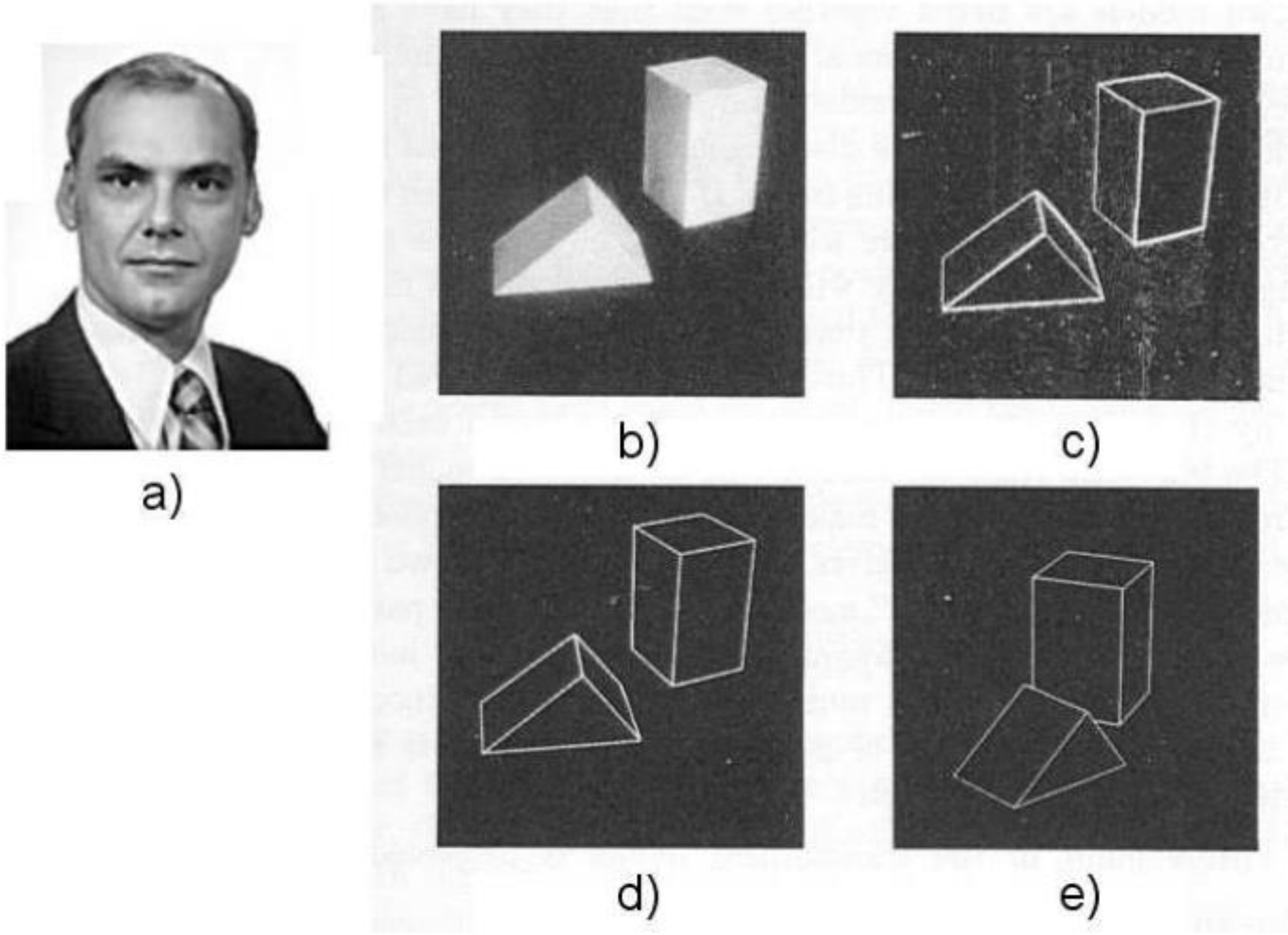
Blocks World: A Simple Vision System



Alexei Efros
CS280, Spring 2024

With slides from Torralba textbook

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

PROJECT MAC

Artificial Intelligence Group
Vision Memo. No. 100.

July 7, 1966

THE SUMMER VISION PROJECT

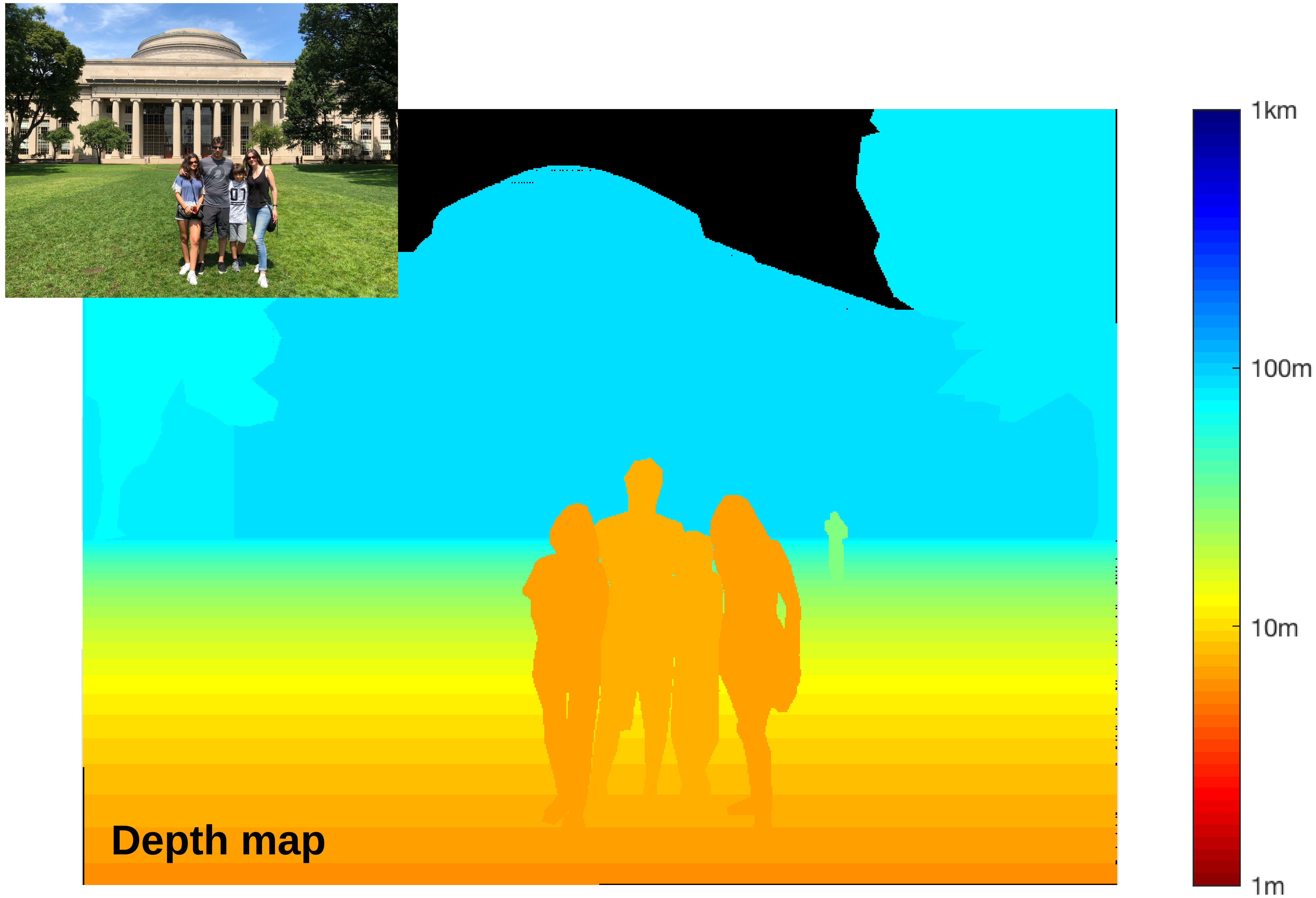
Seymour Papert

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

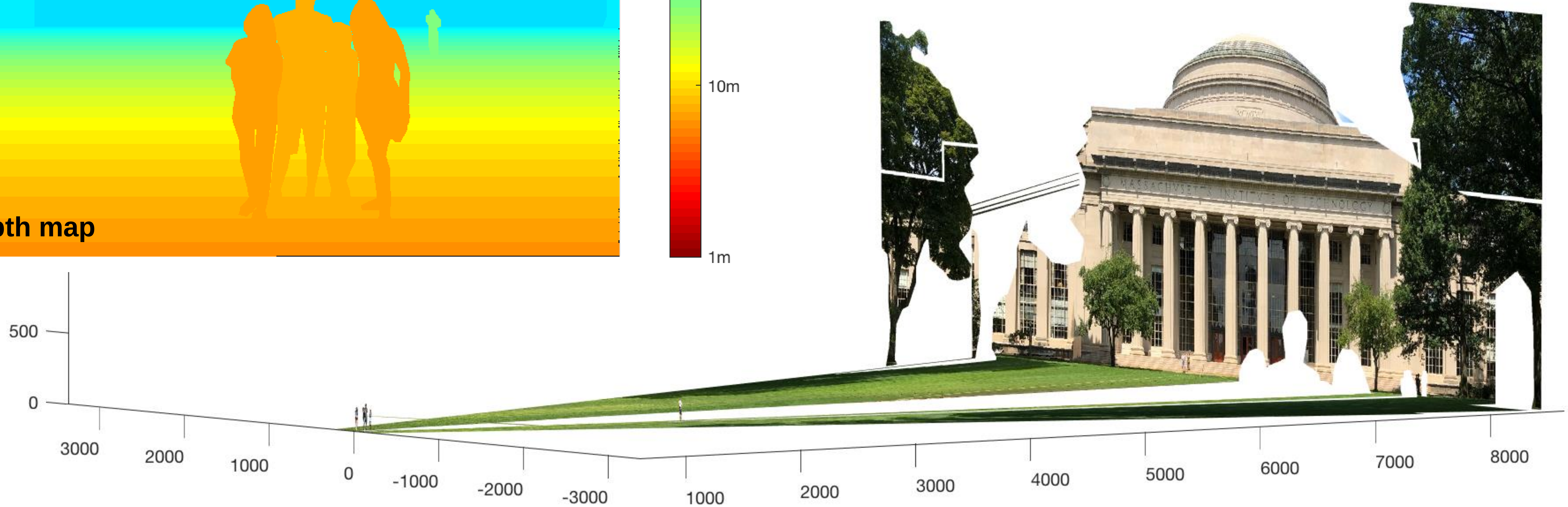
The goal of the first homework is to solve vision!

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

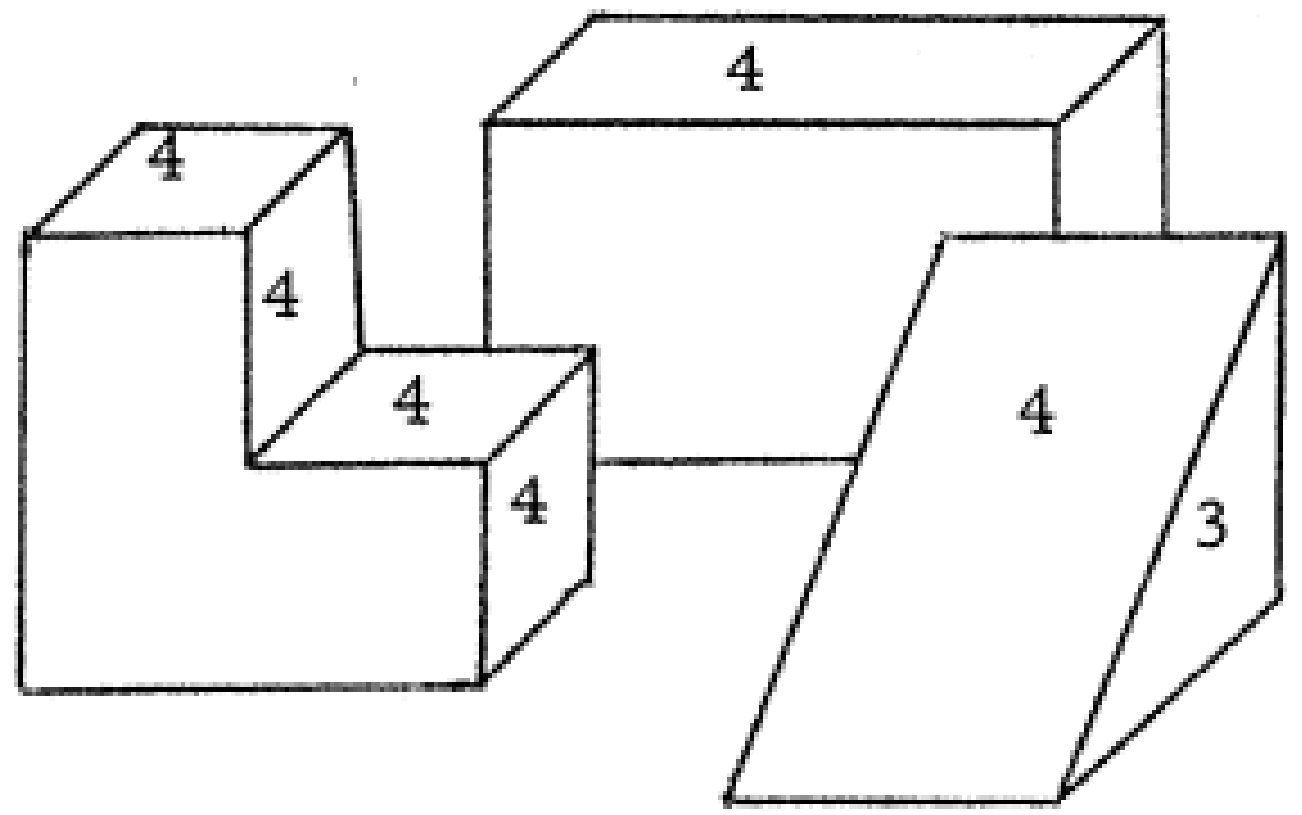
A Simple World

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

Roberts, Blocks world, Copy Demo (1960s)

