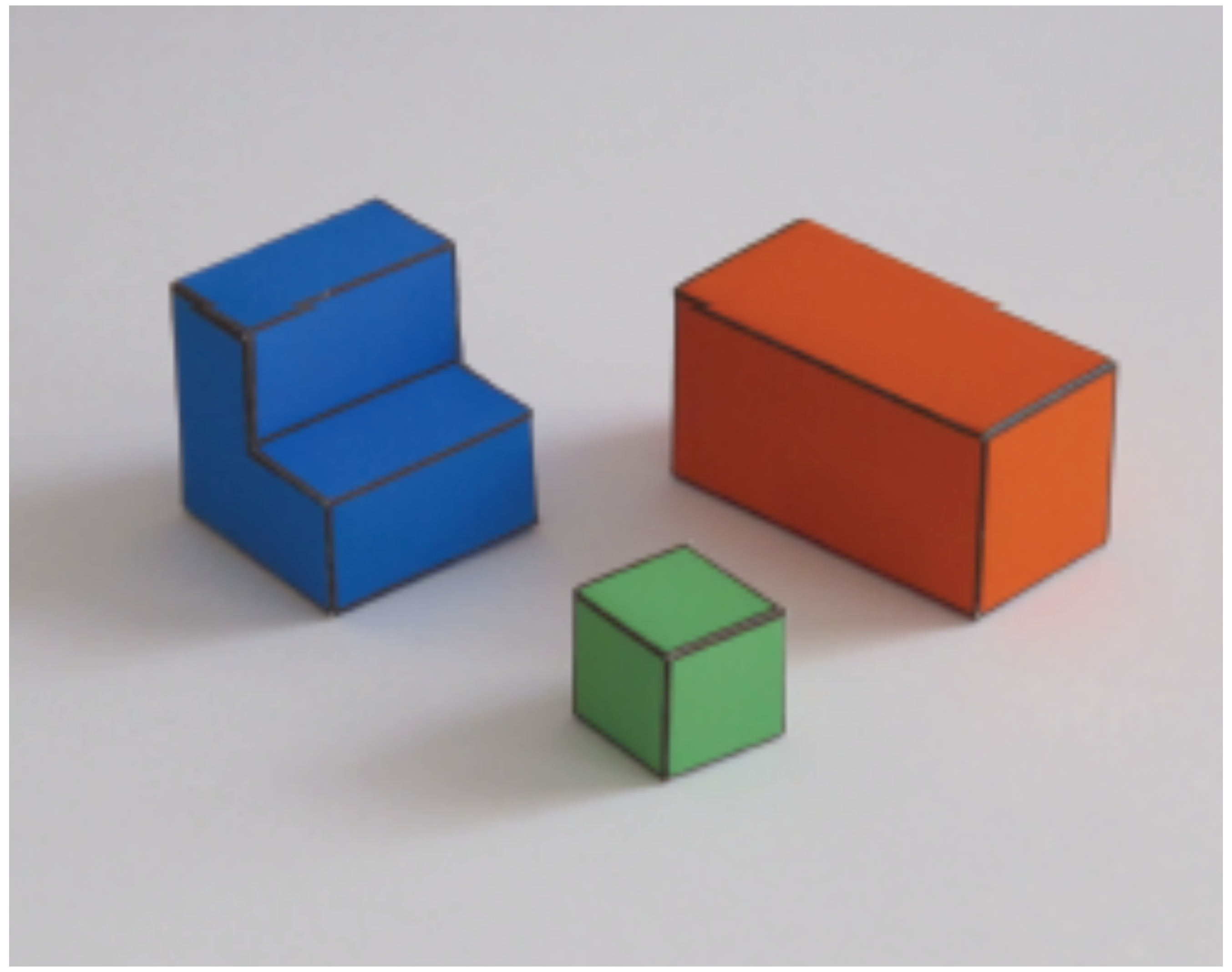
Edge strength $E(x,y) = |\nabla \mathbf{I}(x,y)|$

Edge orientation: $\theta(x,y) = \angle \nabla \mathbf{I} = \arctan \frac{\partial \mathbf{I}/\partial y}{\partial \mathbf{I}/\partial x}$

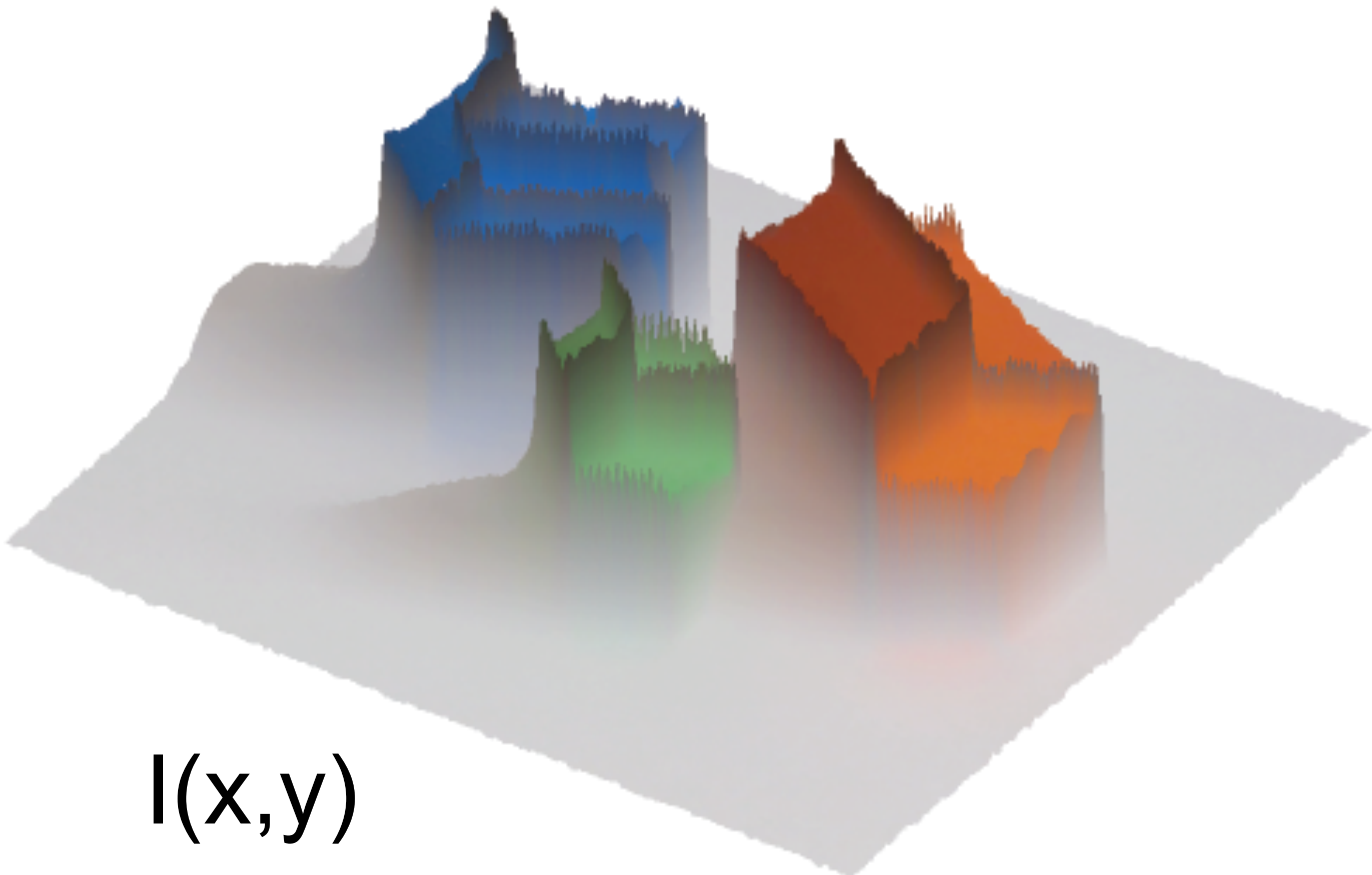
Edge normal: $\mathbf{n} = \frac{\nabla \mathbf{I}}{|\nabla \mathbf{I}|}$

Slide credit: Antonio Torralba

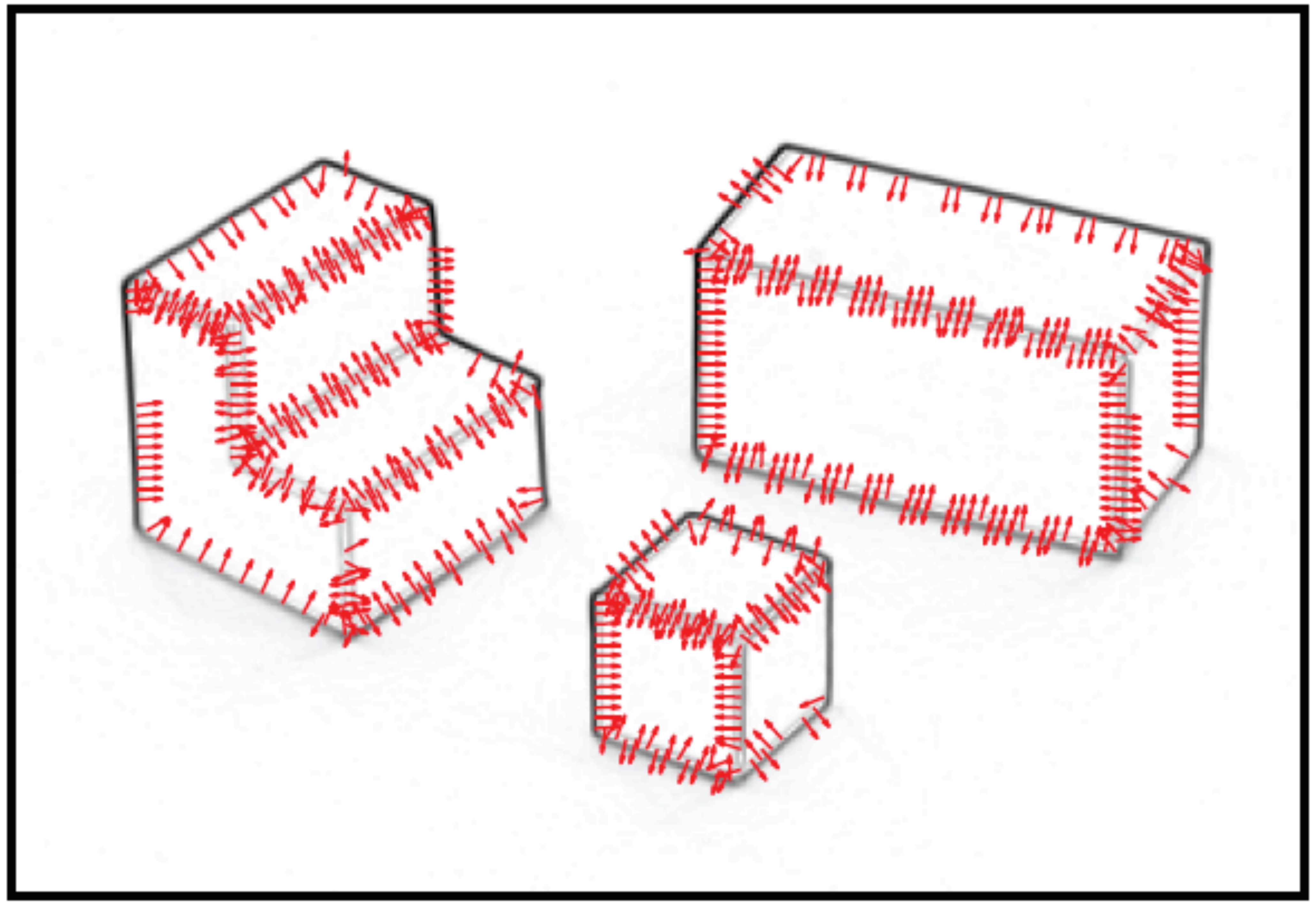
Finding edges in the image



$$\nabla \mathbf{I} = \left(\frac{\partial \mathbf{I}}{\partial x}, \frac{\partial \mathbf{I}}{\partial y}\right) \qquad \mathbf{n} = \frac{\nabla \mathbf{I}}{|\nabla \mathbf{I}|}$$



