Advances in Computer Vision

Lecture TR 1pm - 2:30pm, room 26-100

Vincent Sitzmann, Mina Konaković Luković, Bill Freeman

Undergraduate and Graduate versions of this class share the same lectures.

- Undergraduate version, 6.8301, 15 units:
  - Satisfies MIT's CI-M requirement (Communication Intensive, within the Major).
  - WRAP (Writing, Rhetoric, And Professional communication) staff will offer ~8 recitations and provide coaching on communication aspects of the class. .
  - Their assessments and your participation in the required CI-M recitations will be 10% of your final grade. Please fill-out this Google doc to help schedule the sections.
  - Non-MIT students taking the undergraduate class must still fulfill (and thus benefit from) the CI-M components of the class.
  - Summary: More in-class time that graduate version: about 8 required recitations—coaching related to communication elements of the class. Sometimes shorter problem sets and shorter final project.

- Graduate version, 6.8300, 12 units:
  - Problem sets will usually have one or two problems for the graduate students only (or for extra credit for undergraduates).
  - Final projects will be longer and graded to a higher standard than undergraduate version final projects.
  - We're sorry, but the CI-M recitations and instruction are only available to those enrolled in 6.8301. (But the MIT Writing and Communication Center, not part of this class, is available to all MIT students: http://cmsw.mit.edu/writing-and-communication-center)

Grading

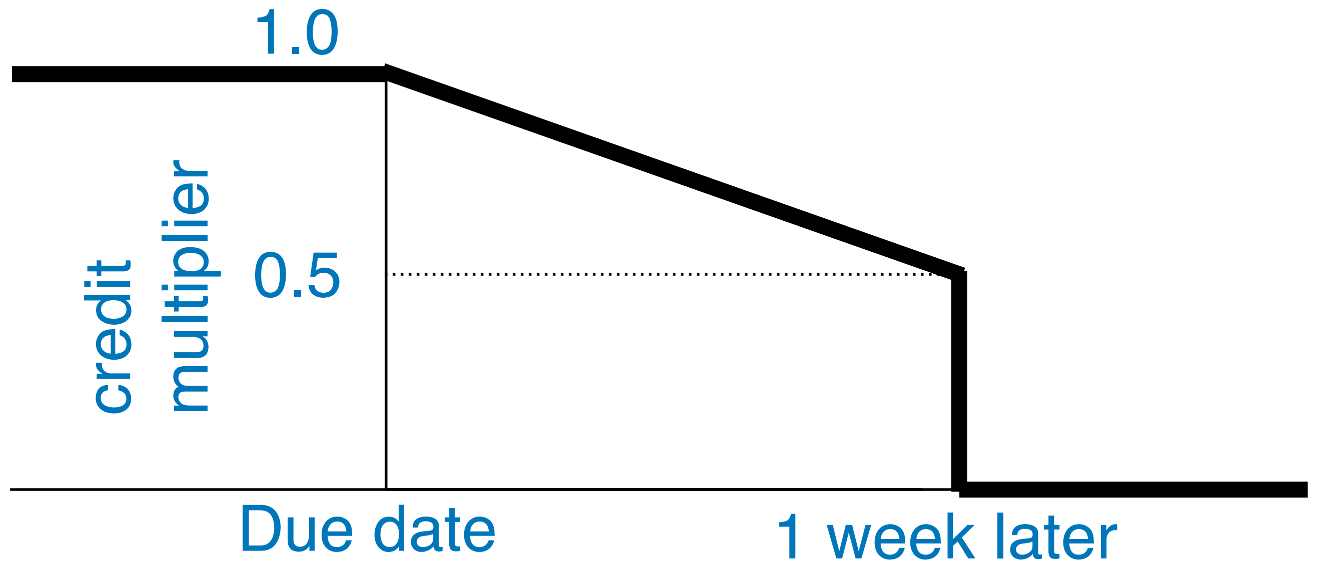
- Problem sets (60%)

- Final Project (40%)

- for 6.8301 students, 10% of course grade will come from classes CI-M components, including required recitations. That will be folded into your final project grade (thus, 1/4 of the final project grade).

- No exams or quizzes

Problem sets



- Problem sets will be posted usually weekly, usually due one week later. See course web page, http://6.8300.csail.mit.edu/sp23/, for the schedule.
- Grades returned two weeks after due date.
- Late penalty: submission deadline is 23:59 on the due date. Late submissions accepted up to one week after deadline, but grade decays linearly down to 1/2 credit over that time (then 0 credit).
- Important-reason grace allowance for late submissions: 3 days, for any important reason. No need to clear it with us, but there's no "saving" it—must be used with any extension. Any request beyond that allowance requires S^3 approval (MIT's student support services).
- Only electronic problem set submissions will be accepted, no hard copies.

- Collaboration policy:
  – You can talk with each other, get advice, and ask questions on Piazza, but **the writing and coding must be done individually and never shared**.

Final Project

- We will provide a list of projects to pick from, or you can propose your own. Can work in pairs, or individually.
- You'll write a final project proposal, and (for 6.8301) a revision of that proposal.
- Every person gives a short presentation of their project during the final week, and submits their written final project.

Additional Information

- For office hours, see course website, http://6.8300.csail.mit.edu/sp23/.
  - Use TA office hours: for psets questions.
  - Use faculty office hours: for questions about lectures or projects.
- Piazza: to ask questions of other students and TA's, use Piazza.
- Textbook: we will post relevant chapters from forthcoming MIT Press computer vision textbook. Other resources are listed on course web page, many of which are free and online.

Course content

- We will cover: Cameras, optics, signals, deep learning, applications, and practical research issues.
- See course web page for schedule/syllabus:

  http://6.8300.csail.mit.edu/sp23/
- Math: Linear algebra, geometry, multivariate calculus, optimization, probabilistic inference, machine learning , deep nets.
- Coding: Python, PyTorch
  - Tutorials in Python and Pytorch will be announced and offered before the assignments that first use them.

Other questions:

We’ll be in the lobby just outside 26-100 after this class for any immediate questions today.

Other mechanisms to answer general questions about the class:

piazza

TA’s or faculty members during their office hours

course web page: http://6.8300.csail.mit.edu/sp23/

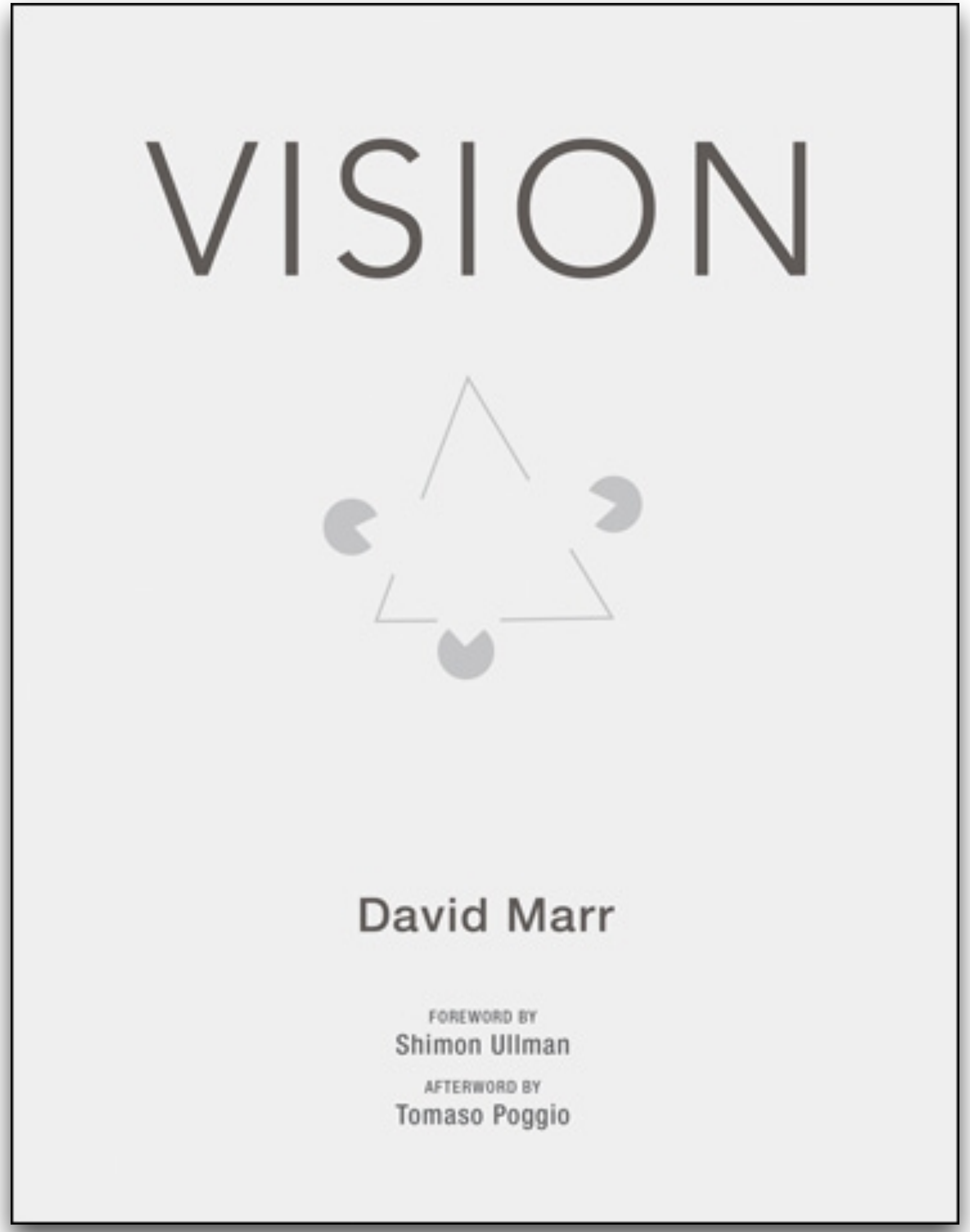
Lecture 1

Introduction to computer vision

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.

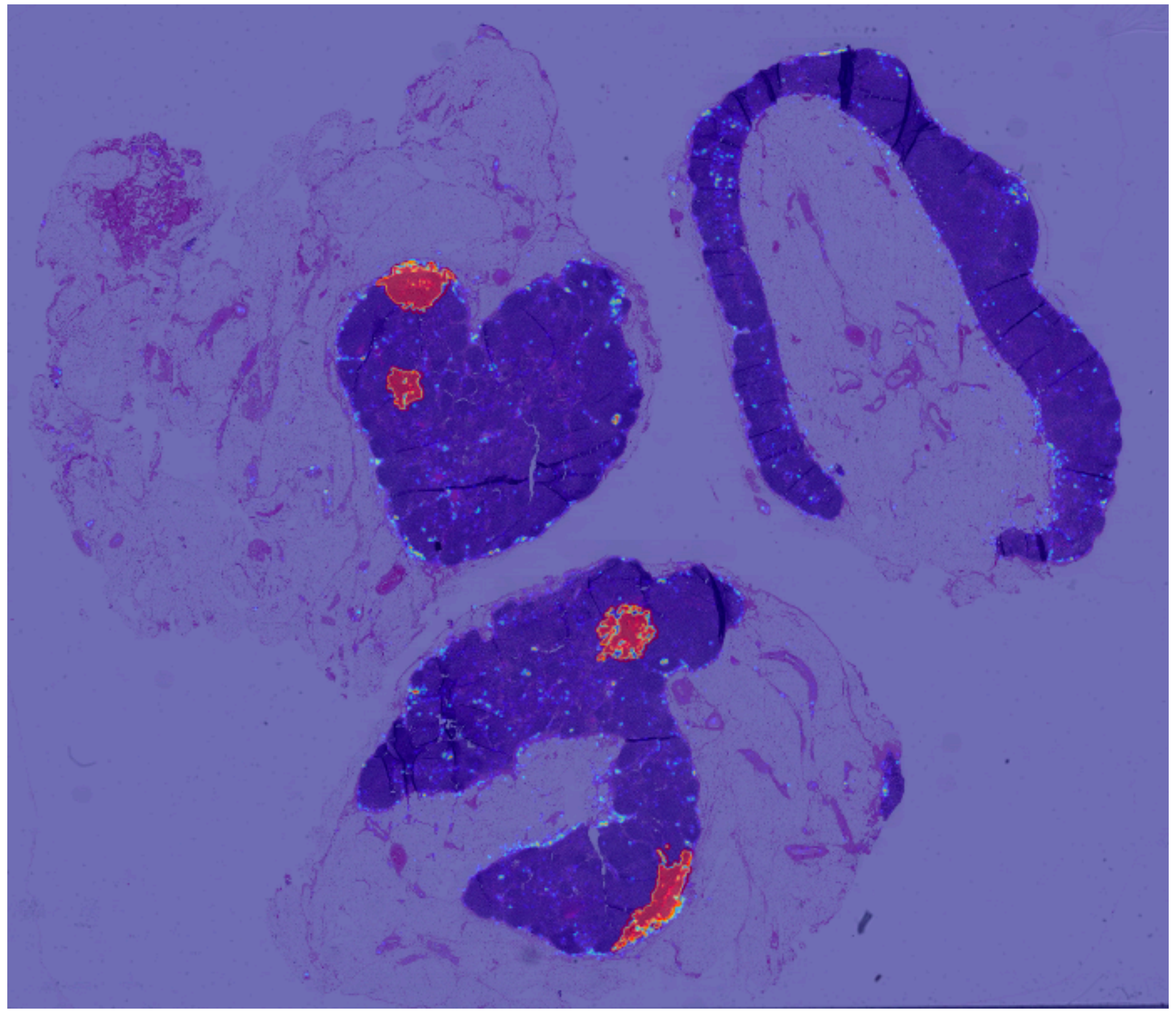


Exciting times in computer vision

Robotics

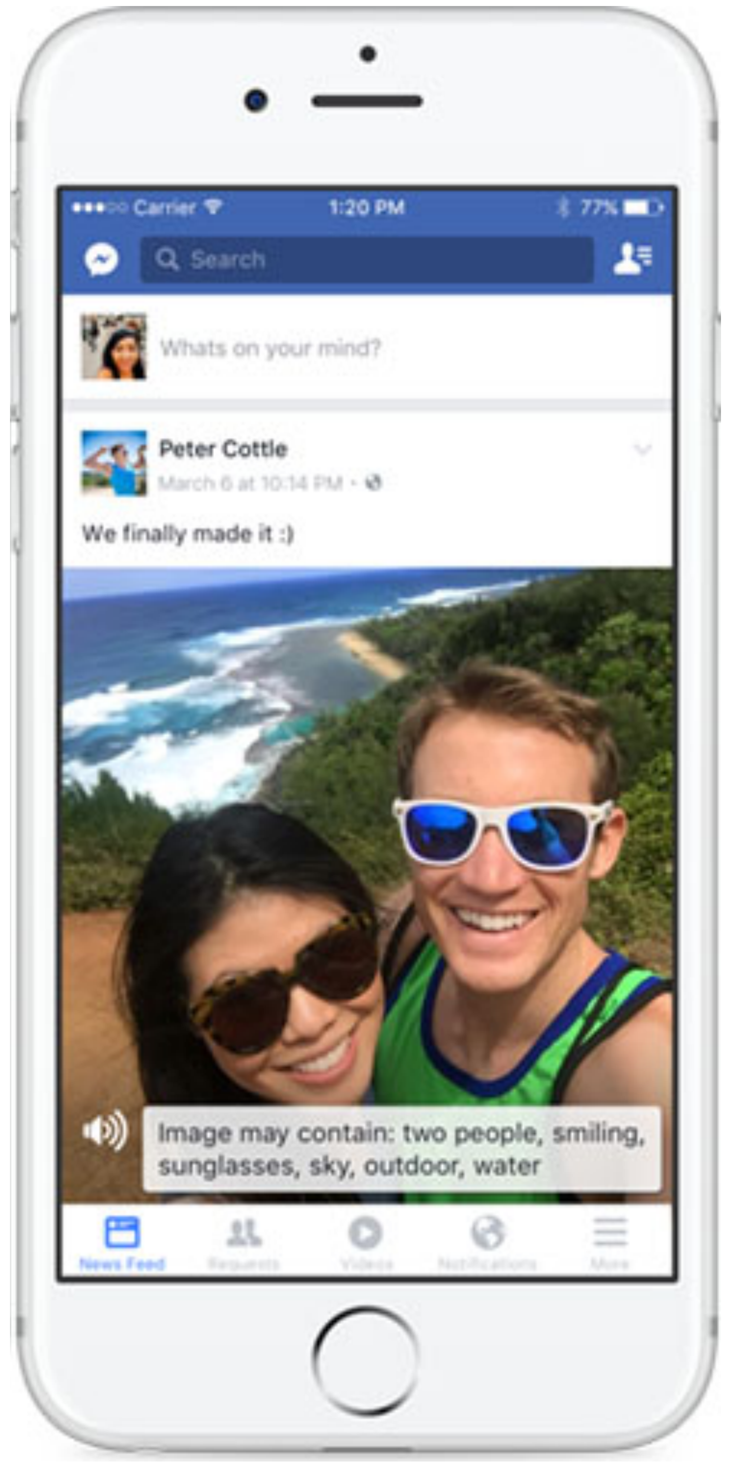
Driving

Medical applications



Gaming

Accessibility



Exciting times in computer vision

"A cup of coffee"

"A cat"

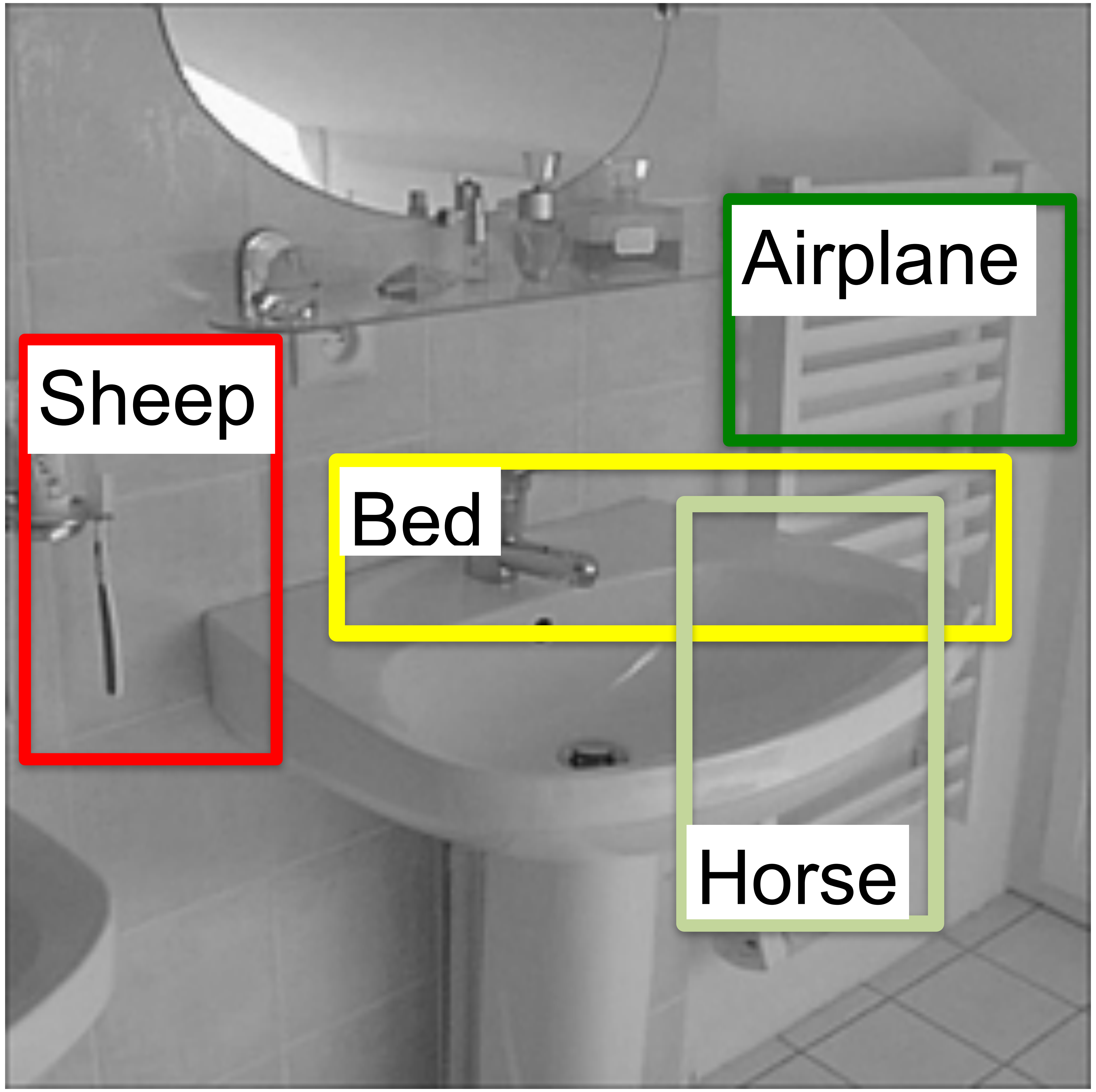
"A cup of cat"

DALL-E 2 (Open AI)

https://www.reddit.com/r/dalle2/comments/y4mygn/a_cup_of_cat/

Slide credit: Shuang Li

When some of us started…



Why is vision hard?

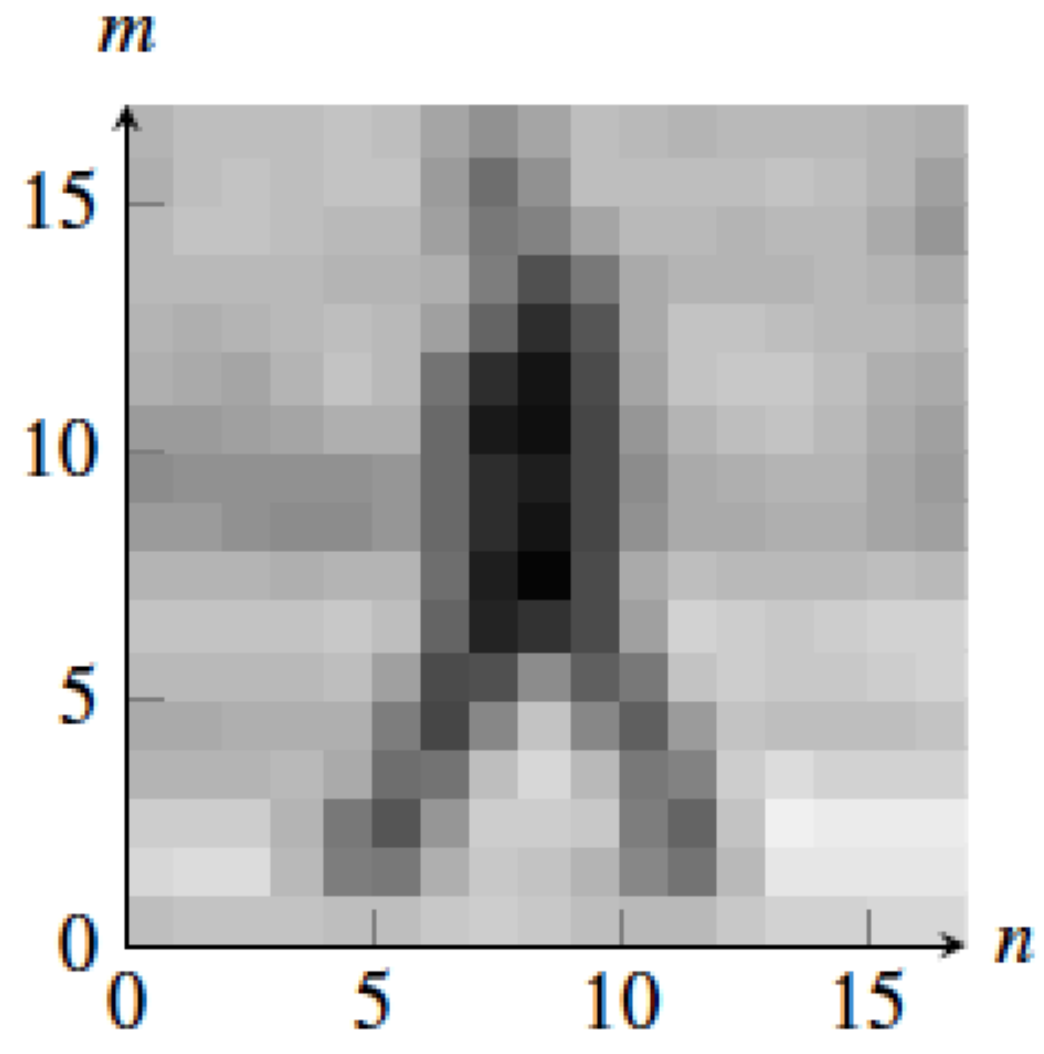
The input

What the machine gets

$$
\mathbf{I} = \begin{bmatrix}
160 & 175 & 171 & 168 & 168 & 172 & 164 & 158 & 167 & 173 & 167 & 163 & 162 & 164 & 160 & 159 & 163 & 162 \\
149 & 164 & 172 & 175 & 178 & 179 & 176 & 118 & 97 & 168 & 175 & 171 & 169 & 175 & 176 & 177 & 165 & 152 \\
161 & 166 & 182 & 171 & 170 & 177 & 175 & 116 & 109 & 169 & 177 & 173 & 168 & 175 & 175 & 159 & 153 & 123 \\
171 & 174 & 177 & 175 & 167 & 161 & 157 & 138 & 103 & 112 & 157 & 164 & 159 & 160 & 165 & 169 & 148 & 144 \\
163 & 163 & 162 & 165 & 167 & 164 & 178 & 167 & 77 & 55 & 134 & 170 & 167 & 162 & 164 & 175 & 168 & 160 \\
173 & 164 & 158 & 165 & 180 & 180 & 150 & 89 & 61 & 34 & 137 & 186 & 186 & 182 & 175 & 165 & 160 & 164 \\
152 & 155 & 146 & 147 & 169 & 180 & 163 & 51 & 24 & 32 & 119 & 163 & 175 & 182 & 181 & 162 & 148 & 153 \\
134 & 135 & 147 & 149 & 150 & 147 & 148 & 62 & 36 & 46 & 114 & 157 & 163 & 167 & 169 & 163 & 146 & 147 \\
135 & 132 & 131 & 125 & 115 & 129 & 132 & 74 & 54 & 41 & 104 & 156 & 152 & 156 & 164 & 156 & 141 & 144 \\
151 & 155 & 151 & 145 & 144 & 149 & 143 & 71 & 31 & 29 & 129 & 164 & 157 & 155 & 159 & 158 & 156 & 148 \\
172 & 174 & 178 & 177 & 177 & 181 & 174 & 54 & 21 & 29 & 136 & 190 & 180 & 179 & 176 & 184 & 187 & 182 \\
177 & 178 & 176 & 173 & 174 & 180 & 150 & 27 & 101 & 94 & 74 & 189 & 188 & 186 & 183 & 186 & 188 & 187 \\
160 & 160 & 163 & 163 & 161 & 167 & 100 & 45 & 169 & 166 & 59 & 136 & 184 & 176 & 175 & 177 & 185 & 186 \\
147 & 150 & 153 & 155 & 160 & 155 & 56 & 111 & 182 & 180 & 104 & 84 & 168 & 172 & 171 & 164 & 168 & 167 \\
184 & 182 & 178 & 175 & 179 & 133 & 86 & 191 & 201 & 204 & 191 & 79 & 172 & 220 & 217 & 205 & 209 & 200 \\
184 & 187 & 192 & 182 & 124 & 32 & 109 & 168 & 171 & 167 & 163 & 51 & 105 & 203 & 209 & 203 & 210 & 205 \\
191 & 198 & 203 & 197 & 175 & 149 & 169 & 189 & 190 & 173 & 160 & 145 & 156 & 202 & 199 & 201 & 205 & 202 \\
153 & 149 & 153 & 155 & 173 & 182 & 179 & 177 & 182 & 177 & 182 & 185 & 179 & 177 & 167 & 176 & 182 & 180
\end{bmatrix}
$$

The input

What we see

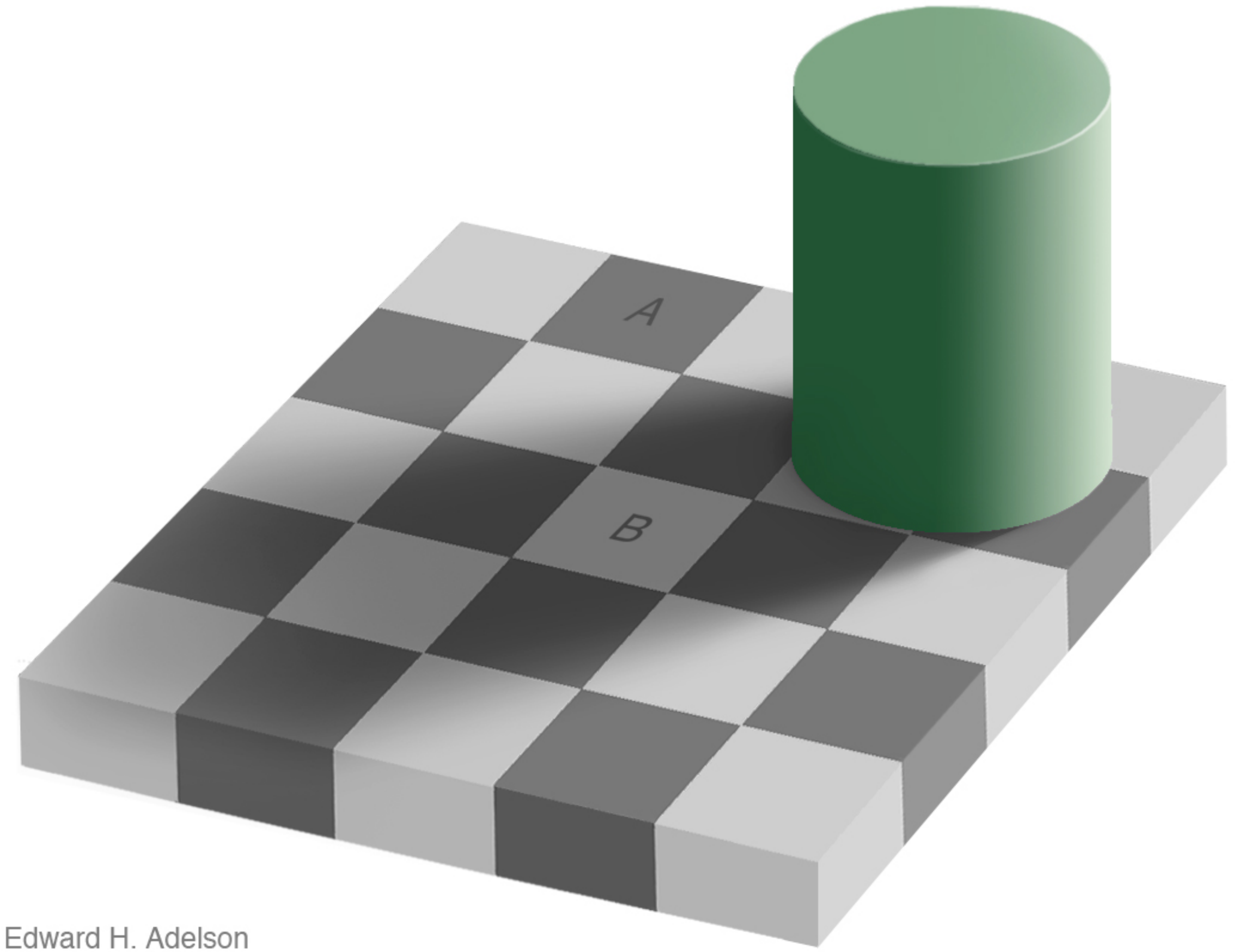


What the machine gets

$$\mathbf{I} = \begin{bmatrix}
160 & 175 & 171 & 168 & 168 & 172 & 164 & 158 & 167 & 173 & 167 & 163 & 162 & 164 & 160 & 159 & 163 & 162 \\
149 & 164 & 172 & 175 & 178 & 179 & 176 & 118 & 97 & 168 & 175 & 171 & 169 & 175 & 176 & 177 & 165 & 152 \\
161 & 166 & 182 & 171 & 170 & 177 & 175 & 116 & 109 & 169 & 177 & 173 & 168 & 175 & 175 & 159 & 153 & 123 \\
171 & 174 & 177 & 175 & 167 & 161 & 157 & 138 & 103 & 112 & 157 & 164 & 159 & 160 & 165 & 169 & 148 & 144 \\
163 & 163 & 162 & 165 & 167 & 164 & 178 & 167 & 77 & 55 & 134 & 170 & 167 & 162 & 164 & 175 & 168 & 160 \\
173 & 164 & 158 & 165 & 180 & 180 & 150 & 89 & 61 & 34 & 137 & 186 & 186 & 182 & 175 & 165 & 160 & 164 \\
152 & 155 & 146 & 147 & 169 & 180 & 163 & 51 & 24 & 32 & 119 & 163 & 175 & 182 & 181 & 162 & 148 & 153 \\
134 & 135 & 147 & 149 & 150 & 147 & 148 & 62 & 36 & 46 & 114 & 157 & 163 & 167 & 169 & 163 & 146 & 147 \\
135 & 132 & 131 & 125 & 115 & 129 & 132 & 74 & 54 & 41 & 104 & 156 & 152 & 156 & 164 & 156 & 141 & 144 \\
151 & 155 & 151 & 145 & 144 & 149 & 143 & 71 & 31 & 29 & 129 & 164 & 157 & 155 & 159 & 158 & 156 & 148 \\
172 & 174 & 178 & 177 & 177 & 181 & 174 & 54 & 21 & 29 & 136 & 190 & 180 & 179 & 176 & 184 & 187 & 182 \\
177 & 178 & 176 & 173 & 174 & 180 & 150 & 27 & 101 & 94 & 74 & 189 & 188 & 186 & 183 & 186 & 188 & 187 \\
160 & 160 & 163 & 163 & 161 & 167 & 100 & 45 & 169 & 166 & 59 & 136 & 184 & 176 & 175 & 177 & 185 & 186 \\
147 & 150 & 153 & 155 & 160 & 155 & 56 & 111 & 182 & 180 & 104 & 84 & 168 & 172 & 171 & 164 & 168 & 167 \\
184 & 182 & 178 & 175 & 179 & 133 & 86 & 191 & 201 & 204 & 191 & 79 & 172 & 220 & 217 & 205 & 209 & 200 \\
184 & 187 & 192 & 182 & 124 & 32 & 109 & 168 & 171 & 167 & 163 & 51 & 105 & 203 & 209 & 203 & 210 & 205 \\
191 & 198 & 203 & 197 & 175 & 149 & 169 & 189 & 190 & 173 & 160 & 145 & 156 & 202 & 199 & 201 & 205 & 202 \\
153 & 149 & 153 & 155 & 173 & 182 & 179 & 177 & 182 & 177 & 182 & 185 & 179 & 177 & 167 & 176 & 182 & 180
\end{bmatrix}$$