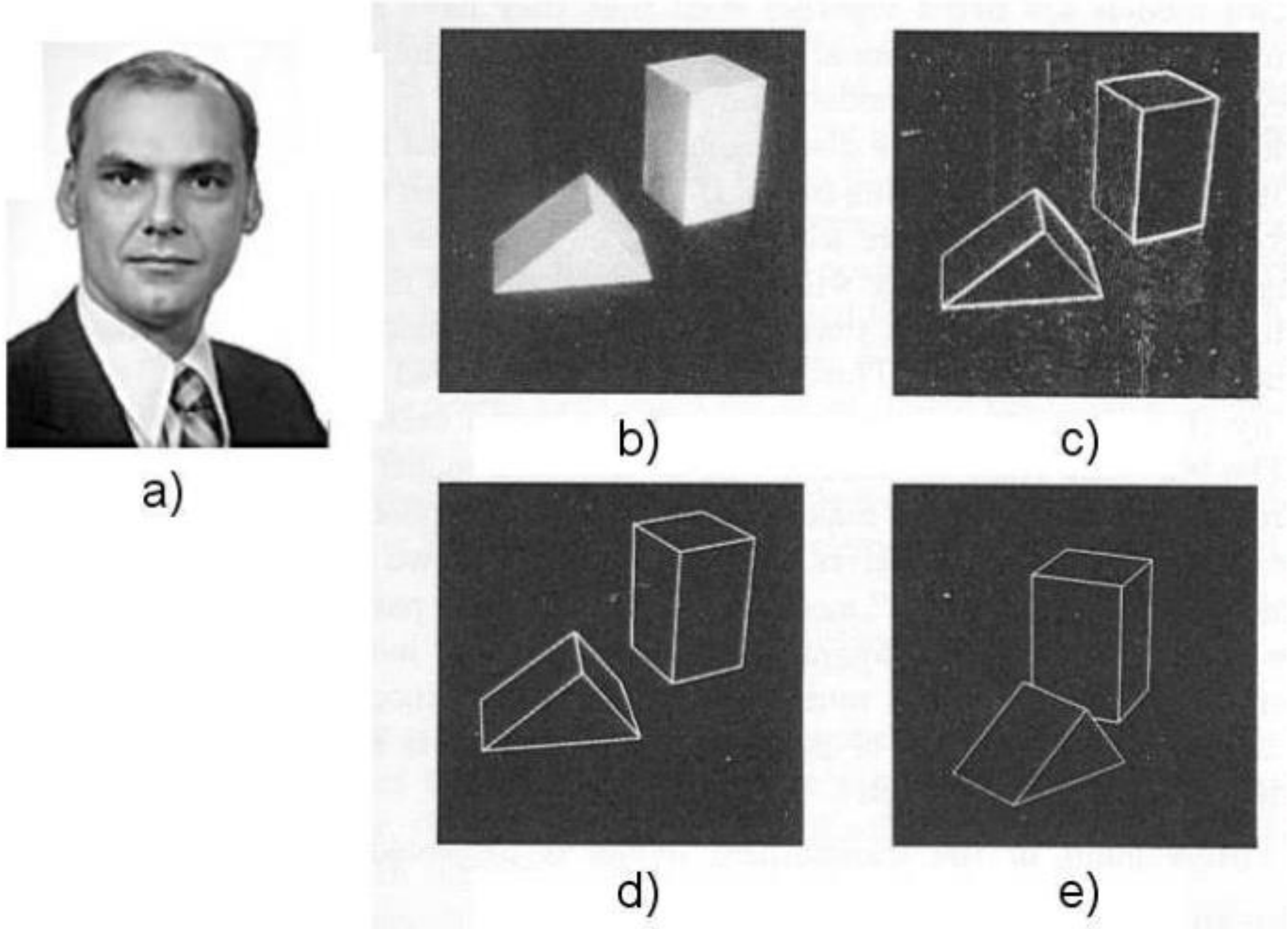
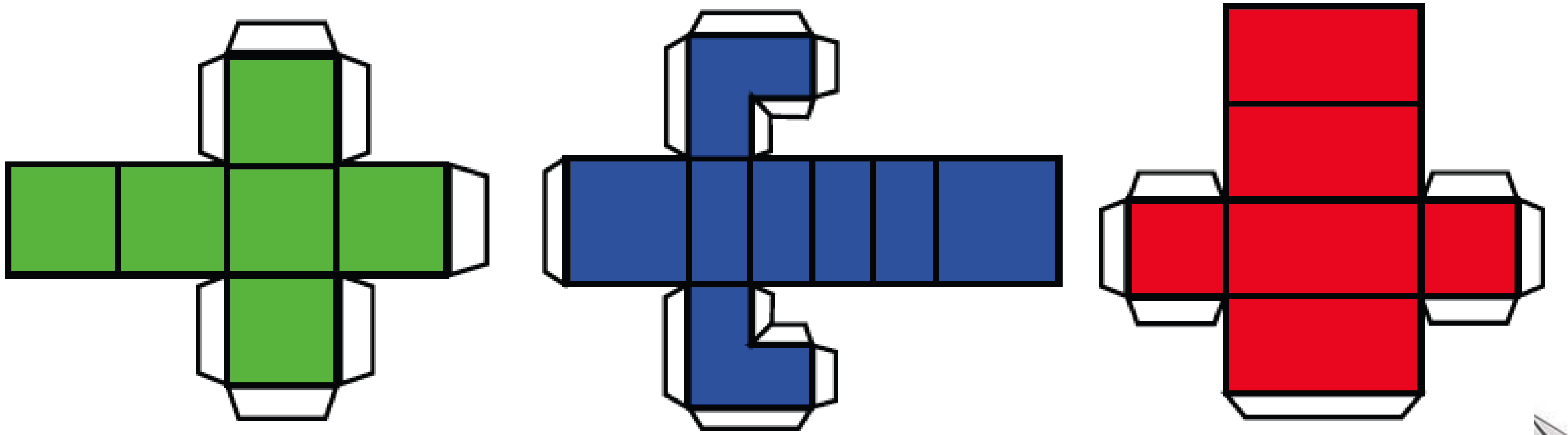
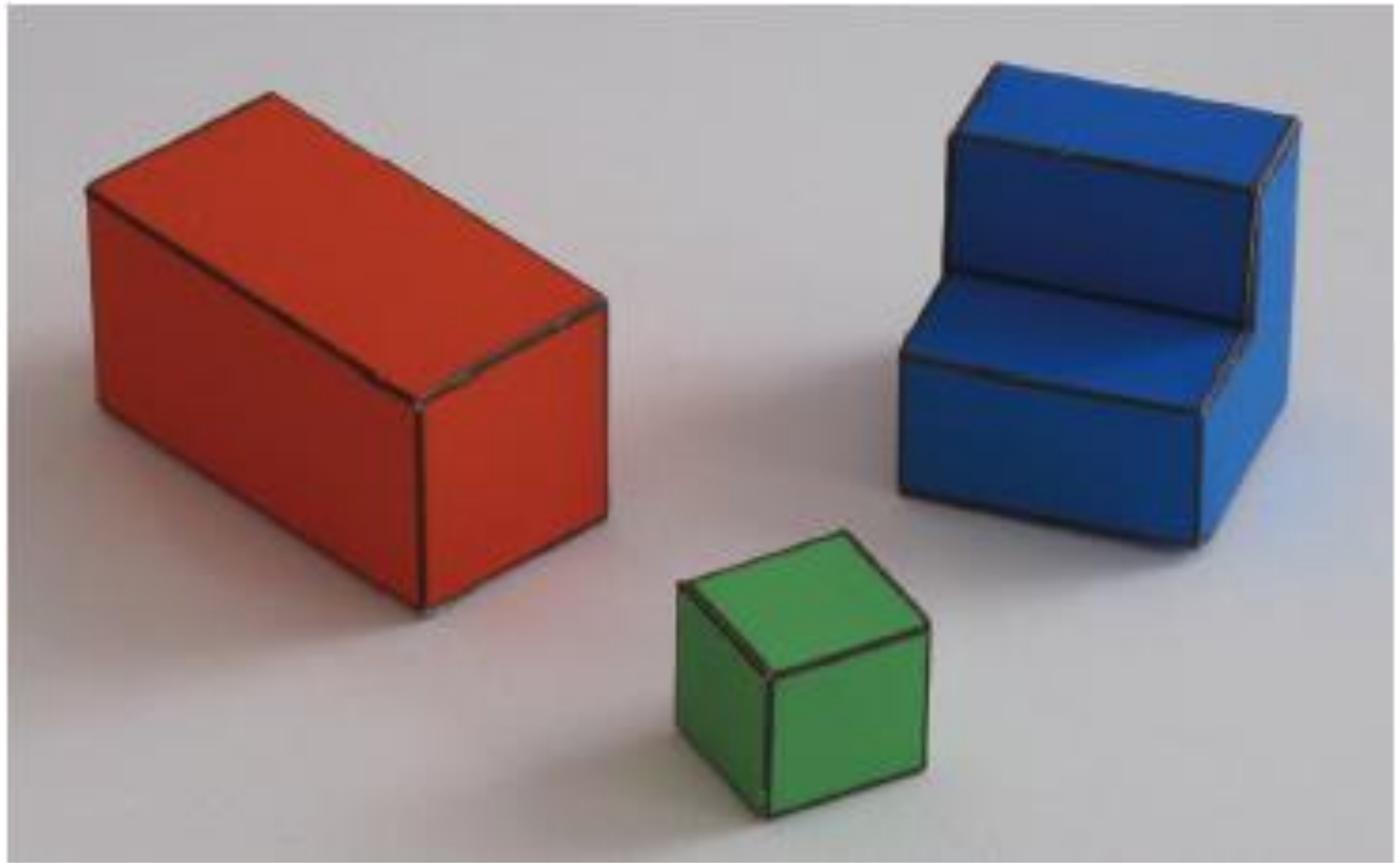


Fig. 1. A system for recognizing 3-d polyhedral scenes. a) L.G. Roberts. b)A blocks world scene. c)Detected edges using a 2x2 gradient operator. d) A 3-d polyhedral description of the scene, formed automatically from the single image. e) The 3-d scene displayed with a viewpoint different from the original image to demonstrate its accuracy and completeness. (b) - e) are taken from [64] with permission MIT Press.)

Object Recognition in the Geometric Era: a Retrospective. Joseph L. Mundy. 2006

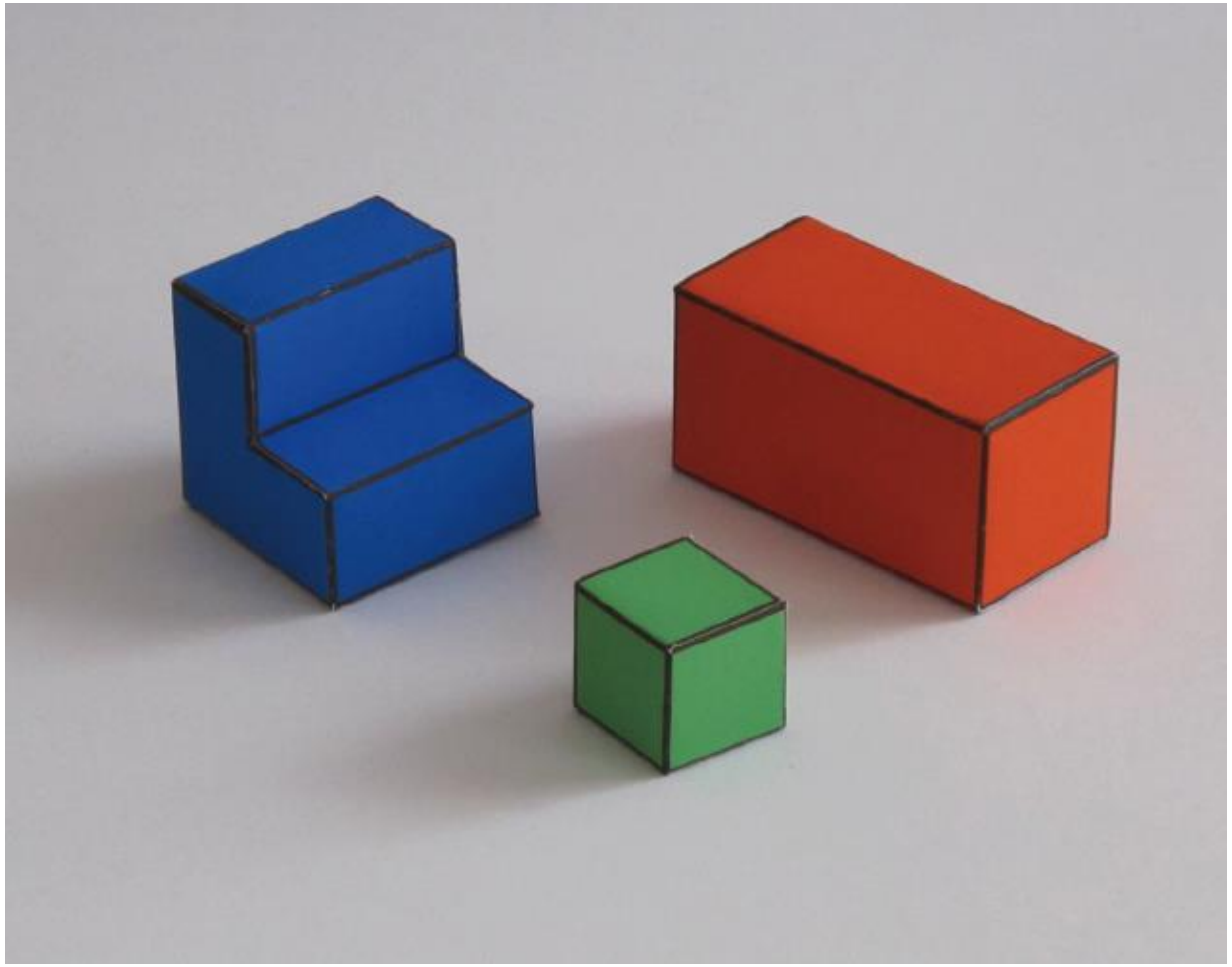
Build your own simple world





A simple goal

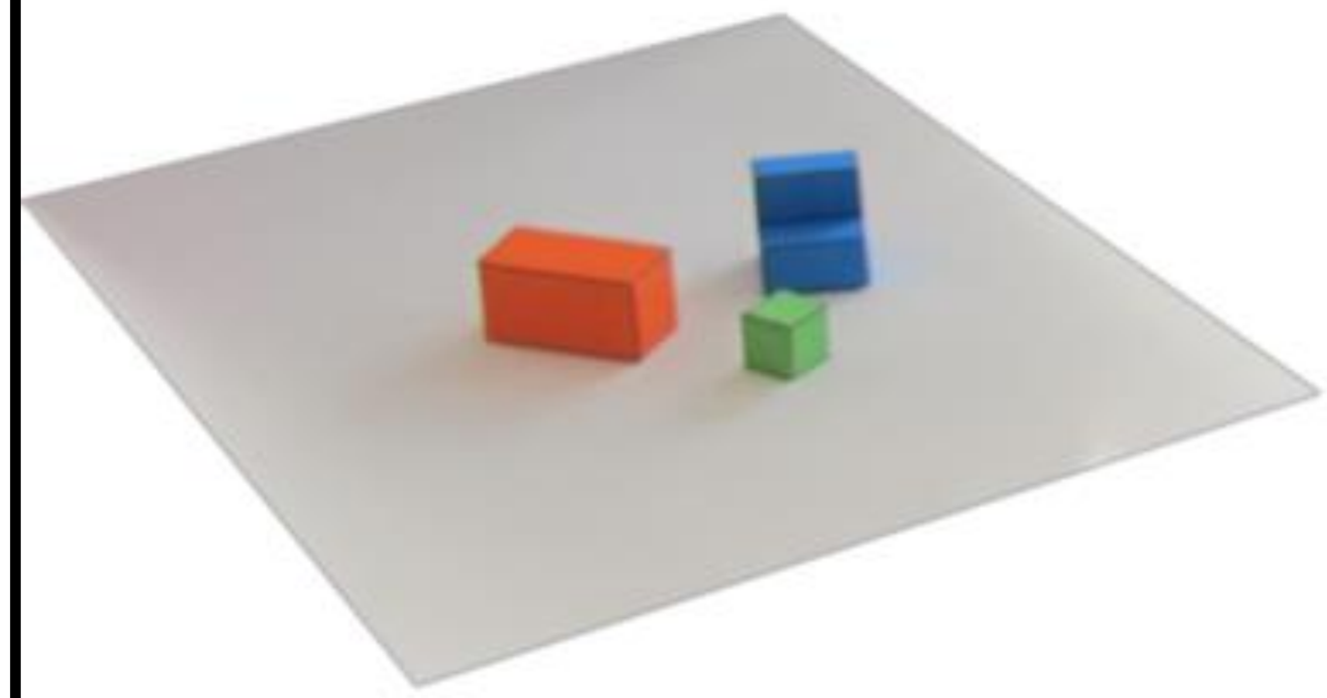
To recover the 3D structure of the world from the 2D image



We will make this goal more explicit later.