I(x,y)

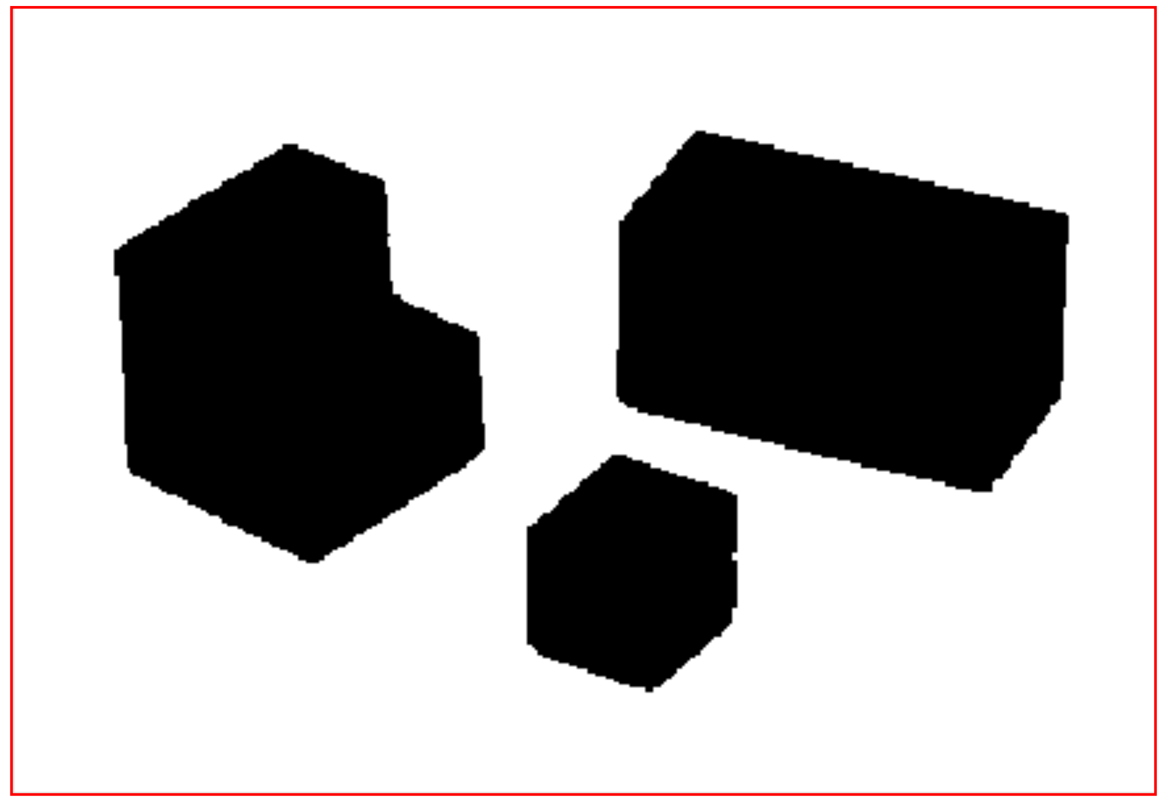


E(x,y) and n(x,y)

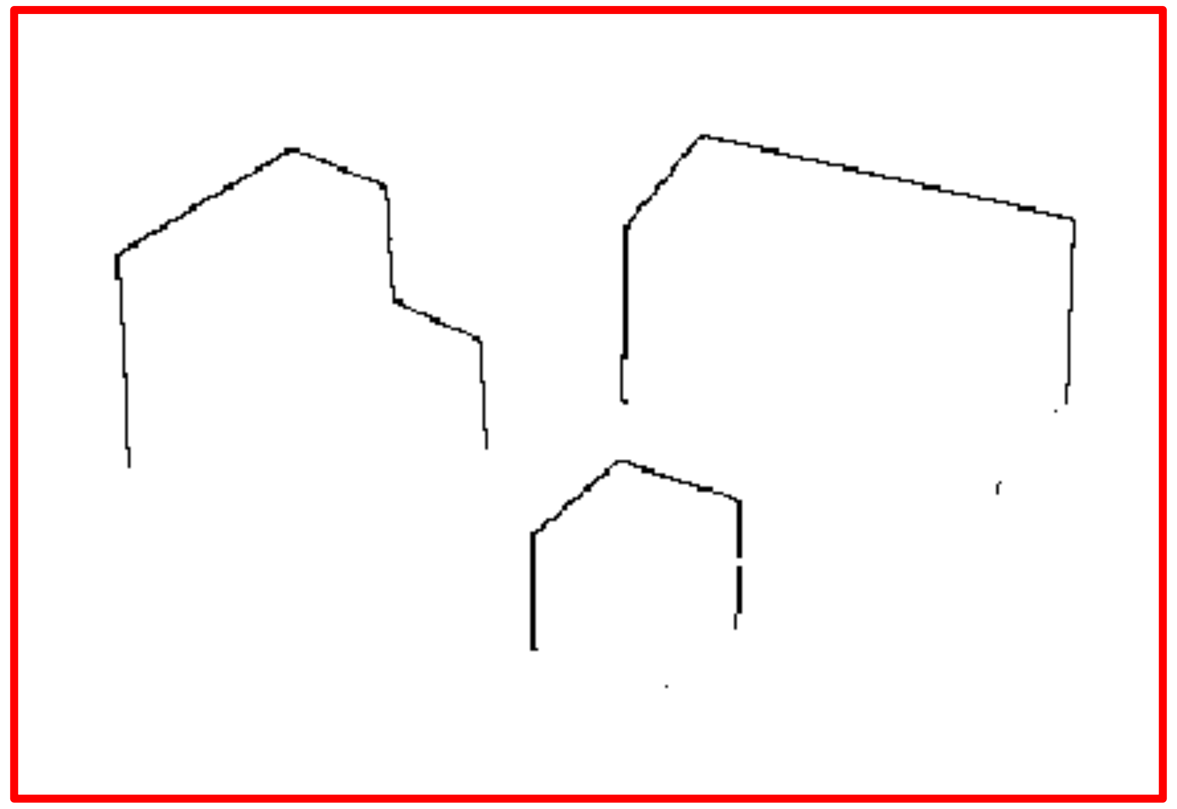
Slide credit: Antonio Torralba

Edge classification

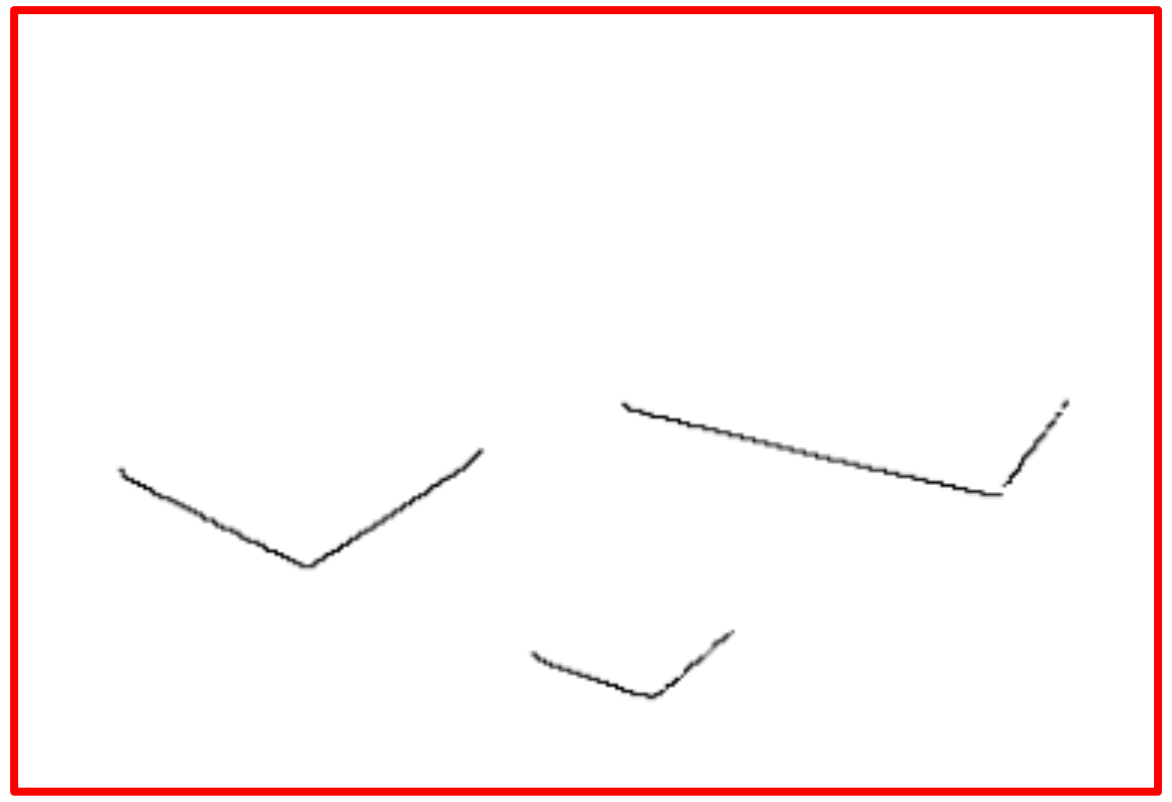
- Figure/ground segmentation
  - Using the fact that objects have color