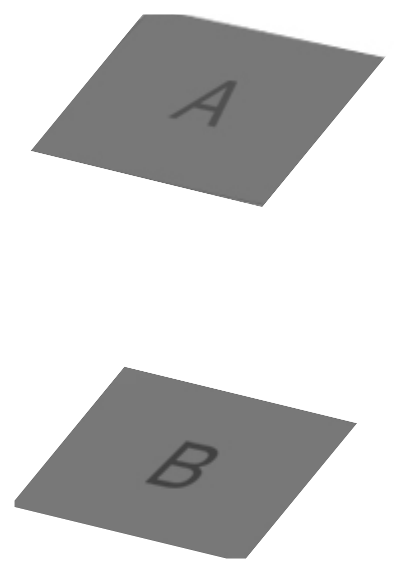
The camera is a measurement device, not a vision system

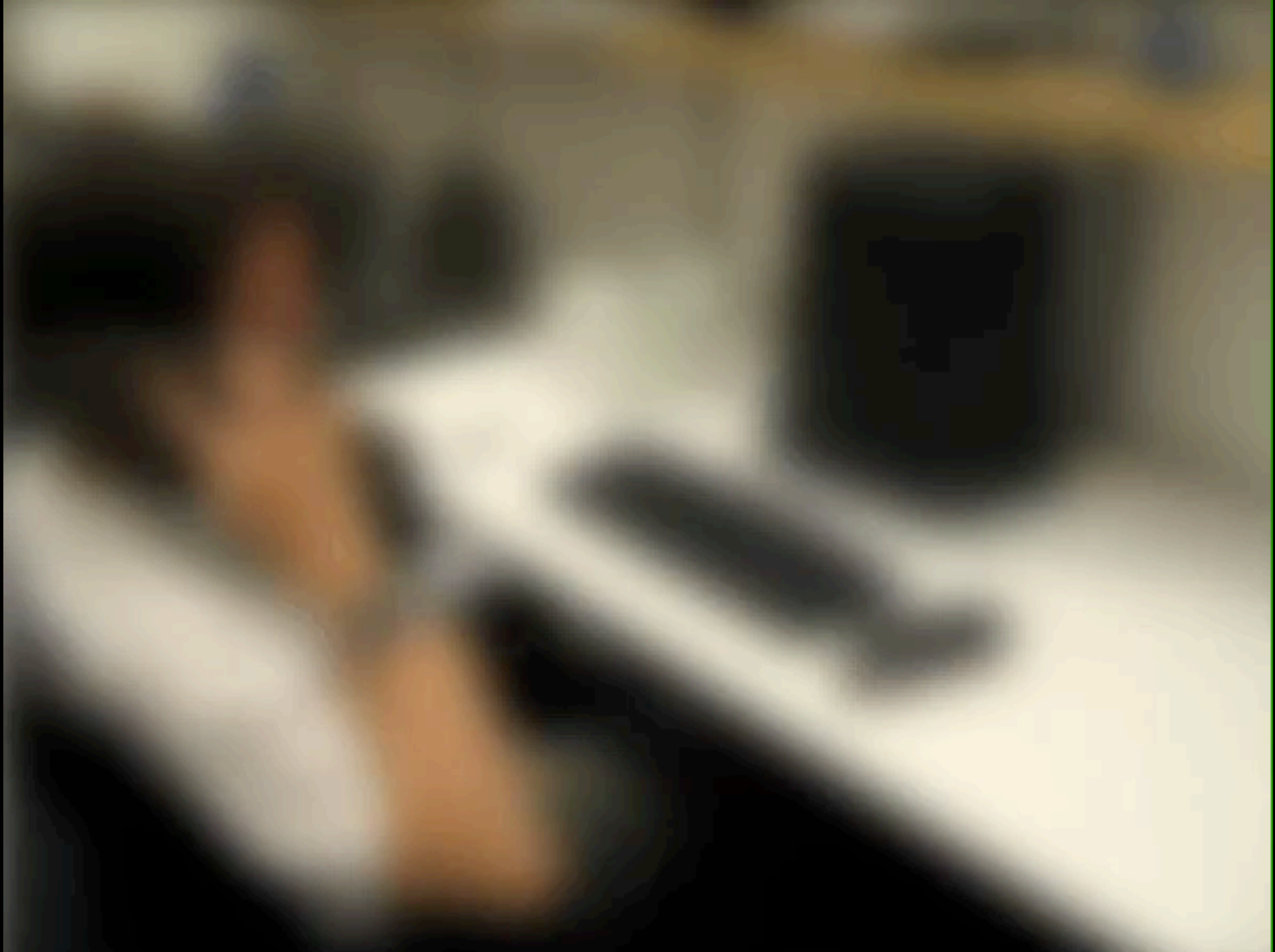
To see: perception vs. measurement



Edward H. Adelson

To see: perception vs. measurement

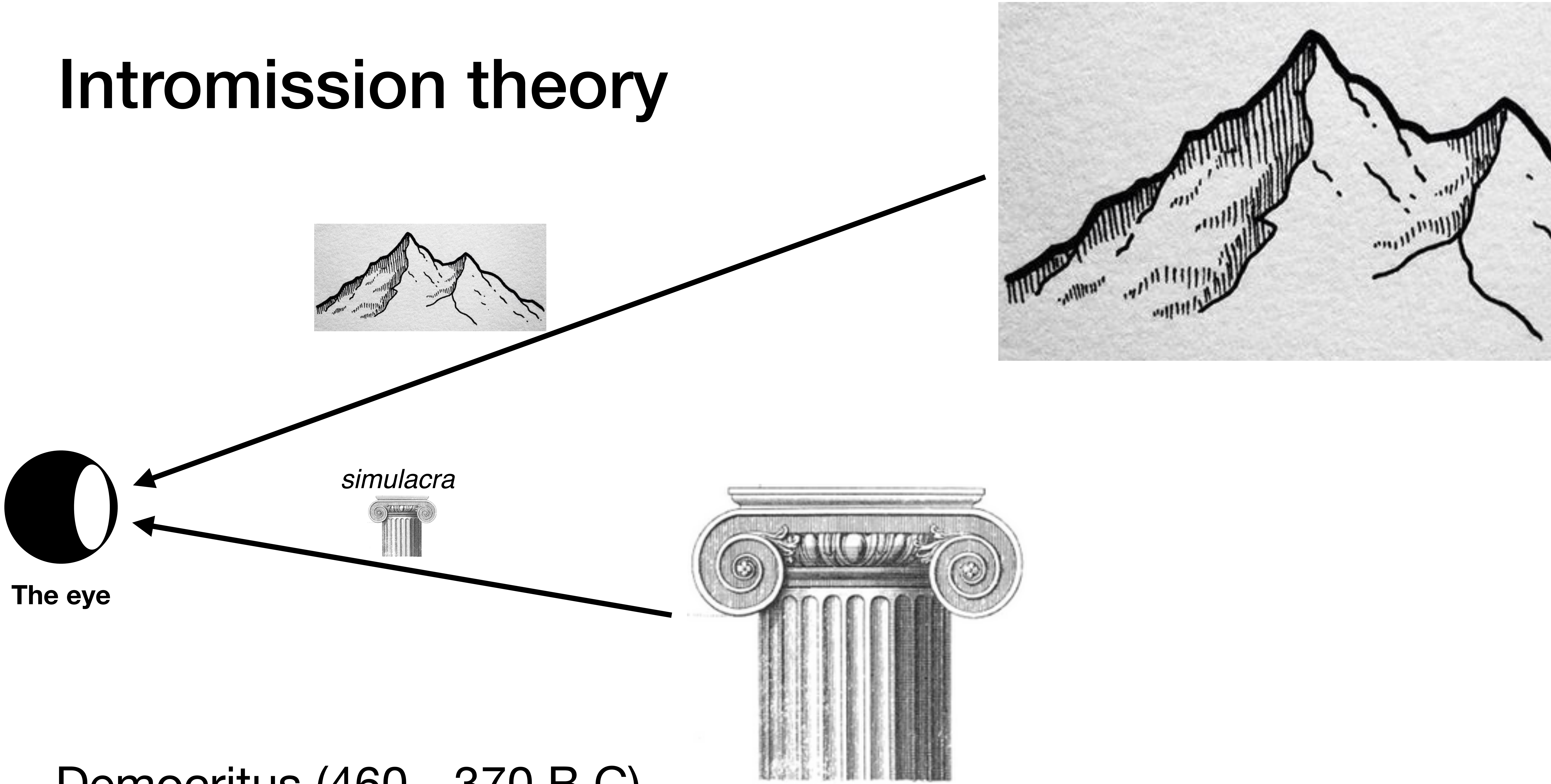




A short story of vision research

The Greeks

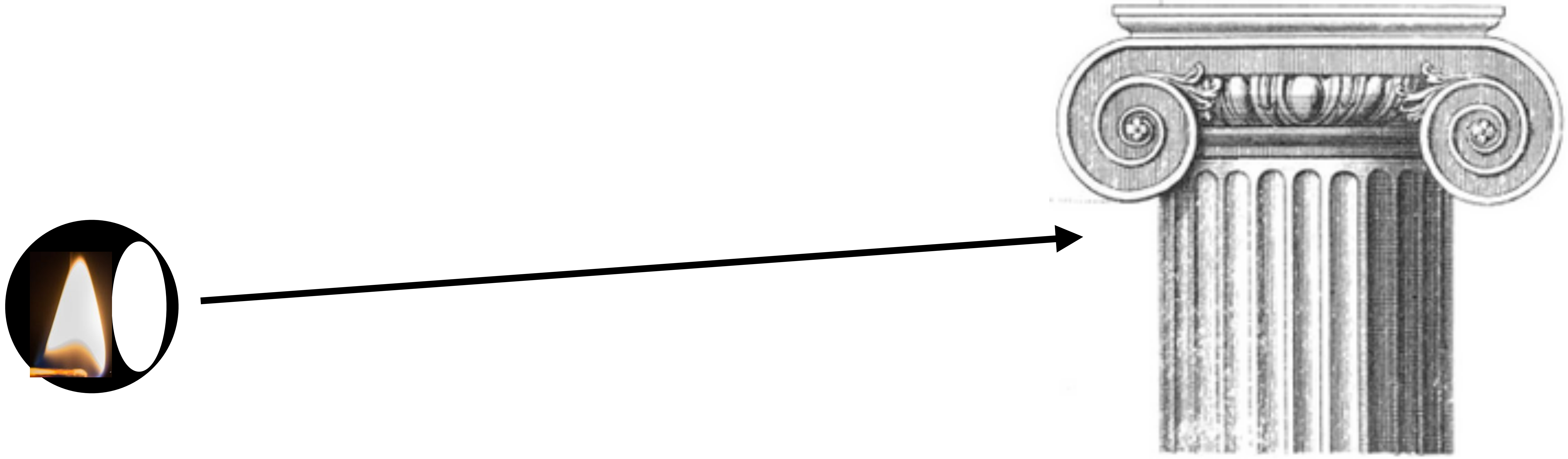
Intromission theory



Democritus (460 - 370 B.C)

The Greeks

Extramission (emission) theory



"So much of fire as would not burn, but gave a gentle light"
Plato

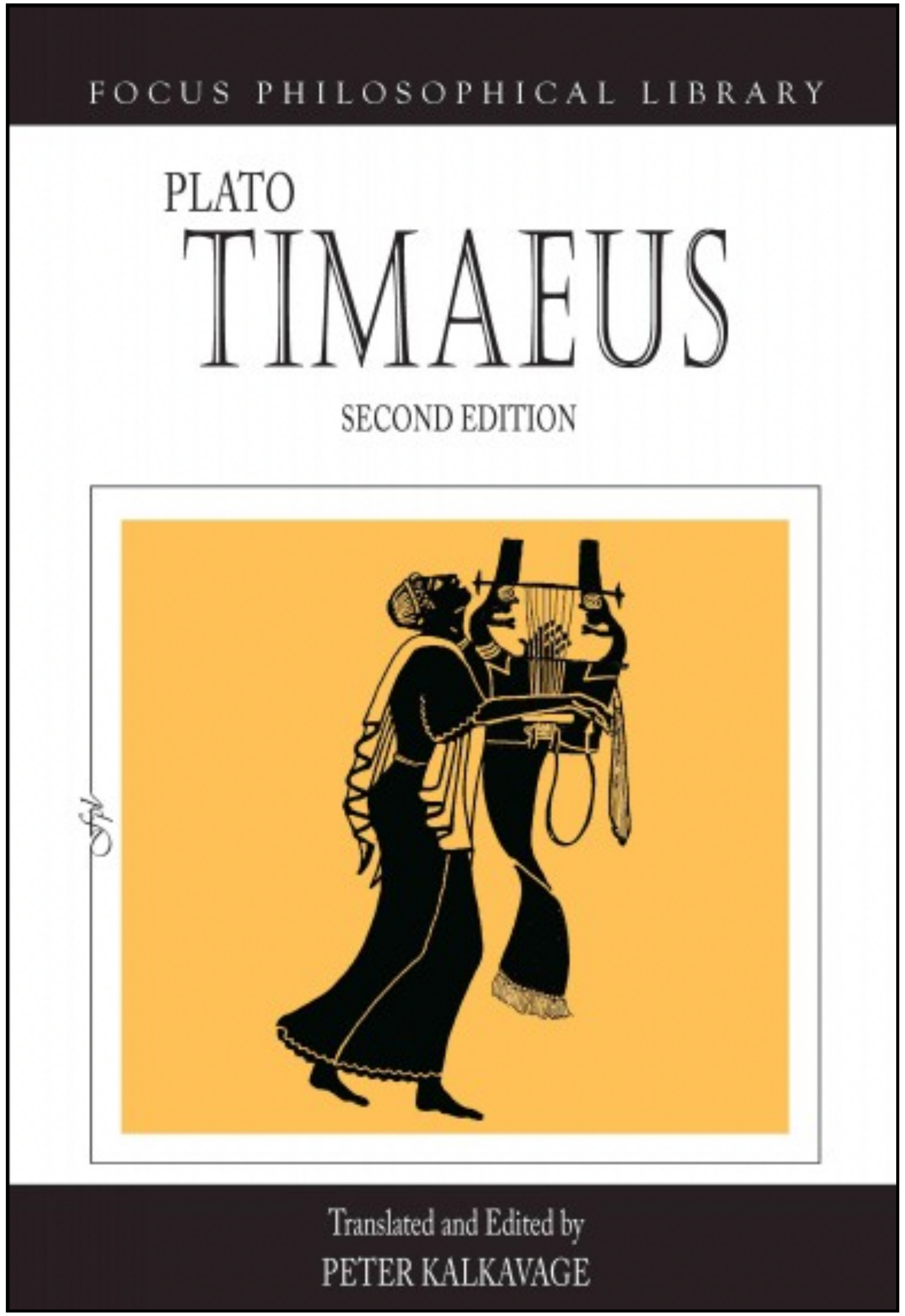
Empedocles (500 BC)

Plato (360 BC)

Extramission theory

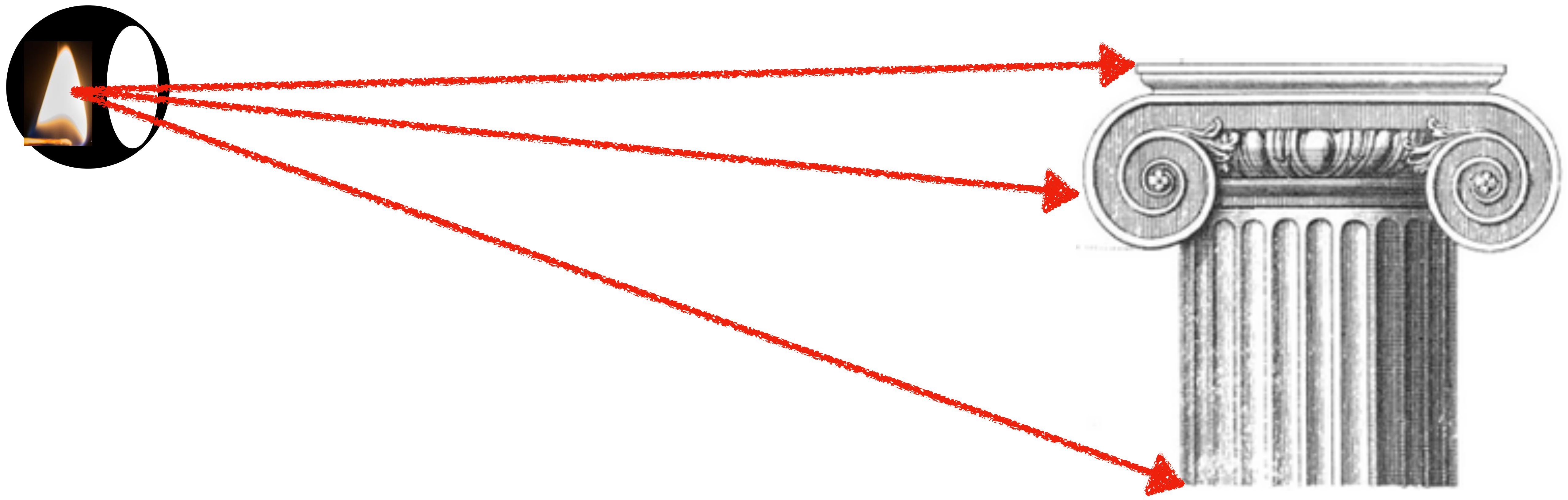
Plato’s theory of vision (427-347 BC)

“And of the organs they first contrived the eyes to give light, and the principle according to which they were inserted was as follows: So much of fire as would not burn, but gave a gentle light, they formed into a substance akin to the light of every-day life; and the pure fire which is within us and related thereto they made to flow through the eyes in a stream smooth and dense, compressing the whole eye, and especially the centre part, so that it kept out everything of a coarser nature, and allowed to pass only this pure element. When the light of day surrounds the stream of vision, then like falls upon like, and they coalesce, and one body is formed by natural affinity in the line of vision, wherever the light that falls from within meets with an external object.”



Euclid (325 BC)

“Let it be assumed that **lines draw directly from the eye** pass through a space of great extent; and that the form of the space included within our **vision is a cone**…” Euclid (translated by Burton)



Remarkable key idea: light travels in straight lines

JOURNAL OF THE OPTICAL SOCIETY OF AMERICA VOLUME 35, NUMBER 5 MAY, 1945

The Optics of Euclid[1]

Translated by Harry Edwin Burton
Dartmouth College, Hanover, New Hampshire

DEFINITIONS

1. Let it be assumed that lines drawn directly from the eye pass through a space of great extent;
2. and that the form of the space included within our vision is a cone, with its apex in the eye and its base at the limits of our vision;
3. and that those things upon which the vision falls are seen, and that those things upon which the vision does not fall are not seen;
4. and that those things seen within a larger angle appear larger, and those seen within a smaller angle appear smaller, and those seen within equal angles appear to be of the same size;
5. and that things seen within the higher visual range appear higher, while those within the lower range appear lower;
6. and, similarly, that those seen within the visual range on the right appear on the right, while those within that on the left appear on the left;
7. but that things seen within several angles appear to be more clear.

Nothing that is seen is seen at once in its entirety. (Fig. 1).

For let the thing seen be AD, and let the eye be B, from which let the rays of vision fall, BA, BG, BK, and BD. So, since the rays of vision, as they fall, diverge from one another, they could not fall in continuous line upon AD; so that there would be spaces also in AD upon which the rays of vision would not fall. So AD will not be seen in its entirety at the same time. But it seems to be seen all at once because the rays of vision shift rapidly.

Objects located nearby are seen more clearly than objects of equal size located at a distance. (Fig. 2.)

Let B represent the eye and let GD and KL represent the objects seen; and we must understand that they are equal and parallel, and let GD be nearer to the eye; and let the rays of vision fall, BG, BD, BK, and BL. For we could not say that the rays falling from the eye upon KL will pass through the points G and D. For in the triangle $BDLKGB$ the line KL would be longer than the line GD; but they are supposed to be of equal length. So GD is seen by more rays of the eye than KL. So GD will appear more clear than KL; for objects seen within more angles appear more clear.

Every object seen has a certain limit of distance, and when this is reached it is seen no longer. (Fig. 3.)

For let the eye be B, and let the object seen be GD. I say that GD, placed at a certain distance, will be seen no longer. For let GD lie midway in the divergence of the rays, at the limit of which is K. So, none of the rays from B will fall upon K. And the thing upon which rays do not fall is not seen. Therefore, every object seen has a certain limit of distance, and, when this is reached, the object is seen no longer.

Of equal spaces located upon the same straight line, those seen from a greater distance appear shorter. (Fig. 4.)

Let AB, BG, and GD represent equal spaces upon one straight line, and let the perpendicular AE be drawn, upon which let E represent the eye. I say that AB will appear longer than BG and BG longer than GD. For let the rays fall, EB, EG, and ED, and through the point B let BZ be drawn parallel to the straight line GE. Now AZ is equal to EZ. For, since parallel to GE, one side of the triangle AEG, the straight line BZ has been drawn, it follows also that EZ is related to ZA as GB to BA. So, as has been said AZ is equal to ZE. But the side BZ is longer than ZA; and so,

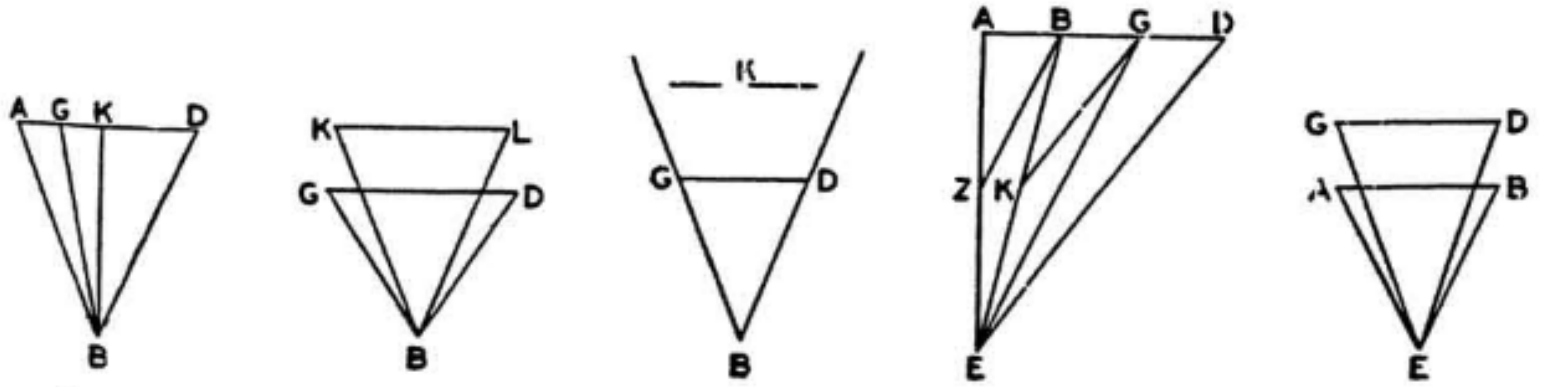


FIG. 1. FIG. 2. FIG. 3. FIG. 4. FIG. 5.

[1] Professor Charles N. Haskins, Professor of Mathematics at Dartmouth College, is largely responsible for this translation of the *Optics* of Euclid. A year ago, when he was doing research for the Dartmouth Eye Institute, he had occasion to use Euclid's essay and asked me if I would translate it. Strangely enough, it had never been translated into English. I agreed to undertake the task. Before the work was finished Professor Haskins died. The Dartmouth Eye Institute decided that the translation should be completed and published, and I wish to express my own gratitude to the Optical Society of America for its cooperation.

Euclid was a teacher of mathematics at Alexandria in the early part of the third century before Christ. Almost nothing is known of his life. He was a voluminous writer on mathematics and kindred subjects, his principal work being the Elements of Geometry in thirteen books. The *Optics* is an essay on the mathematics of optics. It is extant in two forms, one written by Euclid himself, the other a recension by Theon, written in the fourth century after Christ. In its original form, which is here translated, it is the earliest extant work on mathematical optics.

[Professor Burton died on March 20, 1945, before his translation of the *Optics* was set up in galley proof. His colleagues, in seeing the translation through the press, have endeavored only to secure in printed form the exact reproduction of Professor Burton's typewritten manuscript.]

http://philomatica.org/wp-content/uploads/2013/01/Optics-of-Euclid.pdf

MASSACHUSETTS INSTITUTE OF TECHNOLOGY
PROJECT MAC

Artificial Intelligence Group July 7, 1966
Vision Memo. No. 100.

THE SUMMER VISION PROJECT

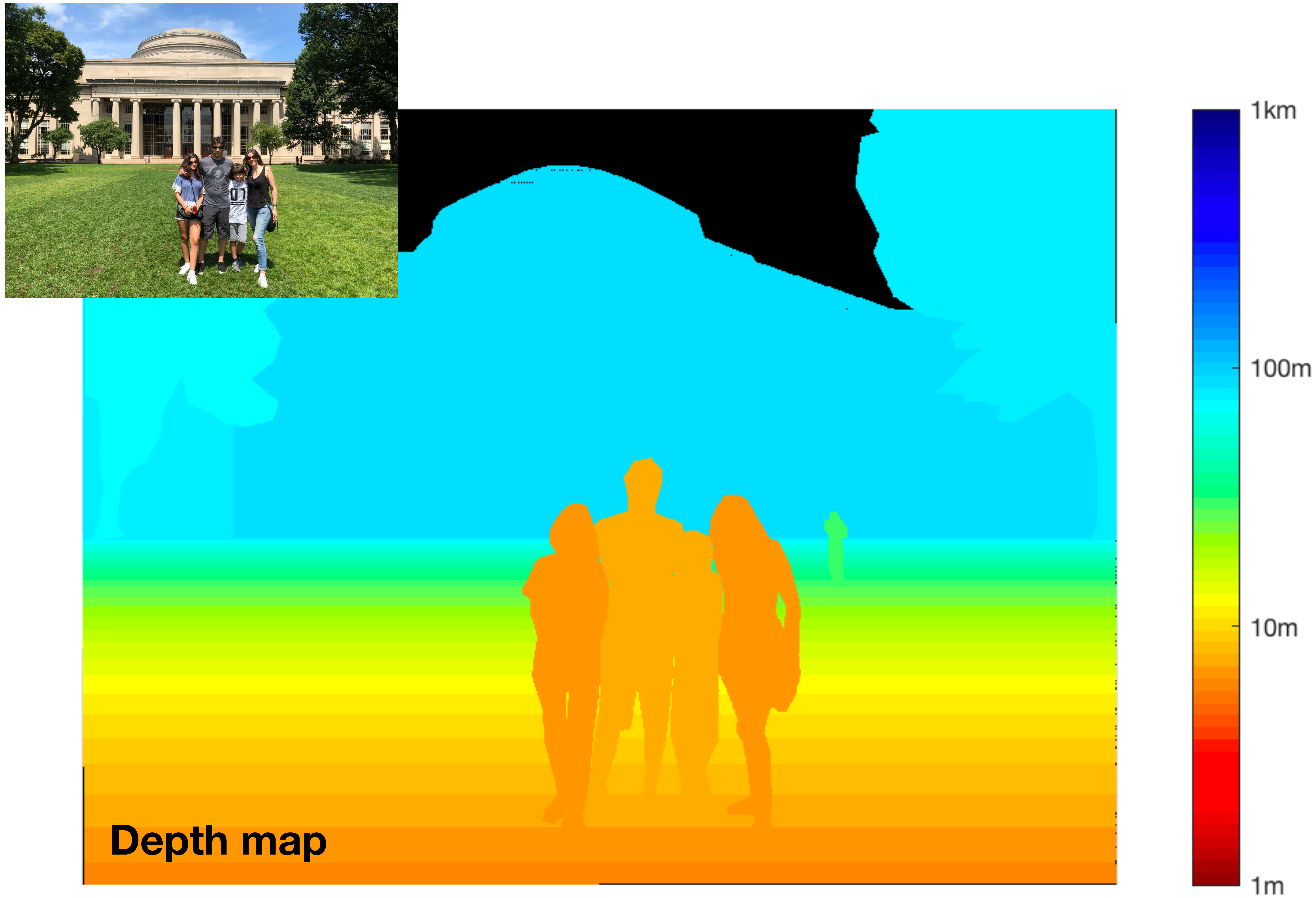
Seymour Papert

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".

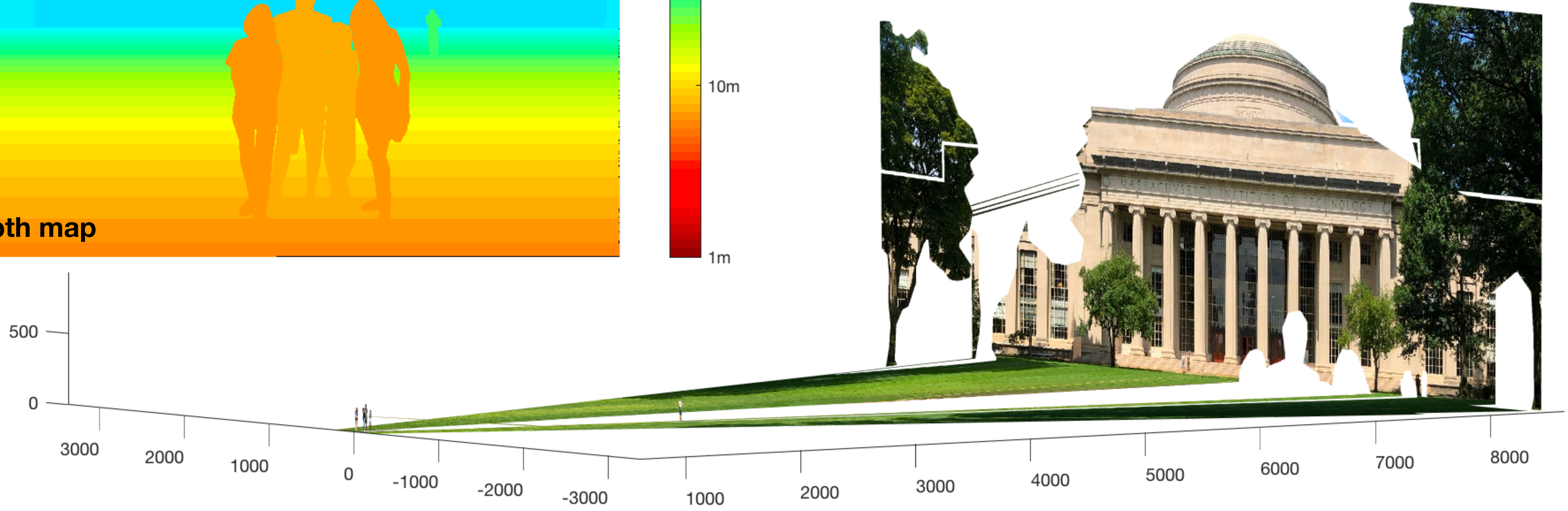
The goal of the first lecture
and pset1 is
to solve vision

Task: given a picture…

… recover the 3D scene structure



3D



A Simple Visual System

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

A Simple World

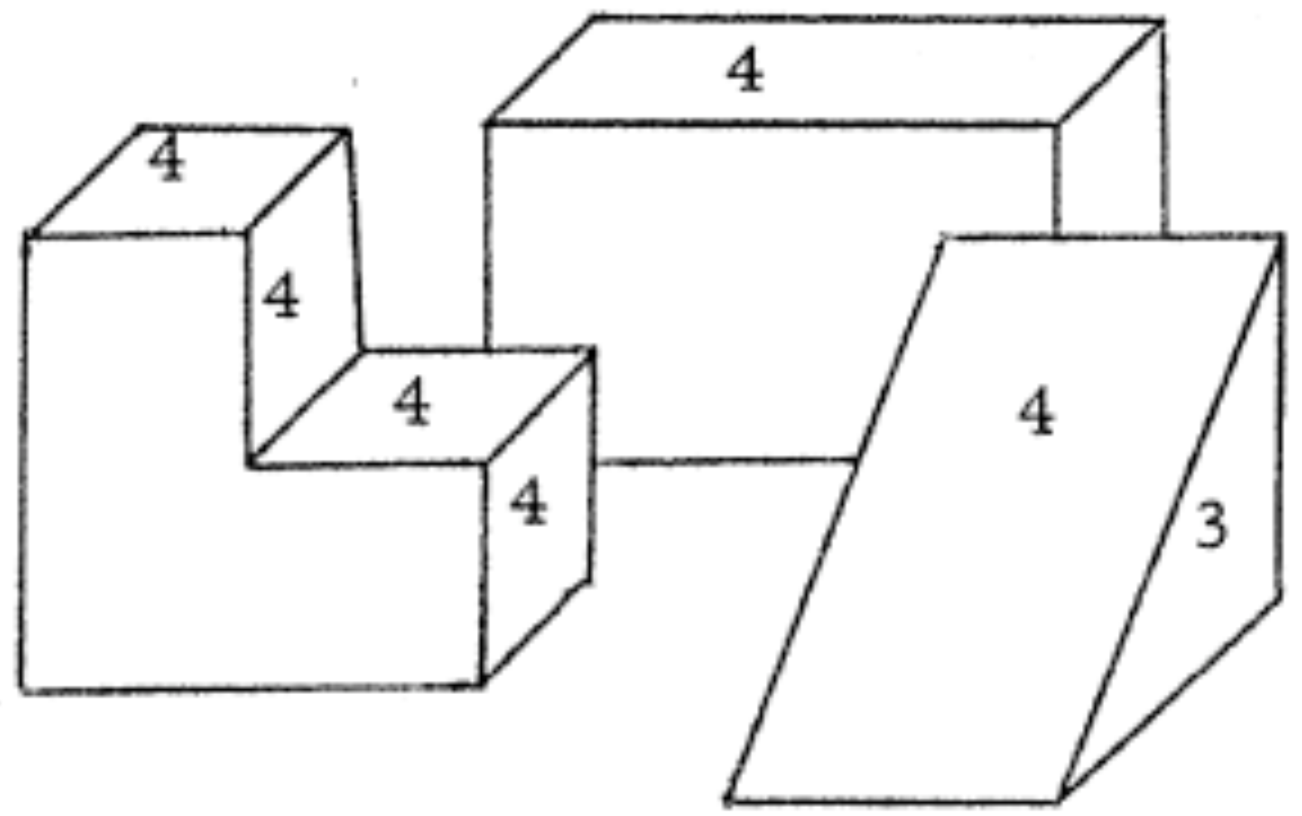
A Simple World

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.