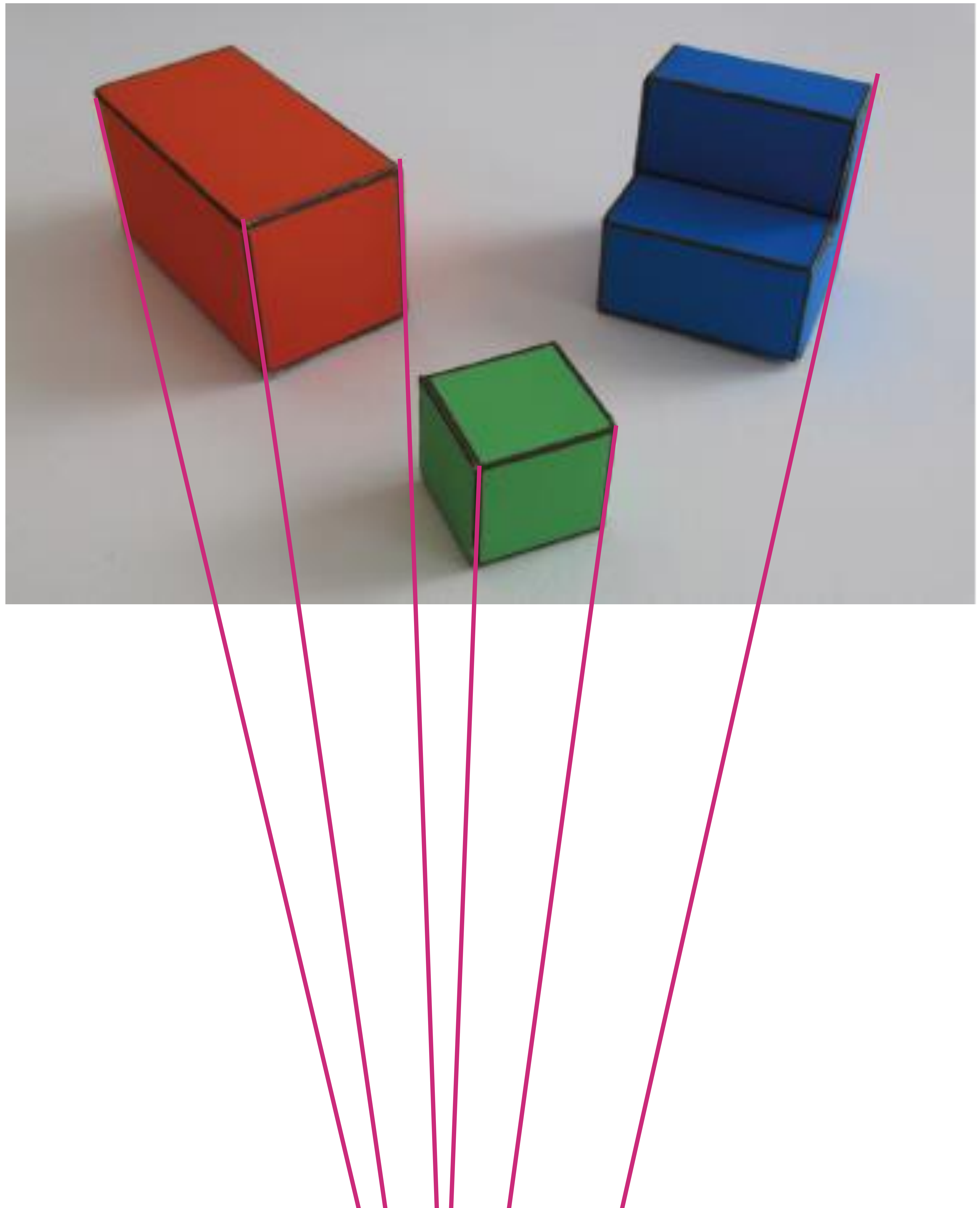
A simple image formation model

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

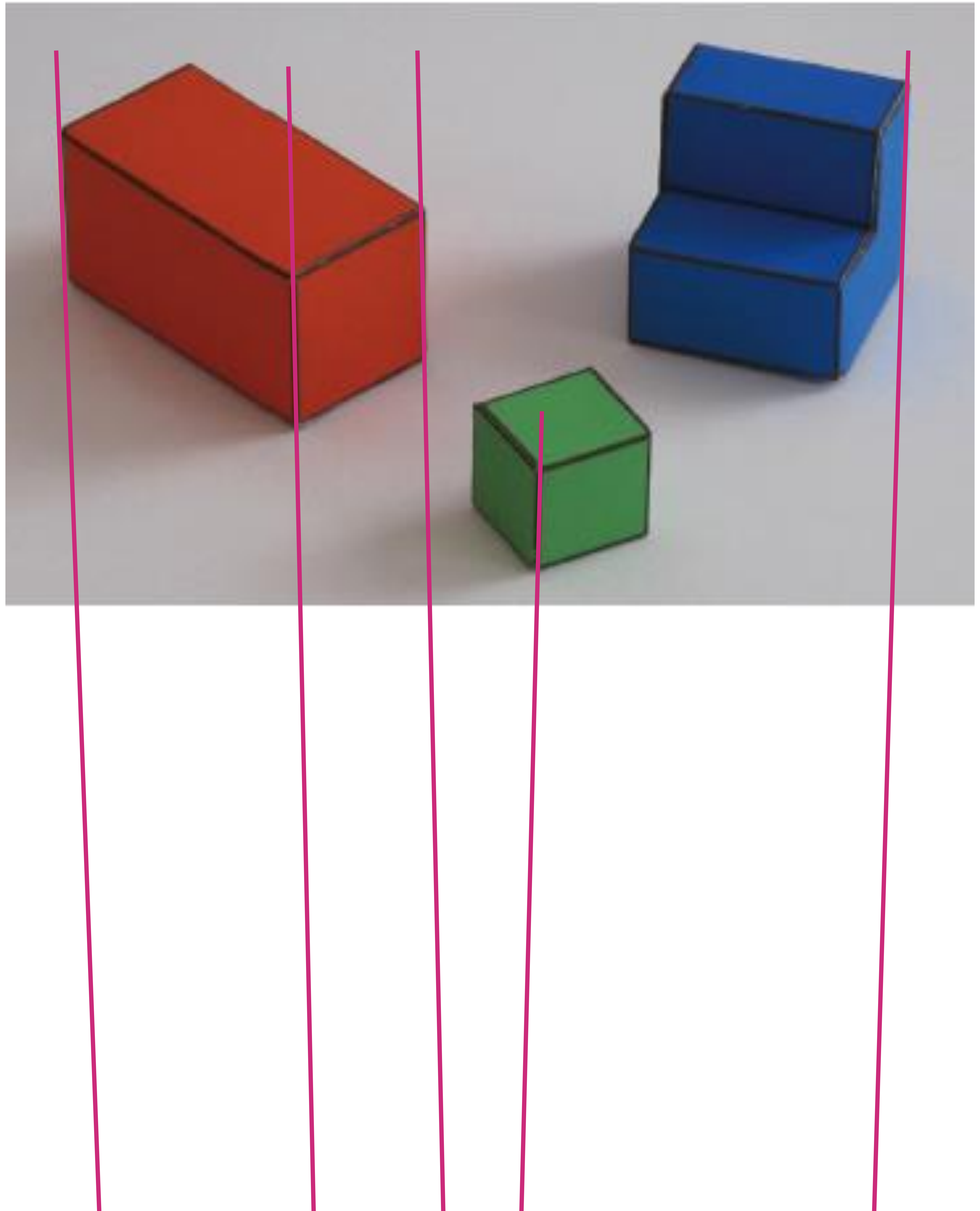


A simple image formation model

Perspective projection

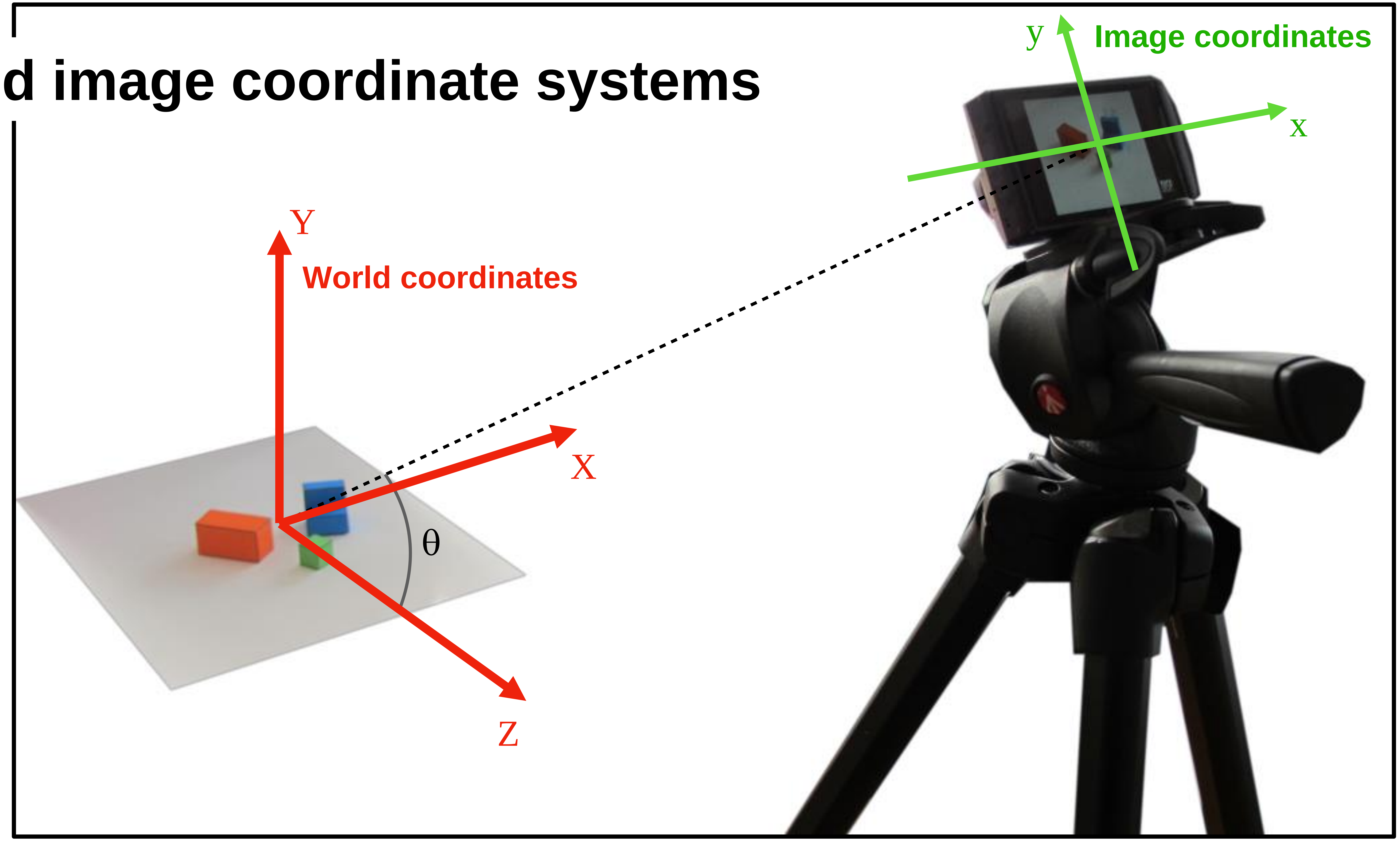


Parallel (orthographic) projection



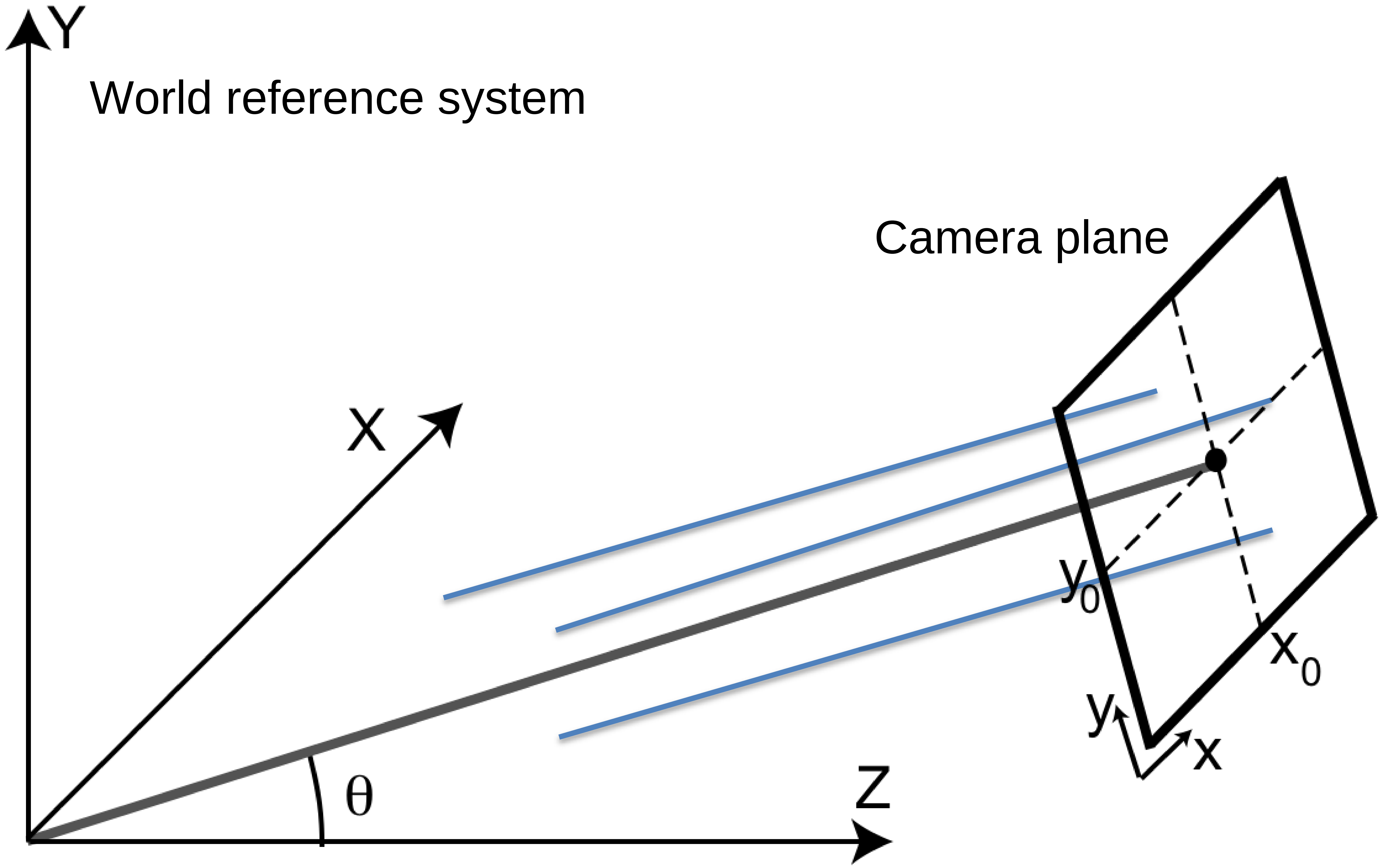
A simple image formation model

World and image coordinate systems



(right-handed reference system)

A simple image formation model



(right-handed reference system)