- Occlusion edges
  - Occlusion edges are owned by the foreground

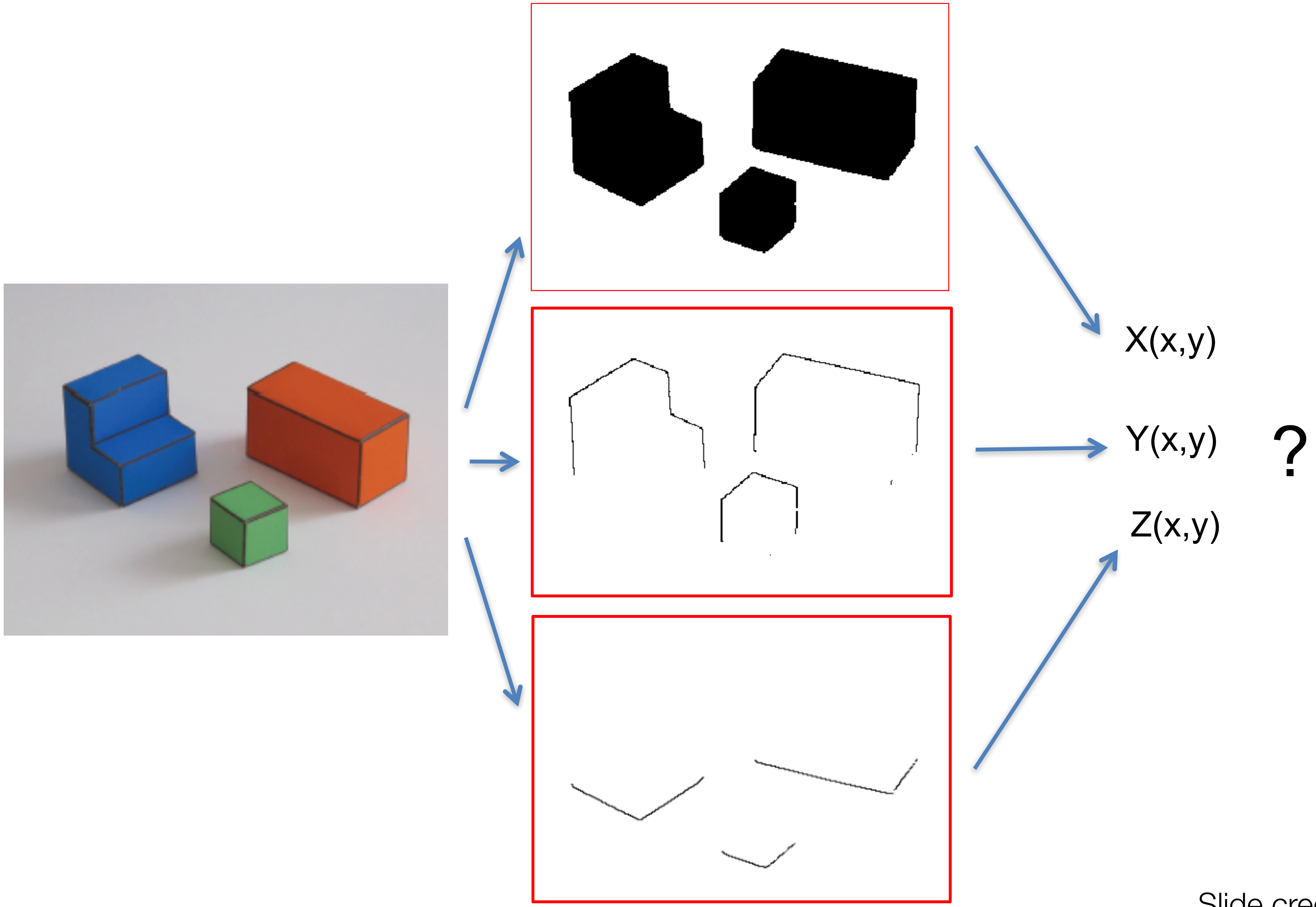


- Contact edges



Slide credit: Antonio Torralba

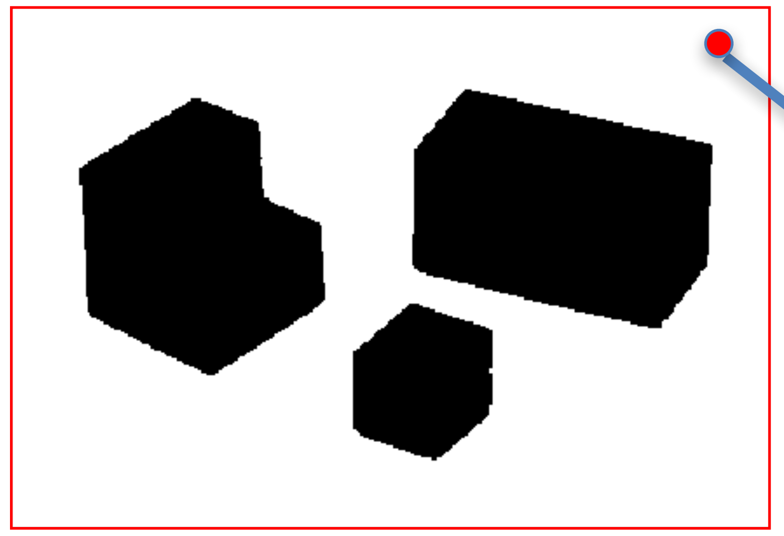
From edges to surface constraints



Slide credit: Antonio Torralba

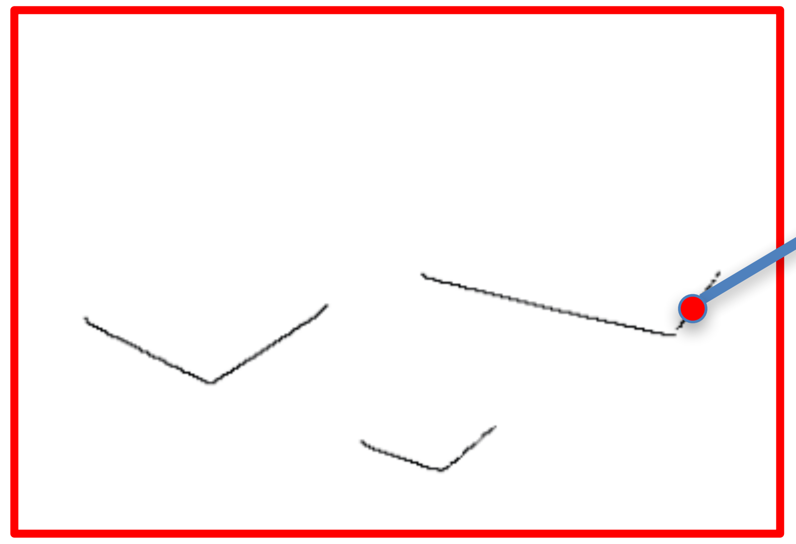
From edges to surface constraints

- Ground



Y(x,y) = 0 if (x,y) belongs to a ground pixel

- Contact edge



Y(x,y) = 0 if (x,y) belongs to foreground and is a contact edge

- What happens inside the objects?

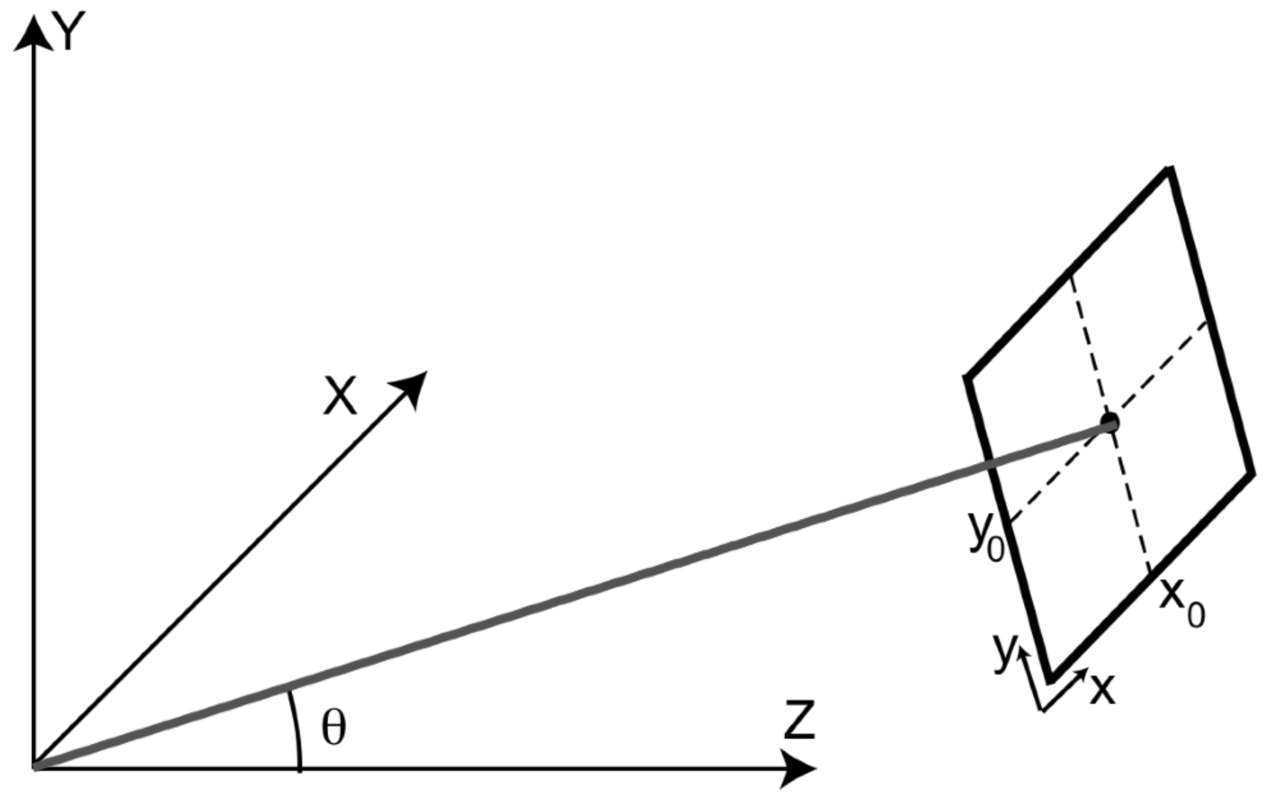
… now things get a bit more complicated.

Slide credit: Antonio Torralba

From edges to surface constraints

How can we relate the information in the pixels with 3D surfaces in the world?

Vertical edges



World coordinates

$$x = X + x_0$$

$$y = \cos(\theta)\, Y - \sin(\theta)\, Z + y_0$$

image coordinates

Given the image, what can we say about X, Y and Z in the pixels that belong to a vertical edge?

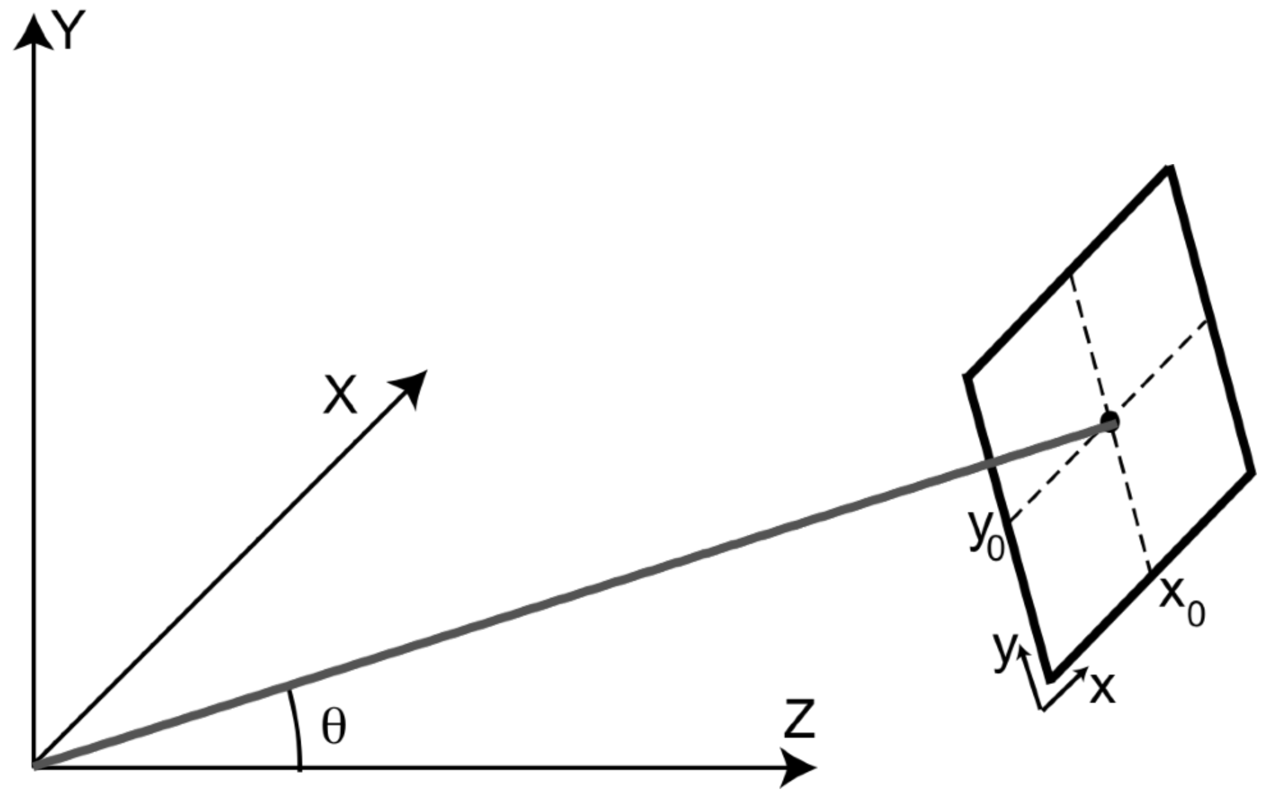
Z = constant along the edge

$$\partial Y / \partial y = 1/\cos(\theta)$$

Slide credit: Antonio Torralba

From edges to surface constraints

- Horizontal edges



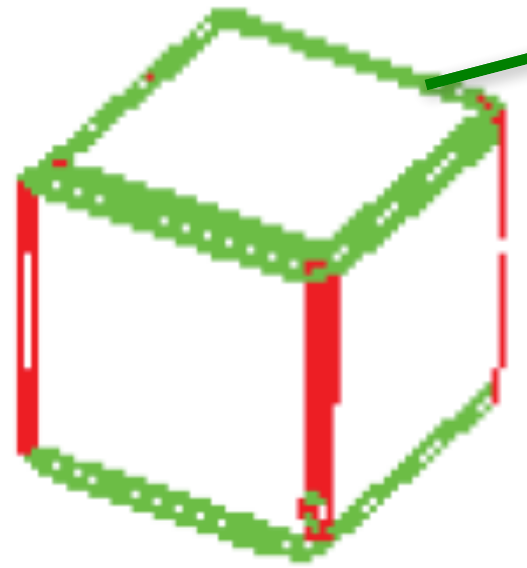
World coordinates

$$x = X + x_0$$

$$y = \cos(\theta)\, Y - \sin(\theta)\, Z + y_0$$

image coordinates

Given the image, what can we say about X, Y and Z in the pixels that belong to an horizontal 3D edge?