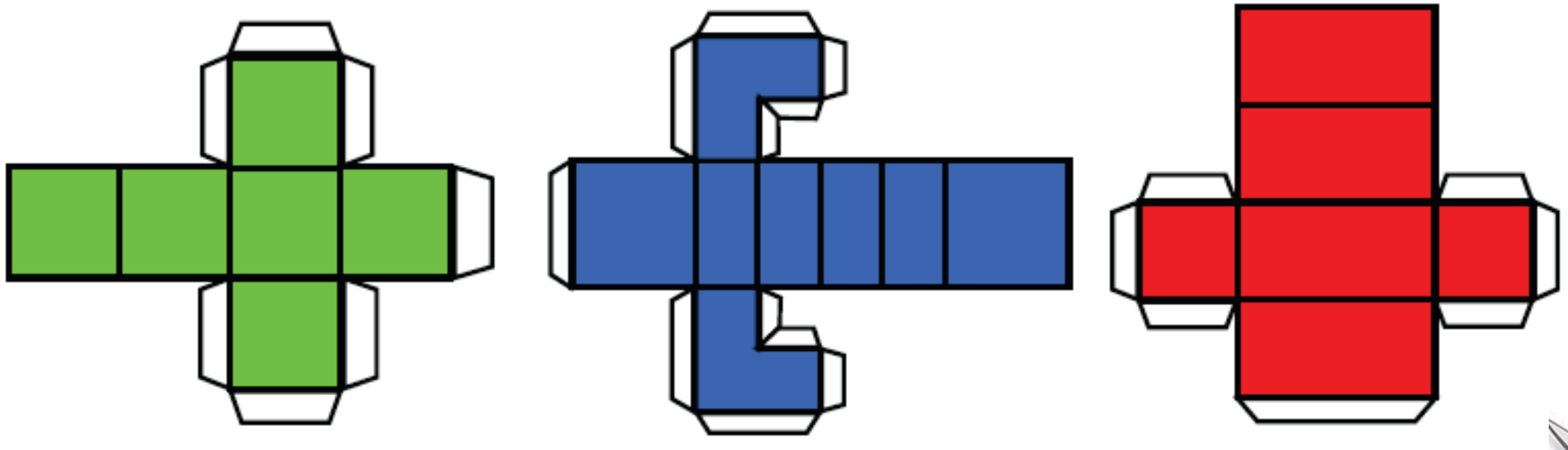
Complete Convex Polygons. The polygon selection procedure would select the numbered polygons as complete and convex. The number indicates the probable number of sides. A polygon is incomplete if one of its points is a collinear joint of another polygon.

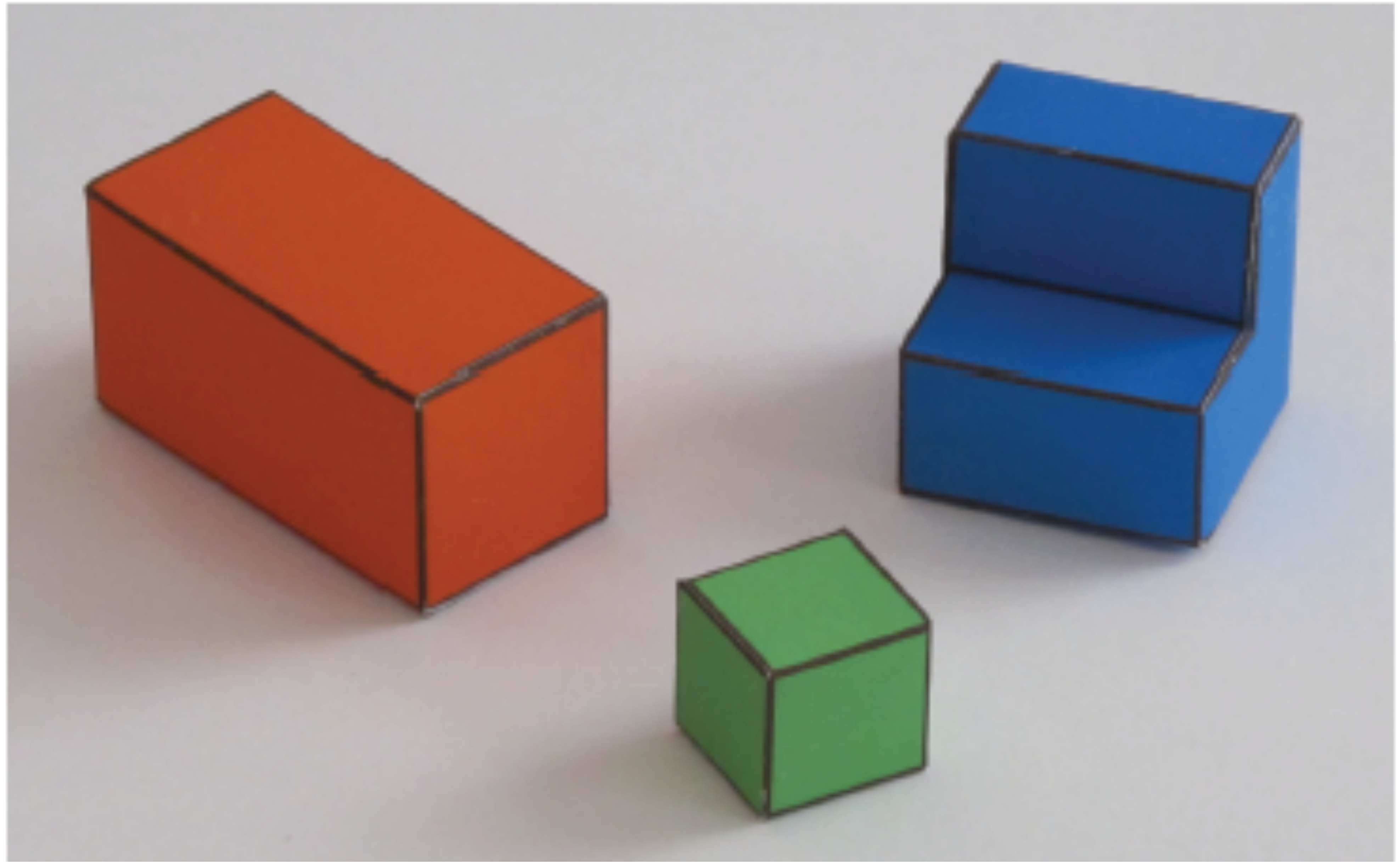
The problem of machine recognition of pictorial data has long been a challenging goal, but has seldom been attempted with anything more complex than alphabetic characters. Many people have felt that research on character recognition would be a first step, leading the way to a more general pattern recognition 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.

… first computer vision PhD

http://www.packet.cc/files/mach-per-3D-solids.html

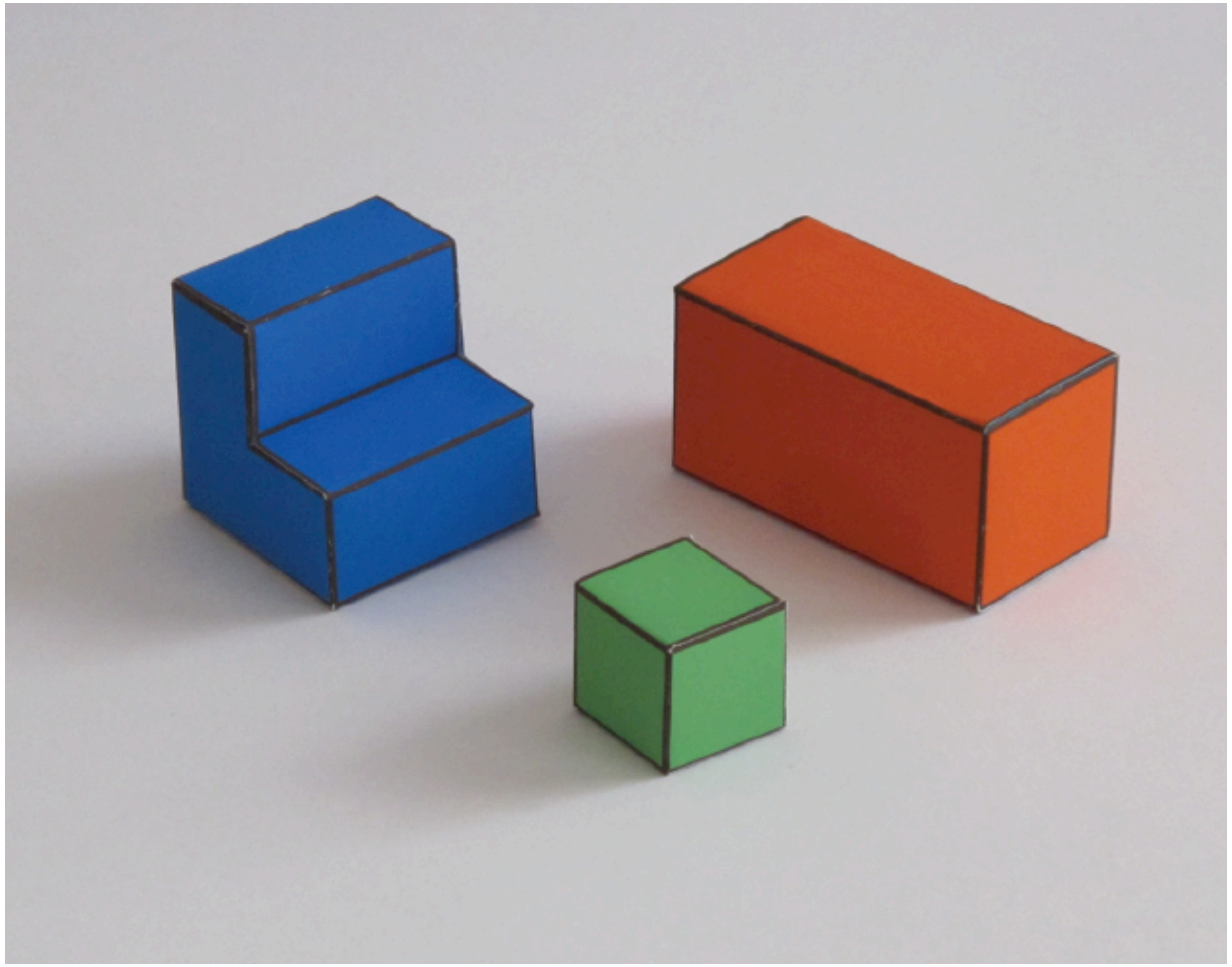
Build your own simple world





A simple goal

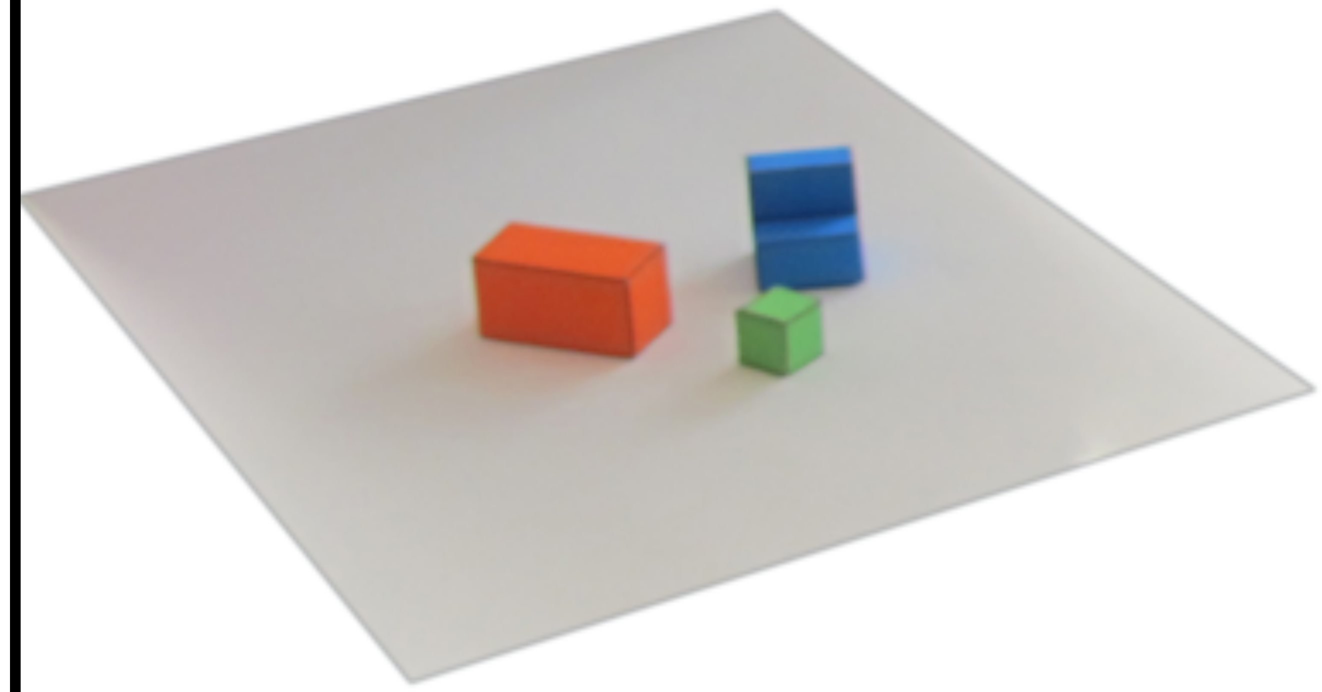
To recover the 3D structure of the world from the 2D image



We will make this goal more explicit later.

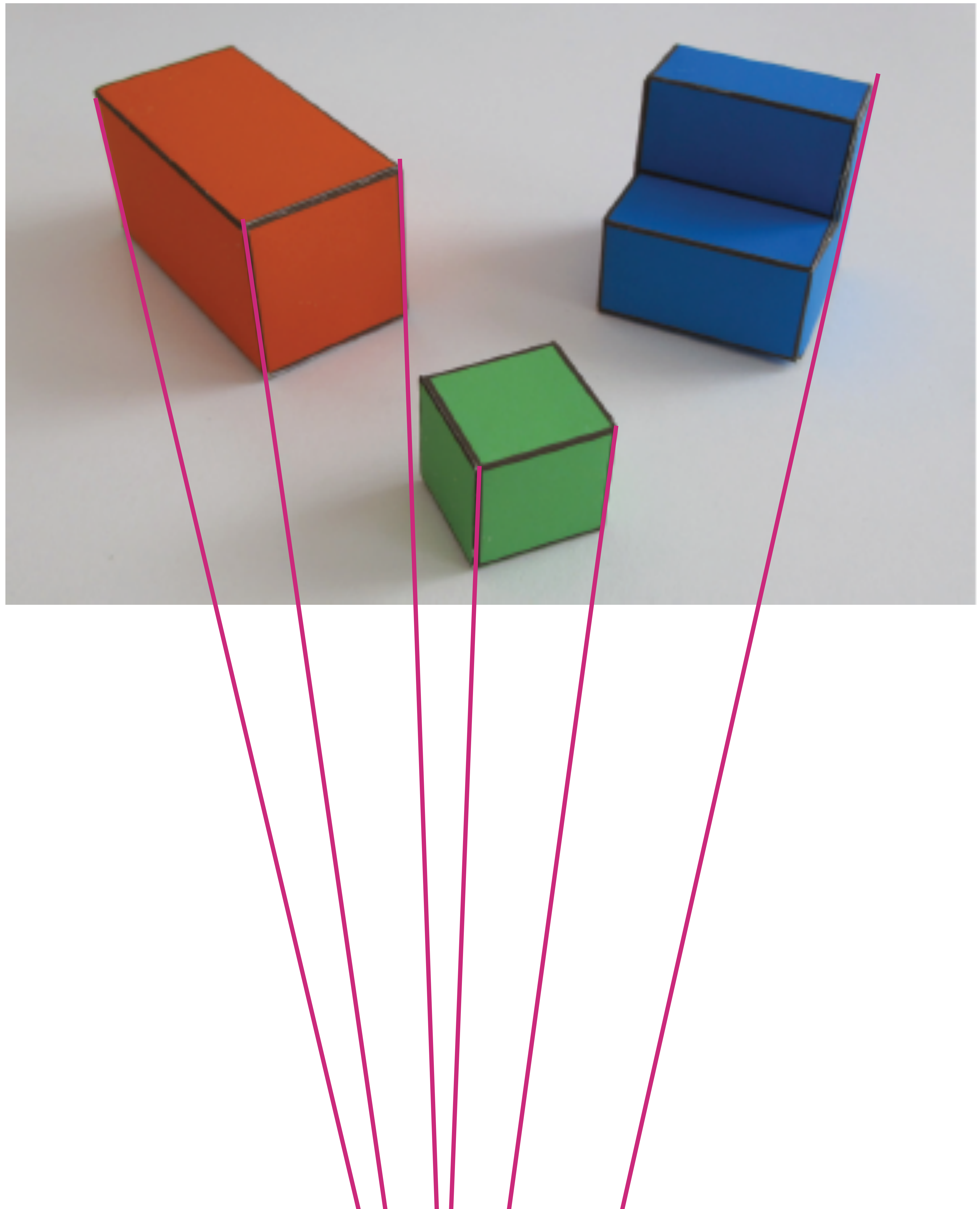
A simple image formation model

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

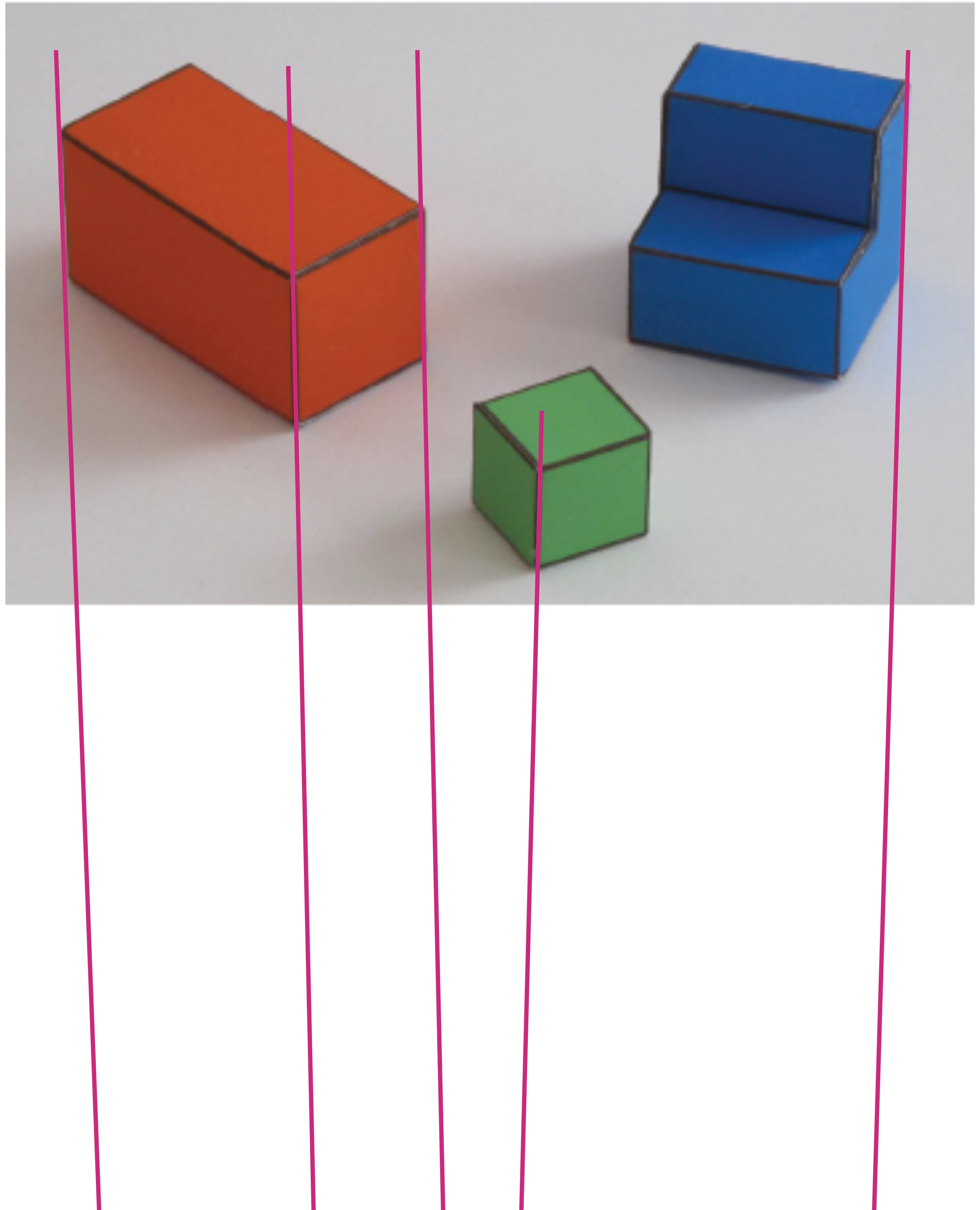


A simple image formation model

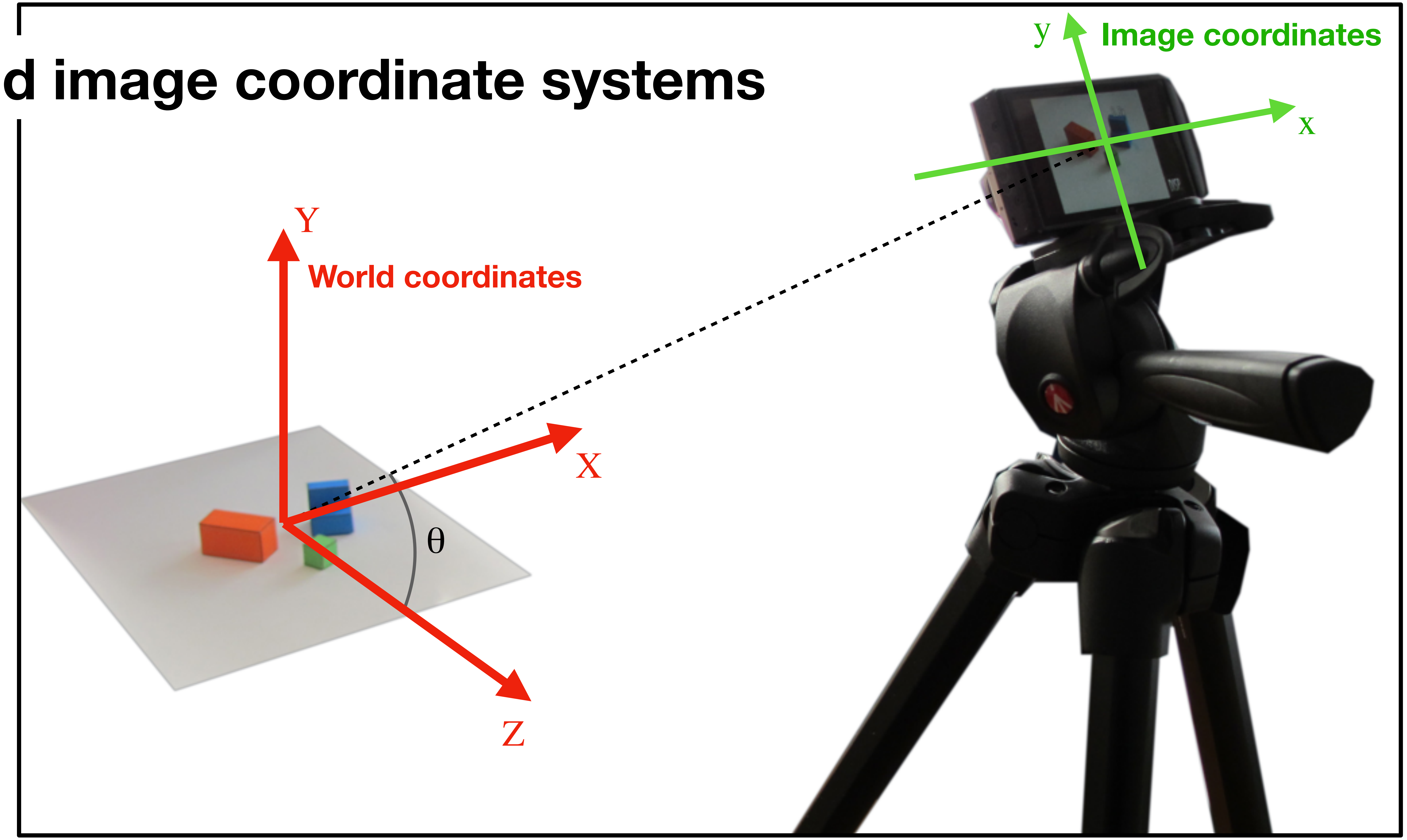
Perspective projection



Parallel (orthographic) projection

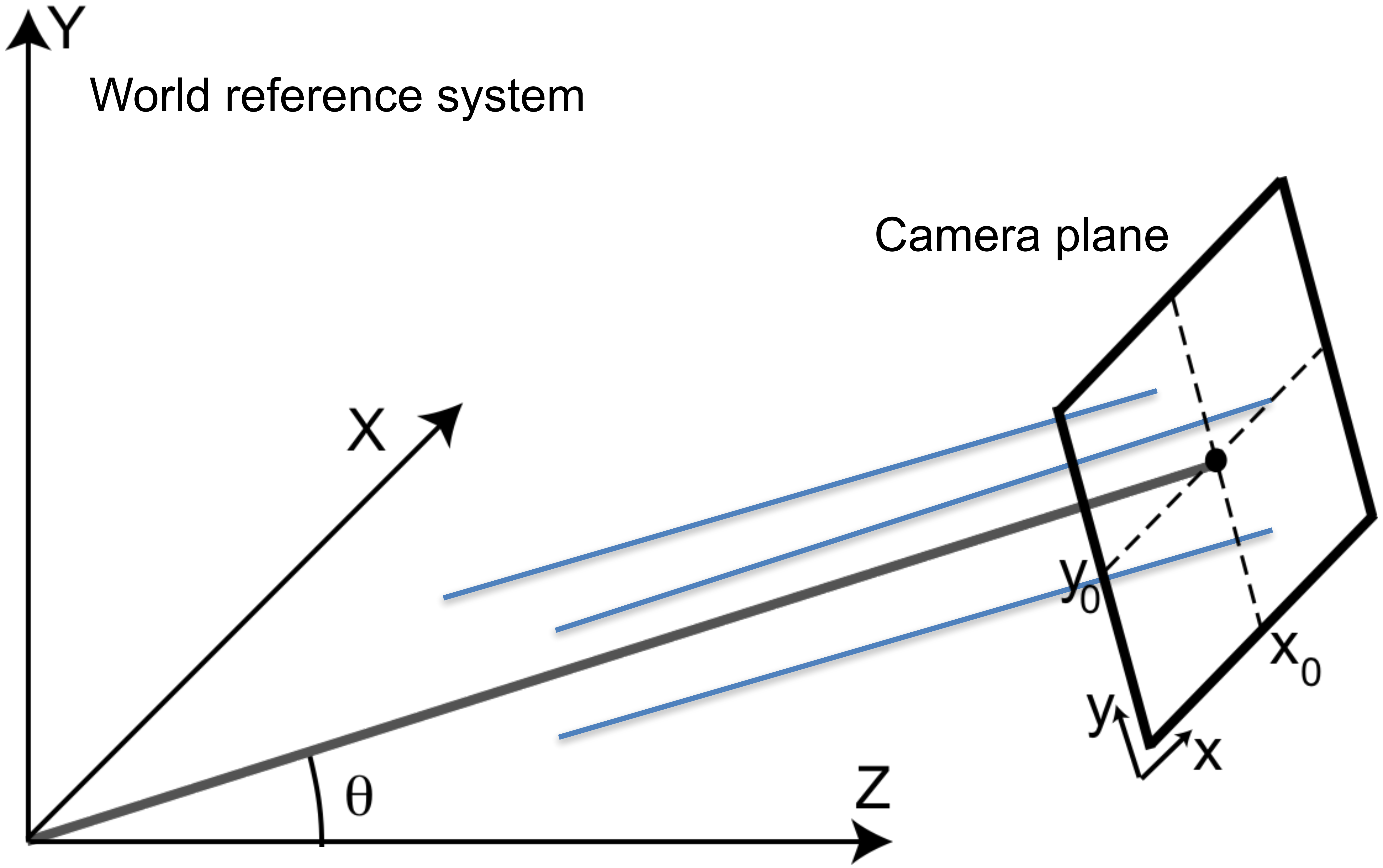


A simple image formation model



(right-handed reference system)

A simple image formation model



(right-handed reference system)

A simple image formation model

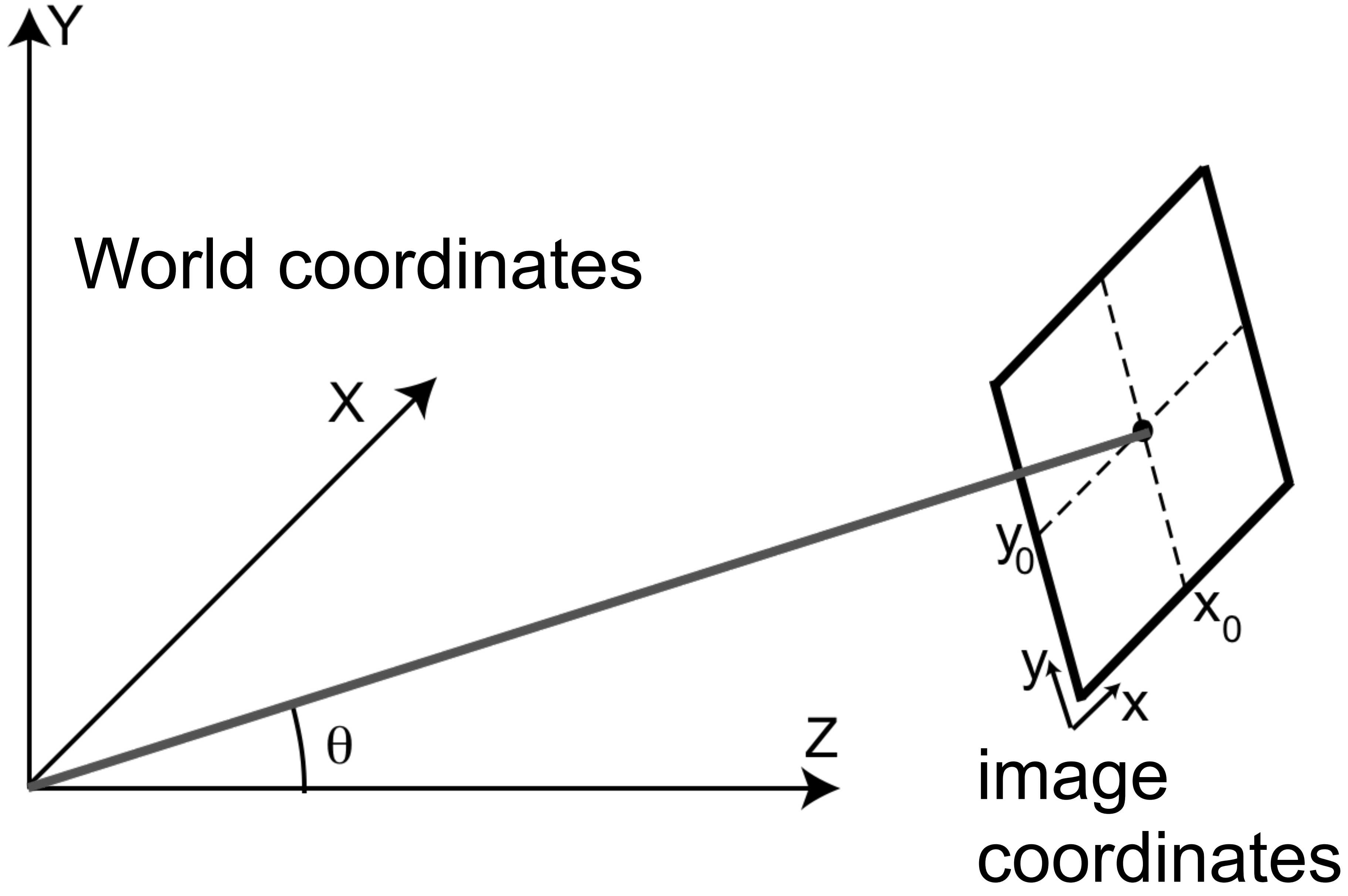
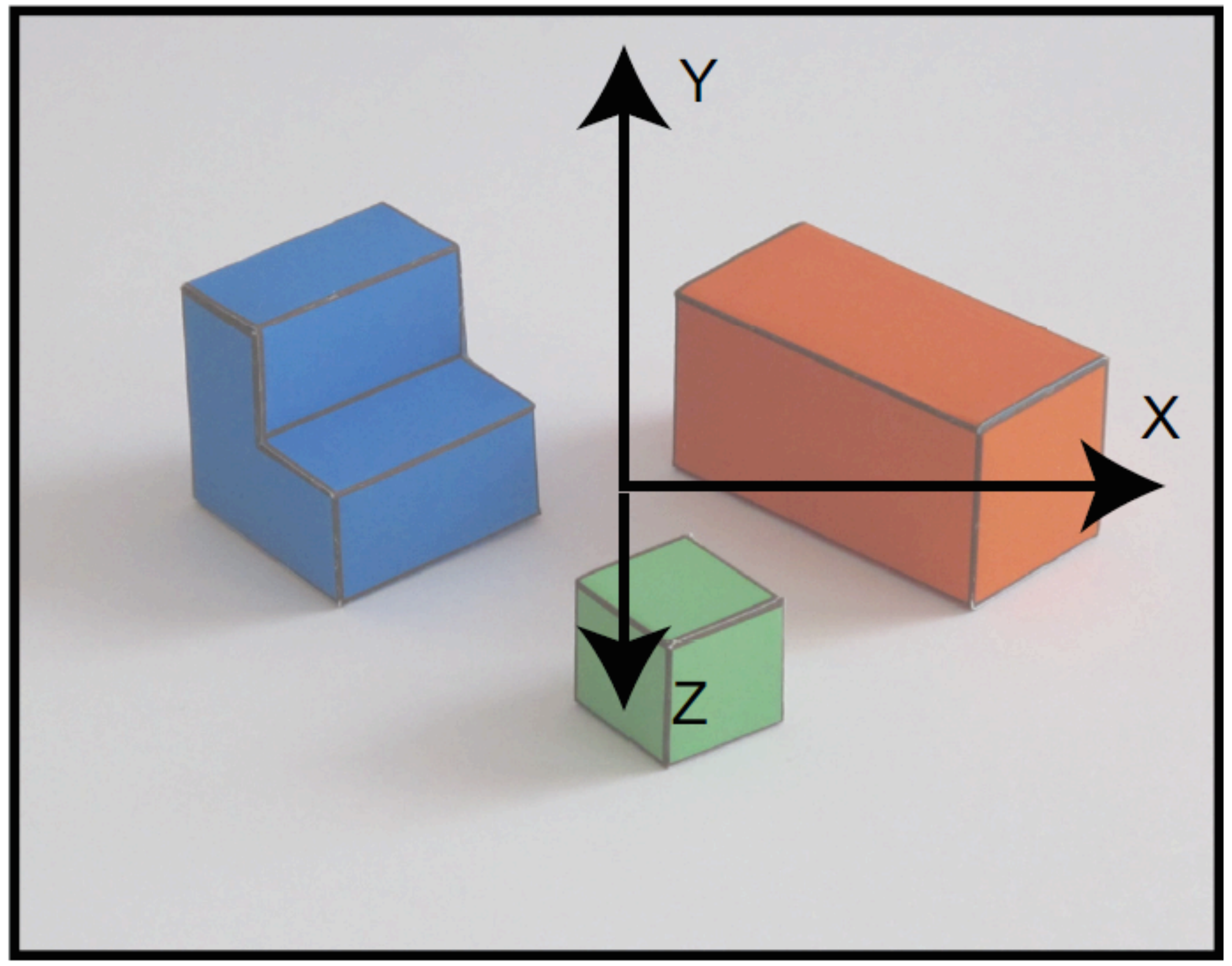