A simple image formation model

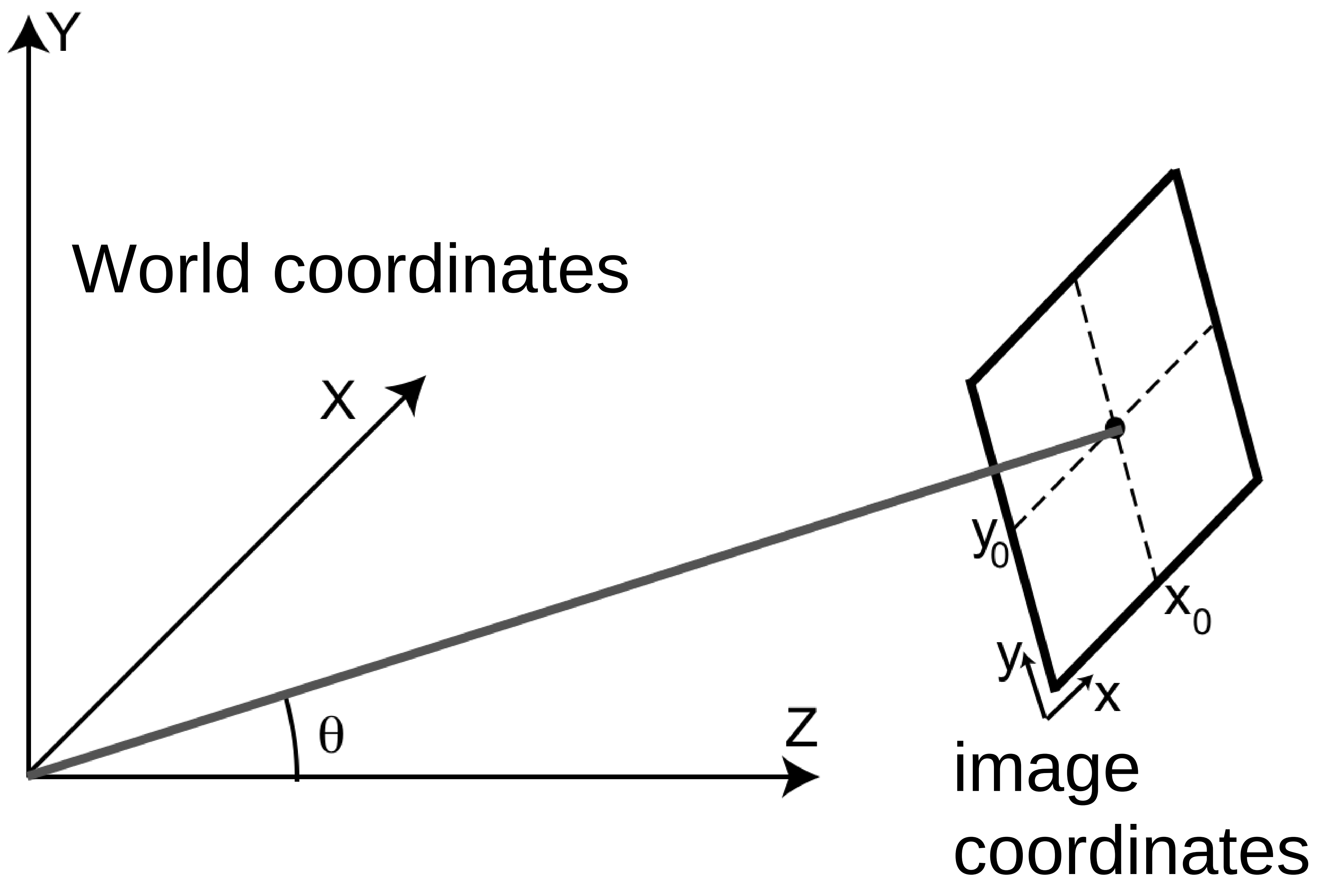
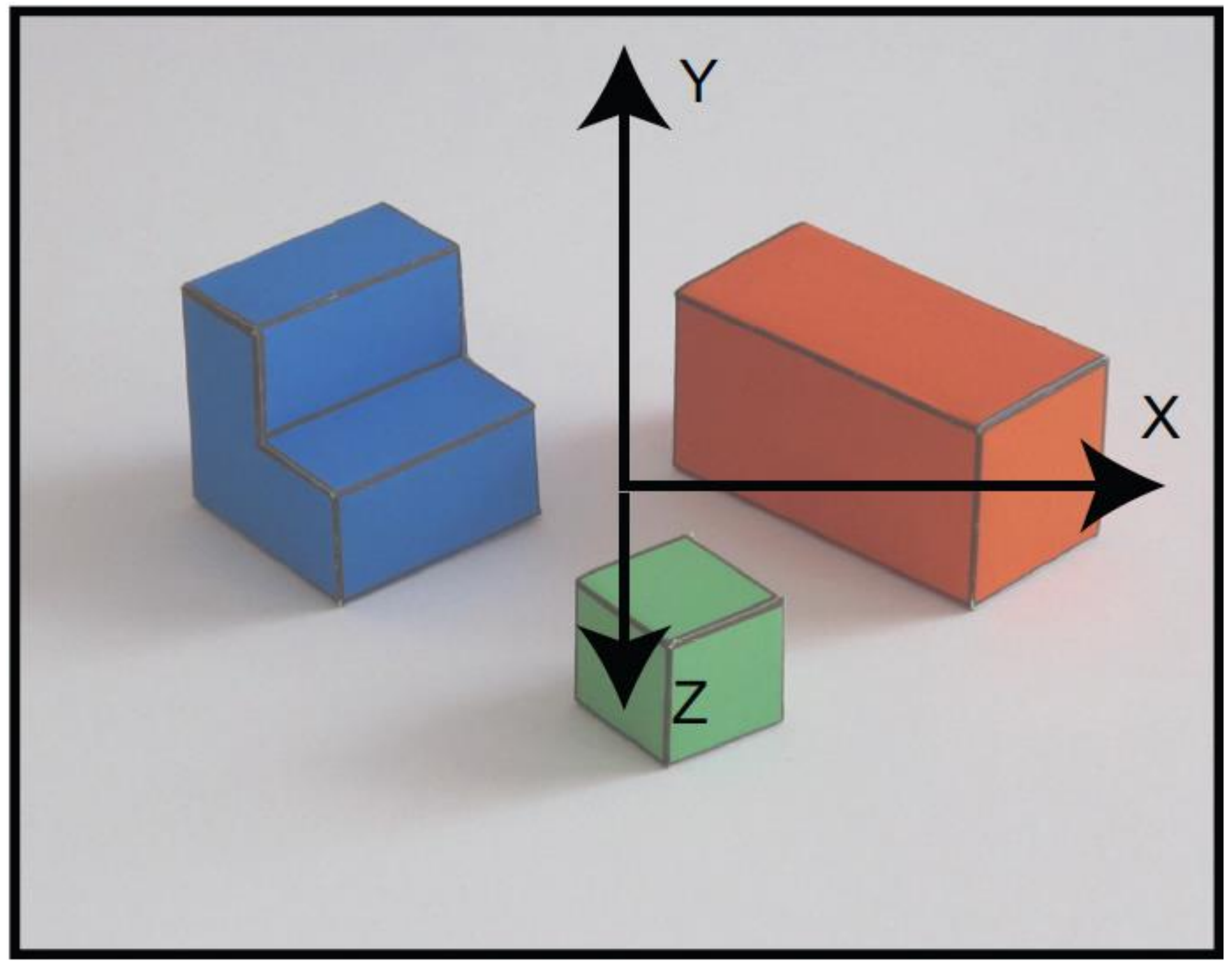