Y = constant along the edge

$$\partial Y / \partial \mathbf{t} \quad = \quad 0$$

Where **t** is the vector parallel to the edge

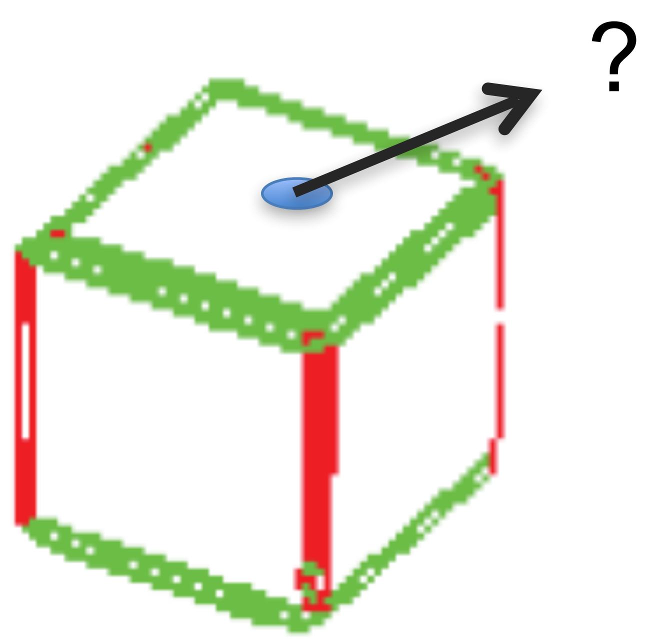
$$\mathbf{t} = (-n_y, n_x)$$

$$\partial Y / \partial \mathbf{t} = -n_y \partial Y / \partial x + n_x \partial Y / \partial y$$

Slide credit: Antonio Torralba

From edges to surface constraints

- What happens where there are no edges?



Assumption of planar faces:

$$
\begin{aligned}
\partial^2 Y/\partial x^2 &= 0 \\
\partial^2 Y/\partial y^2 &= 0 \\
\partial^2 Y/\partial y \partial x &= 0
\end{aligned}
$$

Information has to be propagated from the edges

Slide credit: Antonio Torralba

A simple inference scheme

All the constraints are linear!

$Y(x,y) = 0$ if (x,y) belongs to a ground pixel

$$\boxed{\partial Y/\partial y = 1/\cos(\theta)}$$ if (x,y) belongs to a vertical edge

$$\boxed{\partial Y/\partial \mathbf{t} = 0}$$ if (x,y) belongs to an horizontal edge

$$\boxed{\begin{aligned} \partial^2 Y/\partial x^2 &= 0 \\ \partial^2 Y/\partial y^2 &= 0 \\ \partial^2 Y/\partial y \partial x &= 0 \end{aligned}}$$ if (x,y) is not on an edge

A similar set of constraints could be derived for Z

Slide credit: Antonio Torralba

Discrete approximation

We can transform every differential constraint into a linear constraint on Y(x,y)

Y(x,y)

111	115	113	111	112	111	112	111
135	138	137	139	145	146	149	147
163	168	188	196	206	202	206	207
180	184	206	219	202	200	195	193
189	193	214	216	104	79	83	77
191	201	217	220	103	59	60	68
195	205	216	222	113	68	69	83
199	203	223	228	108	68	71	77

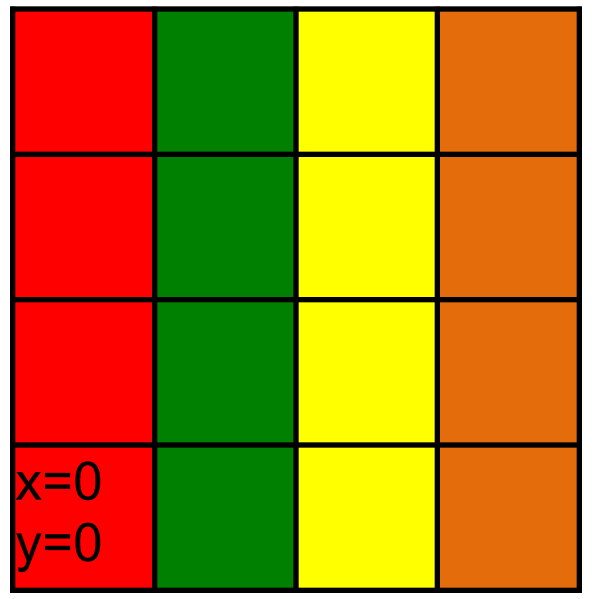
$$\frac{dY}{dx} \approx Y(x,y) - Y(x\text{-}1,y)$$

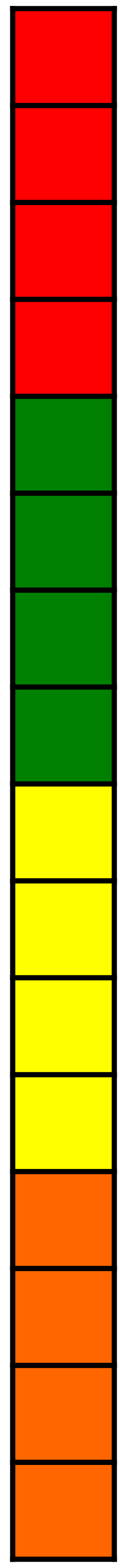
Slide credit: Antonio Torralba

Discrete approximation

Transform the “image” Y(x,y) into a column vector:

Y(x,y)

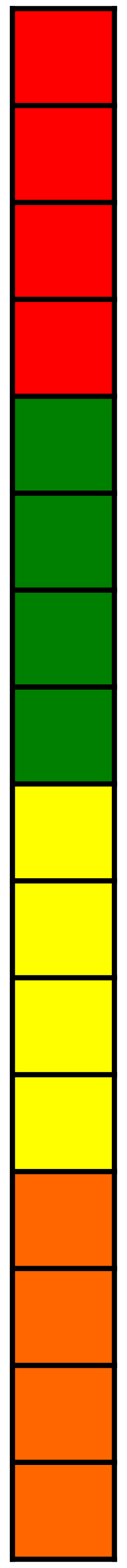




x=2, y=2

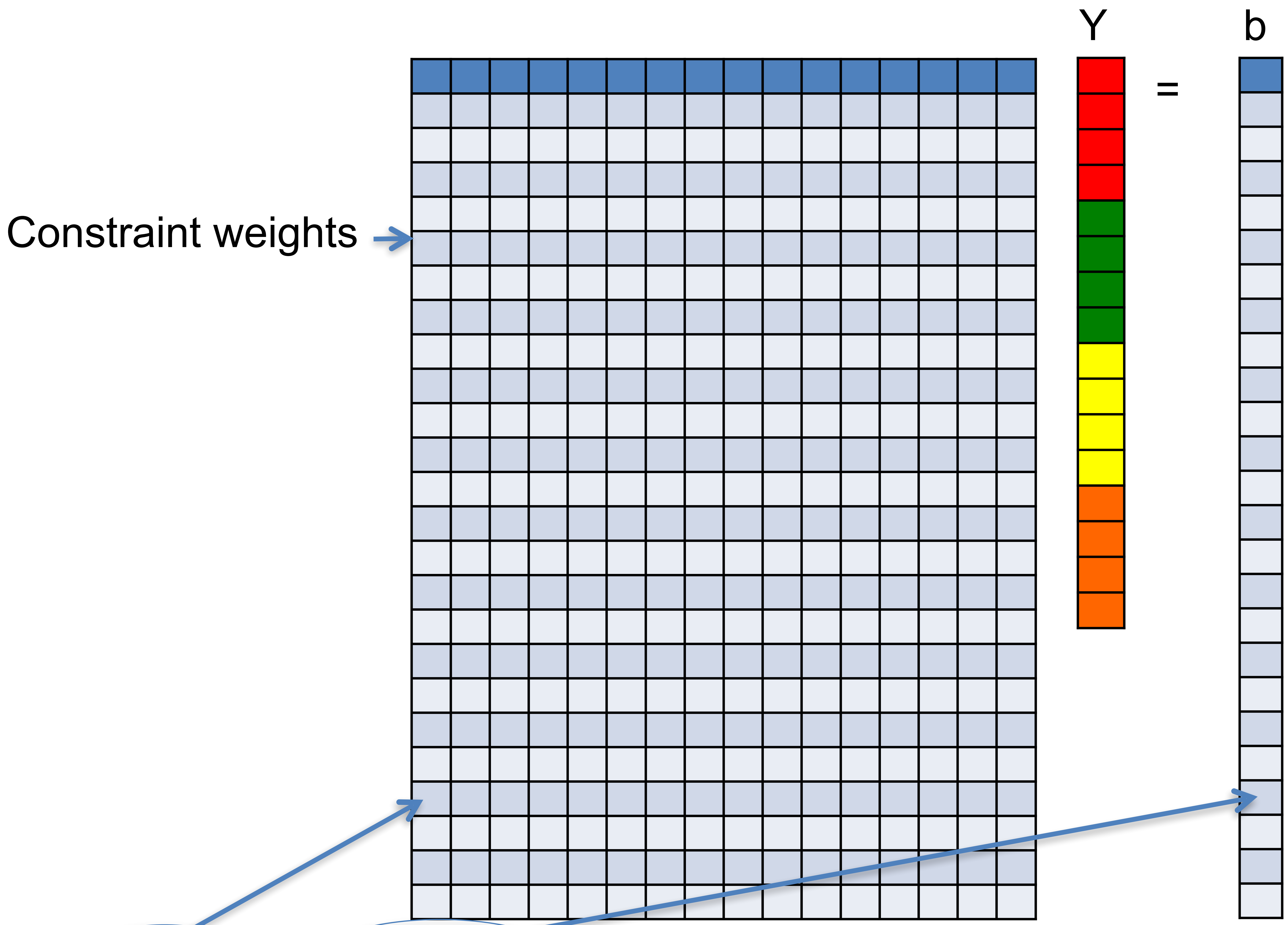
$$\frac{dY}{dx} \approx Y(x,y) - Y(x-1,y) = Y(2,2) - Y(1,2) =$$

0	0	0	0	0	-1	0	0	0	1	0	0	0	0	0	0



Slide credit: Antonio Torralba

A simple inference scheme: solve for Y

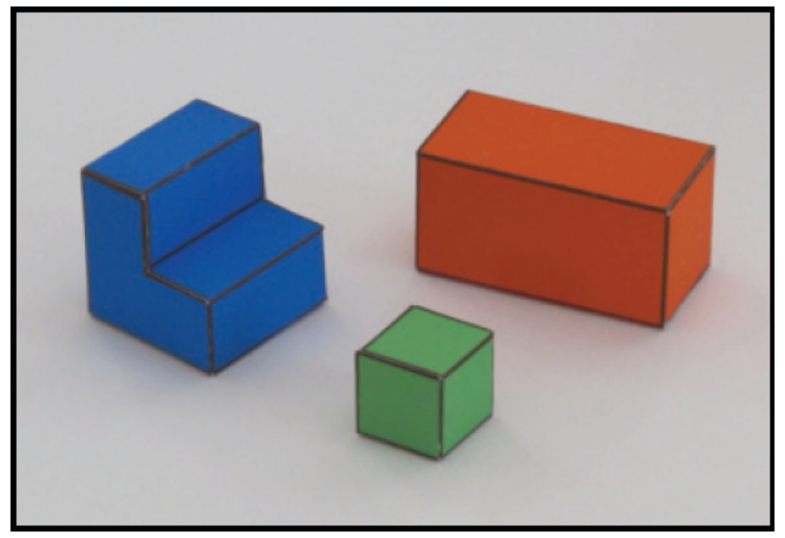


A Y = b

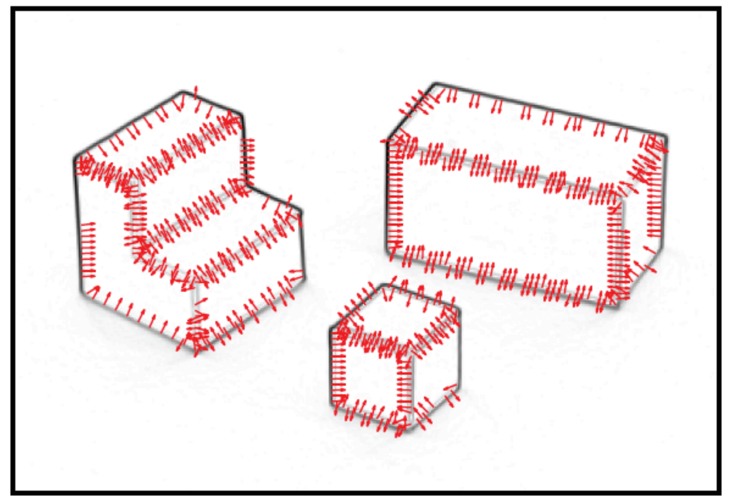
$$\partial Y/\partial y = 1/\cos(\theta)$$