Image and projection of the world coordinate axes into the image plane



World coordinates

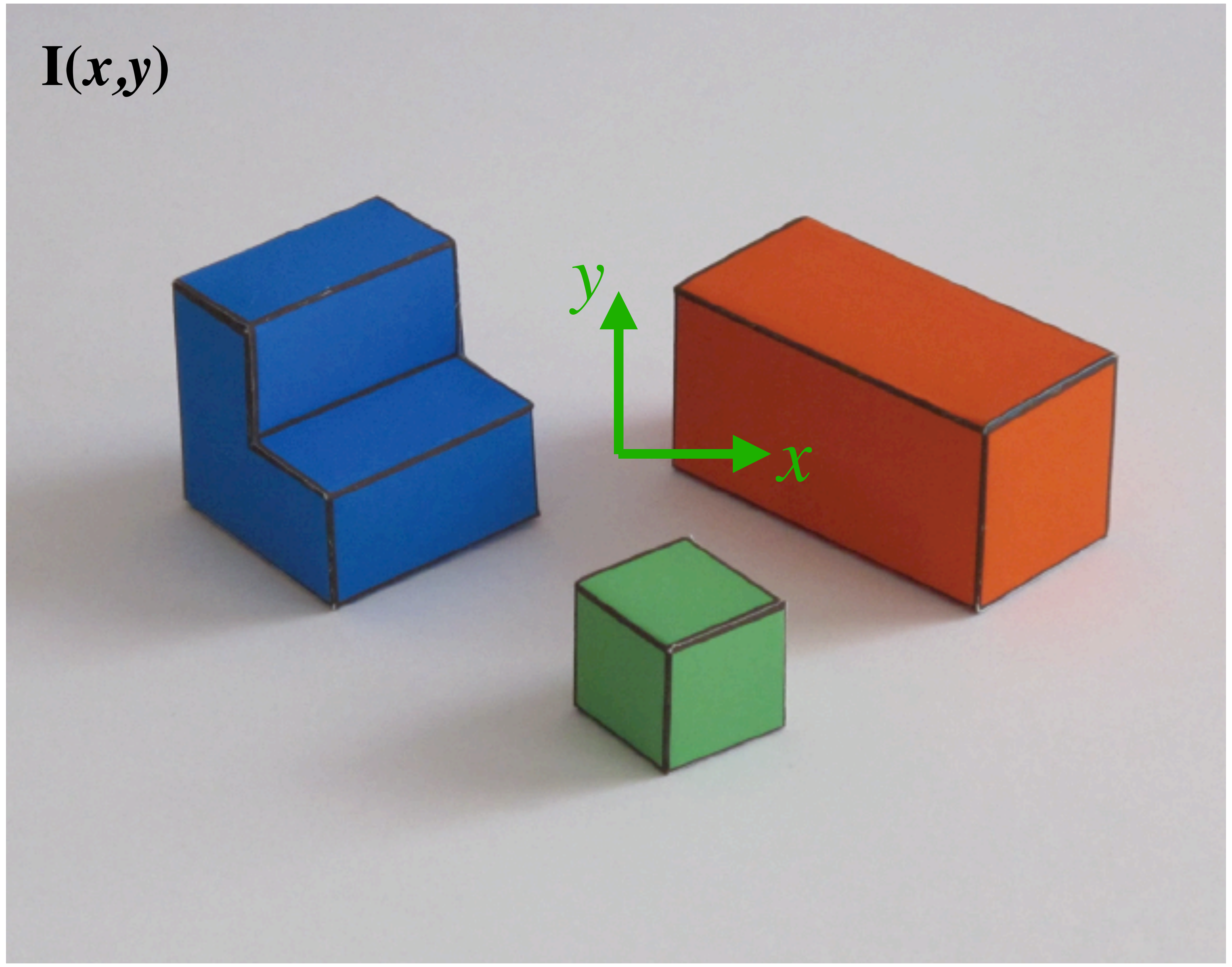
$$x = X + x_0$$

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

image coordinates

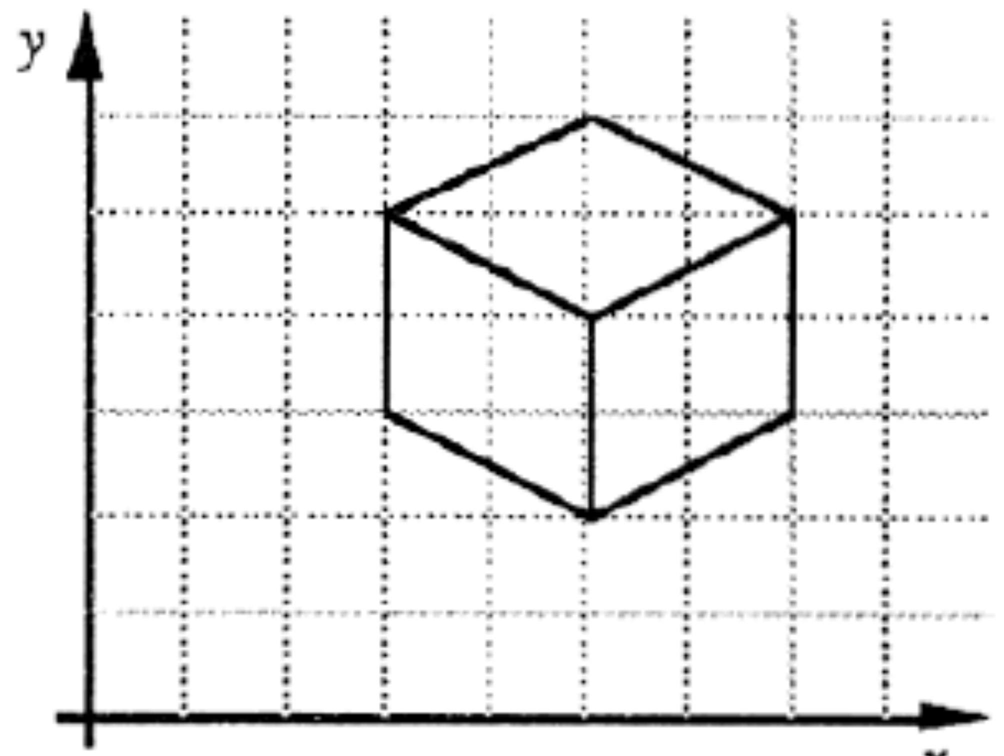
A simple goal

To recover the 3D structure of the world from the 2D image



We want to recover $\mathbf{X}(x,y)$, $\mathbf{Y}(x,y)$, $\mathbf{Z}(\mathbf{x},\mathbf{y})$ using as input $\mathbf{I}(x,y)$

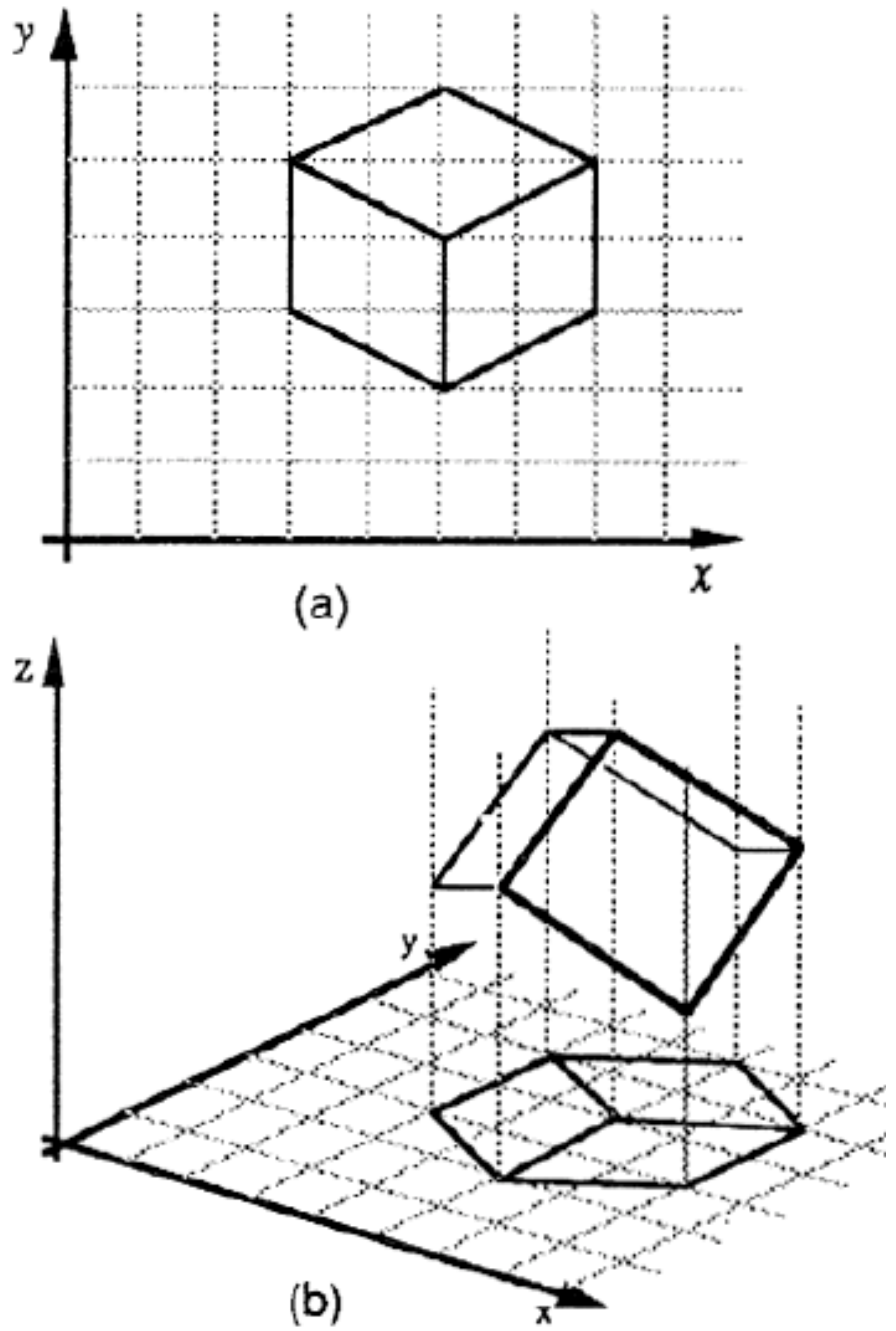
Why is this hard?



(a)

Sinha & Adelson 93

Why is this hard?



Sinha & Adelson 93

Why is this hard?

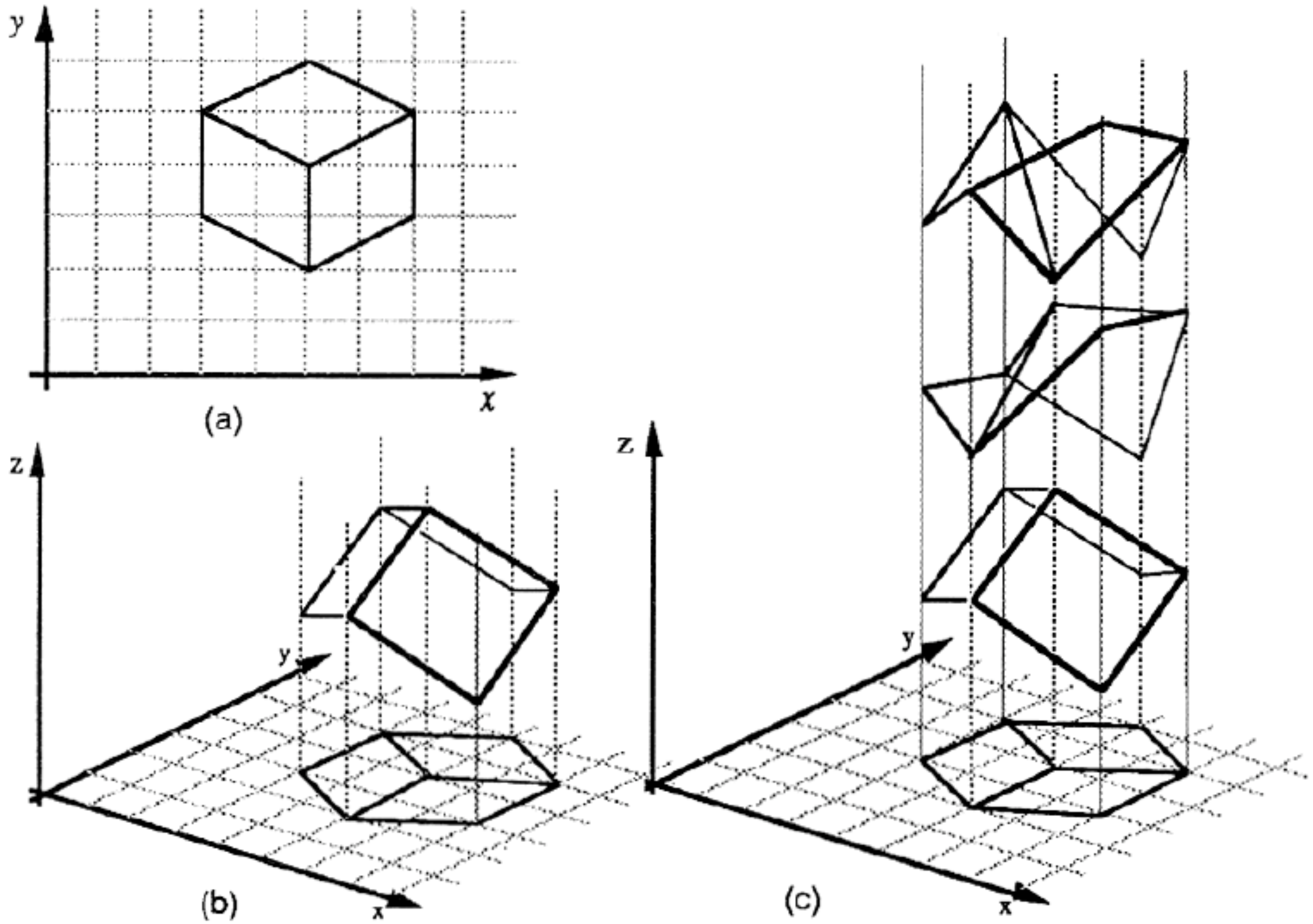
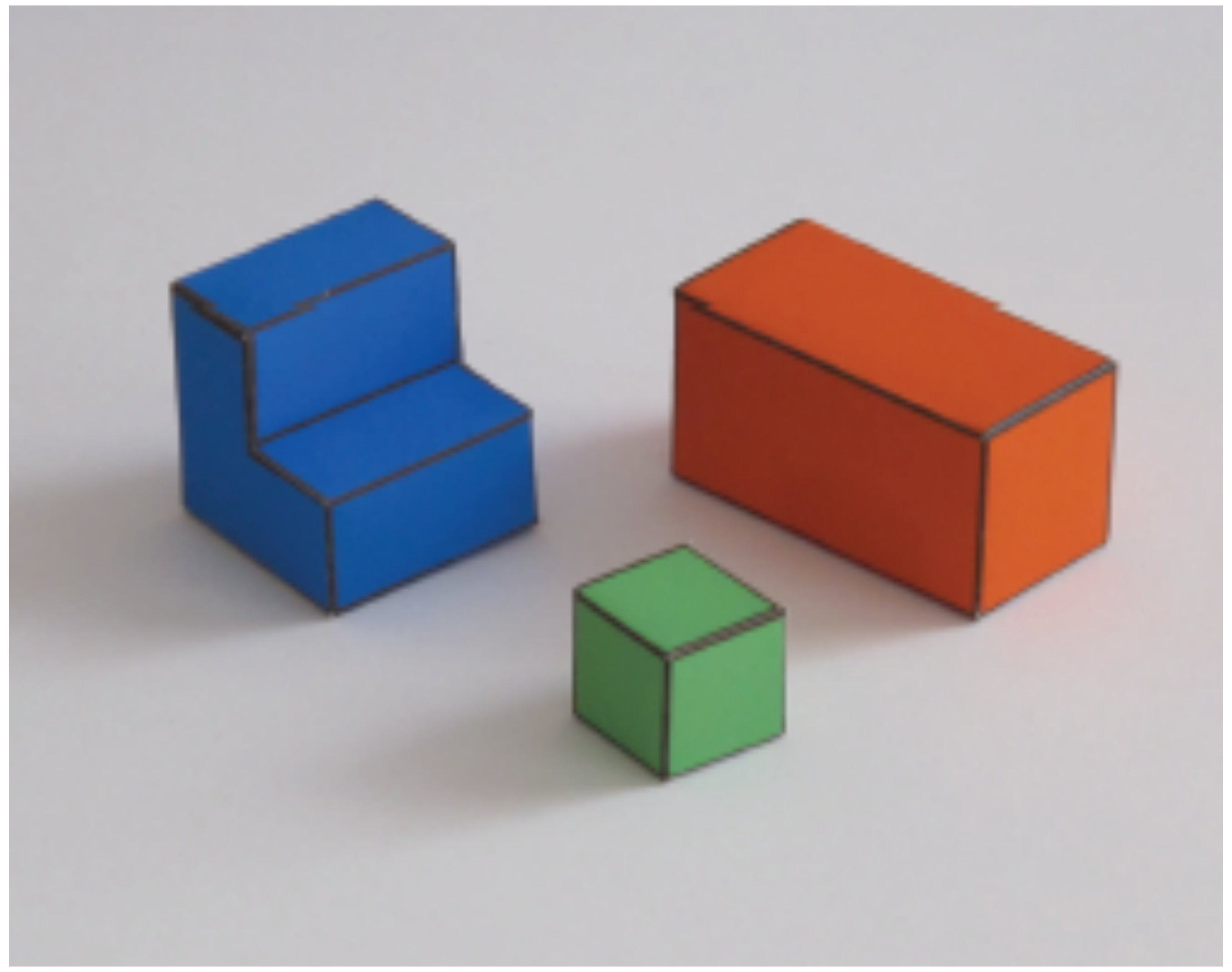
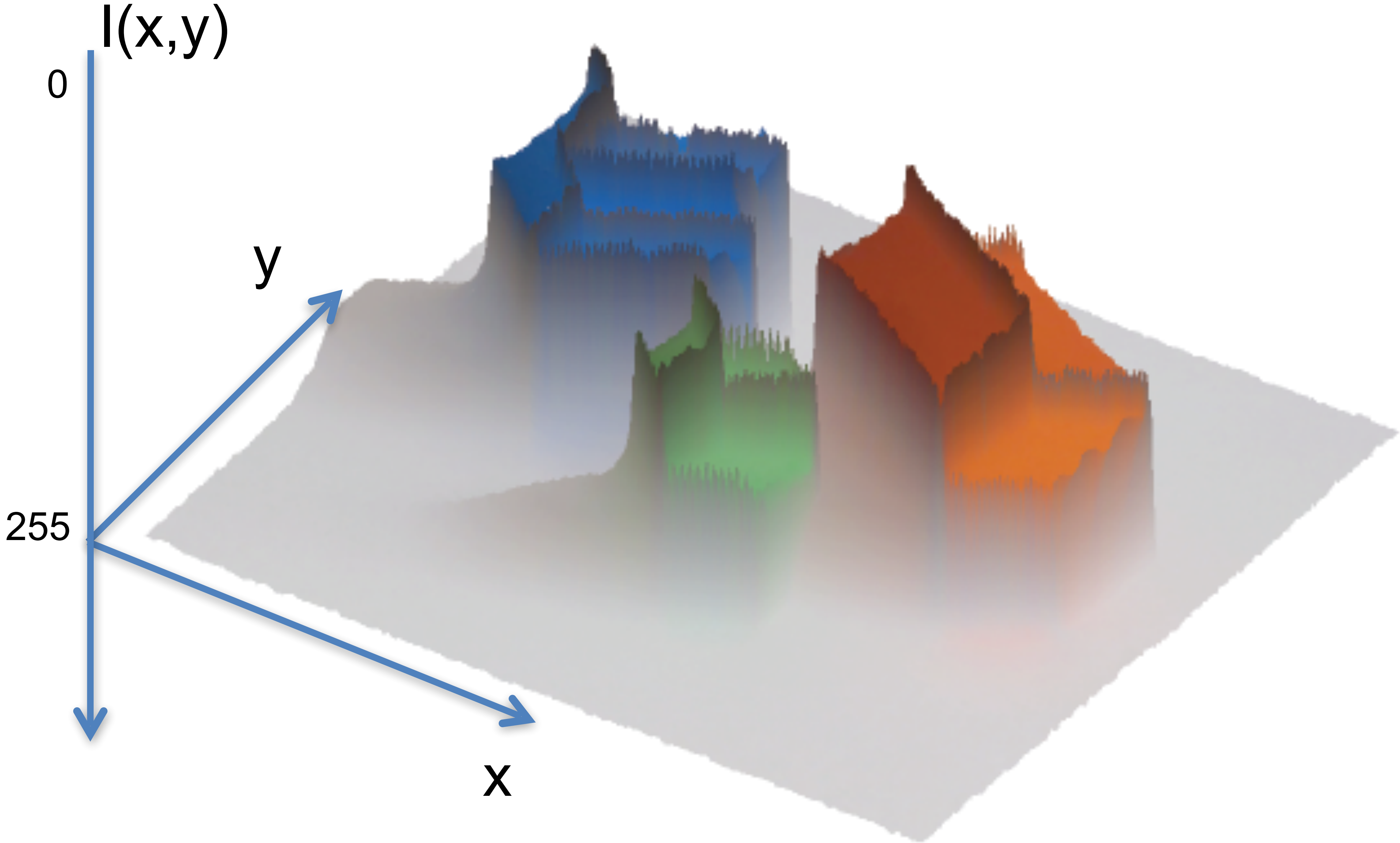


Figure 1. (a) A line drawing provides information only about the x, y coordinates of points lying along the object contours. (b) The human visual system is usually able to reconstruct an object in three dimensions given only a single 2D projection (c) Any planar line-drawing is geometrically consistent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

Sinha & Adelson 93

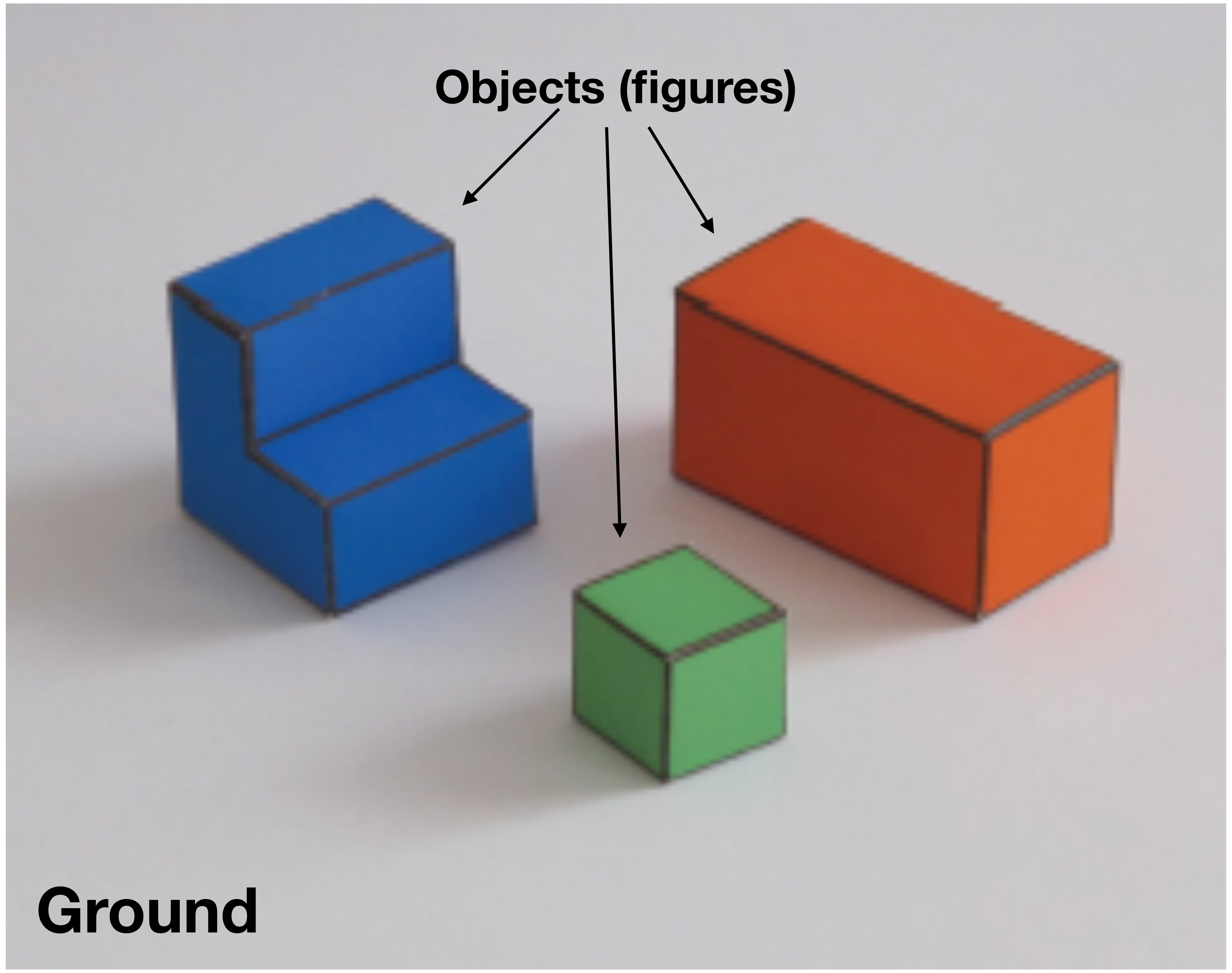
A simple visual system
The input image

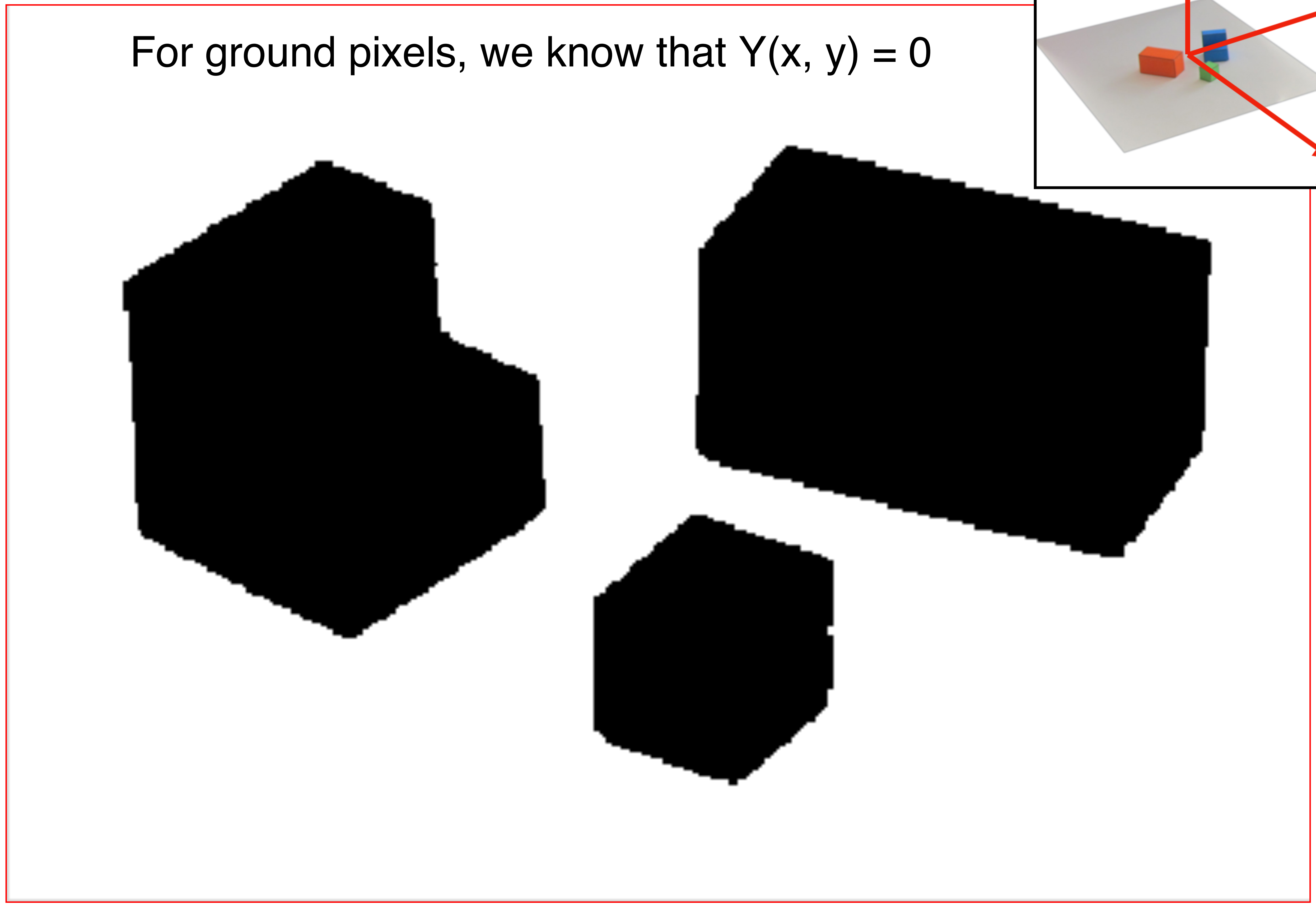


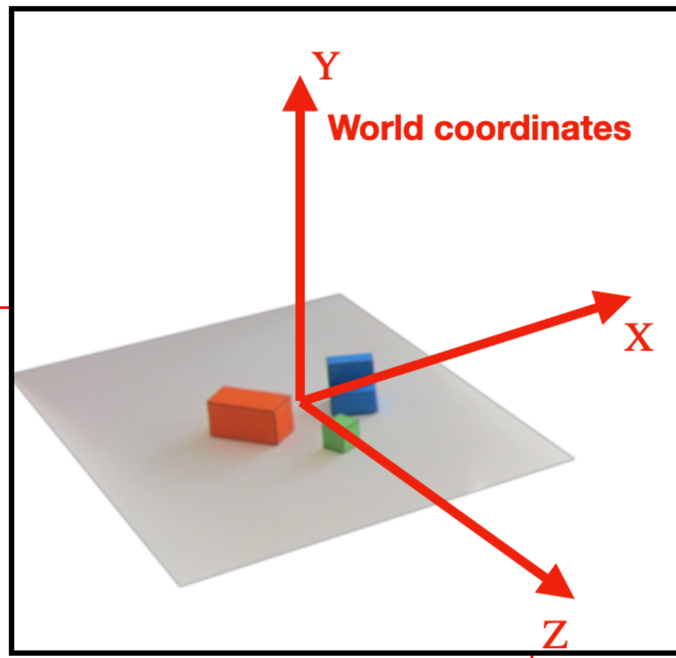


In this **representation**, the image is an array of intensity values (color values) indexed by location.