Image and projection of the world coordinate axes into the image plane



World coordinates

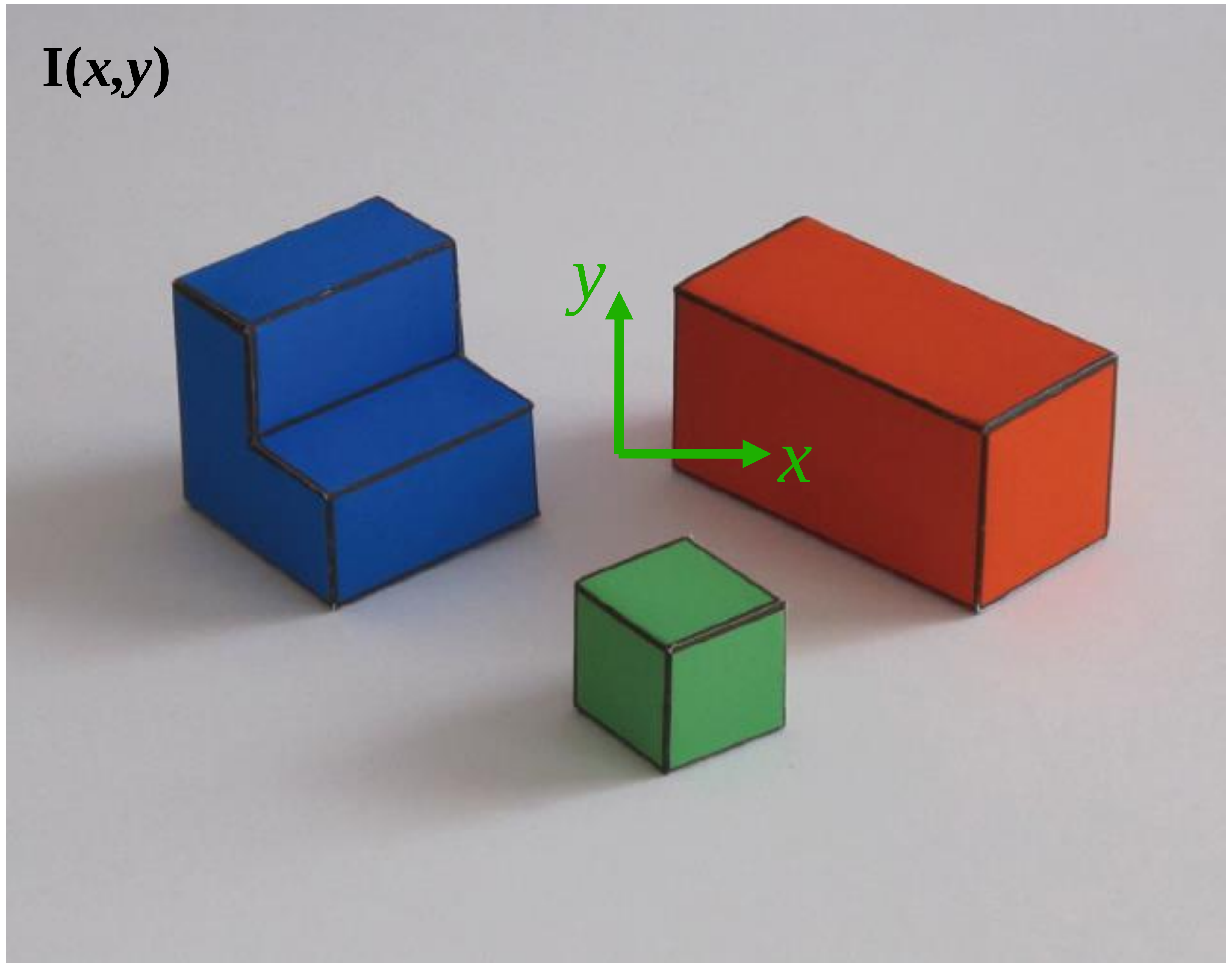
$$x = X + x_0$$

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

image coordinates

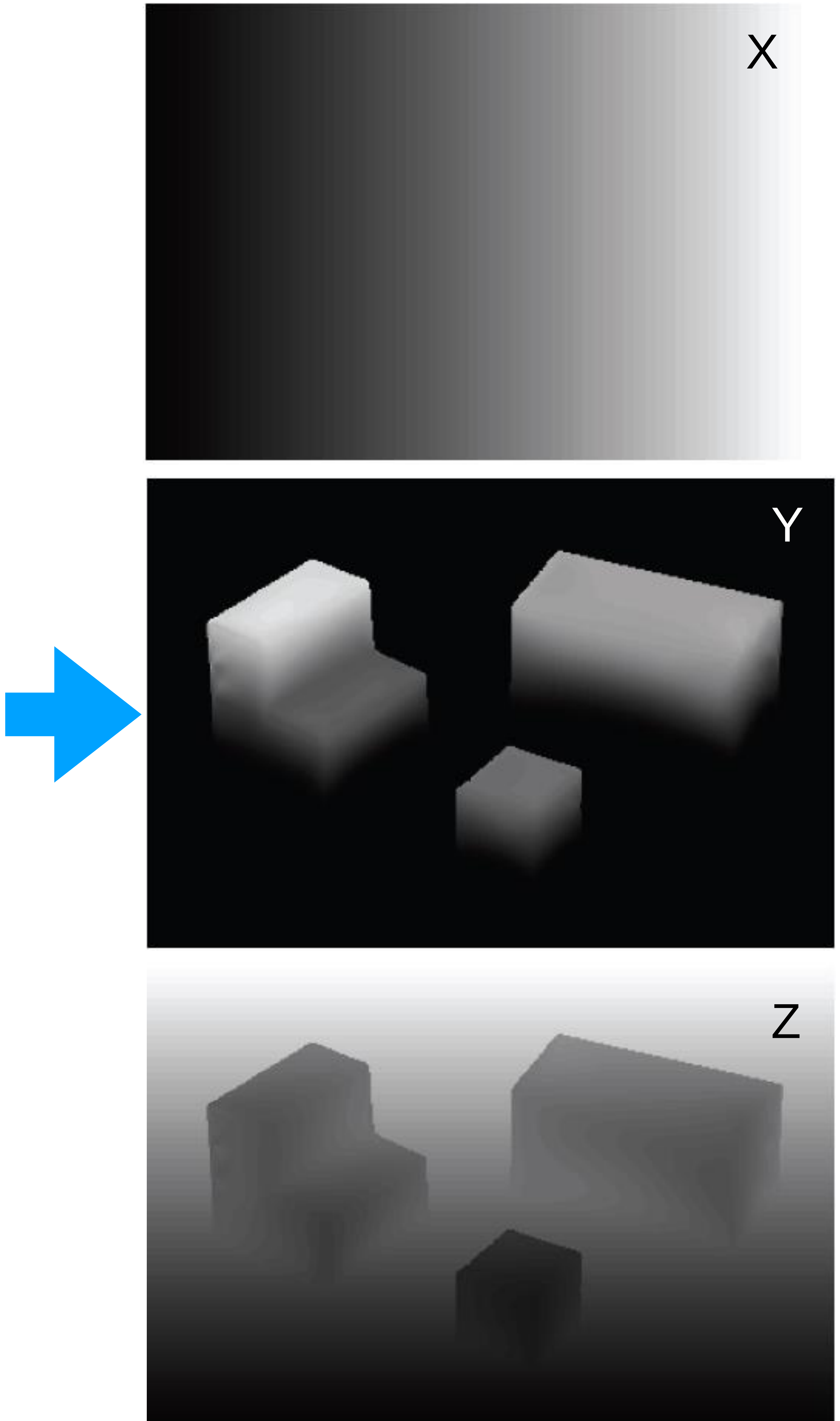
A simple goal

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

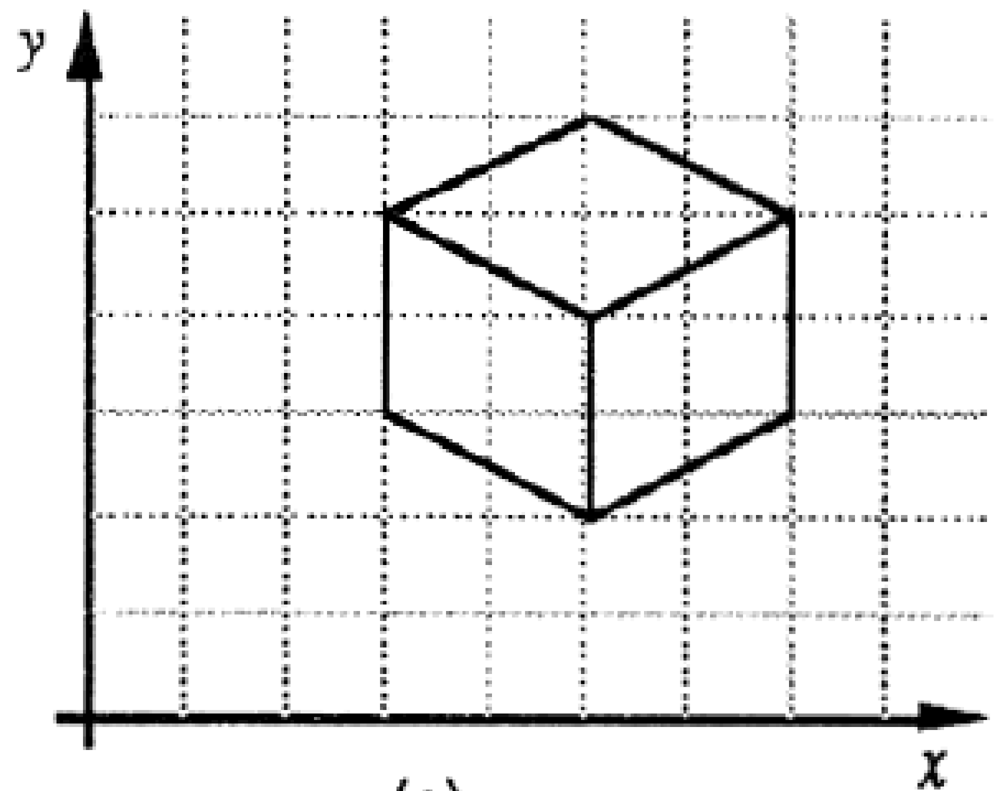


We want to recover **X(*x,y*), Y(*x,y*), Z(x,y)** using as input **I(*x,y*)**

Output



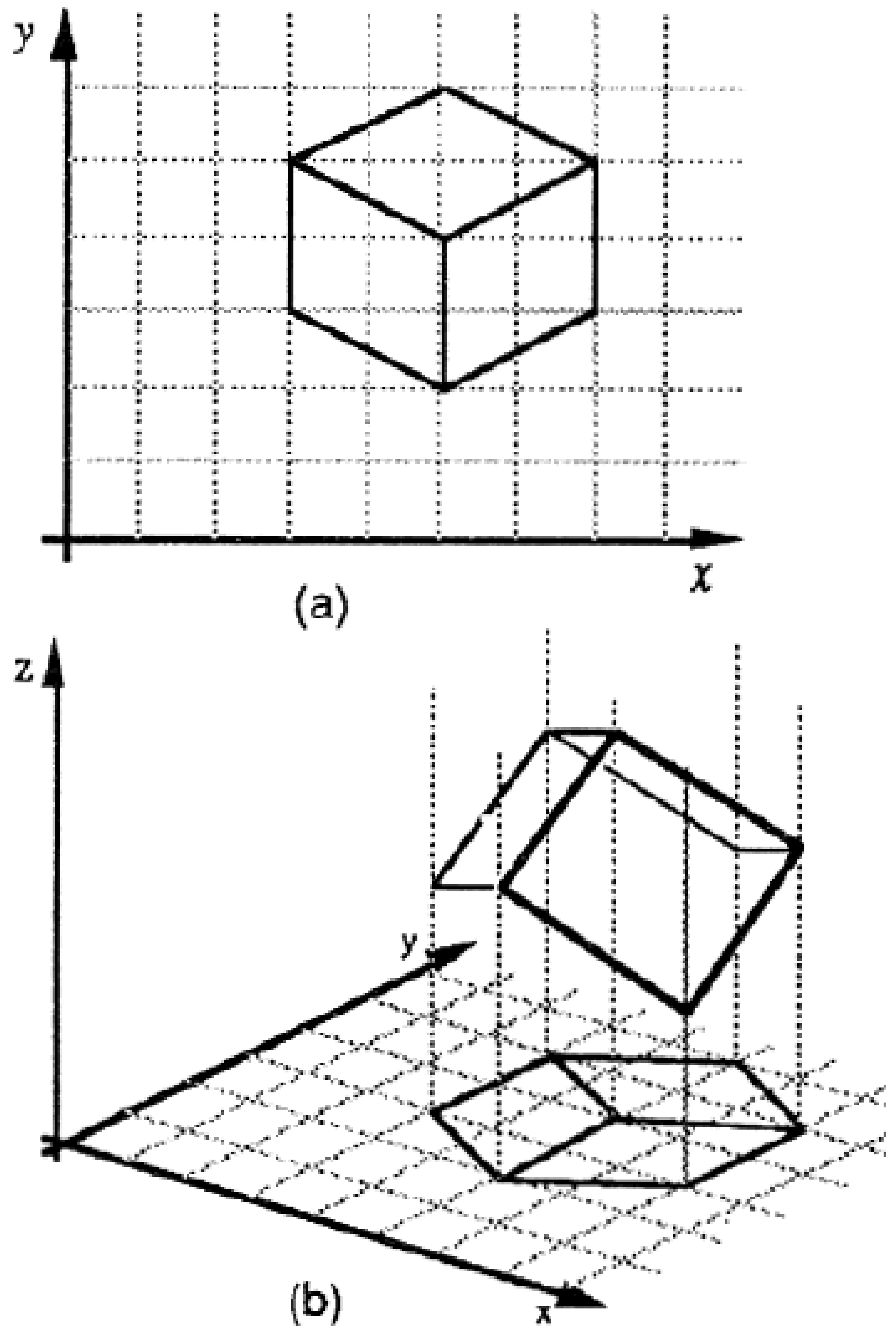
Why is this hard?



(a)

Sinha & Adelson 93

Why is this hard?



Sinha & Adelson 93

Why is this hard?

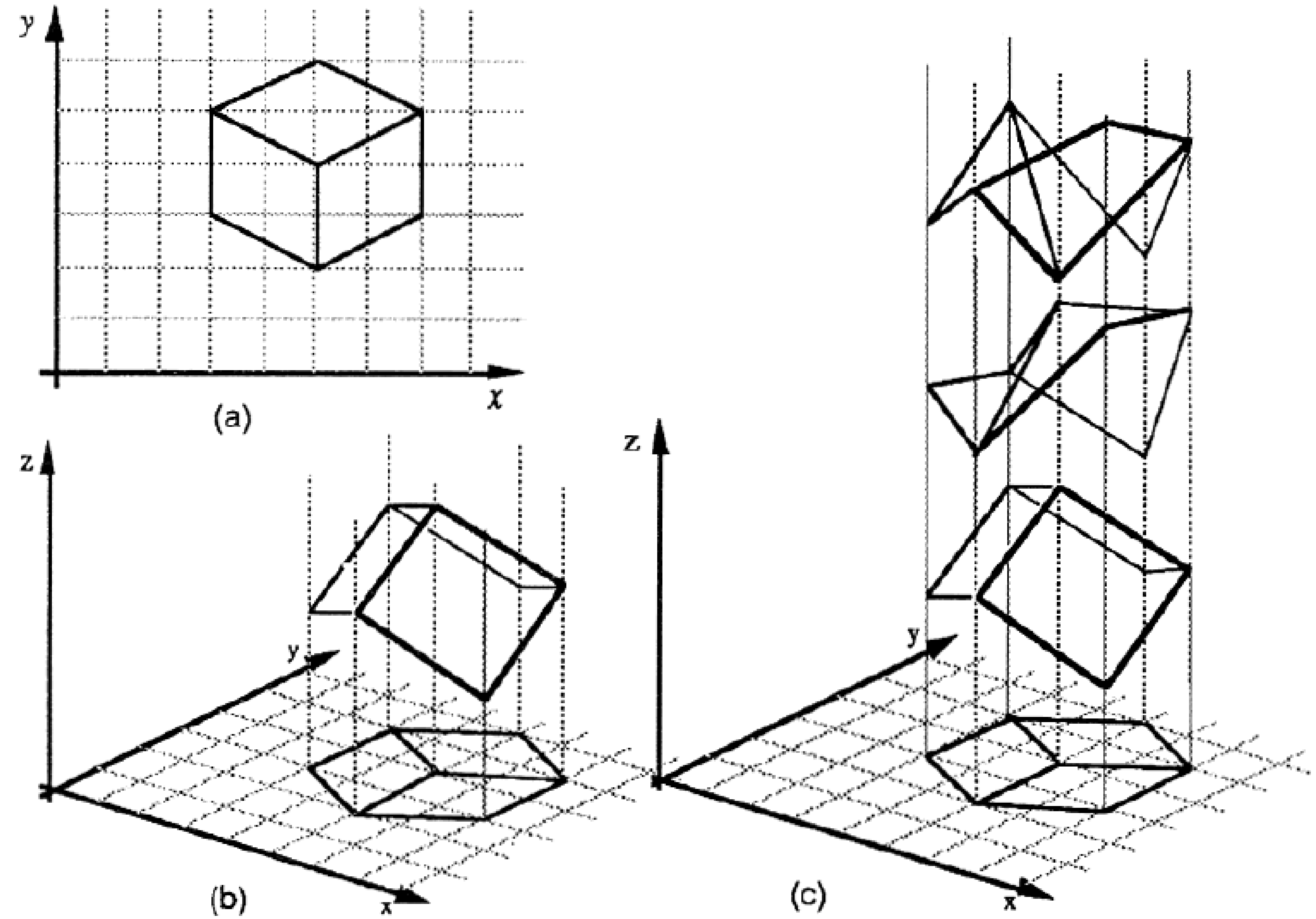
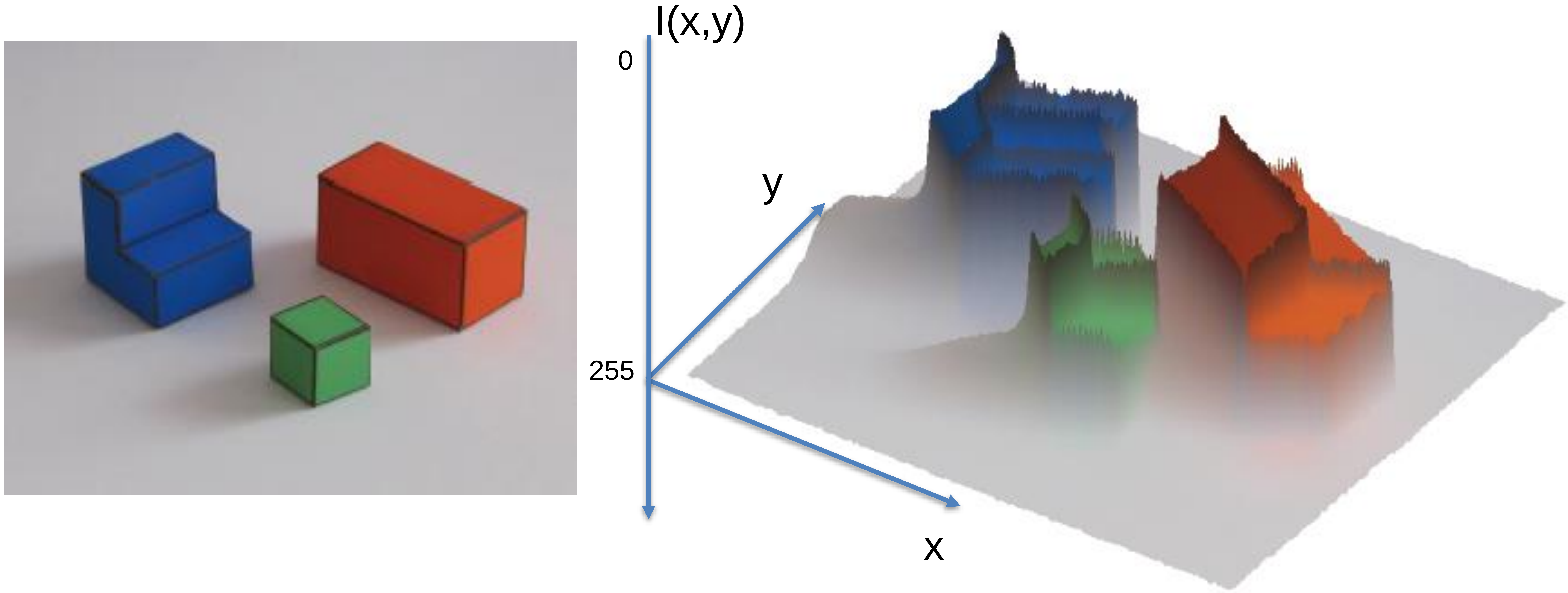


Figure 1. (a) A line drawing provides information only about the x, y coordinates of points lying along the object contours. (b) The human visual system is usually able to reconstruct an object in three dimensions given only a single 2D projection (c) Any planar line-drawing is geometrically consistent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