Slide credit: Antonio Torralba

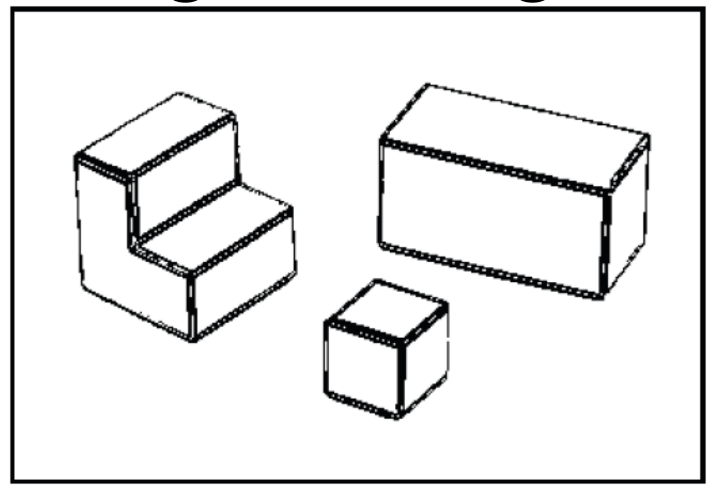
Results



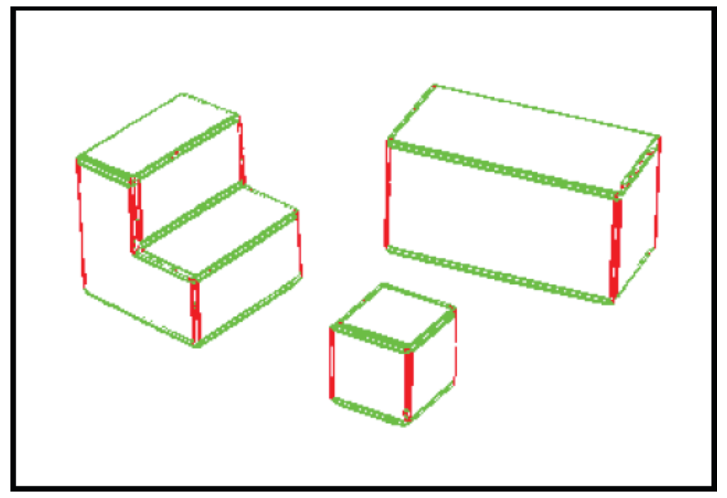
Edge normals



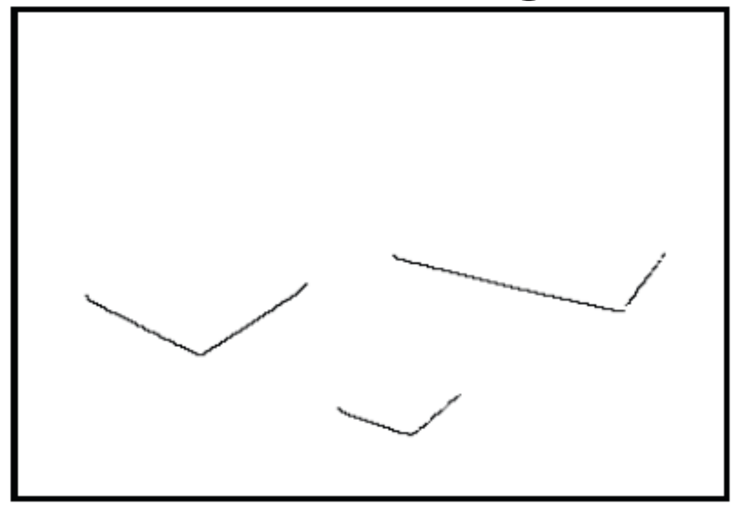
Edge strength



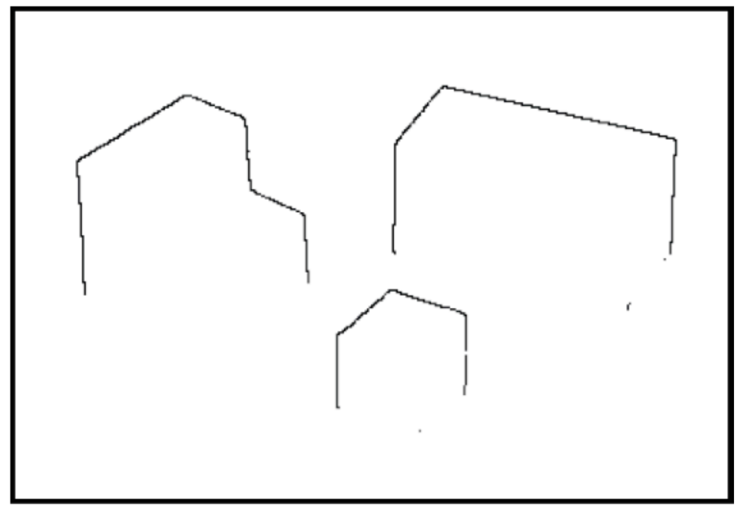
3D orientation

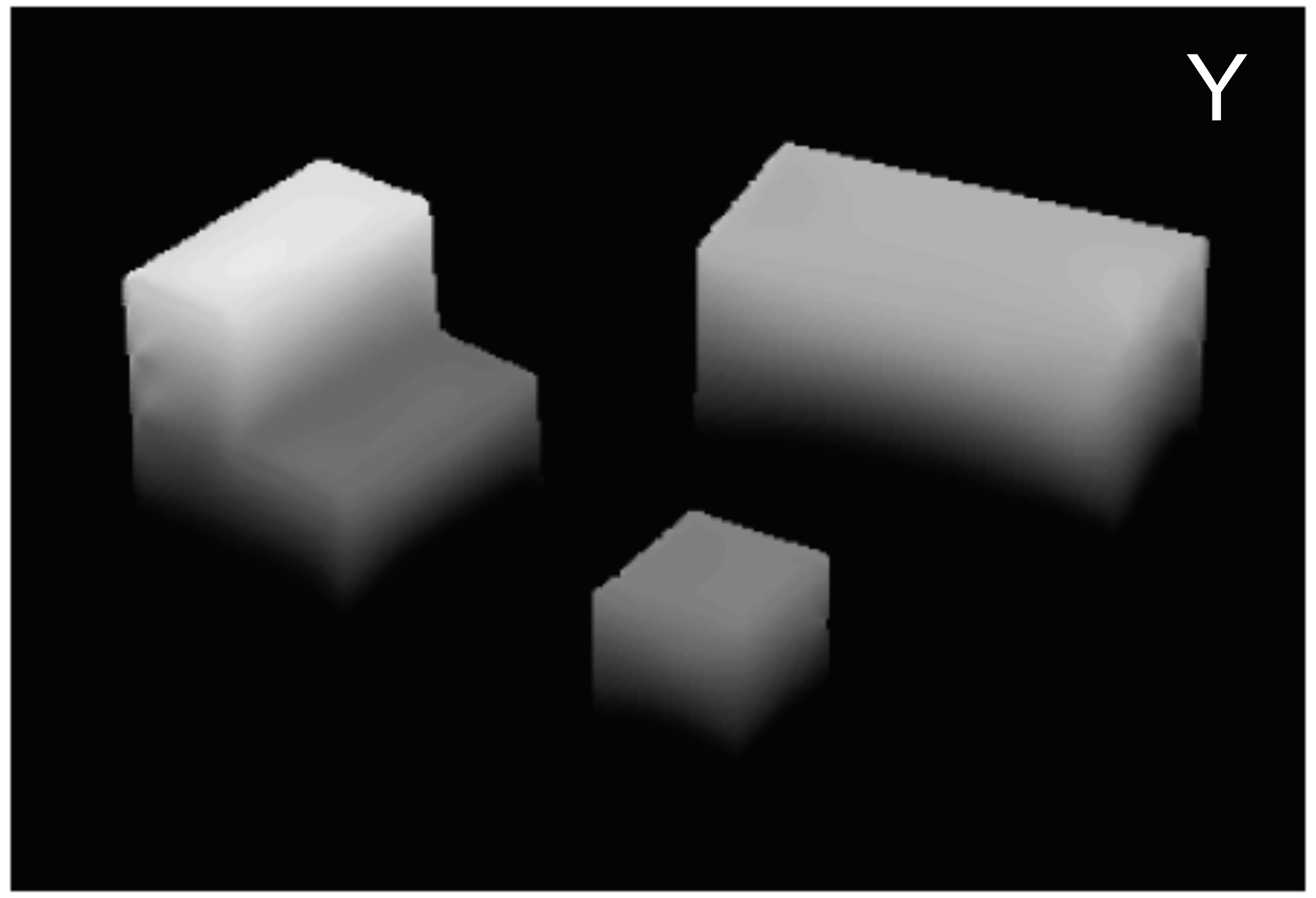