A better representation: Figure/ground





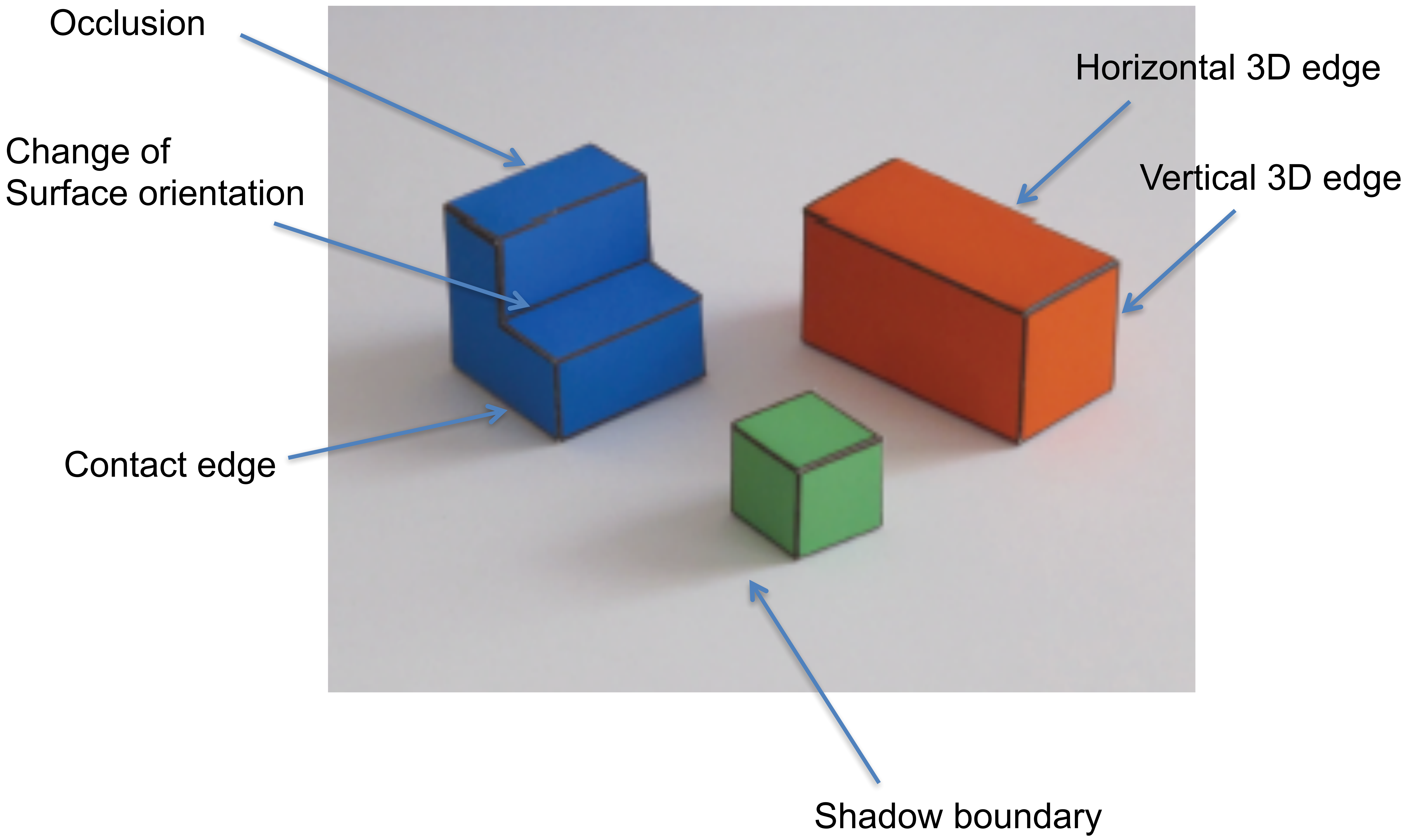


In our simple world:
Using the fact that objects have color and are darker than the ground.

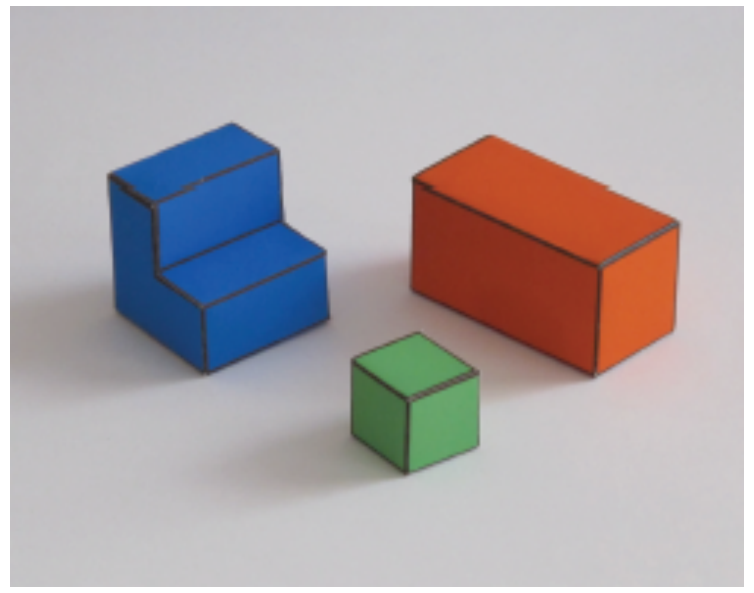
Figure/ground segmentation

classical visual illusion
"*two faces or a vase*"

A better representation: Edges



Finding edges in the image



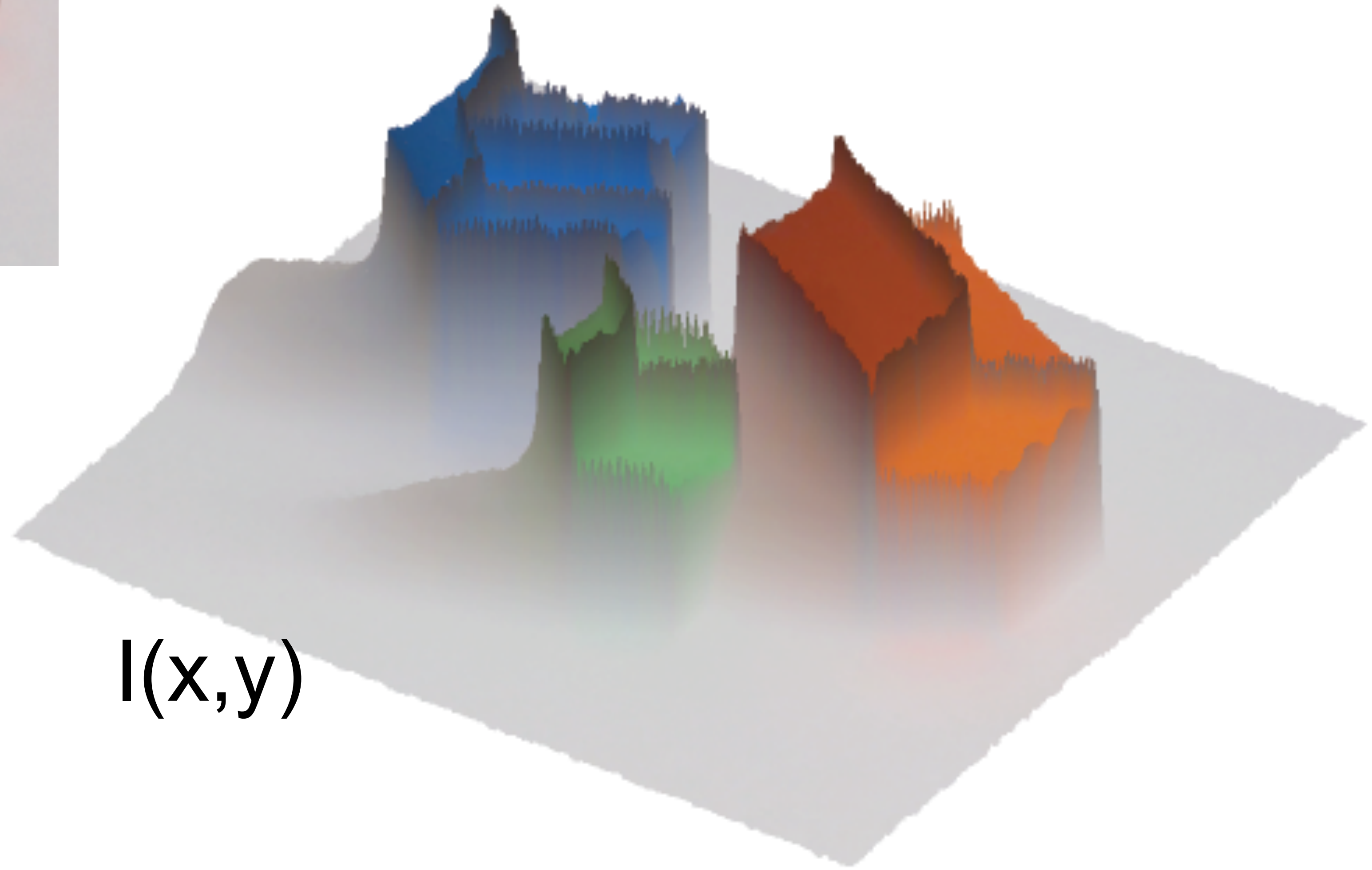


Image gradient:

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

Approximation image derivative:

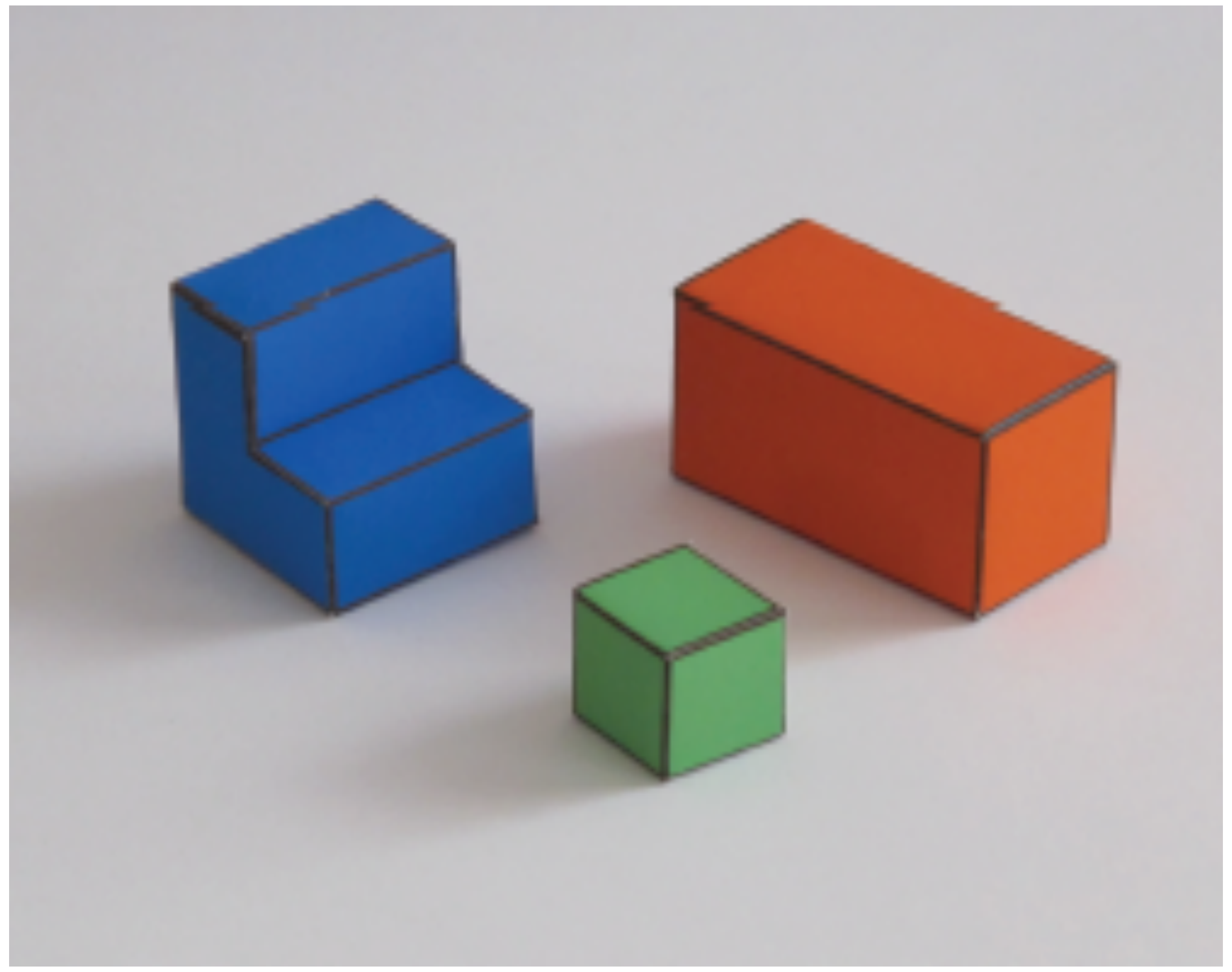
$$\frac{\partial \mathbf{I}}{\partial x} \simeq \mathbf{I}(x, y) - \mathbf{I}(x-1, y)$$

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

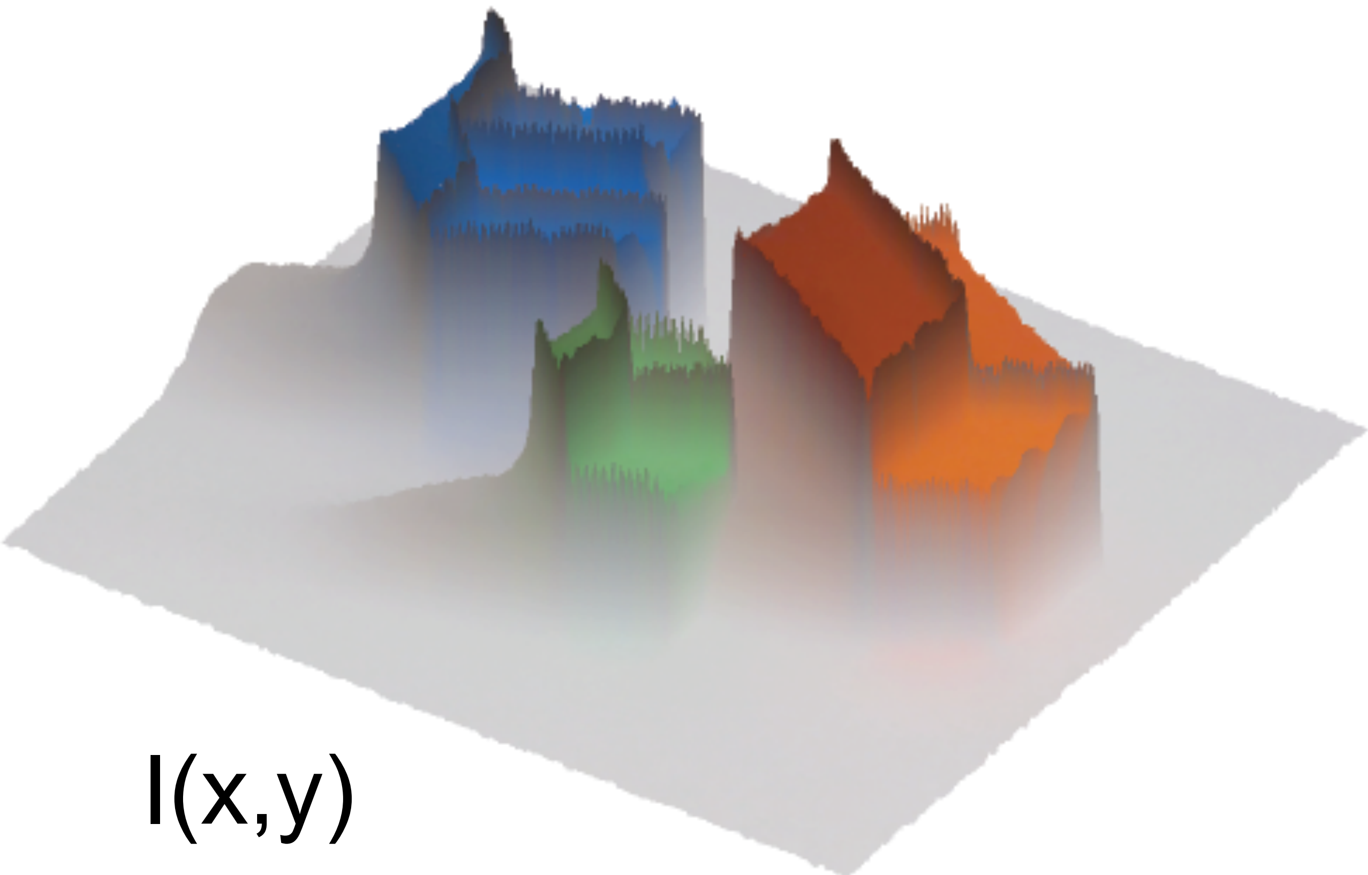
Edge orientation: $\theta(x, y) = \angle \nabla \mathbf{I} = \arctan \frac{\partial \mathbf{I}/\partial y}{\partial \mathbf{I}/\partial x}$

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

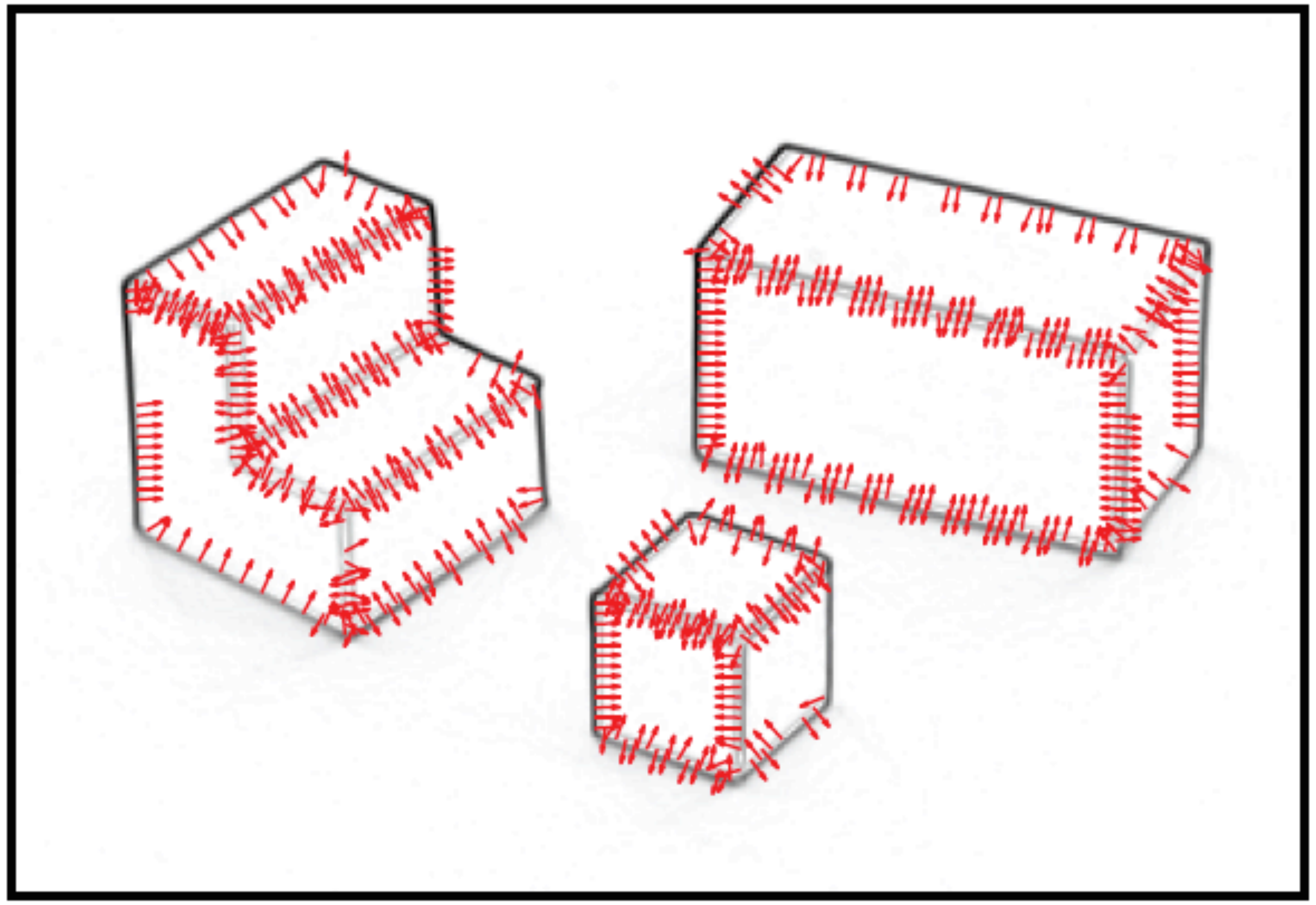
Finding edges in the image



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



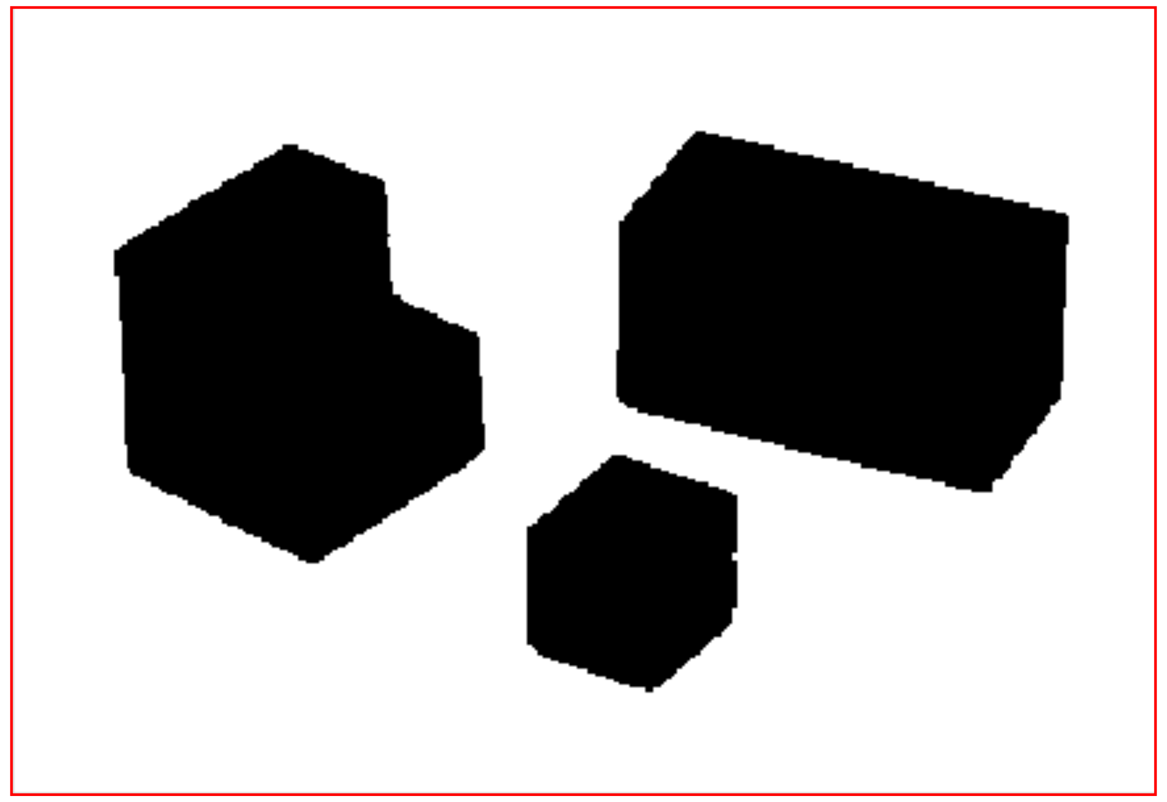
I(x,y)

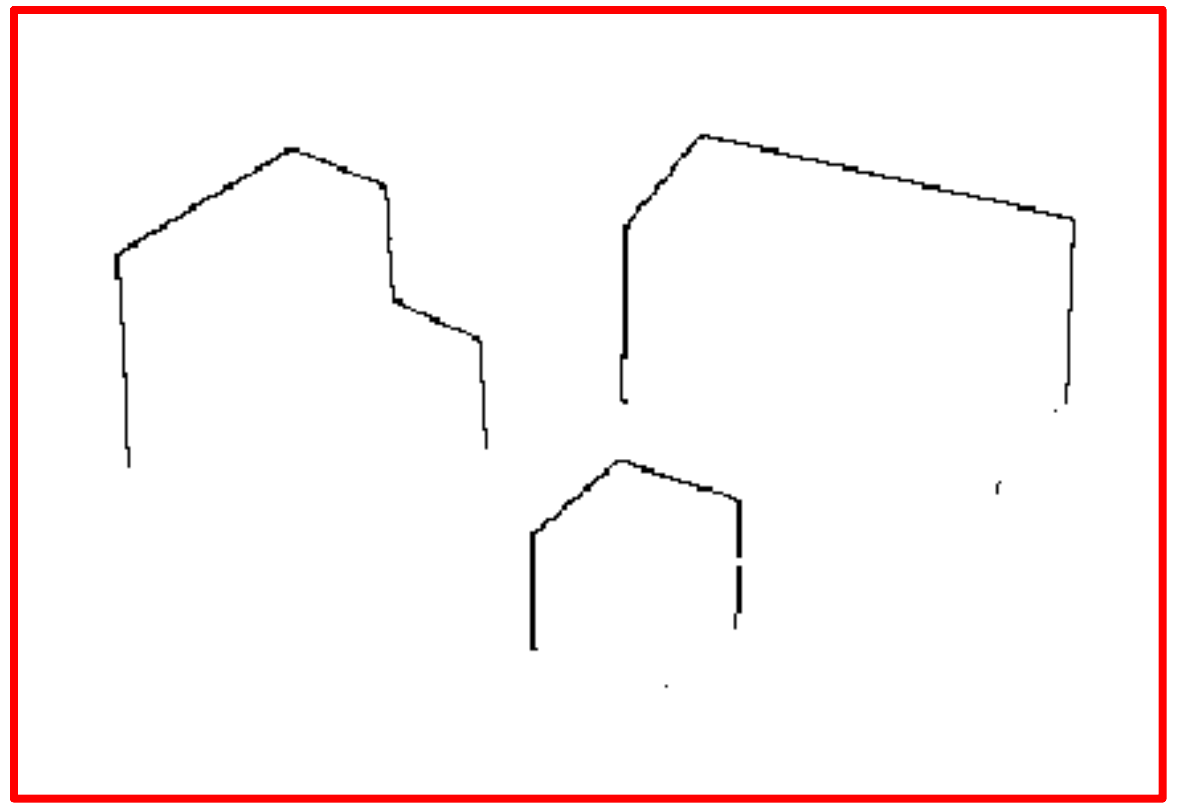


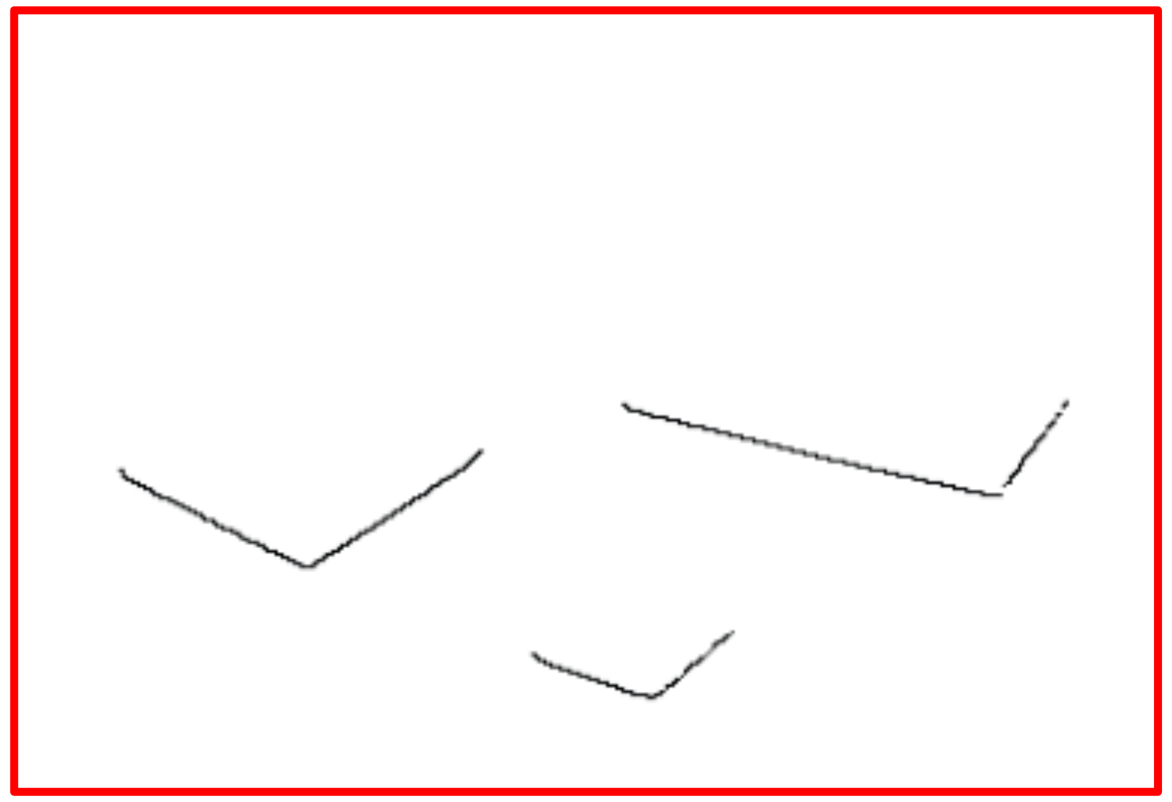
E(x,y) and *n*(x,y)

Edge classification

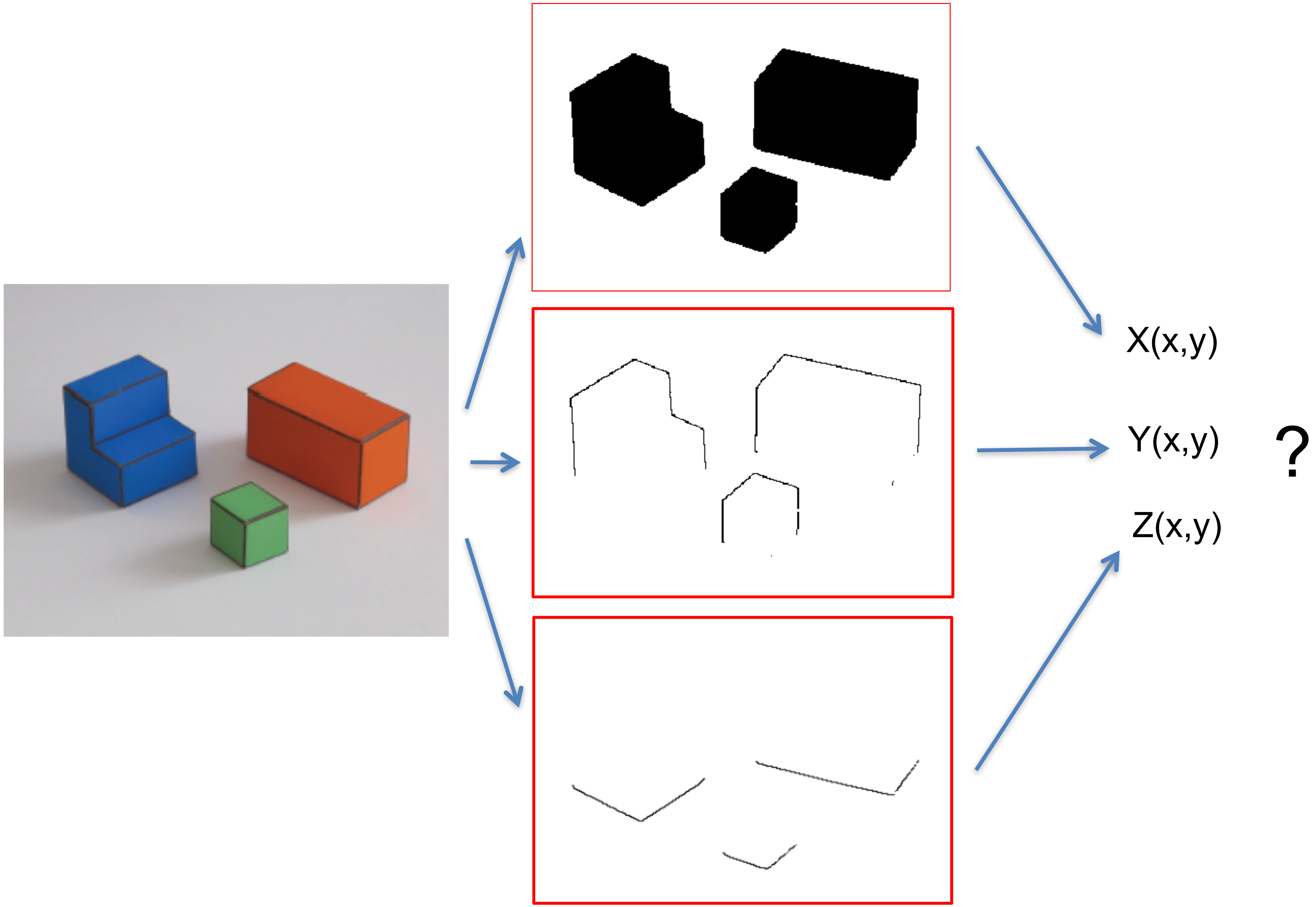
- **Figure/ground segmentation**
  - Using the fact that objects have color
- **Occlusion edges**
  - Occlusion edges are owned by the foreground
- **Contact edges**