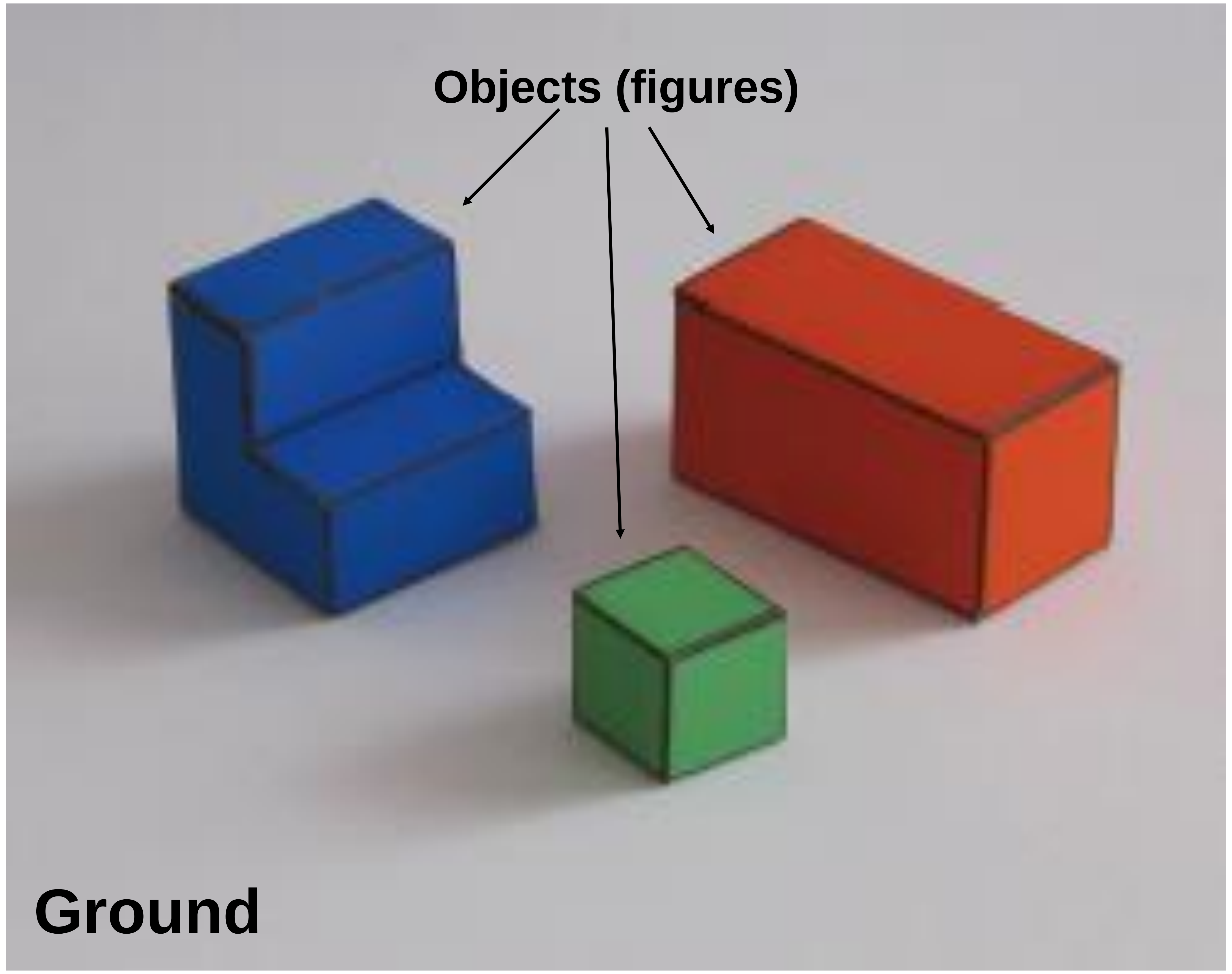
Sinha & Adelson 93

A simple visual system
The input image

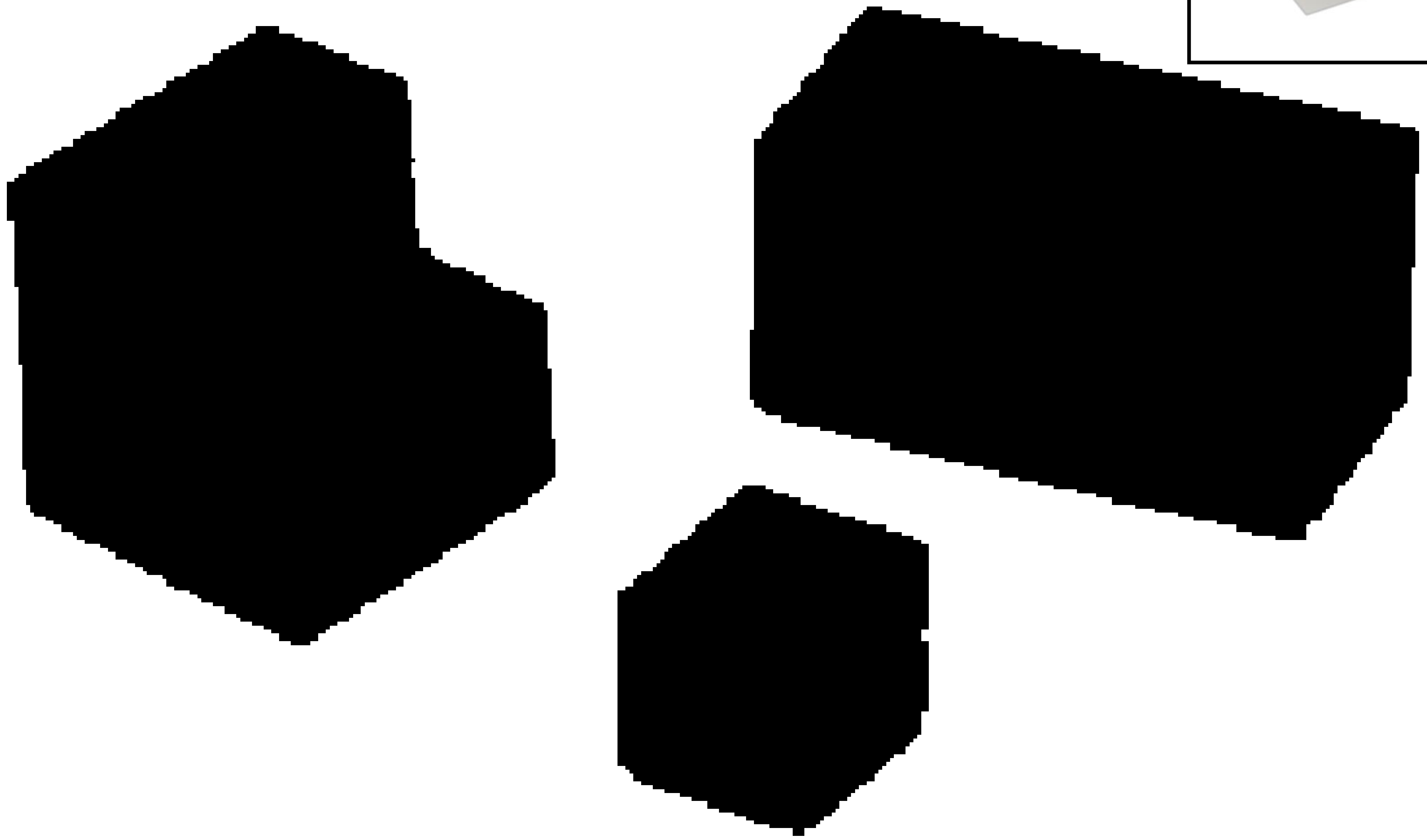


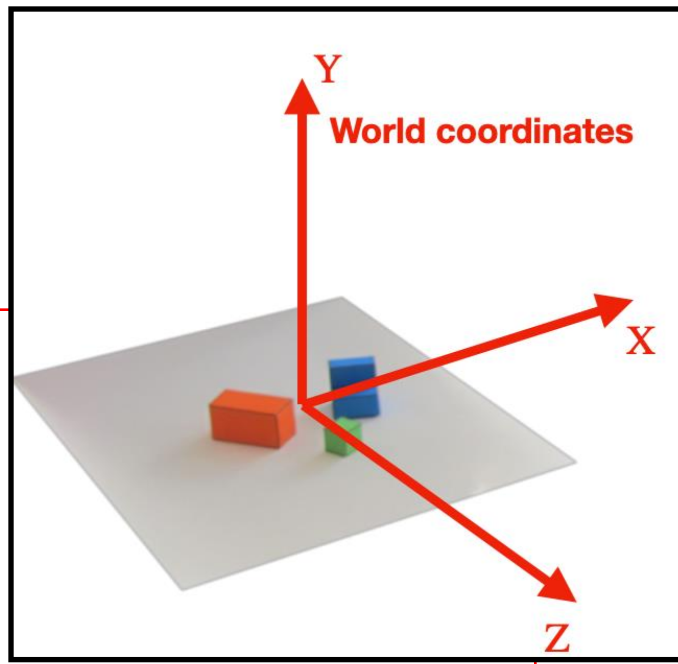
In this **representation**, the image is an array of intensity values (color values) indexed by location.

A better representation: Figure/ground



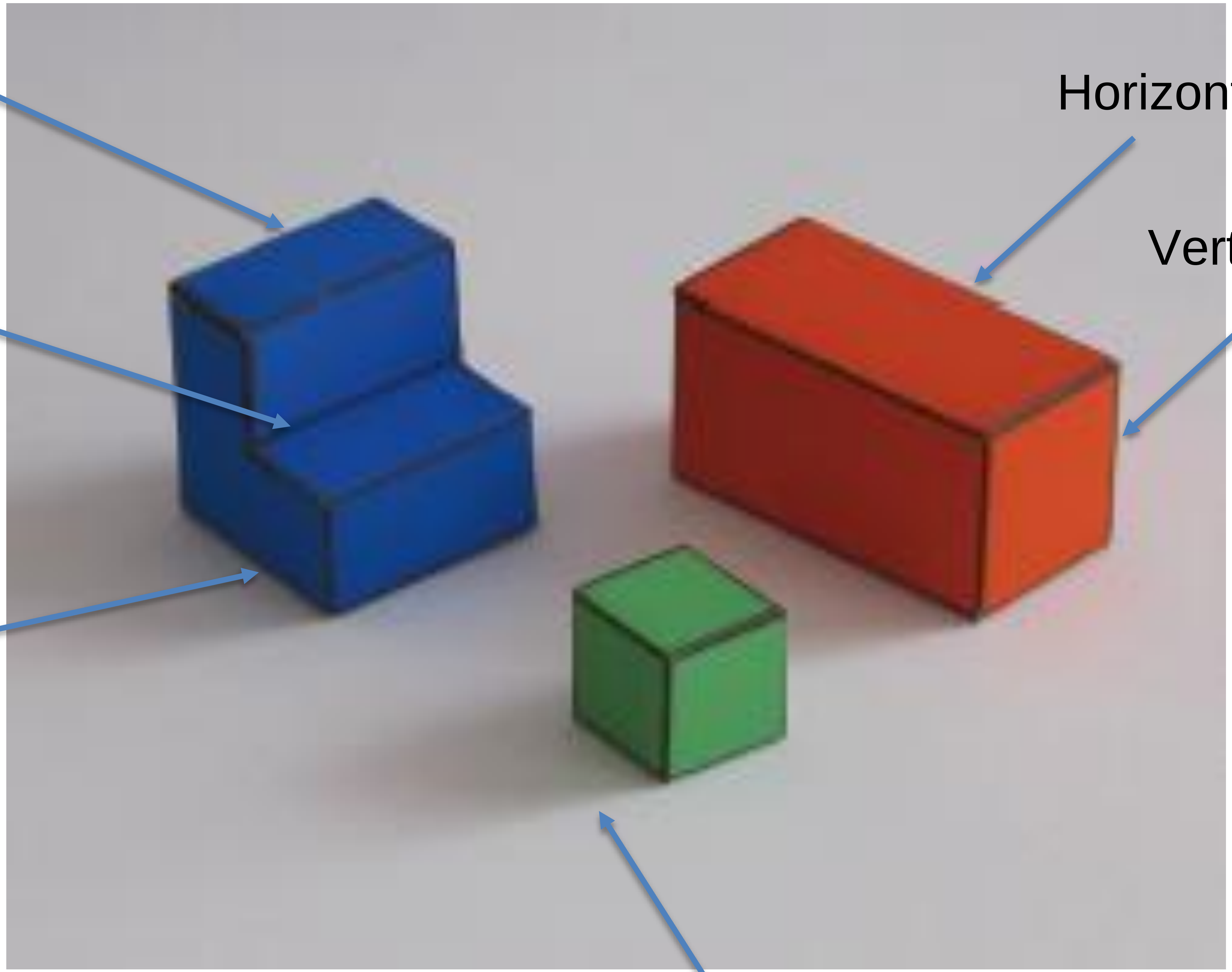
For ground pixels, we know that $Y(x, y) = 0$



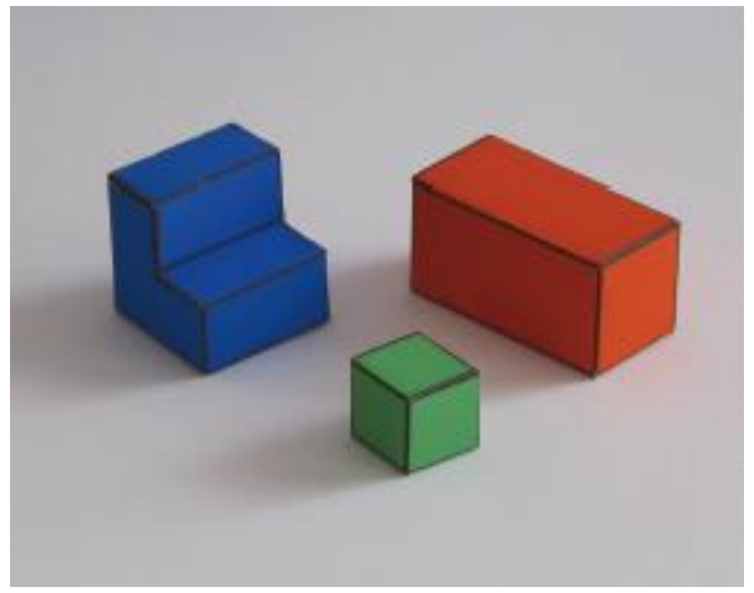


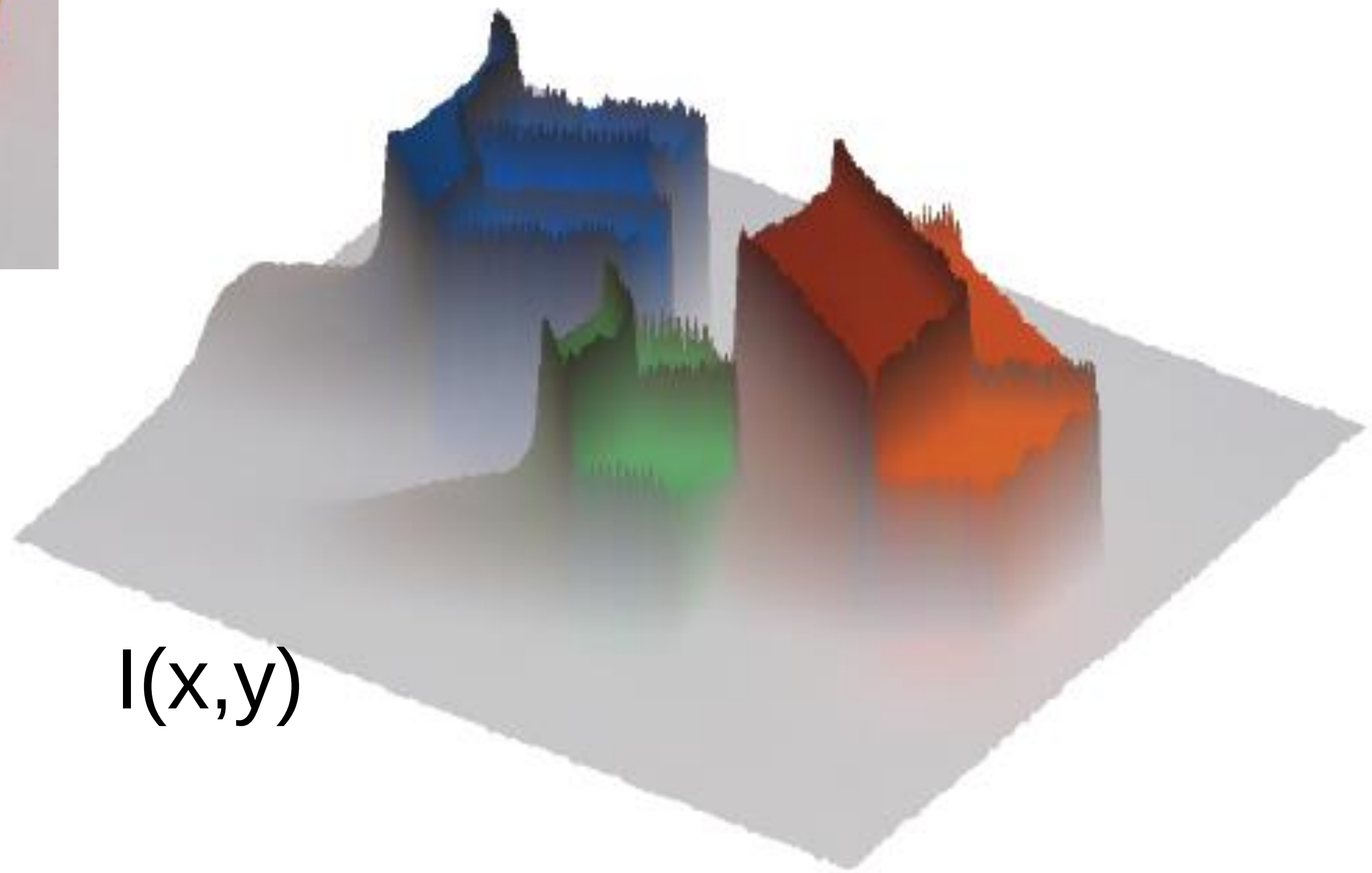
In our simple world:
Using the fact that objects have color and are darker than the ground.