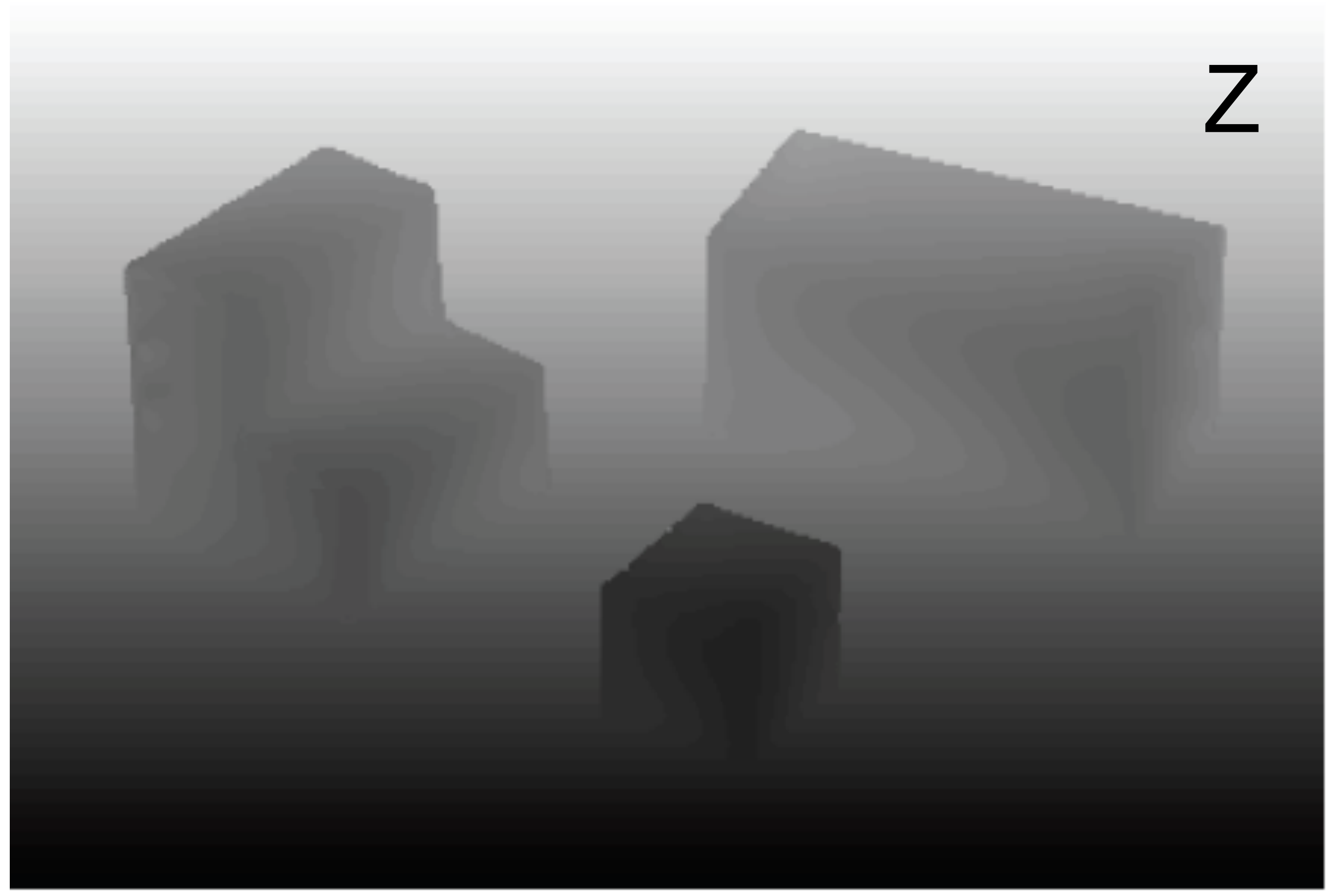
Contact edges



Depth discontinuities



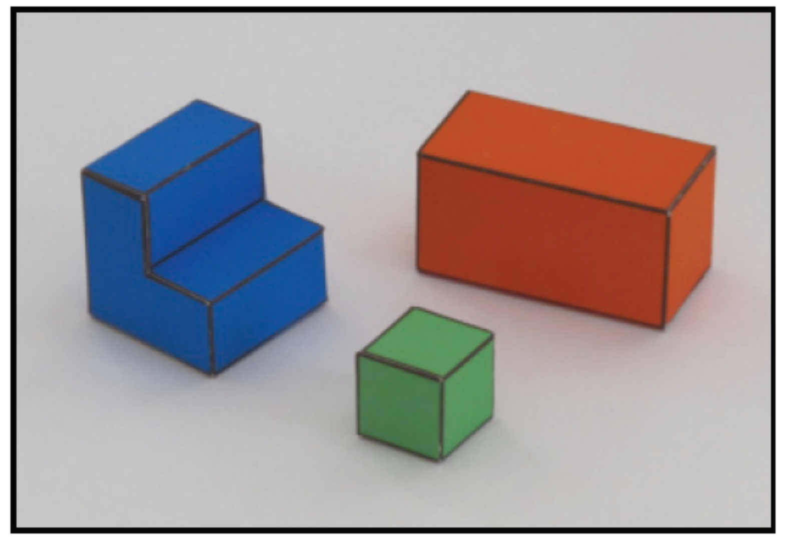




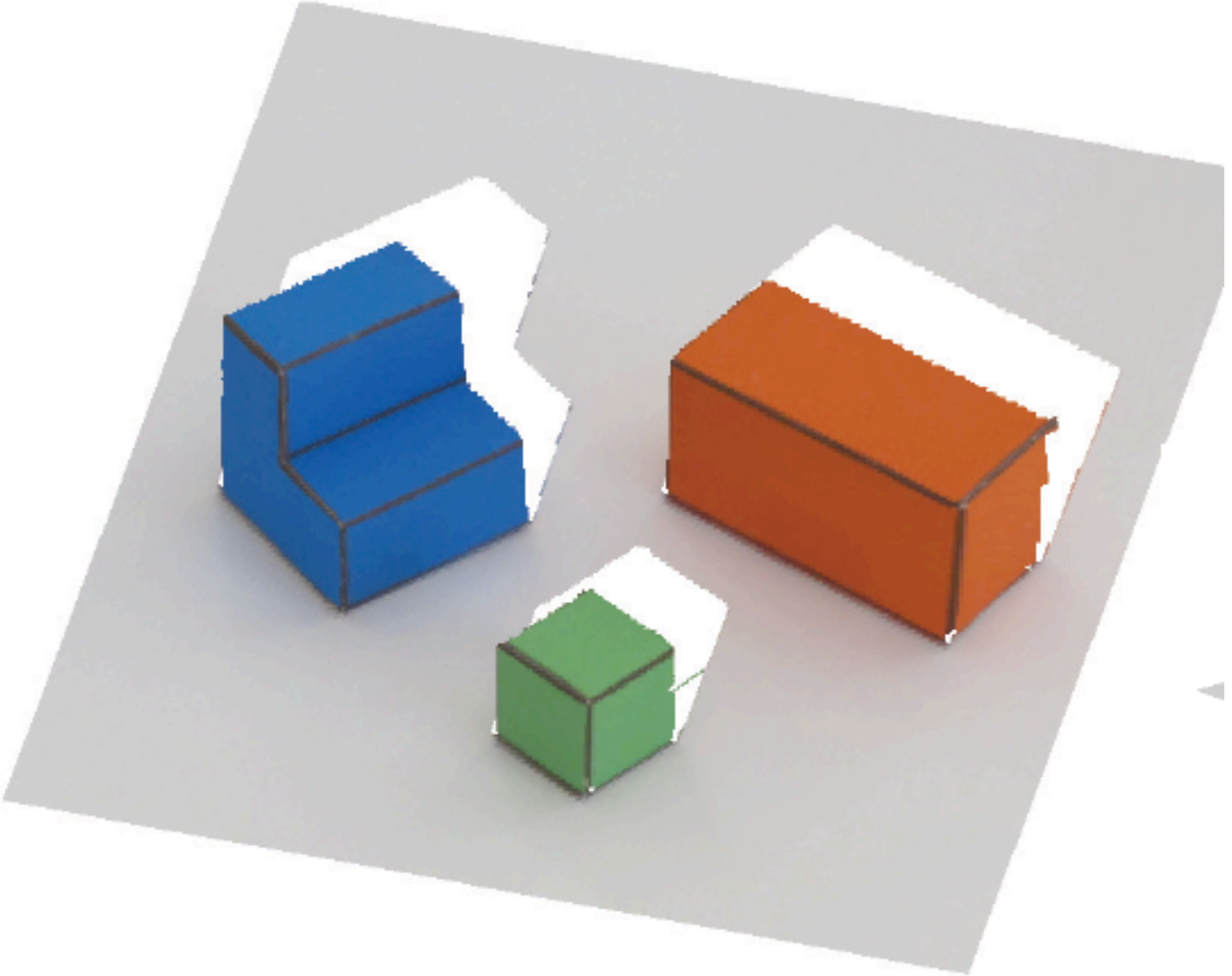
Slide credit: Antonio Torralba

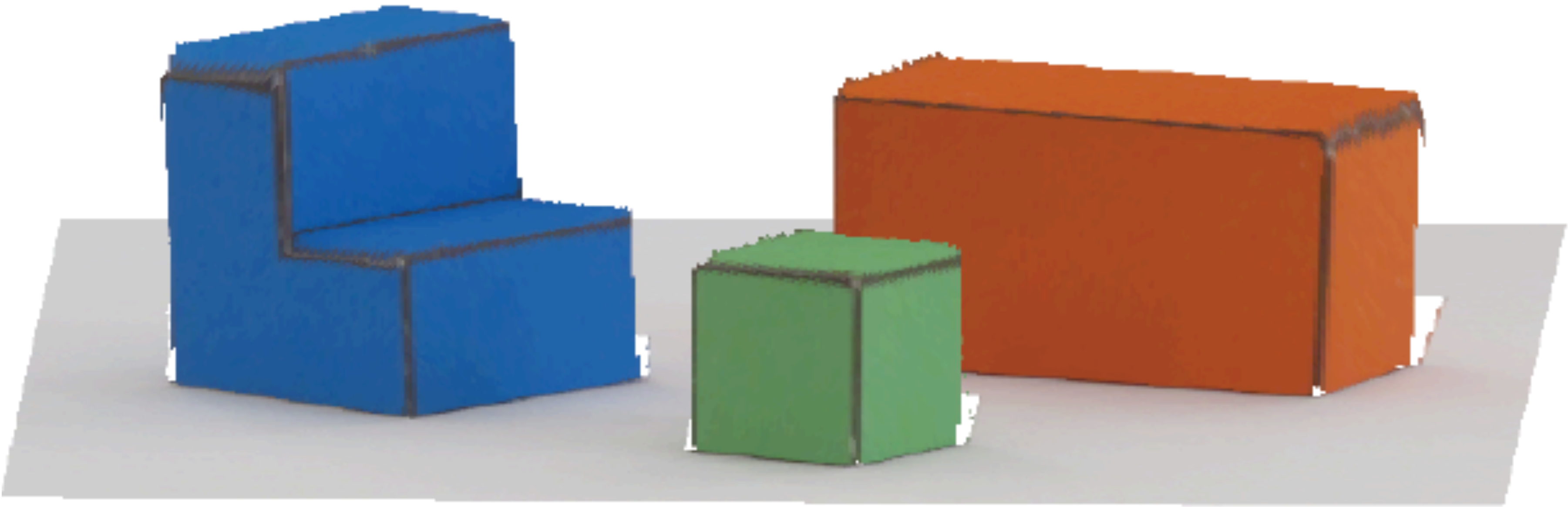
Changing view point

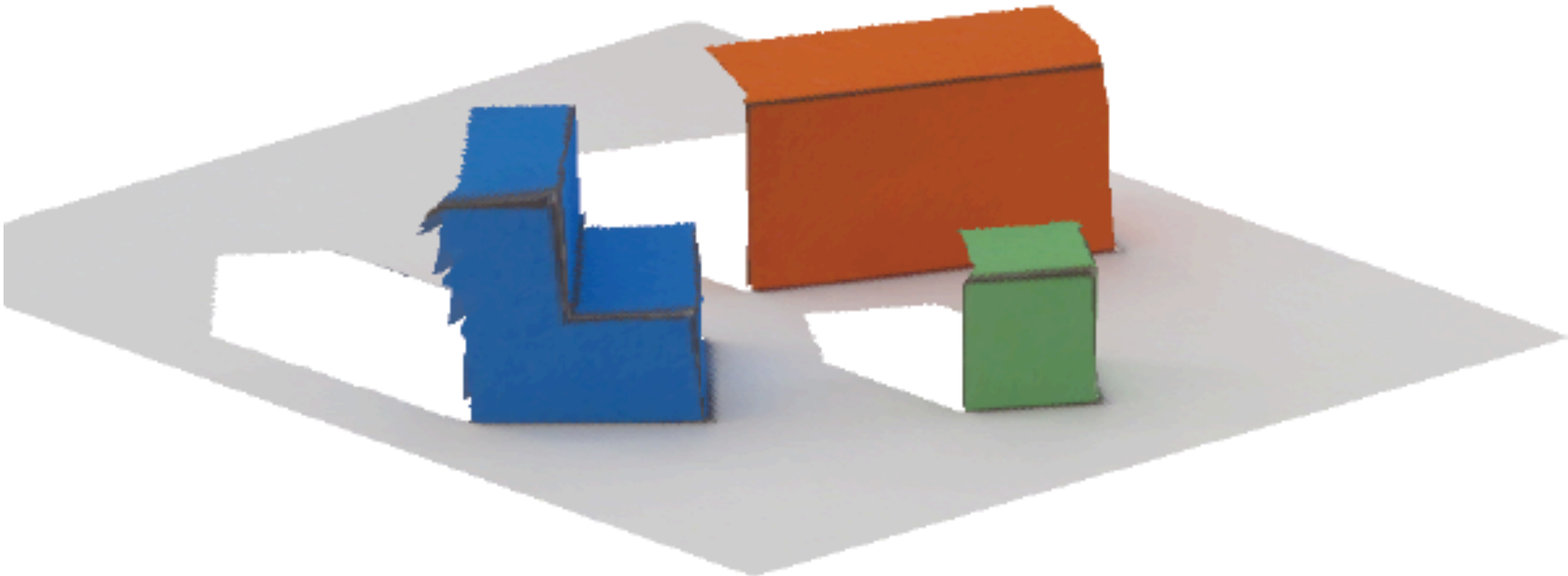
Input



New view points:

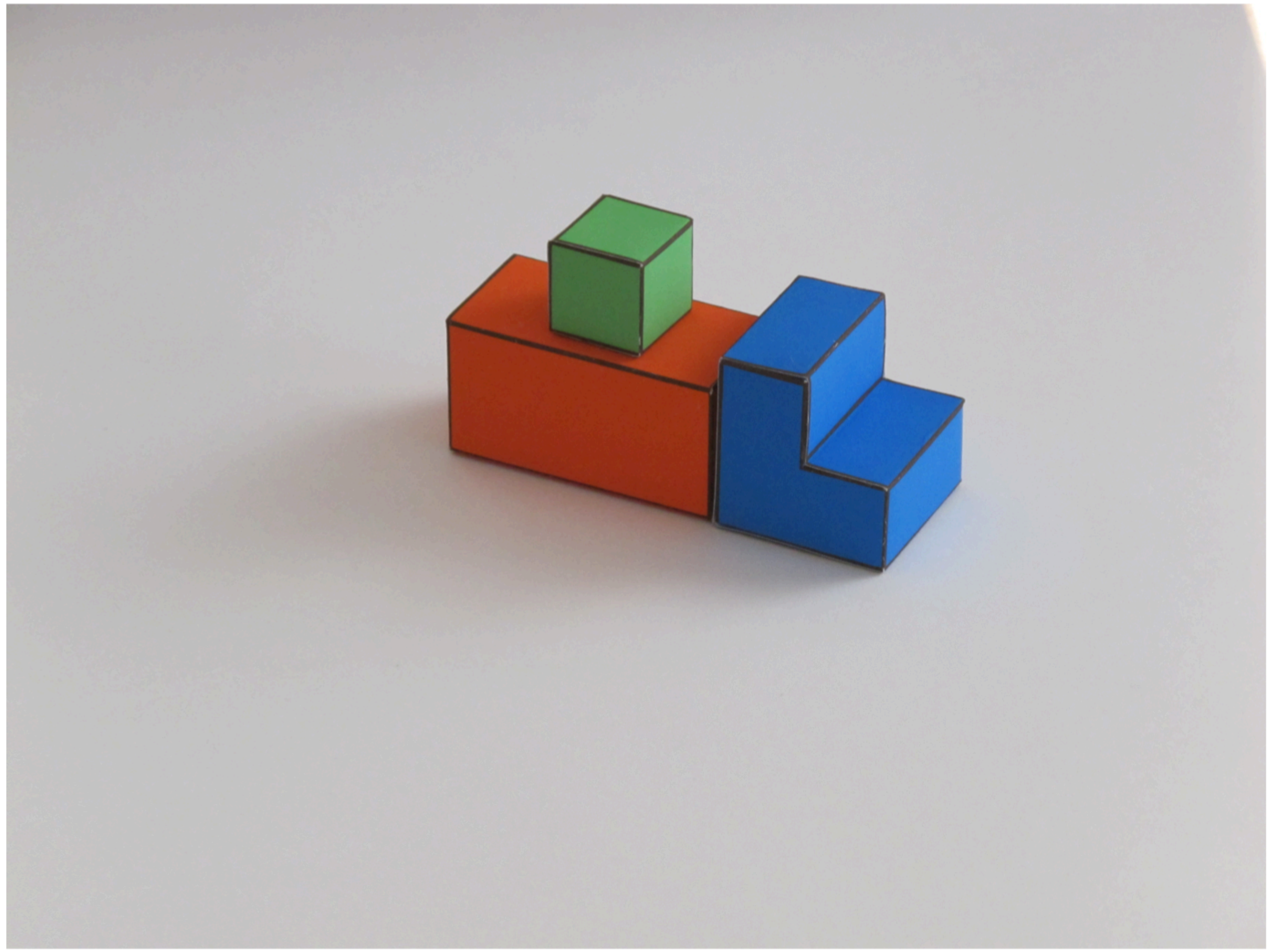


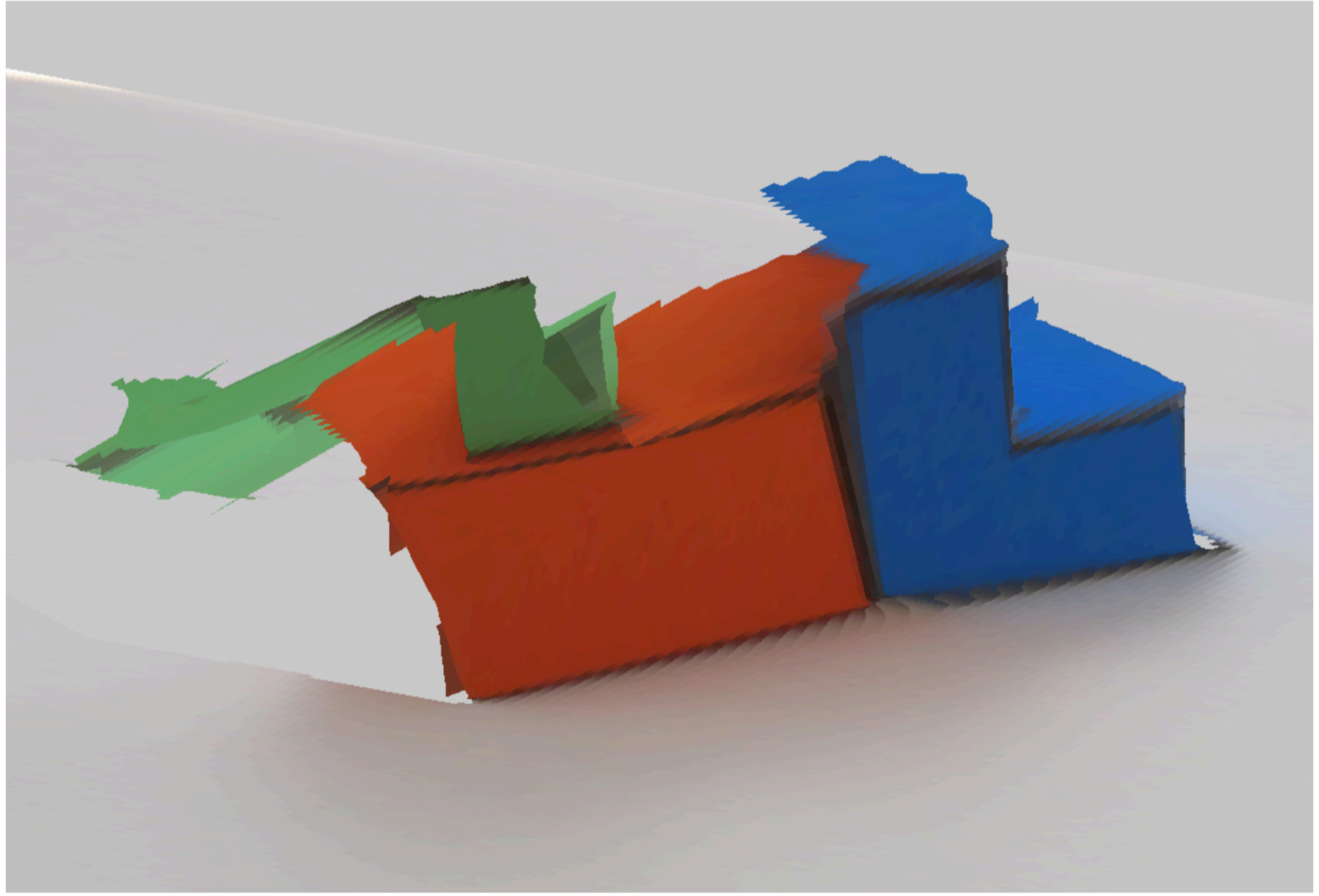




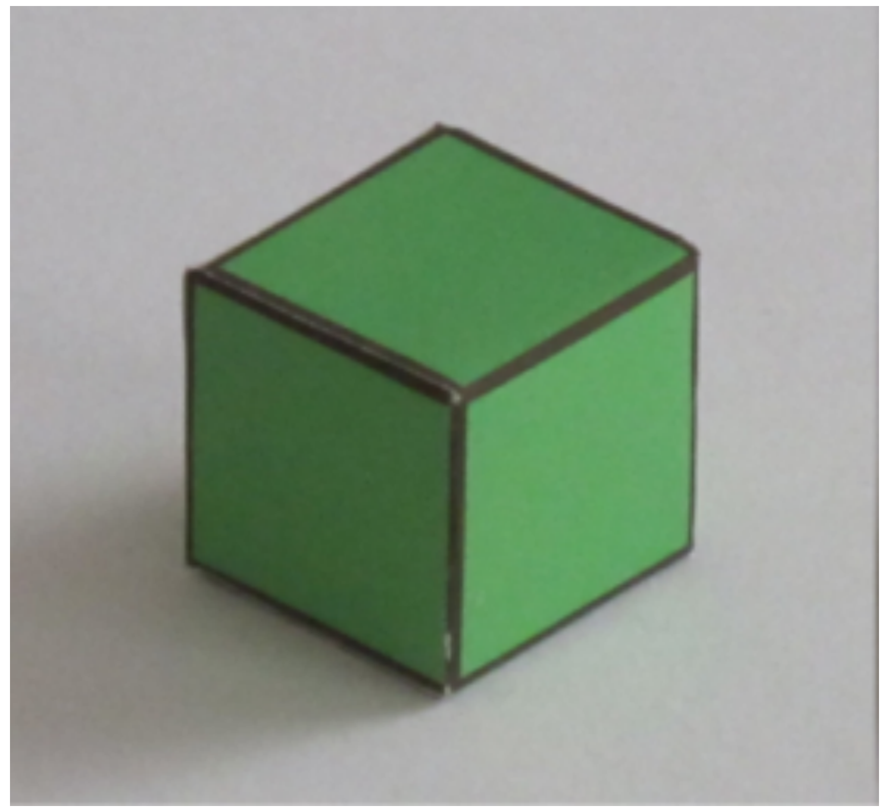
Slide credit: Antonio Torralba

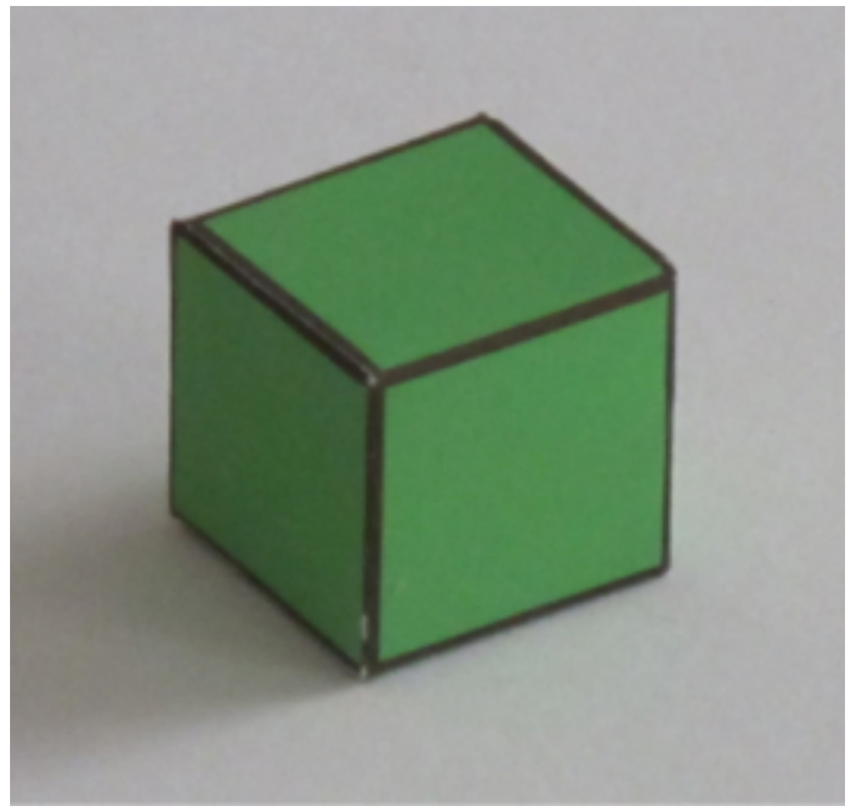
Failure cases… even in a simple world!