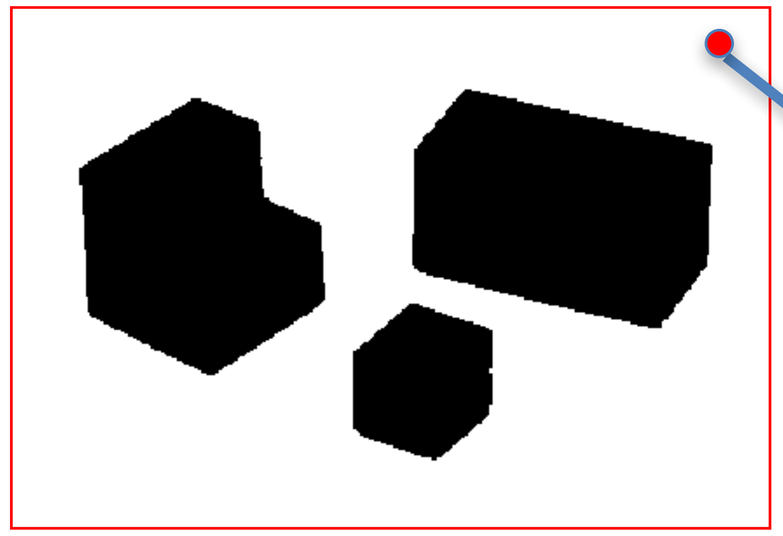


From edges to surface constraints



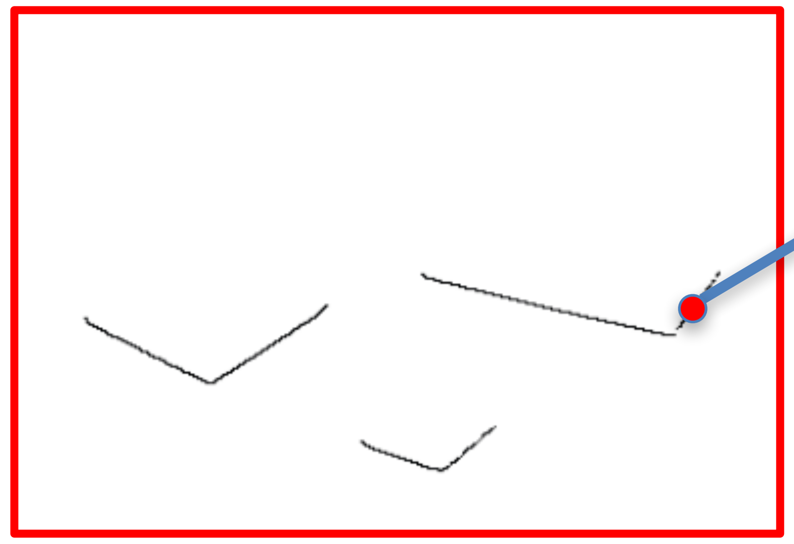
From edges to surface constraints

- Ground



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

- Contact edge

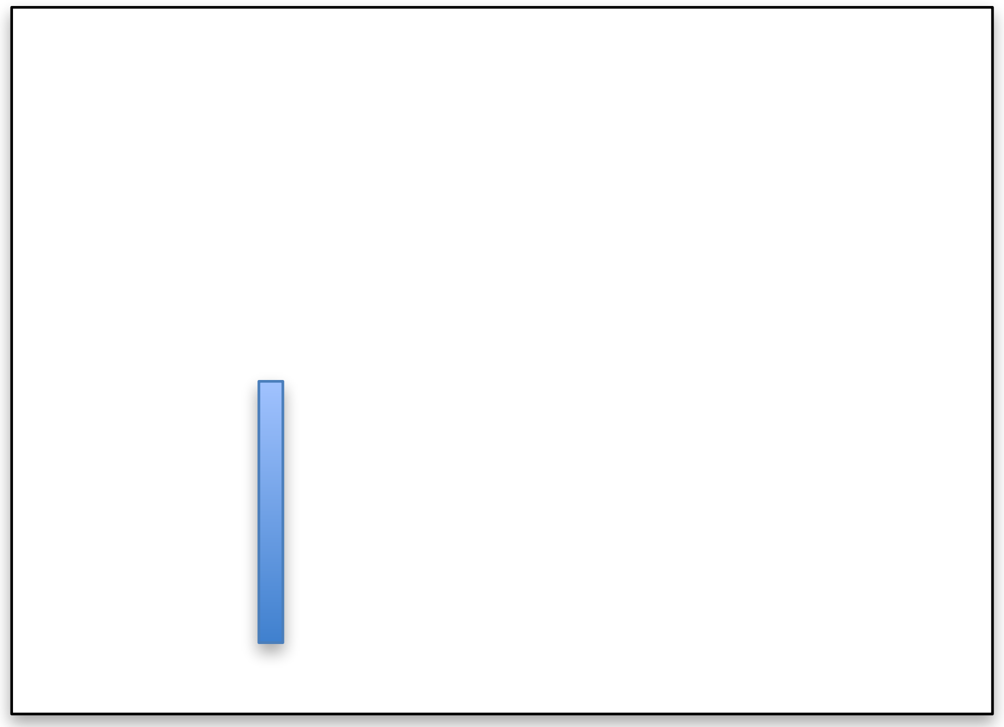


Y(x,y) = 0   if (x,y) belongs to foreground and is a contact edge

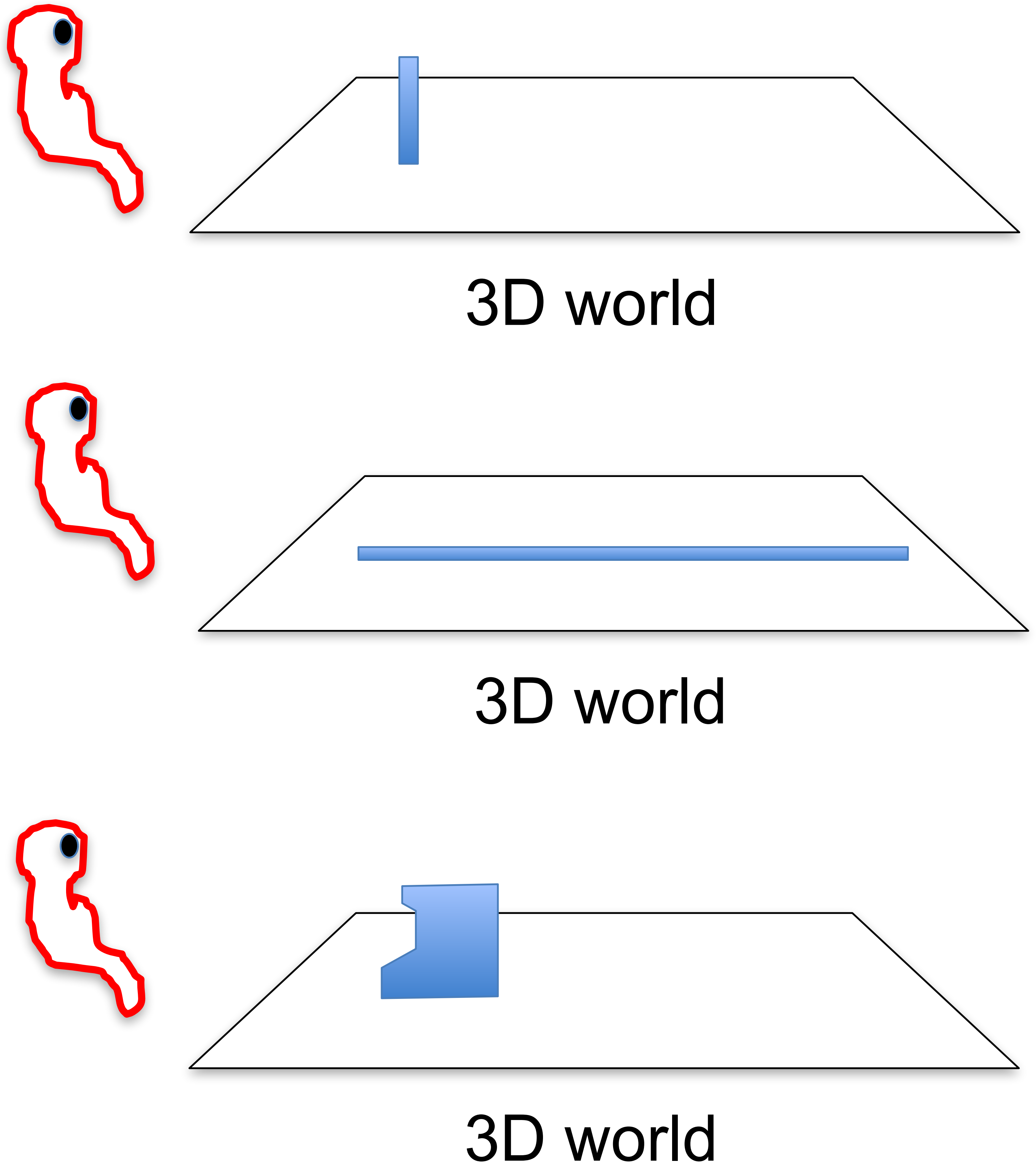
- What happens inside the objects?

… now things get a bit more complicated.

Generic view assumption



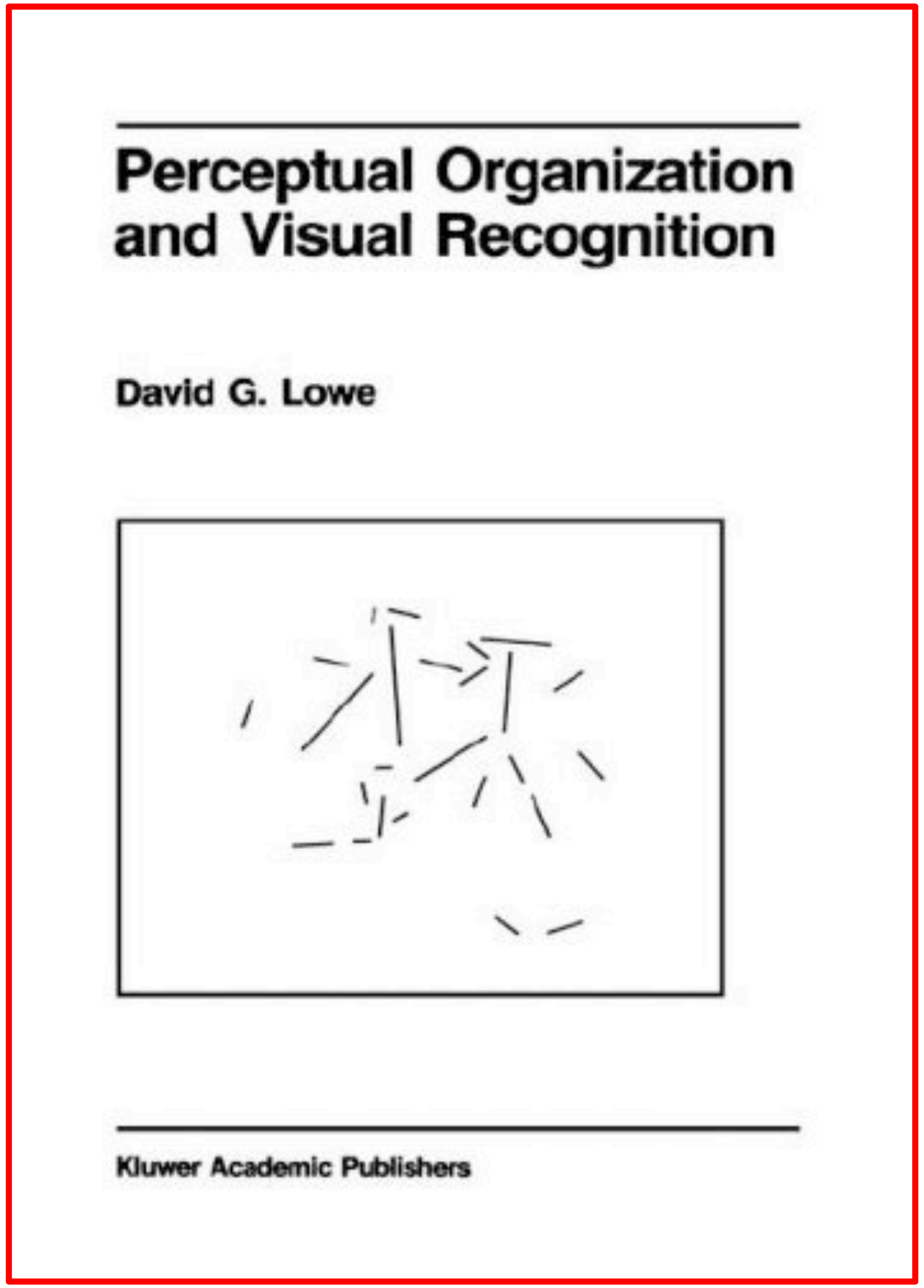
Image



Generic view assumption: the observer should not assume that he has a special position in the world… The most generic interpretation is to see a vertical line as a vertical line in 3D.

Freeman, 93

Non-accidental properties



D. Lowe, 1985

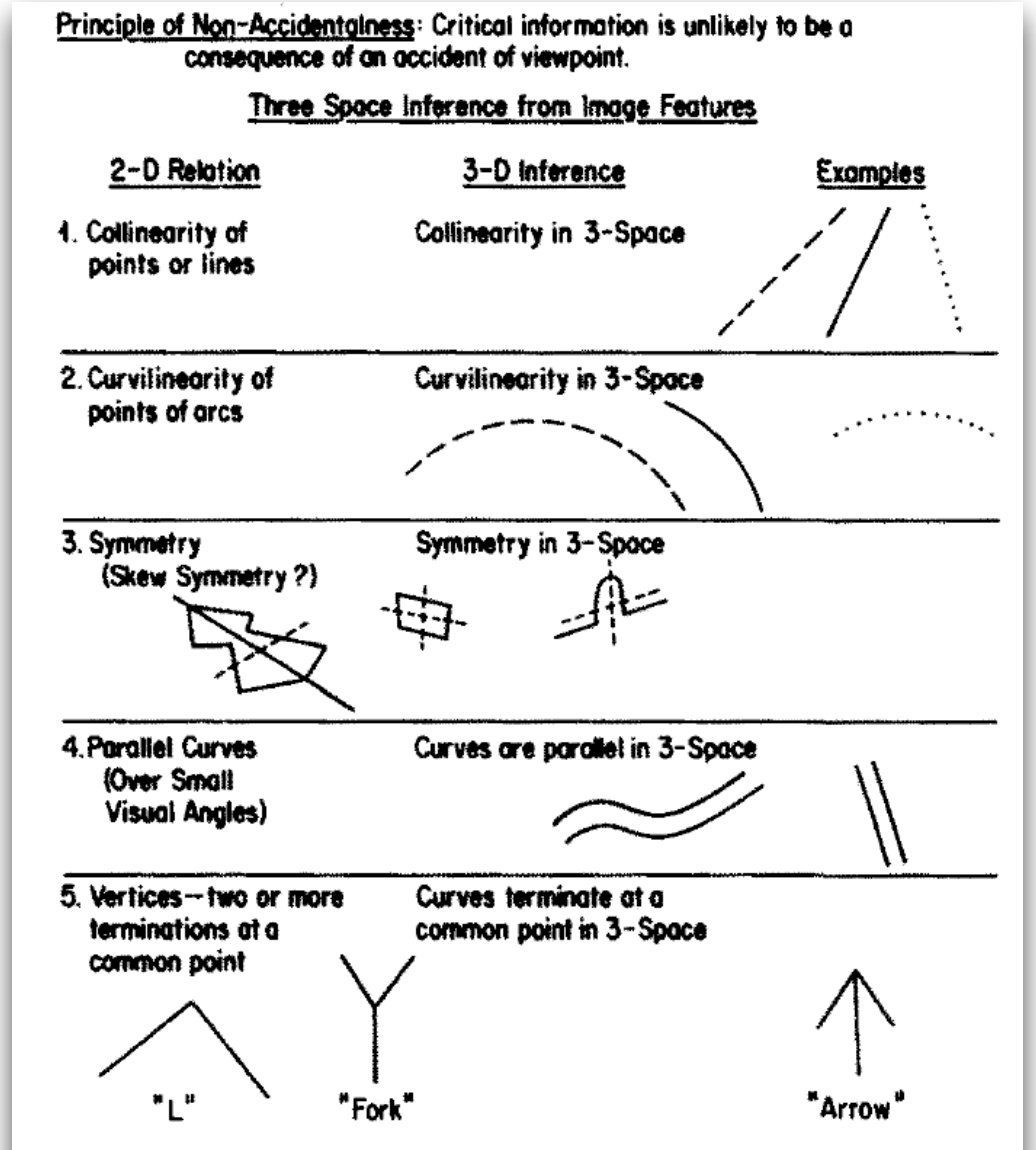
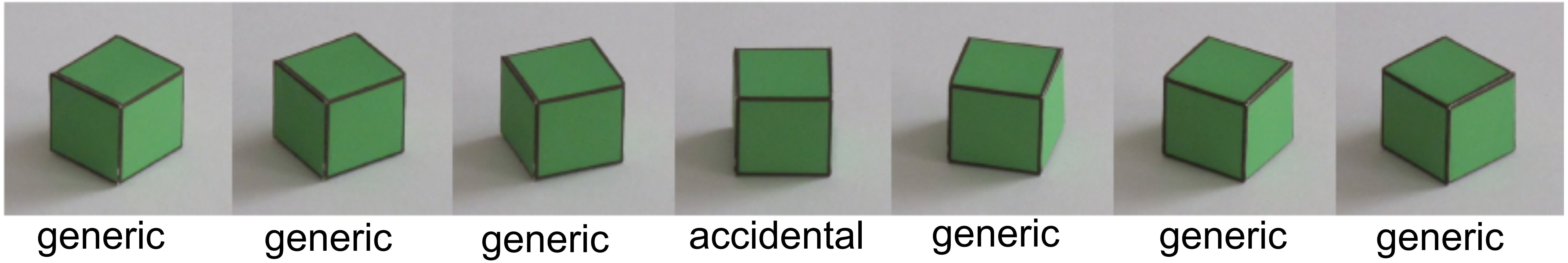
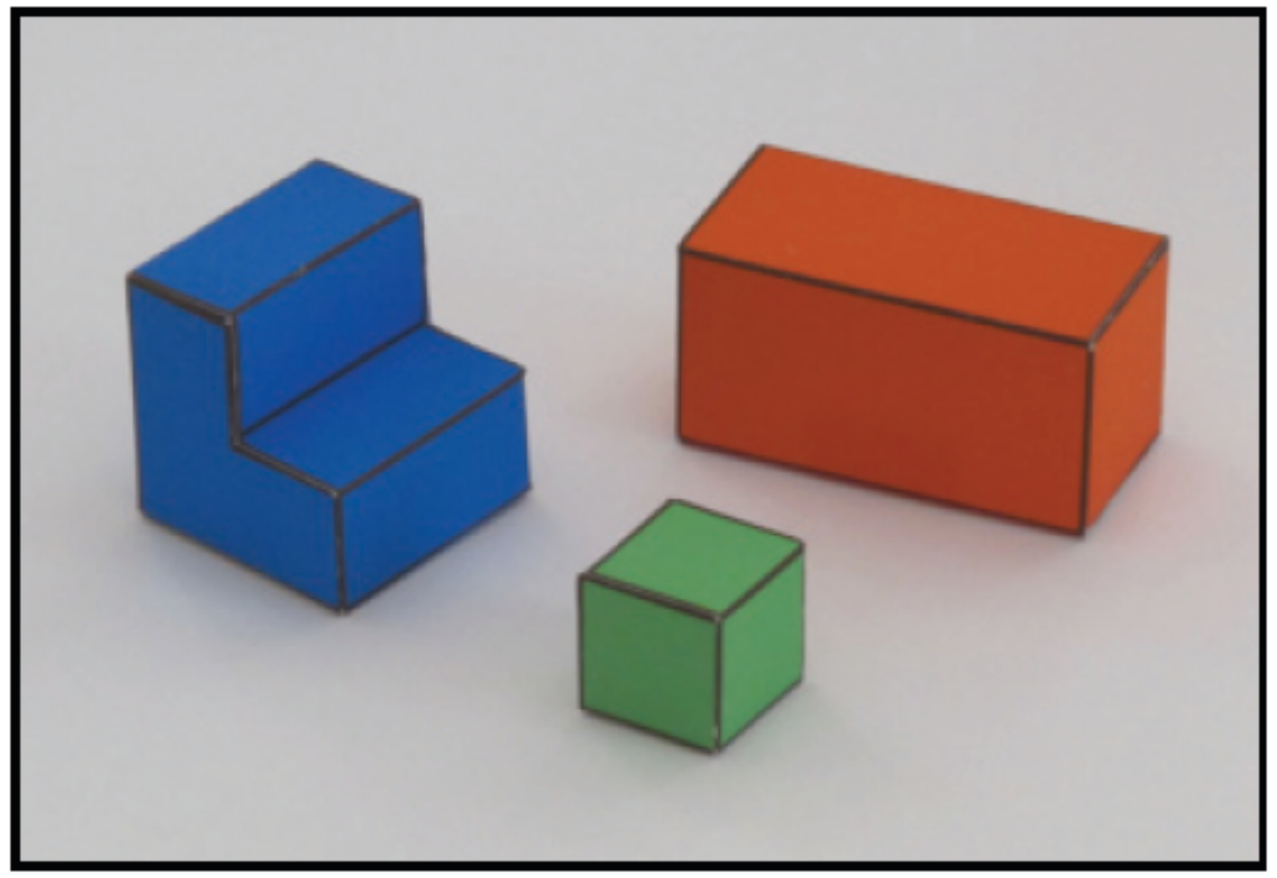


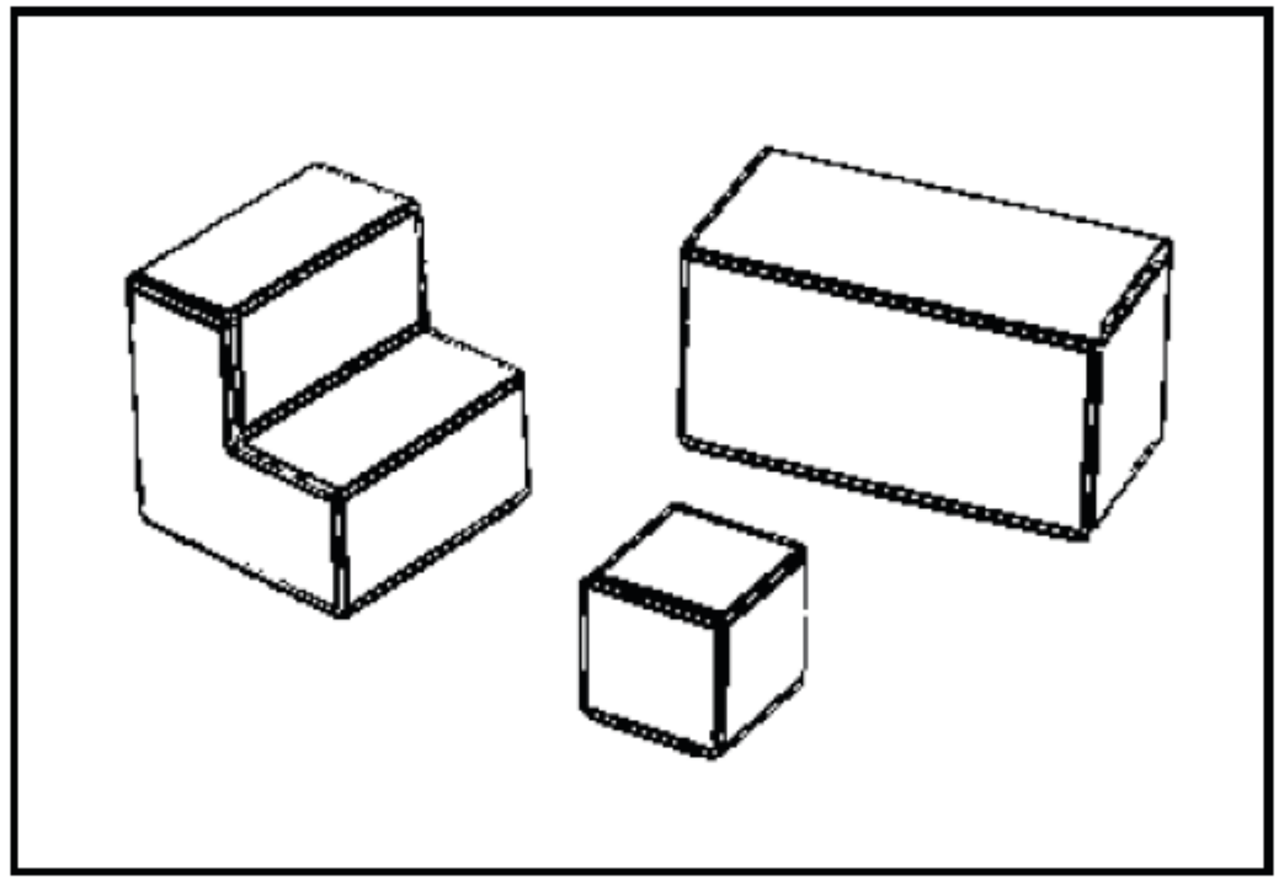
Figure 4. Five nonaccidental relations. (From Figure 5.2, *Perceptual organization and visual recognition* [p. 77] by David Lowe. Unpublished doctorial dissertation, Stanford University. Adapted by permission.)

Biederman_RBC_1987

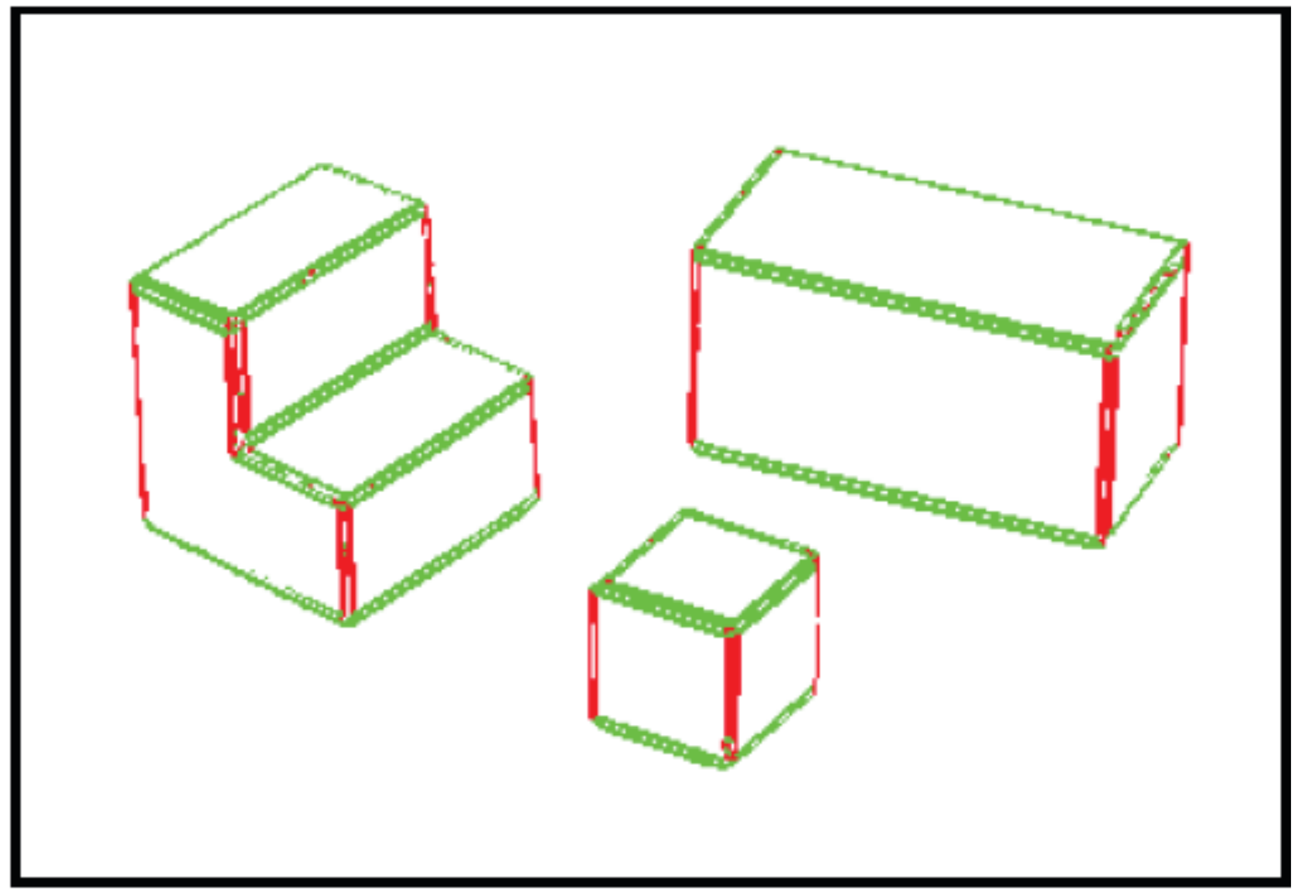
Non-accidental properties in the simple world







Using E(x,y)

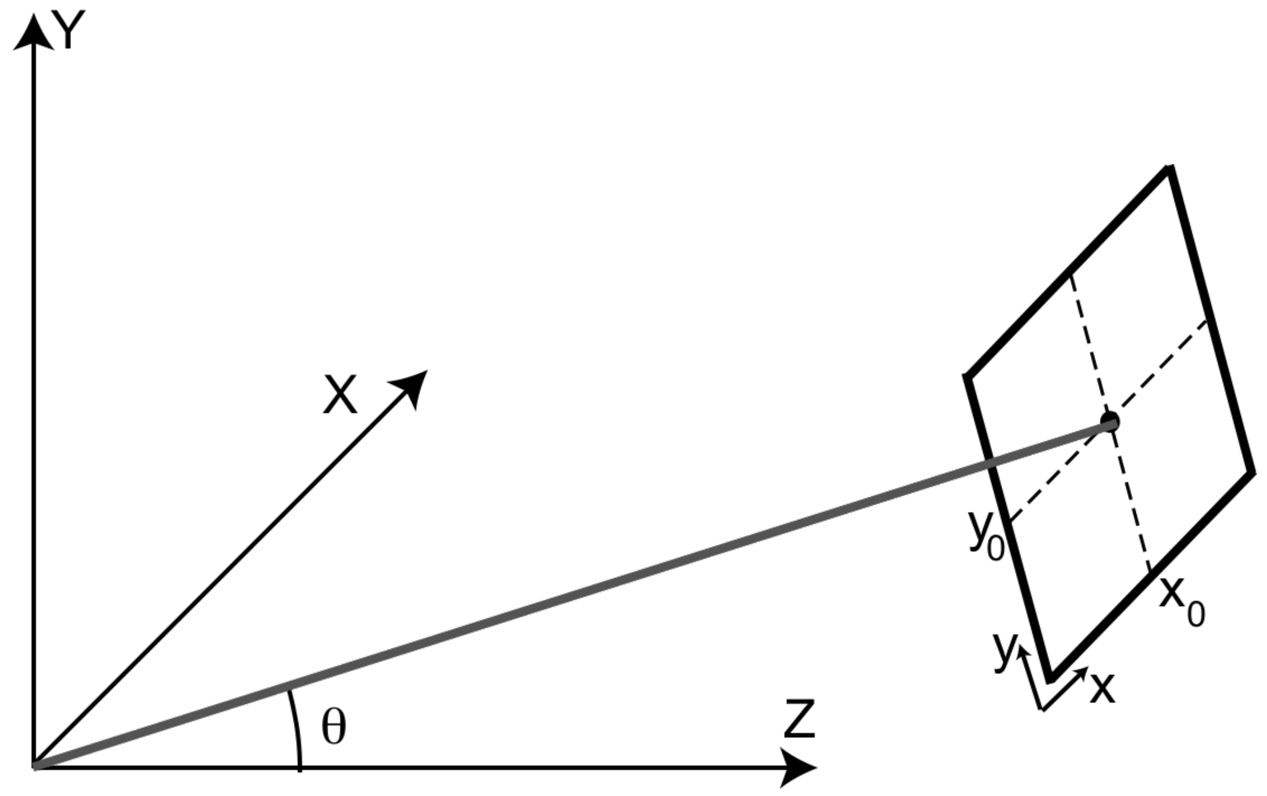


Using $\theta(x,y)$

From edges to surface constraints

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

- Vertical edges are 3D vertical lines



World coordinates

$$x = X + x_0$$

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

image coordinates

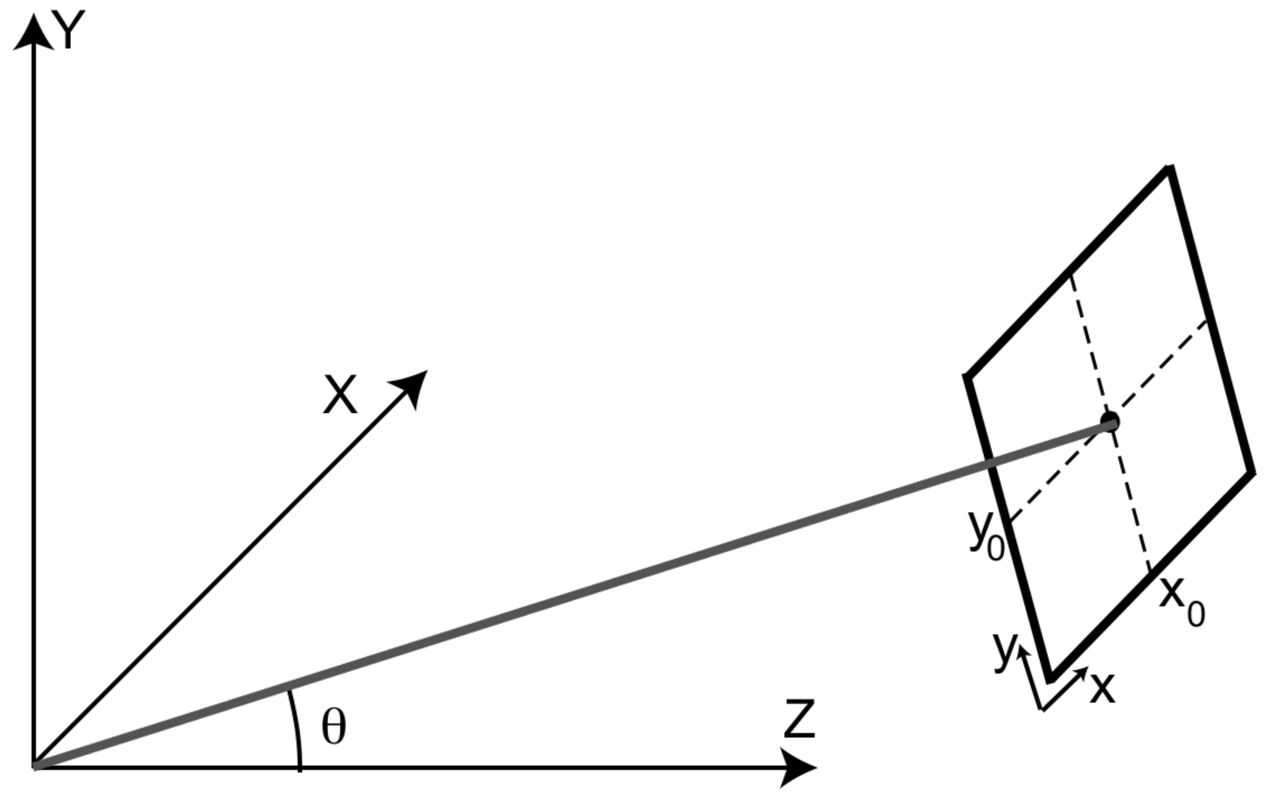
Given the image, what can we say about X, Y and Z in the pixels that belong to a vertical edge?