A better representation: Edges



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:

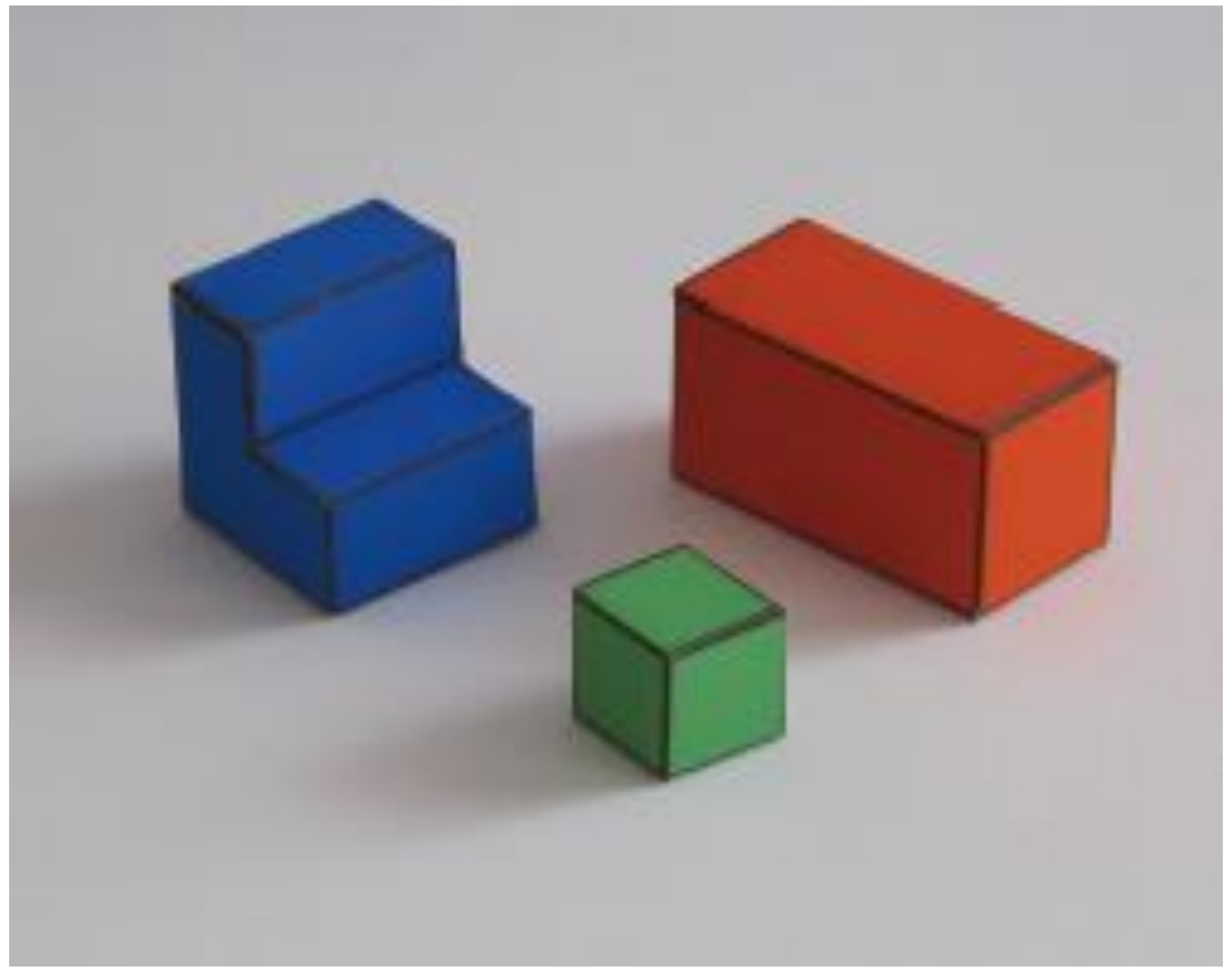
$$\frac{\partial \mathbf{I}}{\partial x} \simeq \mathbf{I}(x,y) - \mathbf{I}(x-1,y)$$

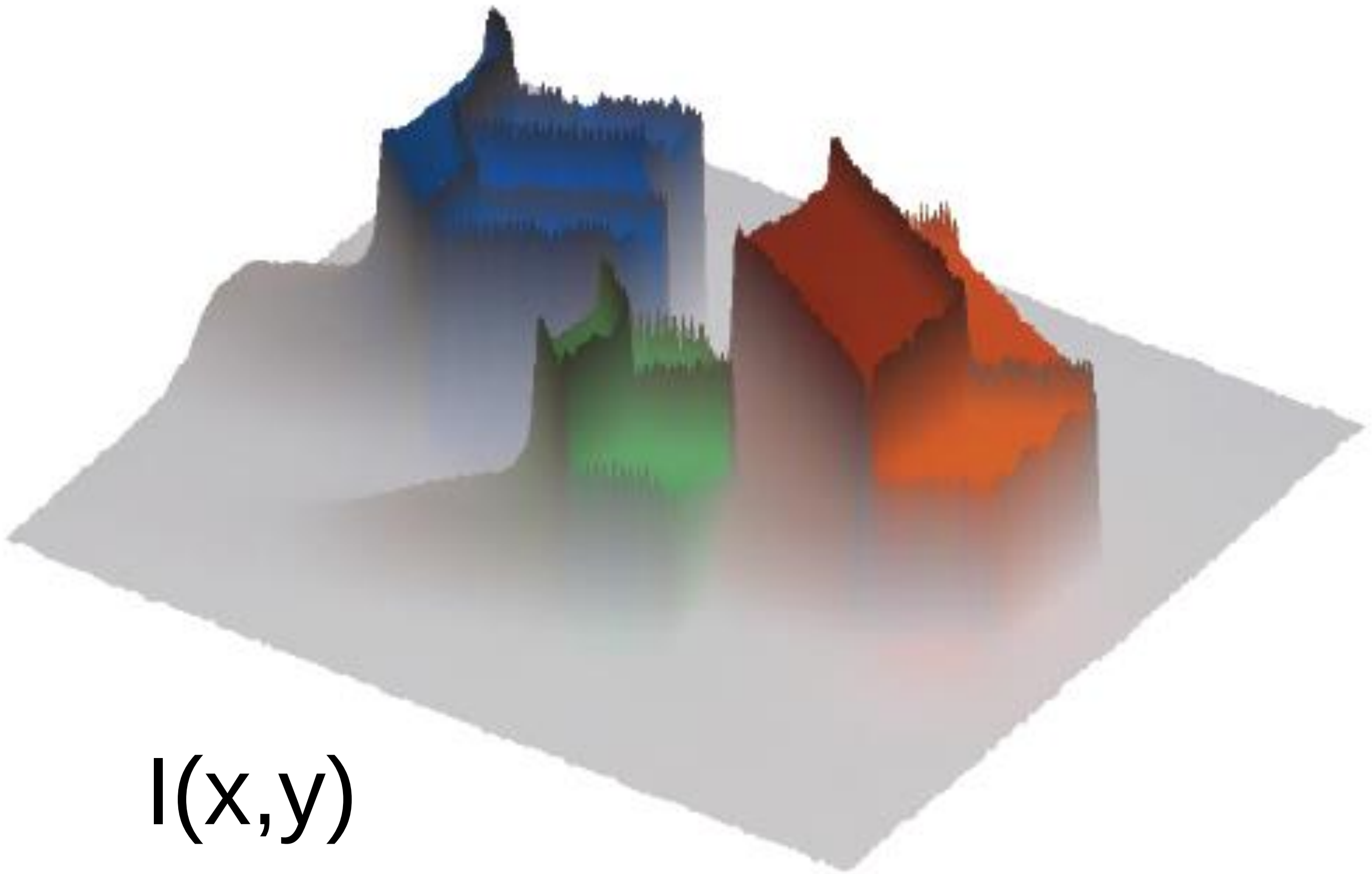
Edge strength $$E(x,y) = |\nabla \mathbf{I}(x,y)|$$

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

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

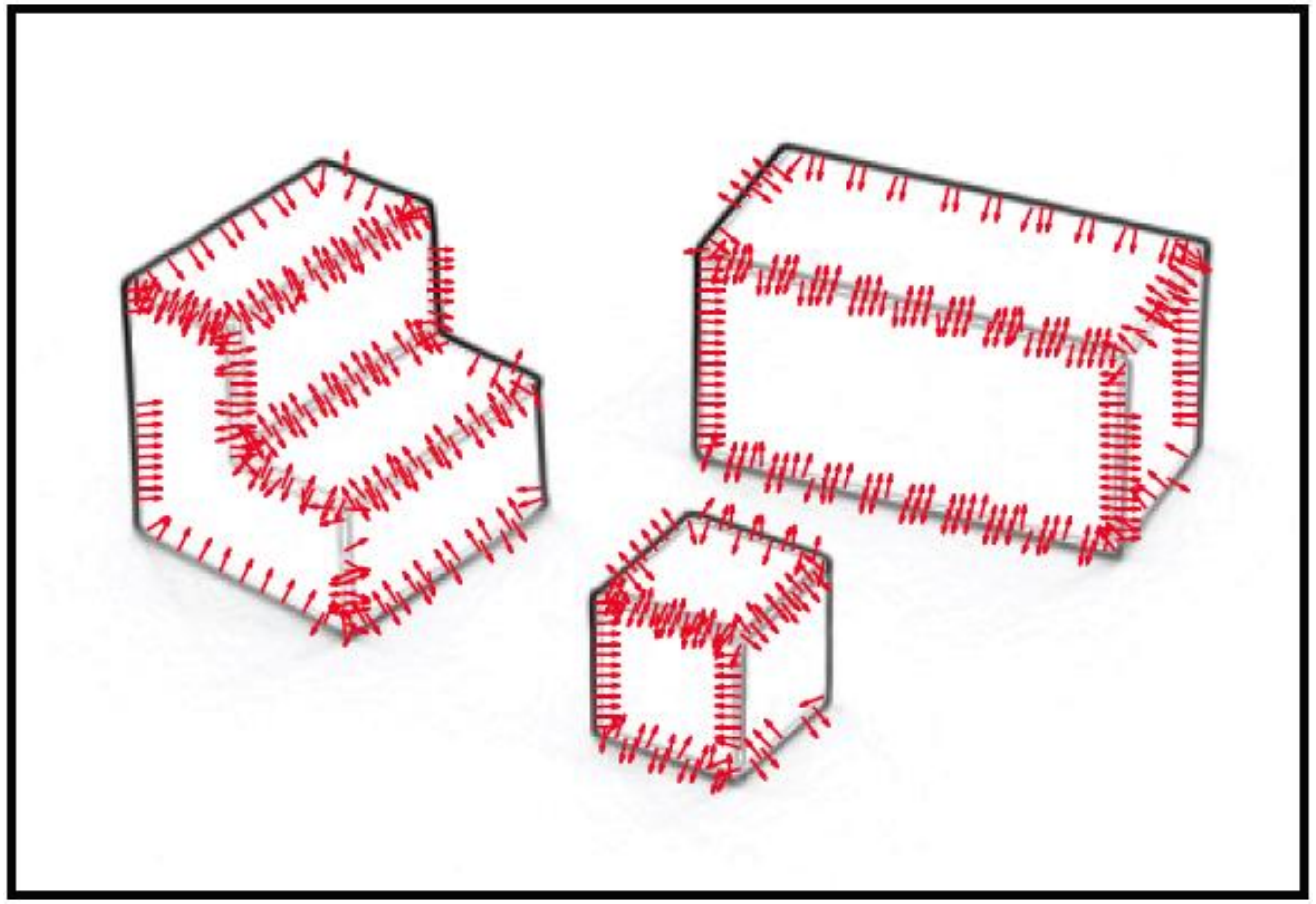
Finding edges in the image





I(x,y)

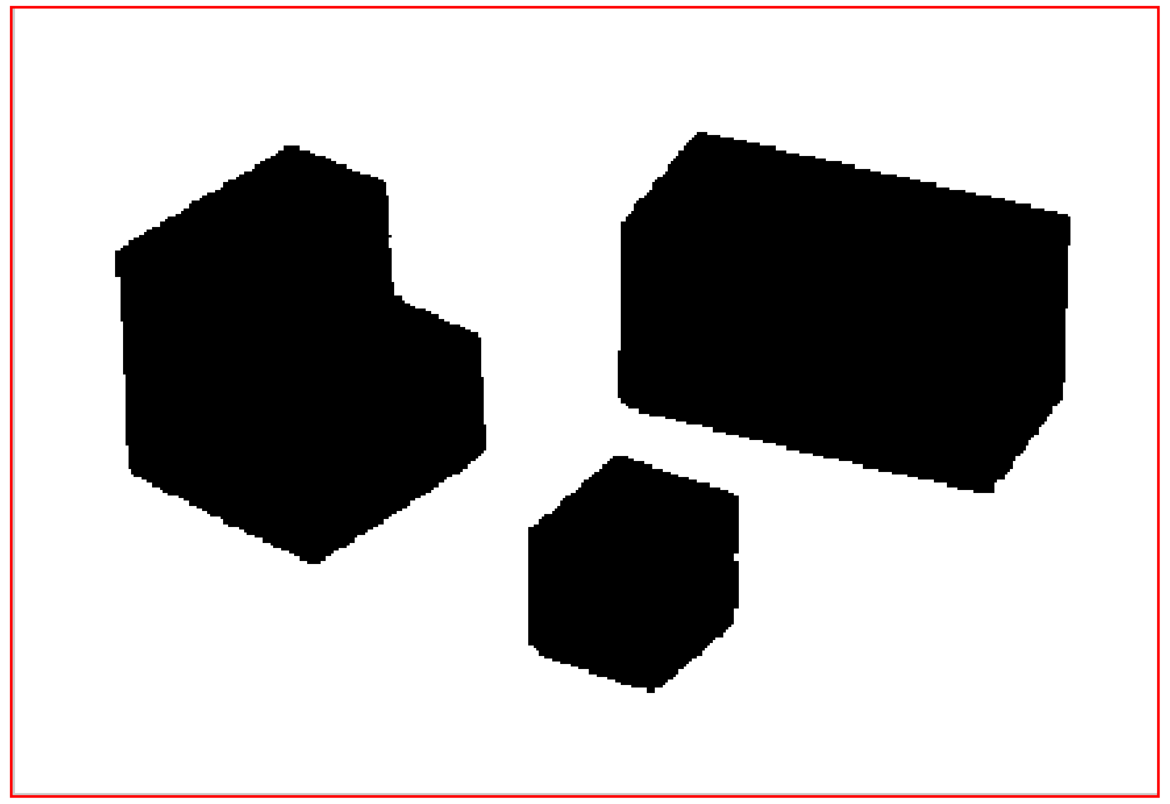
$$\nabla \mathbf{I} = \left(\frac{\partial \mathbf{I}}{\partial x}, \frac{\partial \mathbf{I}}{\partial y} \right) \qquad \mathbf{n} = \frac{\nabla \mathbf{I}}{|\nabla \mathbf{I}|}$$