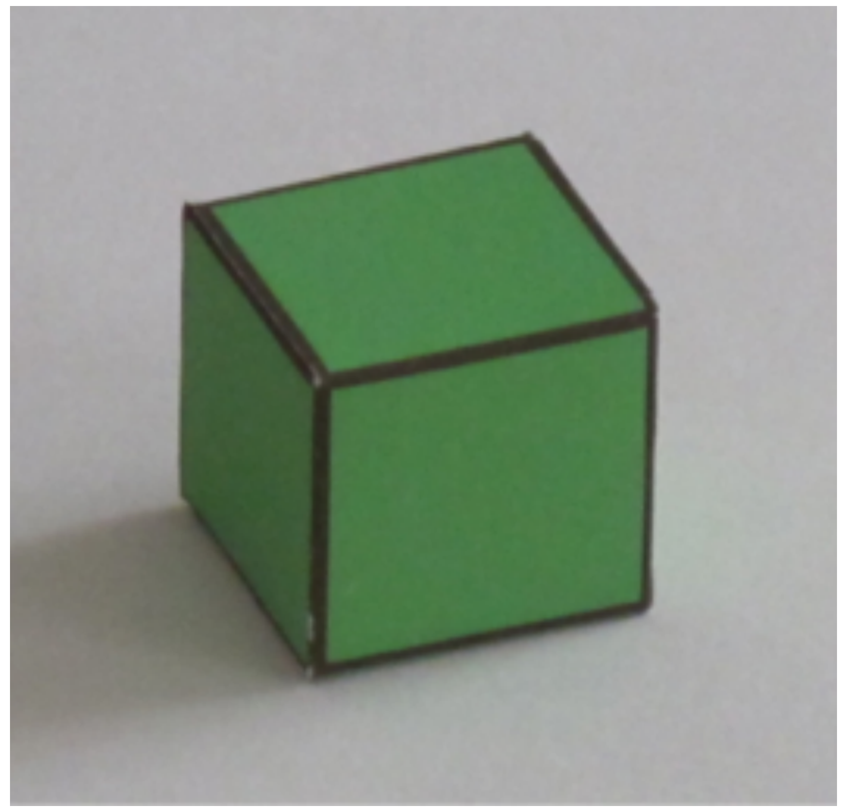


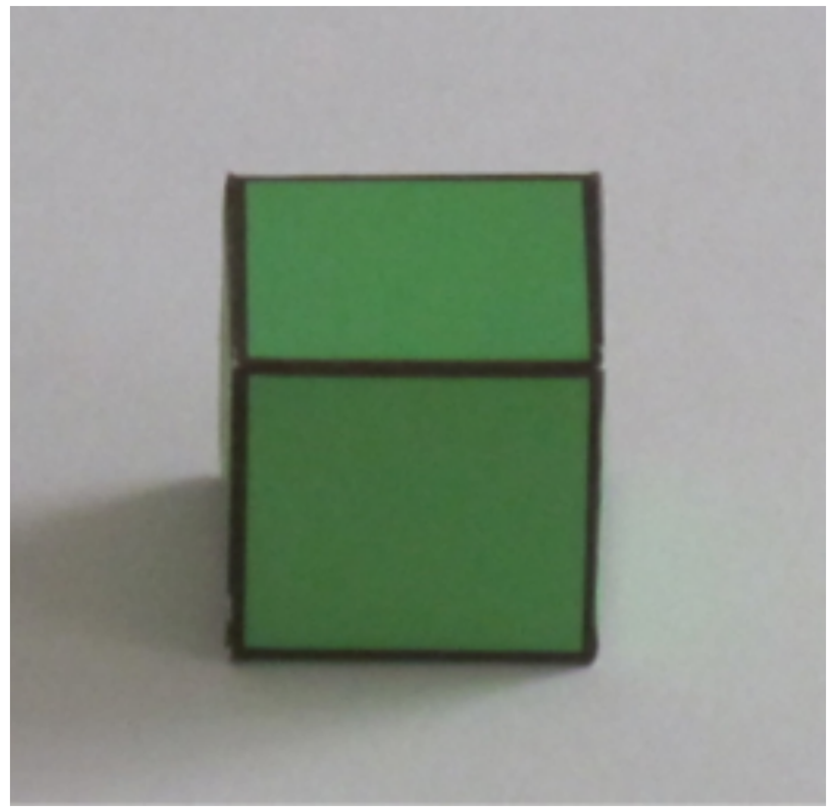


Failure cases… even in a simple world!

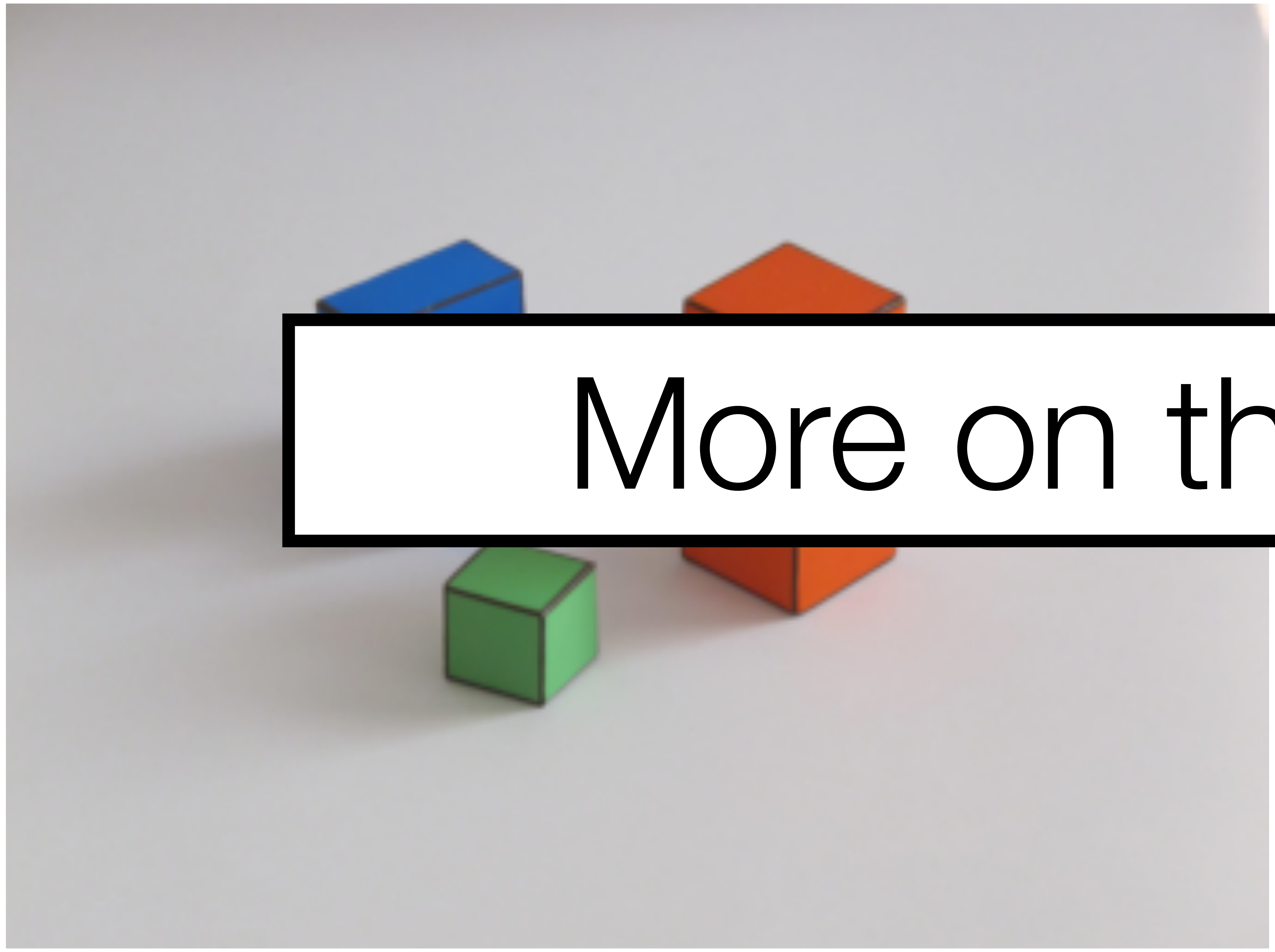




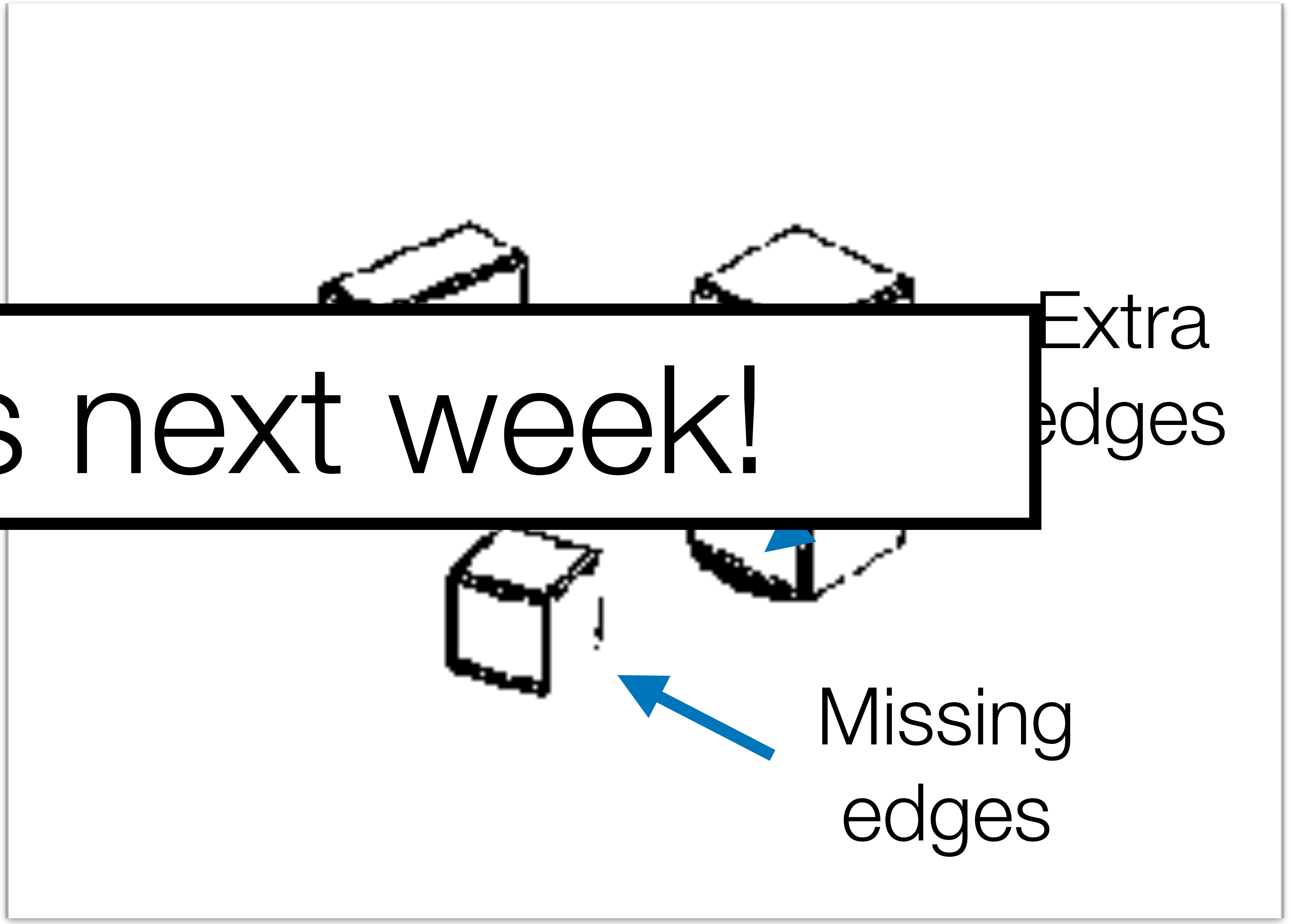




Failure cases… even in a simple world!



Input image



Edges

More on this next week!

Questions about the course?