Z = constant along the edge

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

From edges to surface constraints

- Horizontal edges are 3D horizontal lines



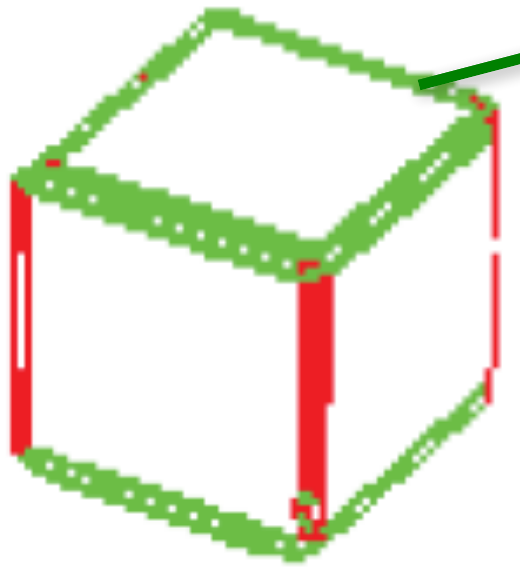
World coordinates

$x = X + x_0$

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

image coordinates

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



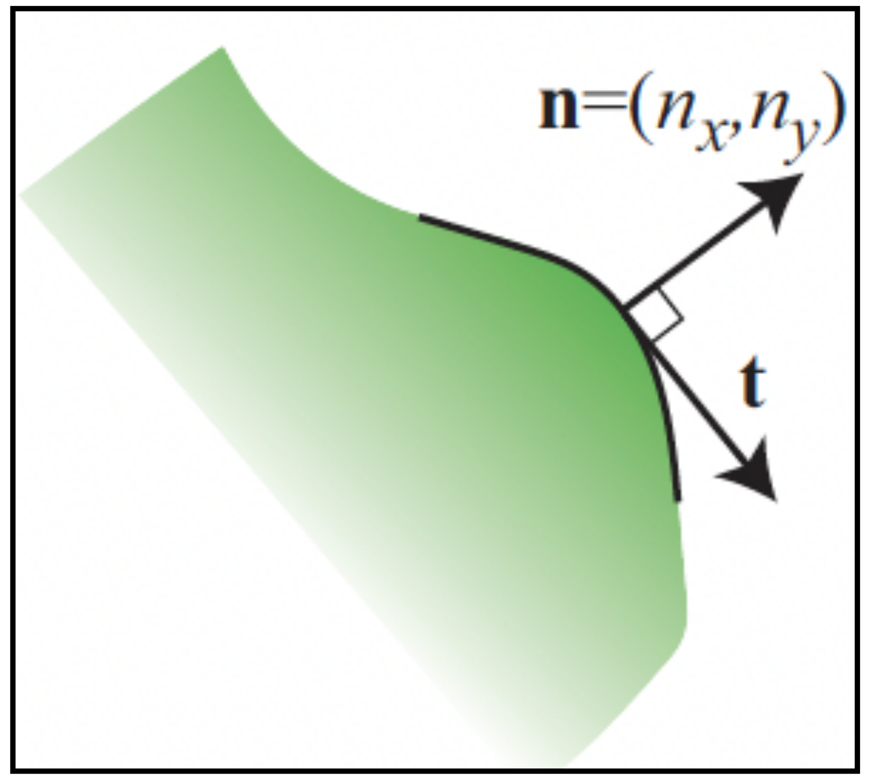
Y = constant along the edge

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

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

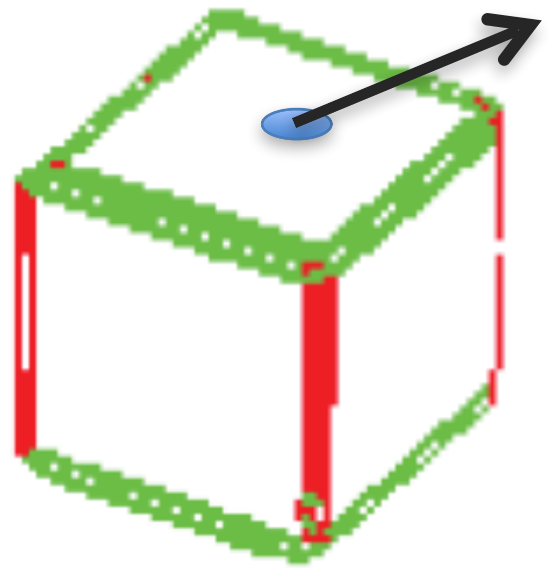
$$\mathbf{t} = (-n_y, n_x)$$

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



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

The “Rule of Nothing” (Ted Adelson): where you see nothing, assume nothing happens, and just propagate information from where something happened.

A simple inference scheme

All the constraints are linear

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

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

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

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

A similar set of constraints could be derived for Z

Discrete approximation

We can transform every differential constrain into a discrete 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

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

-1	1

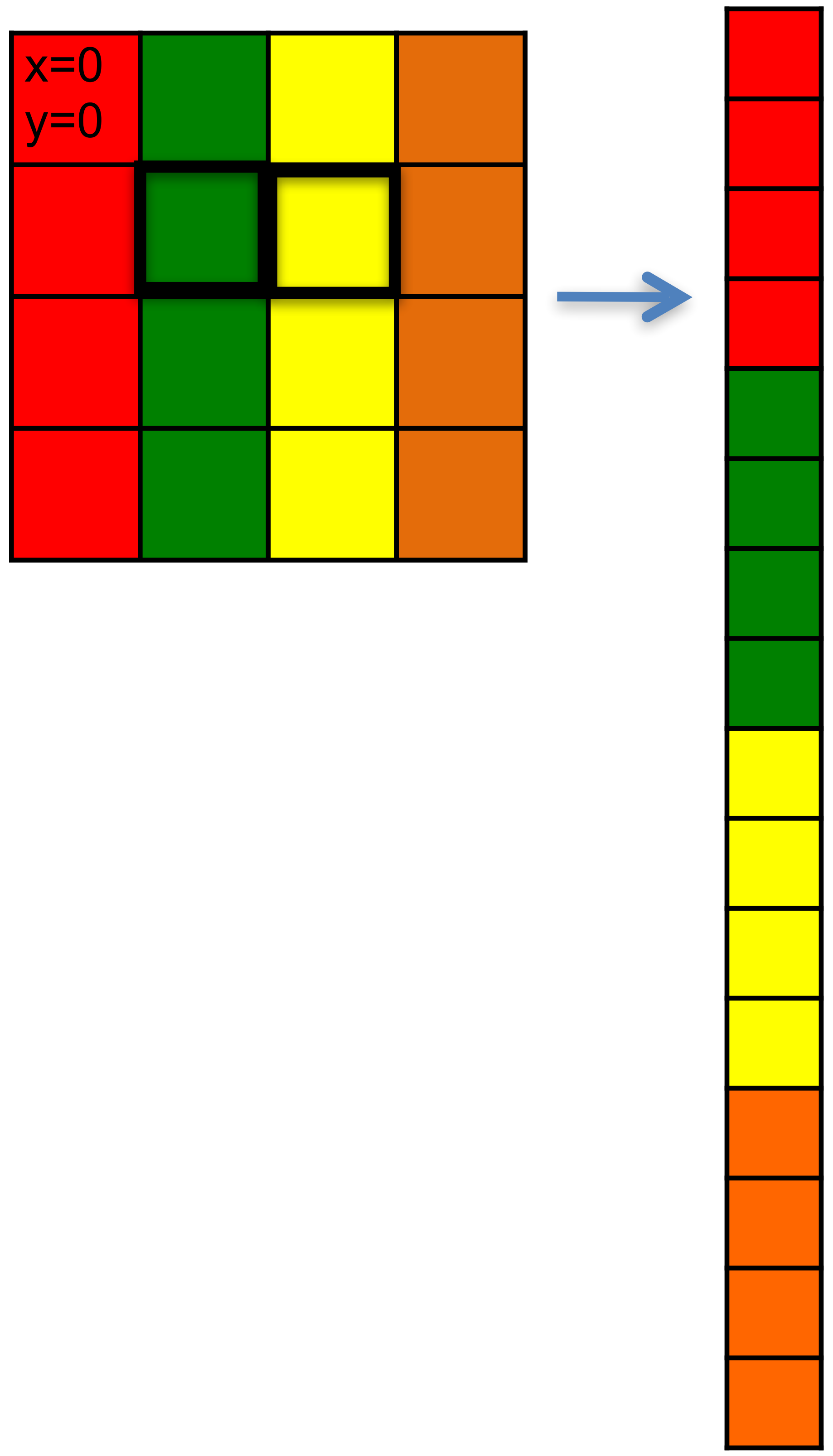
A slightly better approximation
(it is symmetric, and it averages horizontal derivatives over 3 vertical locations)

-1	0	1
-2	0	2
-1	0	1

Discrete approximation

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

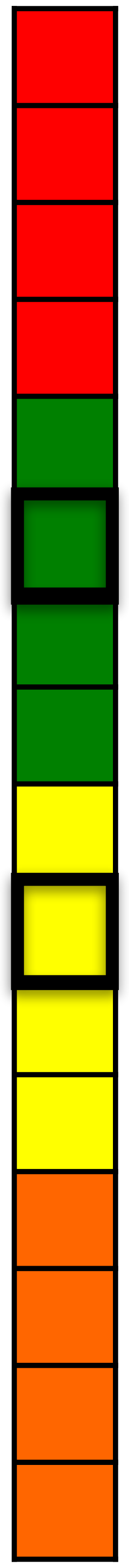
Y(x,y)



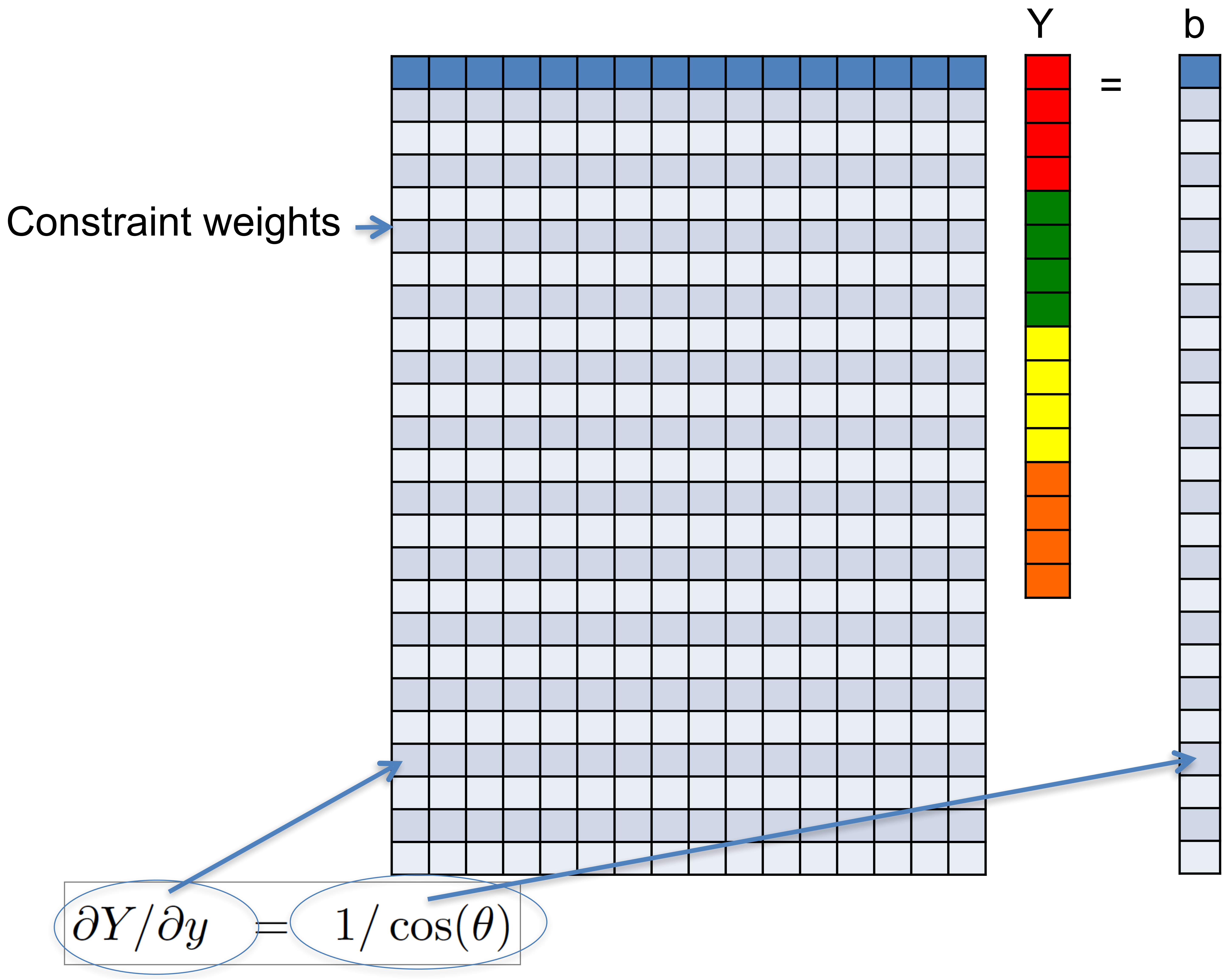
x=2, y=1

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

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



A simple inference scheme



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

$$A\,Y = b$$

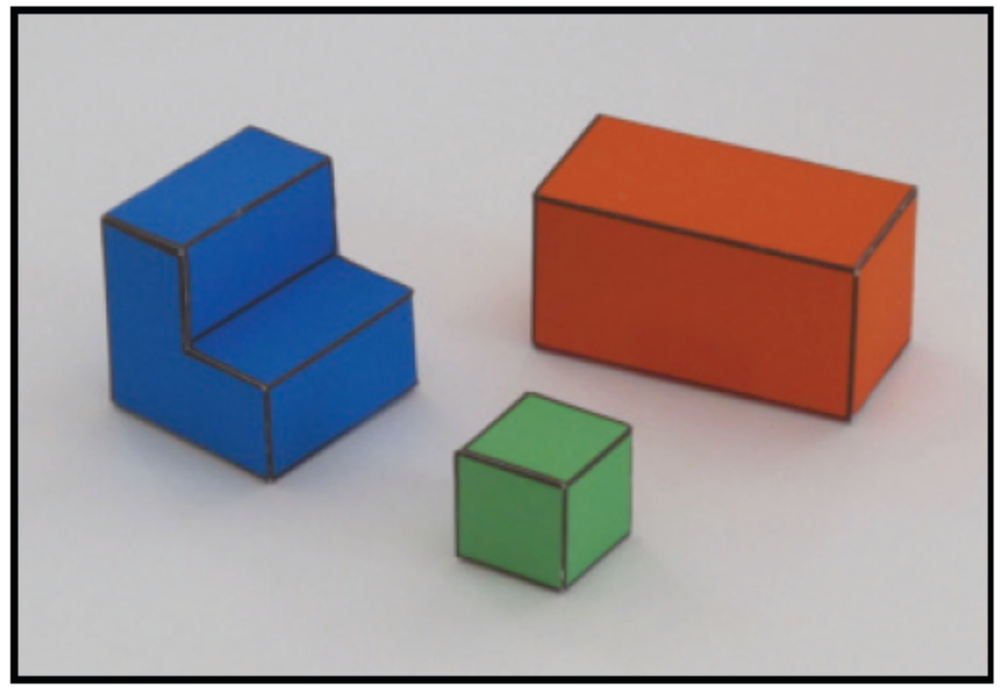
$$Y = (A^T A)^{-1} A^T b$$

Results

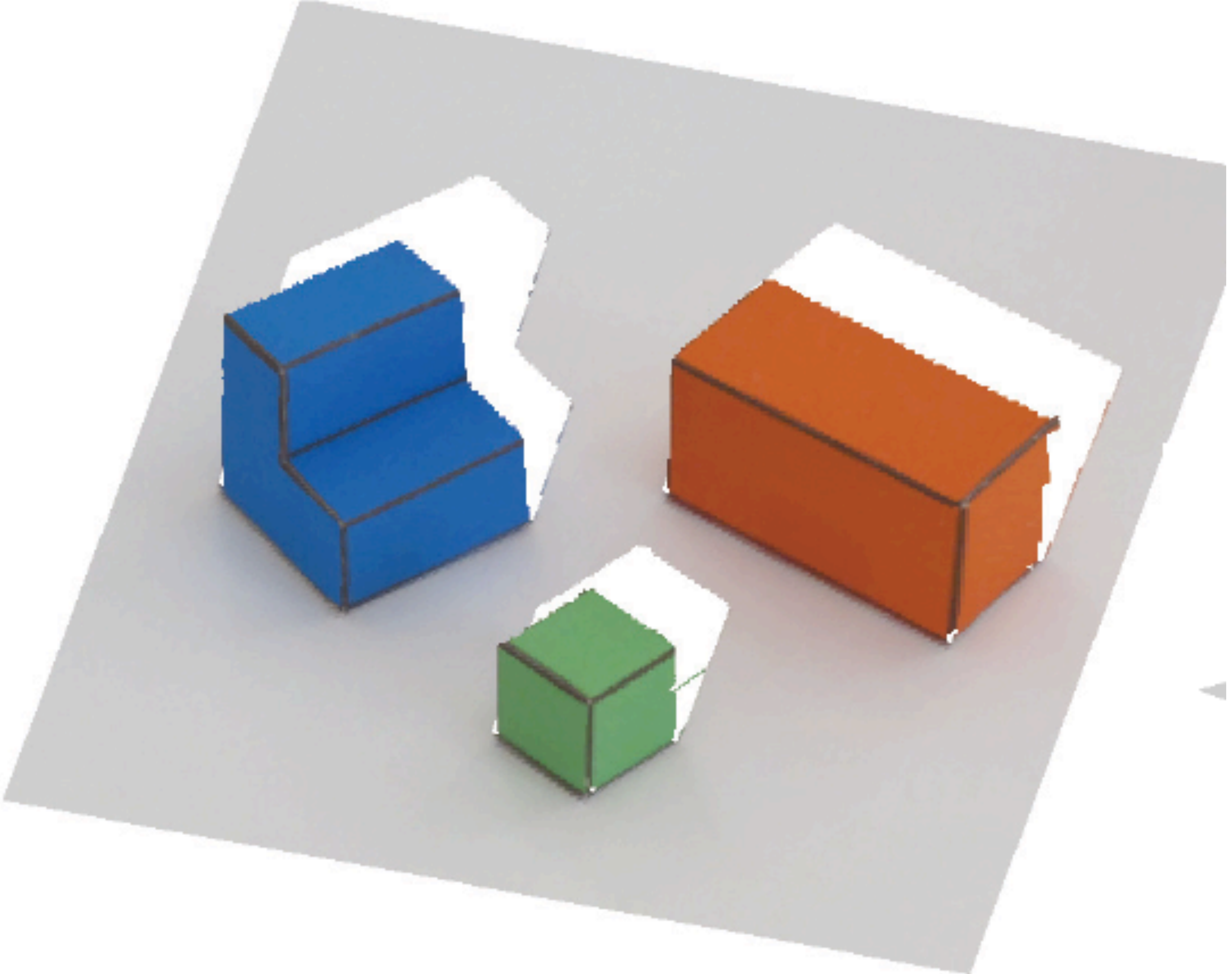

Input
Representation 1
Edge normals
Figure/ground
Representation 2
Edge strength
3D orientation
Contact edges
Depth discontinuities
Linear system
Output
X
Y
Z

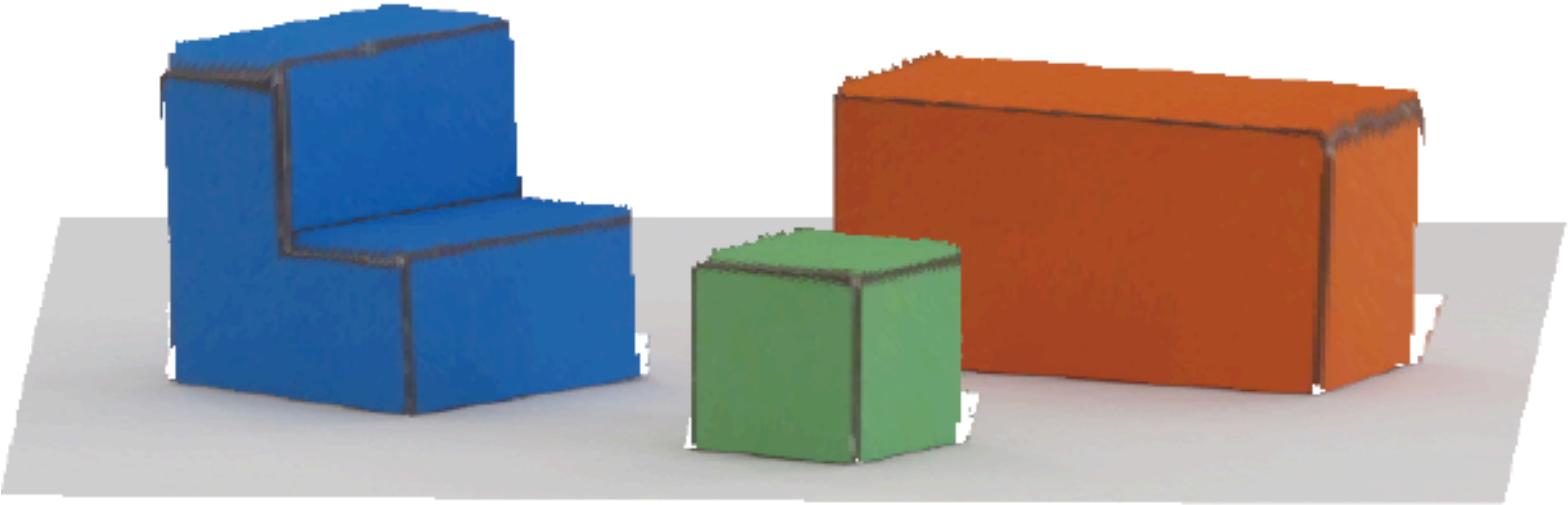
Changing view point

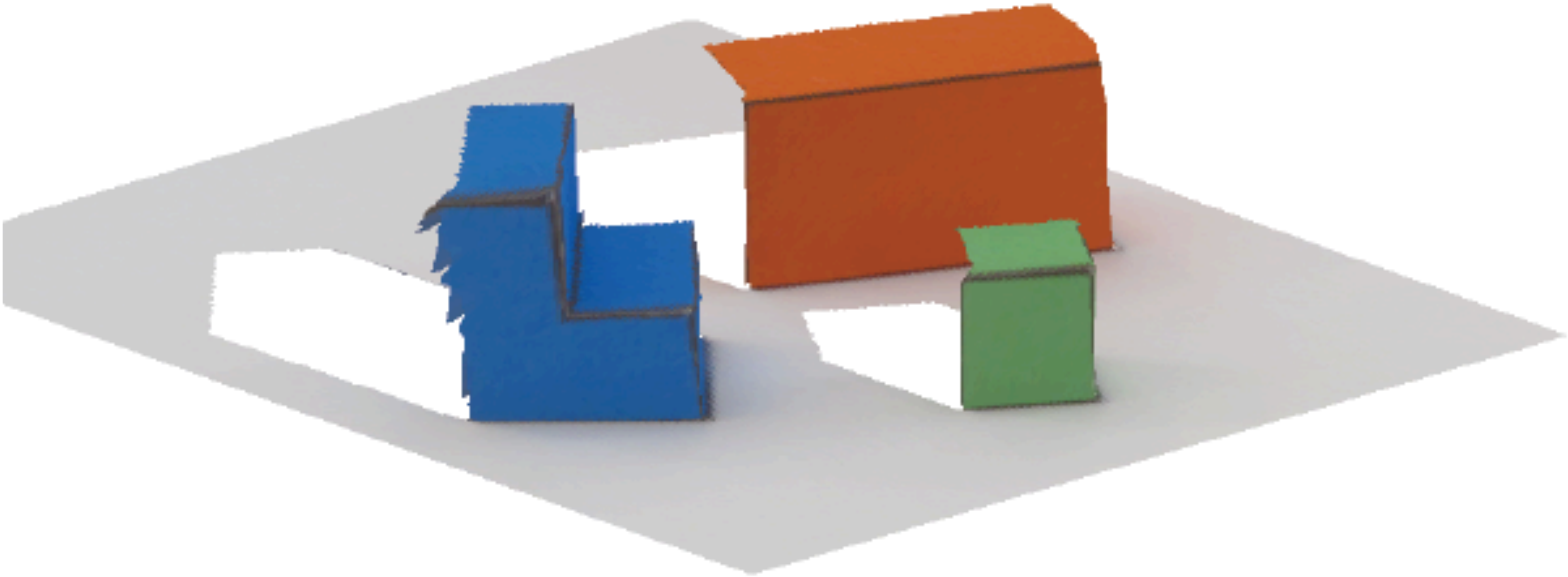
Input



New view points:

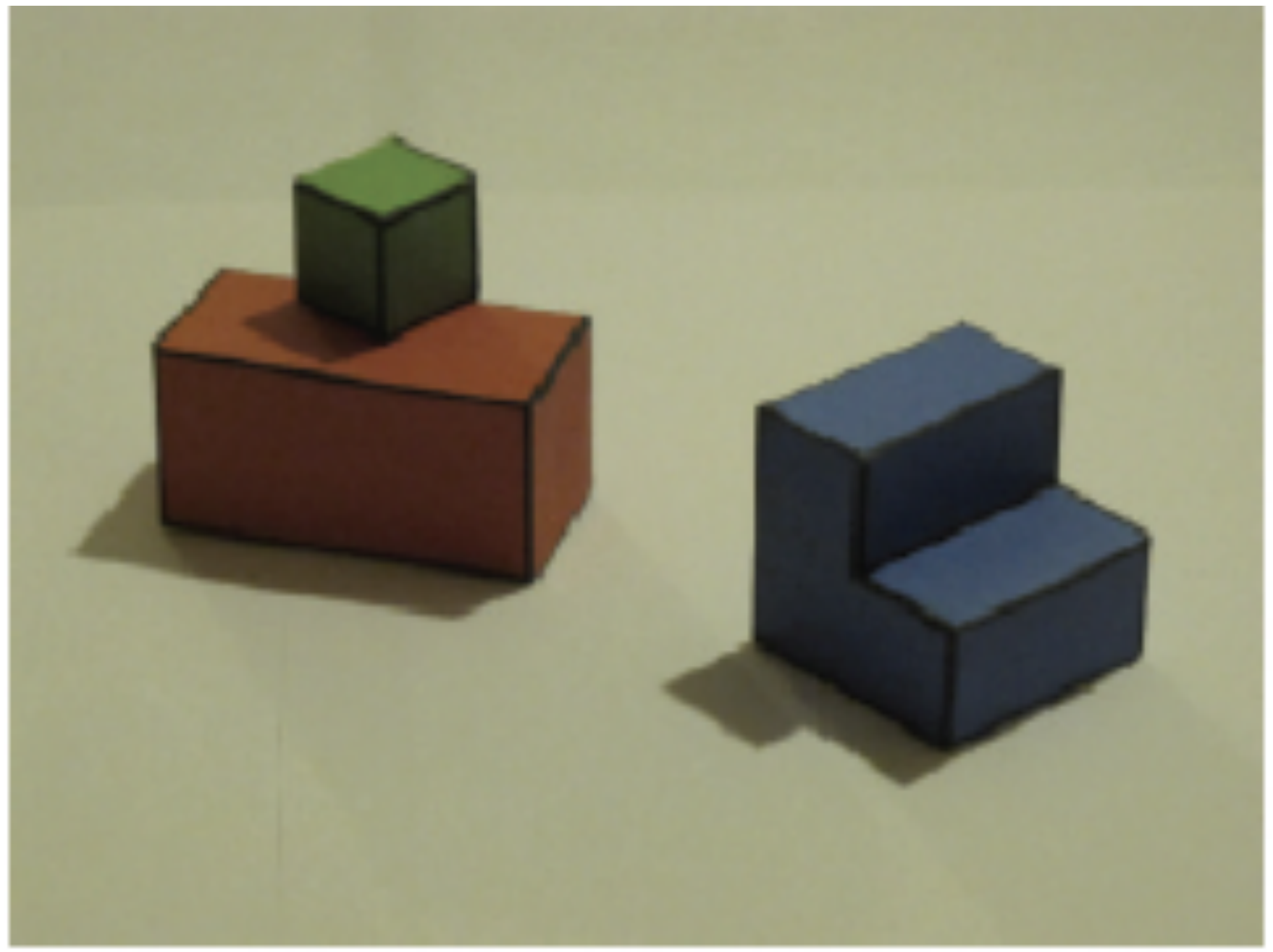




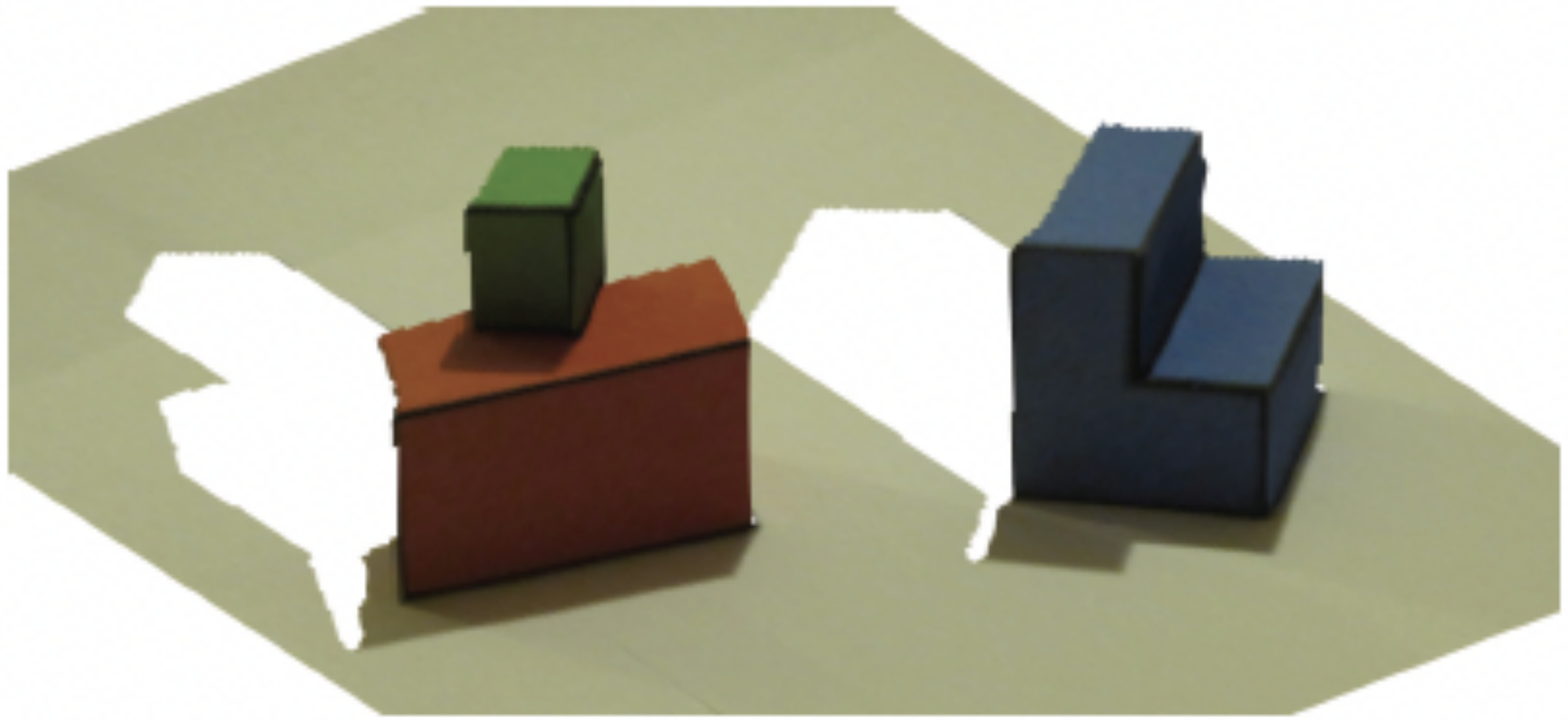


Generalization

Input

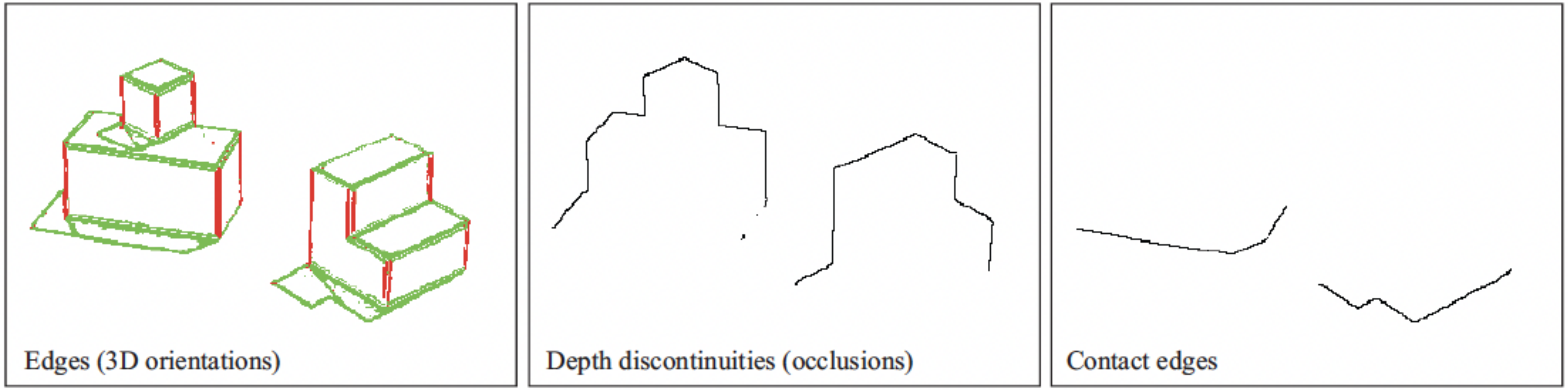


New view point:



It seems to work!

… but the representation is wrong!



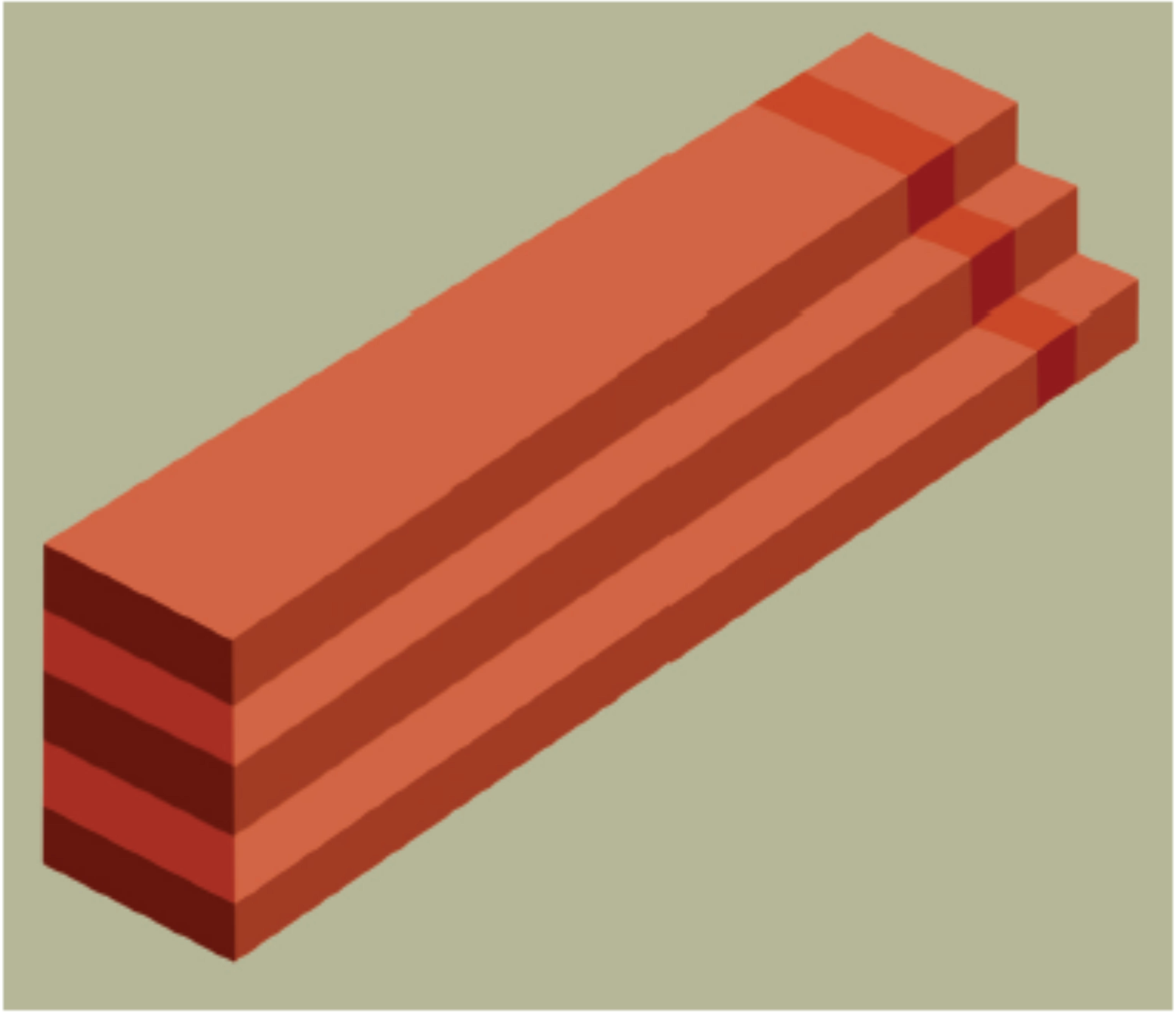
Generalization
2nd test

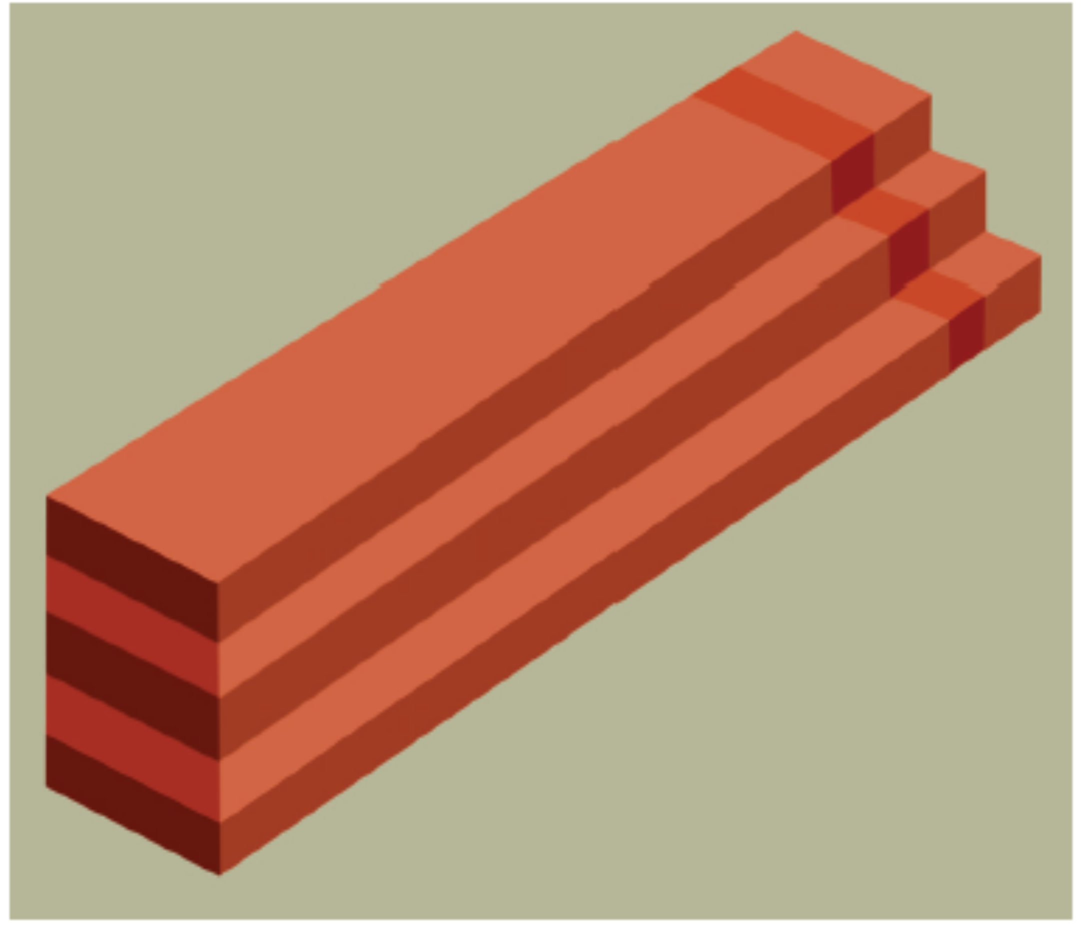
Impossible steps

Adelson, E.H. Lightness Perception and Lightness Illusions. In *The New Cognitive Neurosciences*, 2nd ed., M. Gazzaniga, ed. Cambridge, MA: MIT Press, pp. 339-351, (2000).

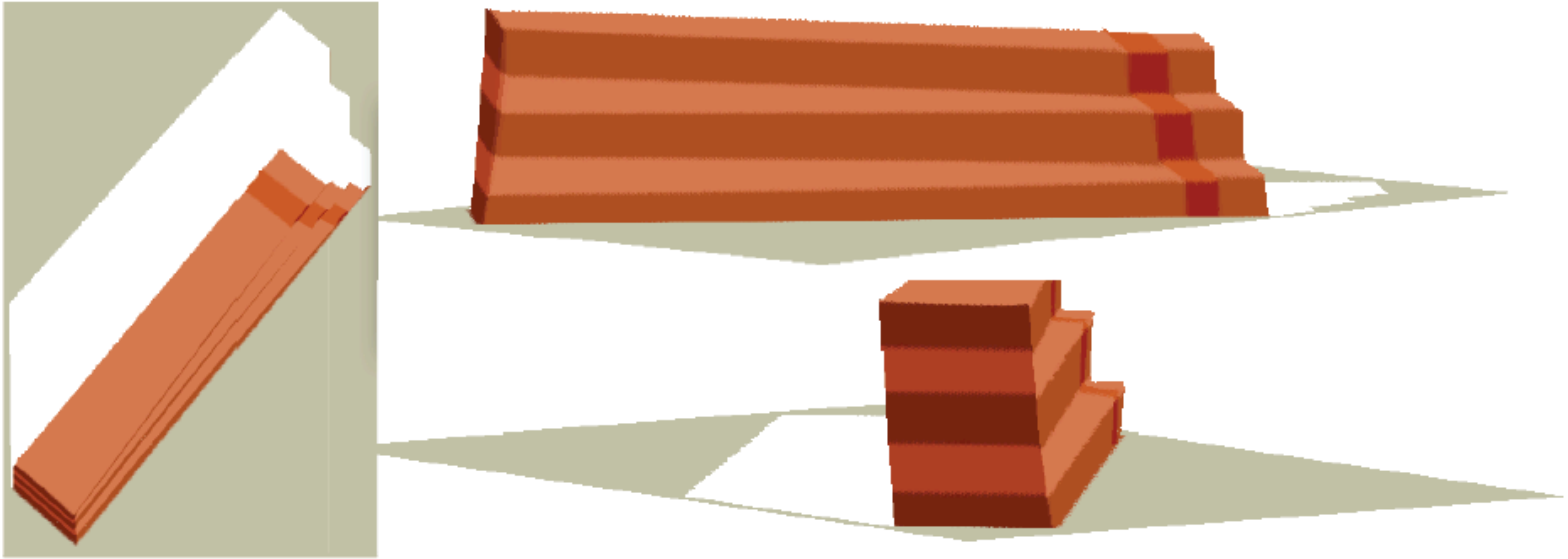
24 Lightness Perception and Lightness Illusions

EDWARD H. ADELSON





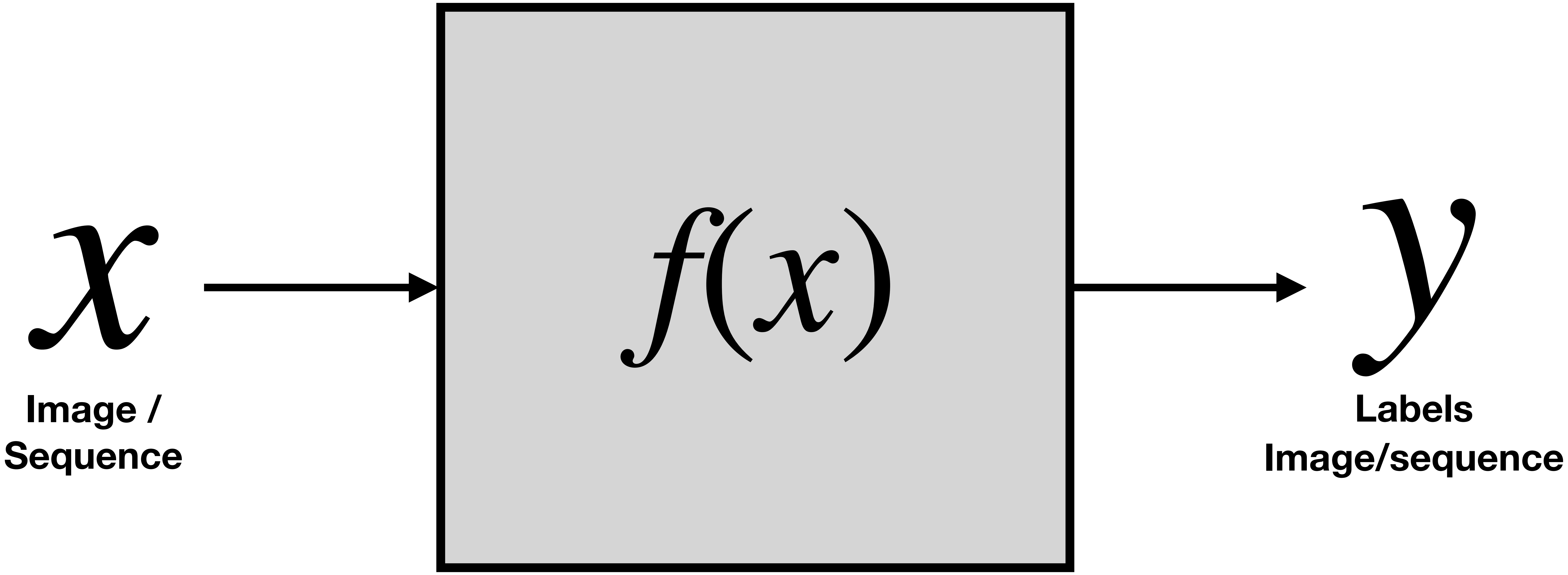
Impossible steps



Some keywords

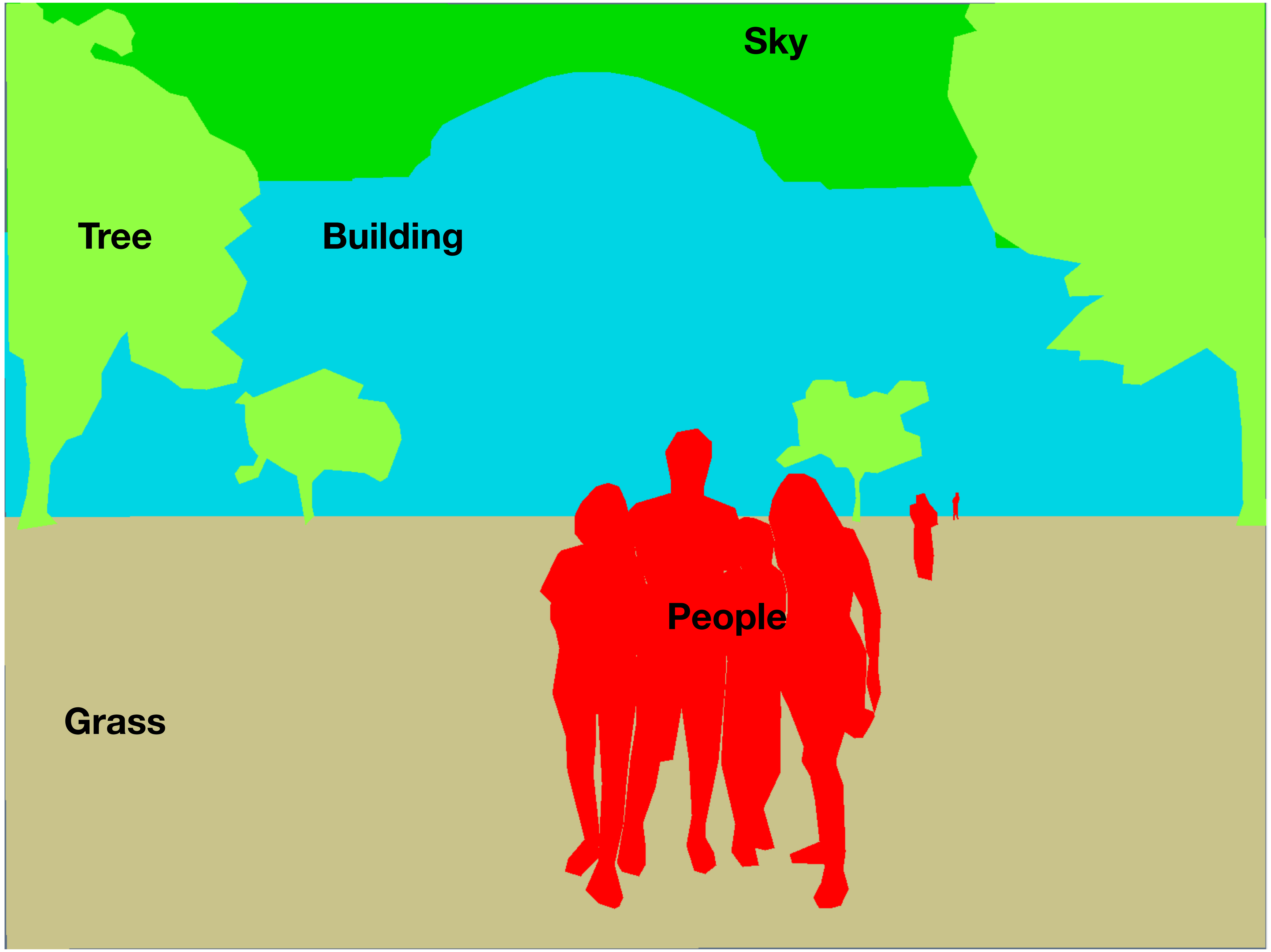
- Light rays
- Image formation, parallel projection
- 3D, World and image coordinates
- Representation
- Figure / ground
- Edges
- Accidental views (generic view assumption)
- Image gradients and discrete approximation
- Inference
- Generalization

Tasks: generic formulation



Tasks: what humans care about

Tasks: what humans care about

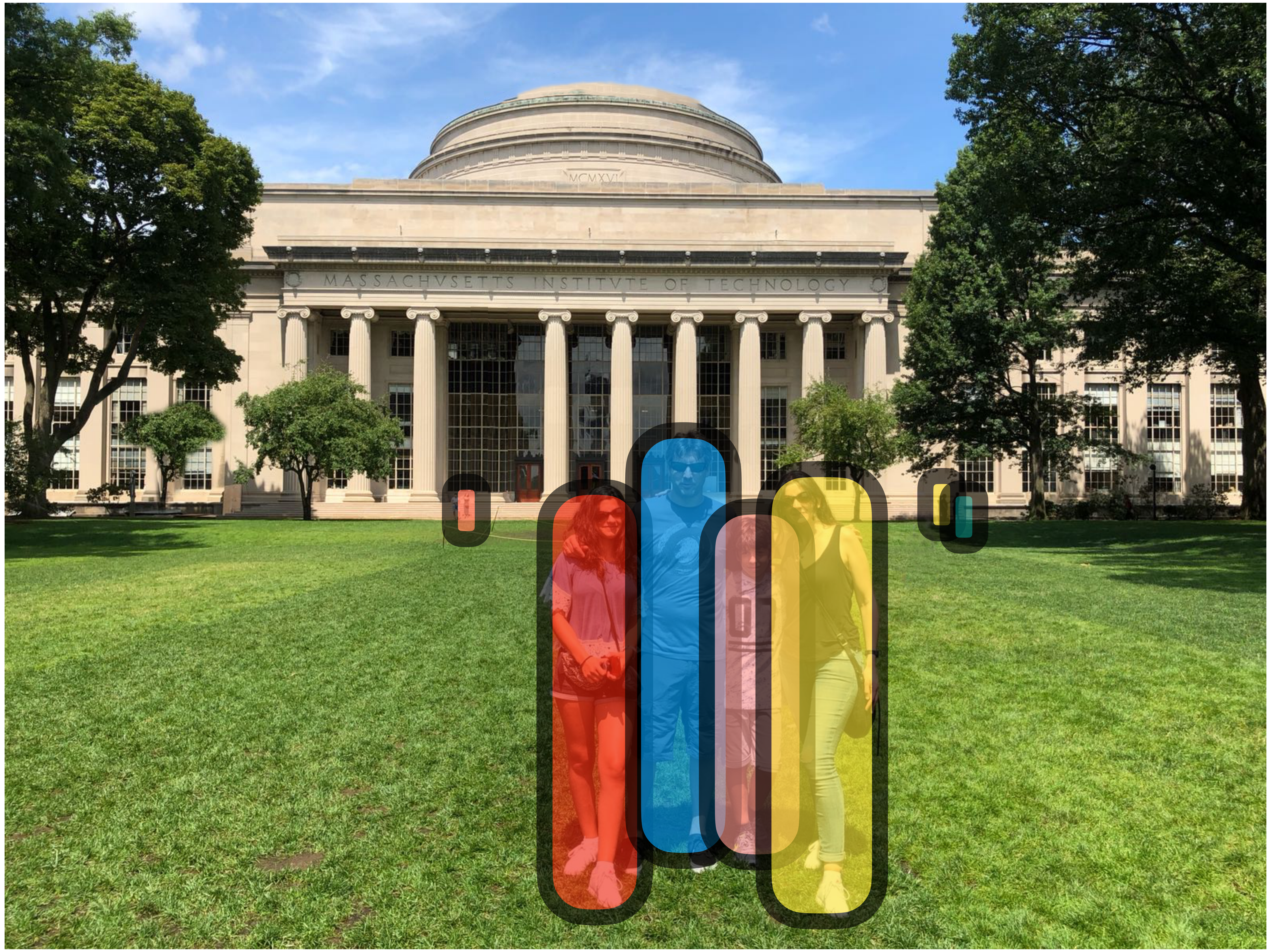


Semantic segmentation:
Assign labels to all the pixels in the image

Related tasks:

- **Semantic segmentation**
- **Object categorization**

Tasks: what humans care about



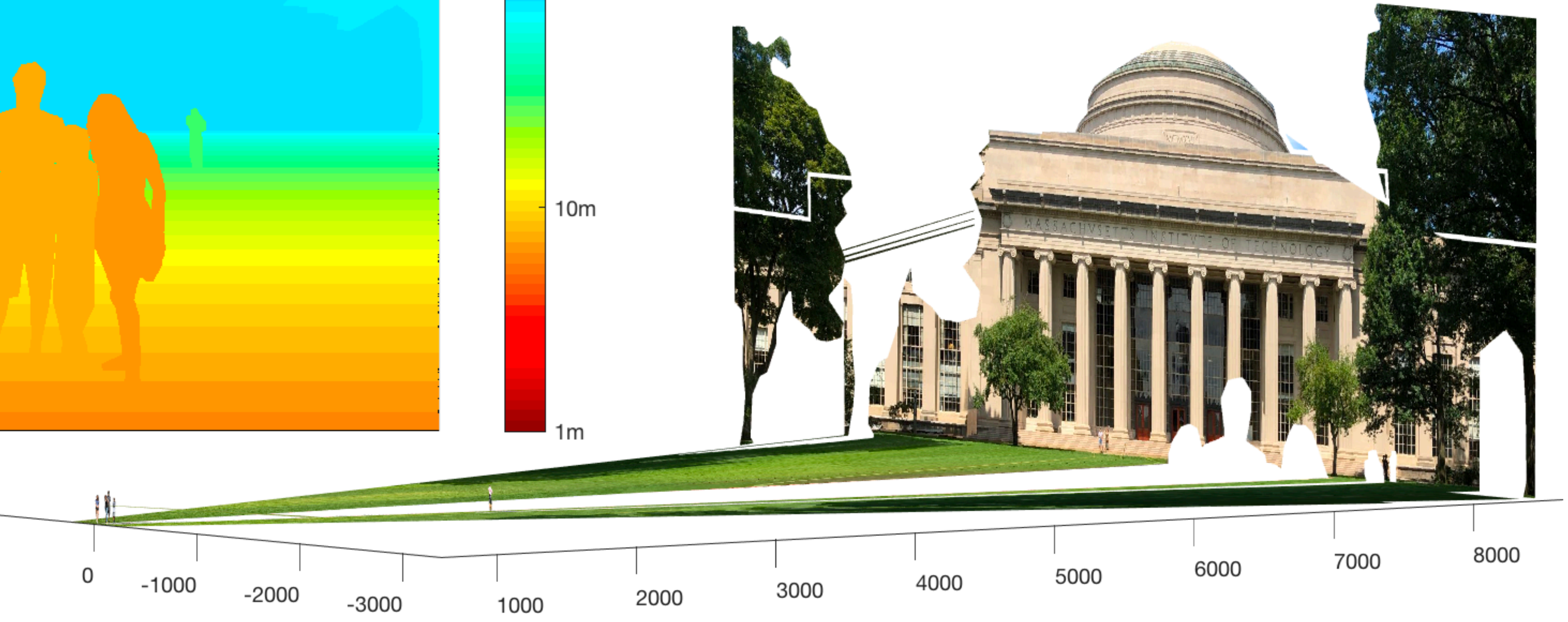
Detection: Locate all the people in this image

Tasks: what humans care about

Recognition: who is this person?

Tasks: what humans care about

Rough 3D layout, depth ordering

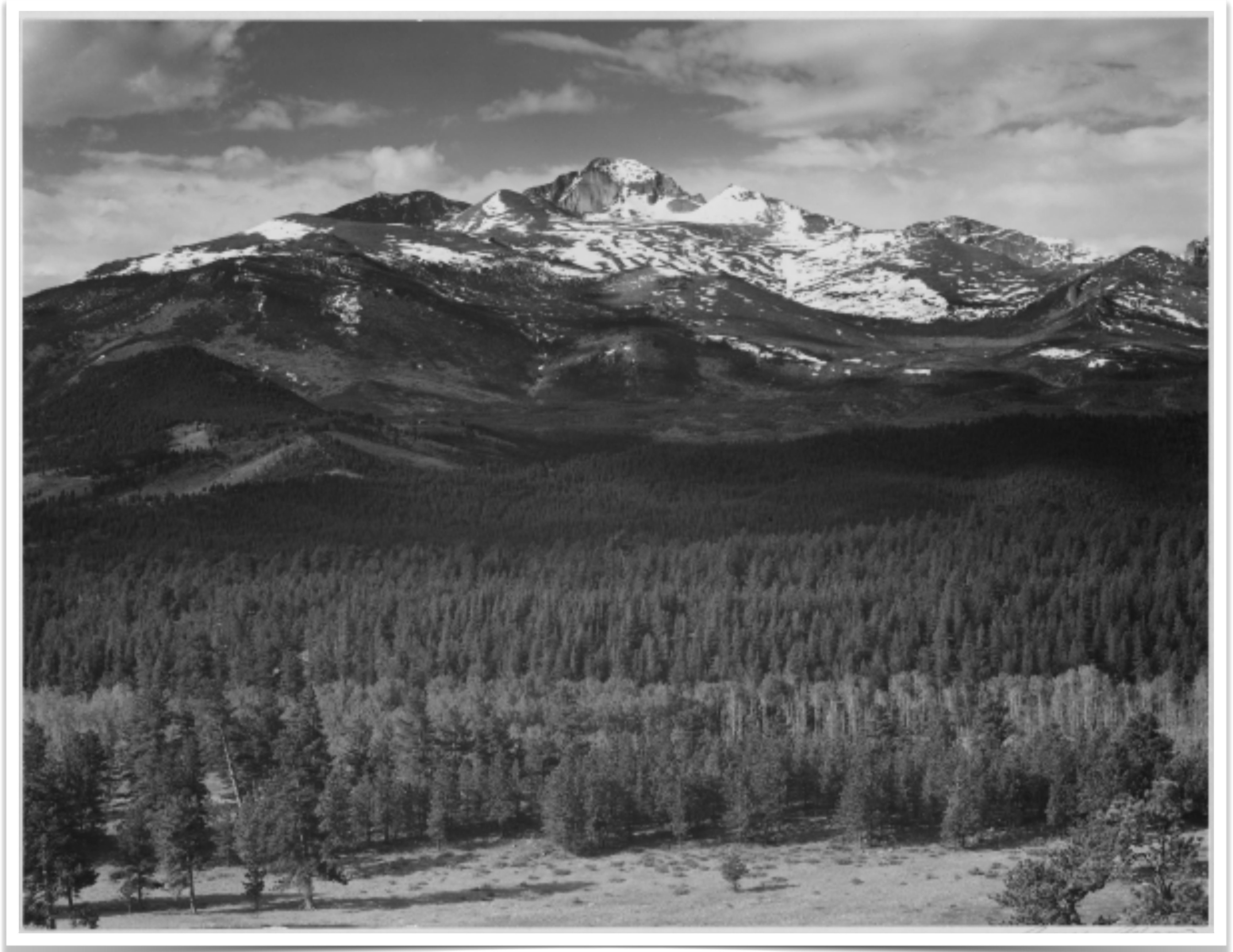


Tasks: what humans care about

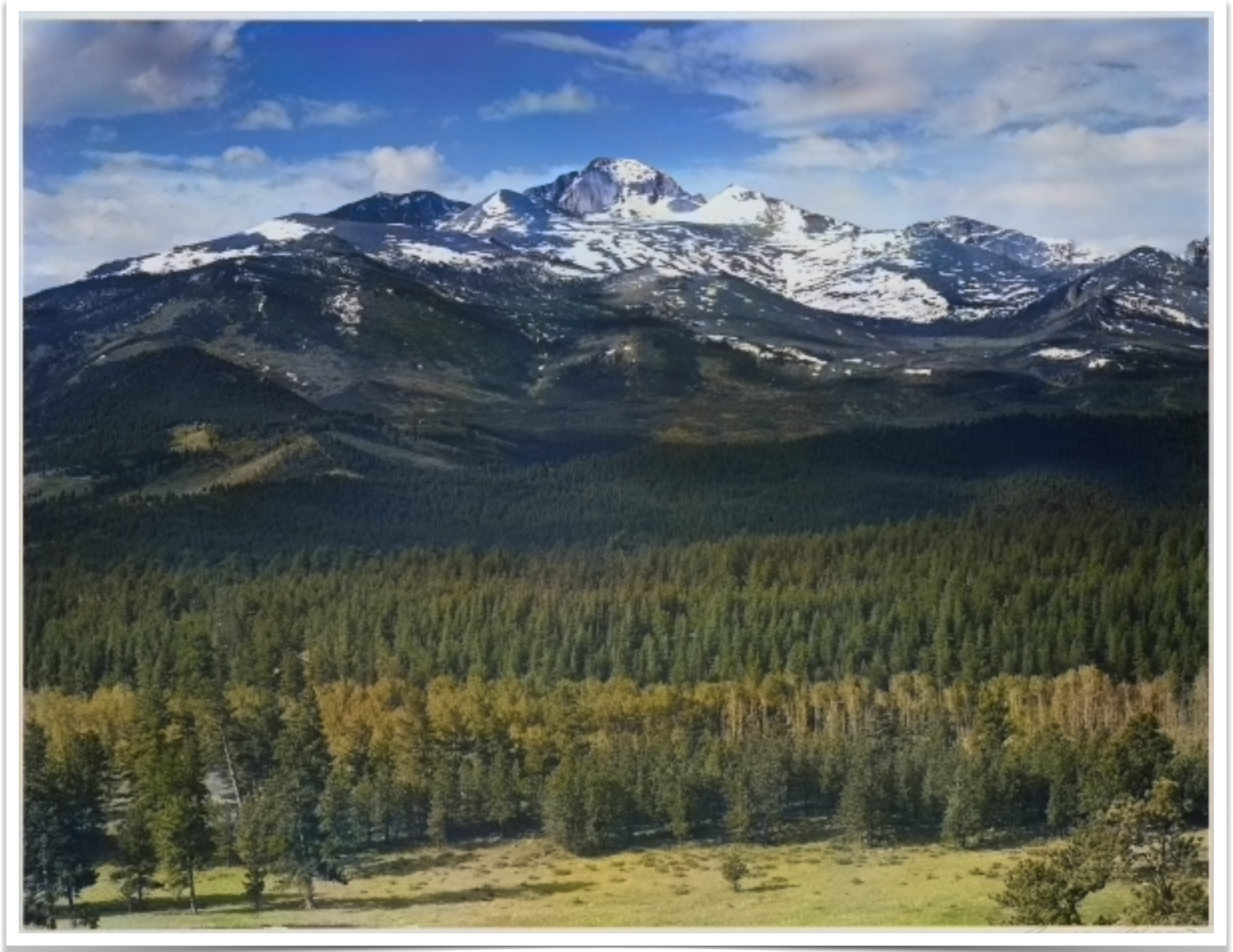
Making new images

Tasks: what humans care about

Adding missing content



Input image



Colorized output

Tasks: what humans care about

Predicting future events

What is going to happen?

1. Introduction to computer vision

- History
- Perception versus measurement
- Simple vision system
- Taxonomy of computer vision tasks