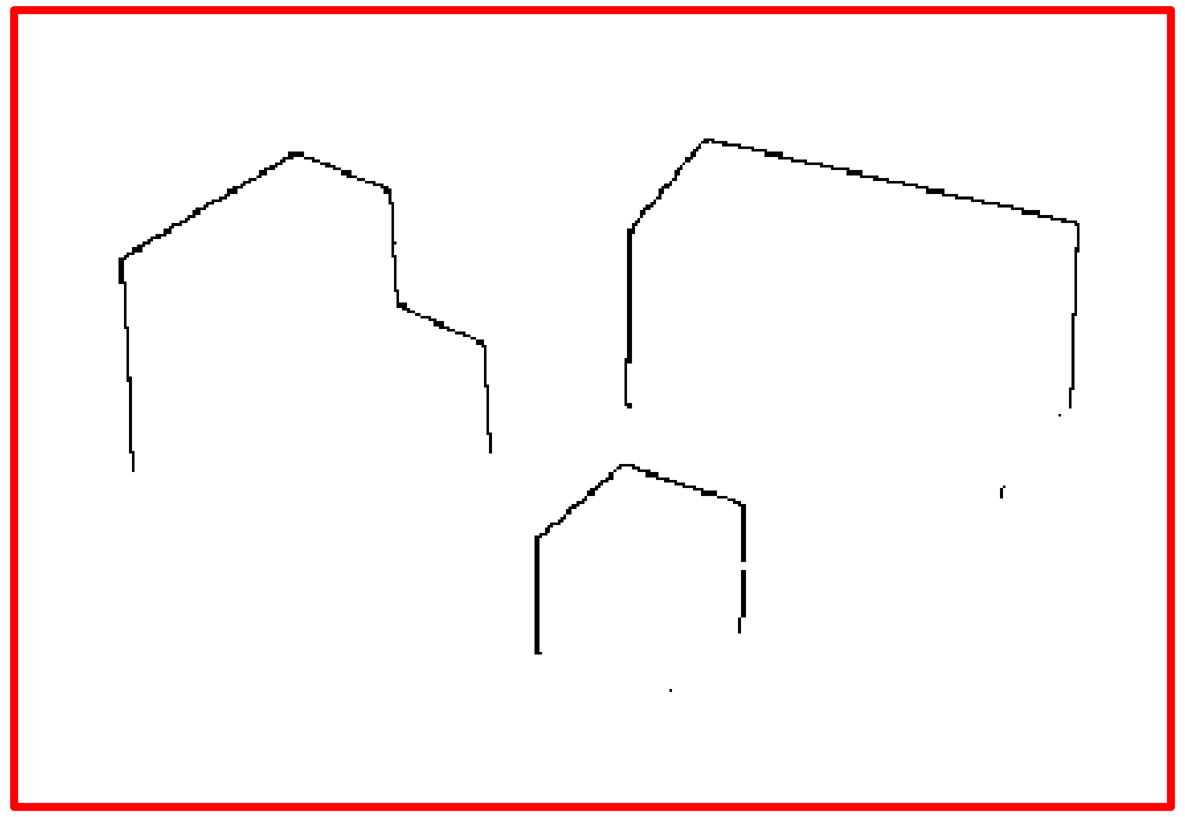
E(x,y) and *n*(x,y)

Edge classification

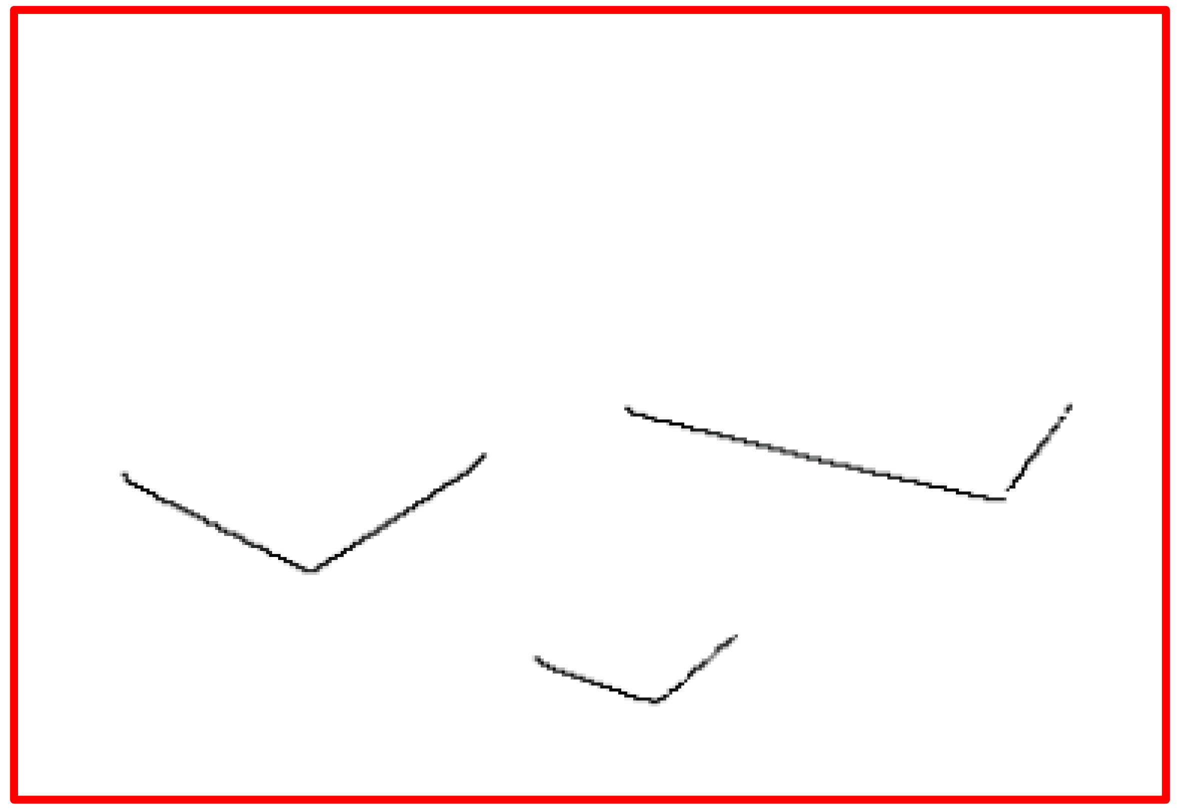
- **Figure/ground segmentation**
  – Using the fact that objects have color



- **Occlusion edges**
  – Occlusion edges are owned by the foreground

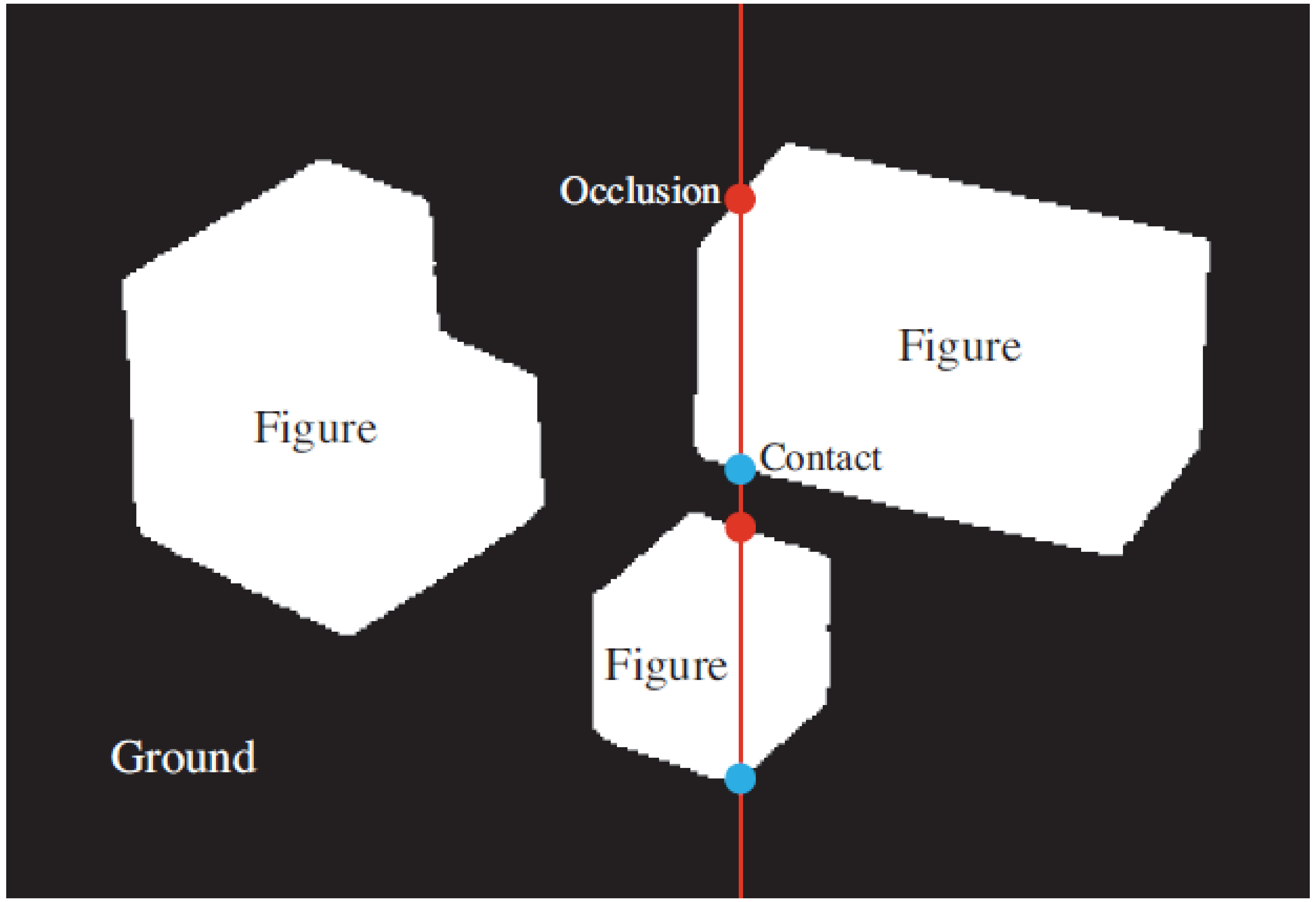


- **Contact edges**

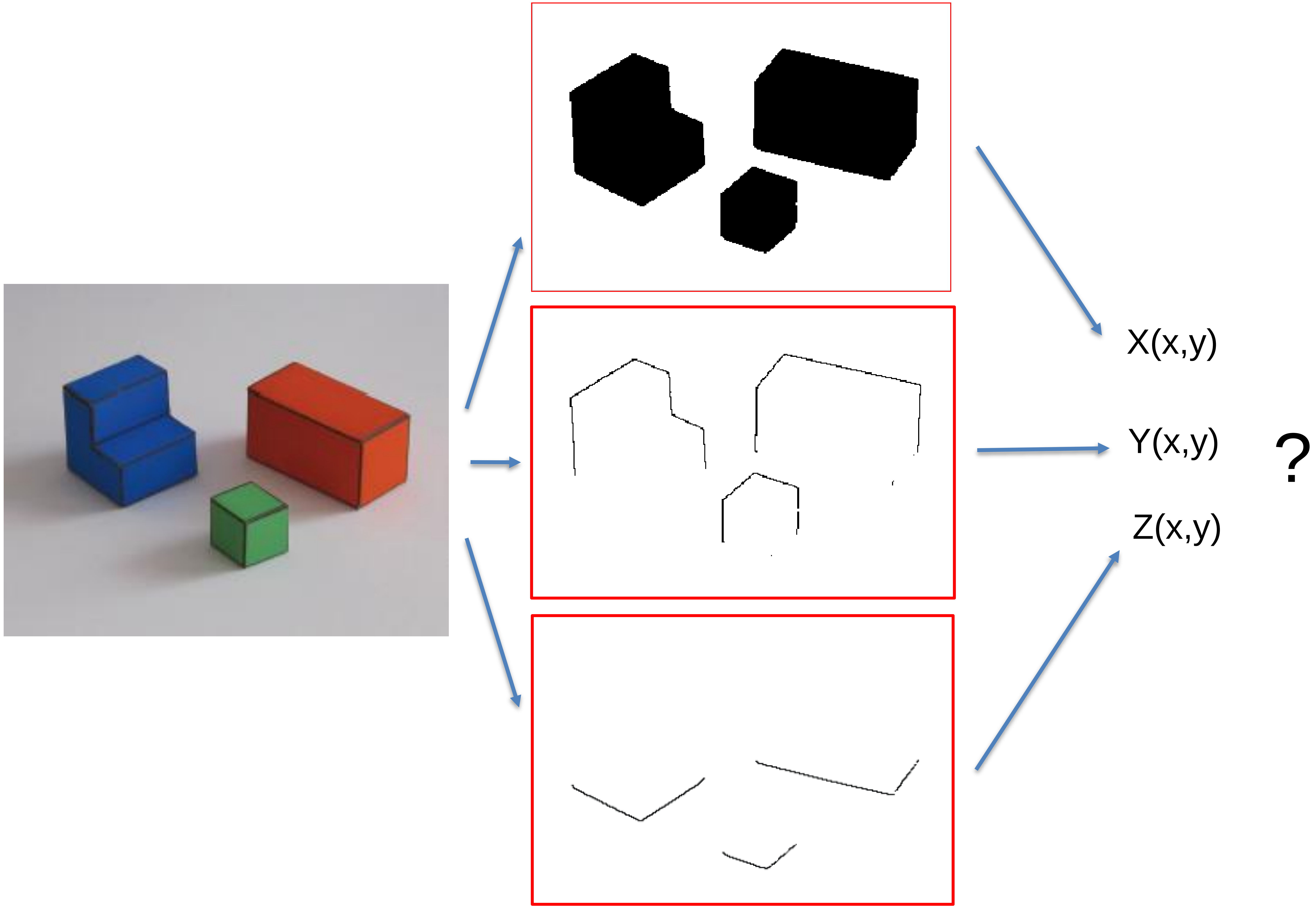


Hack to find contact edges

Figure 2.7: For each vertical line (shown in red), scanning from top to bottom, transitions from ground to figure are occlusion boundaries, and transitions from figure to ground are contact edges. This heuristic will fails when an object occludes another.

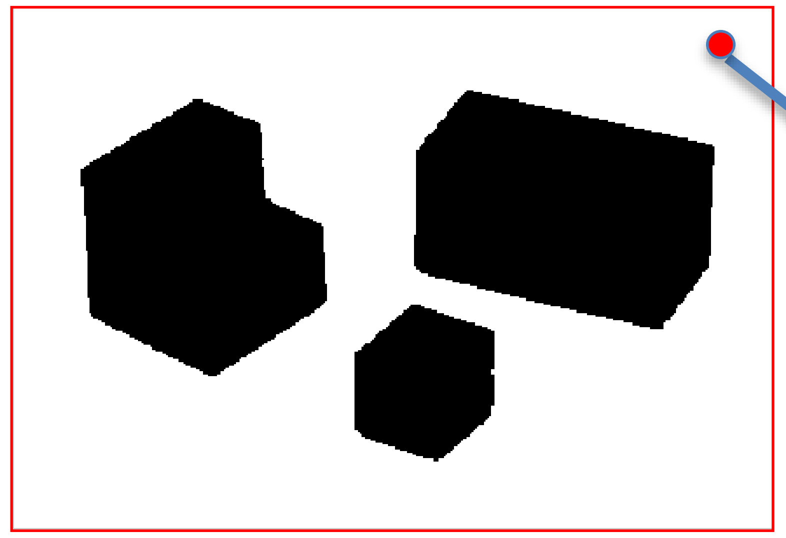


From edges to surface constraints



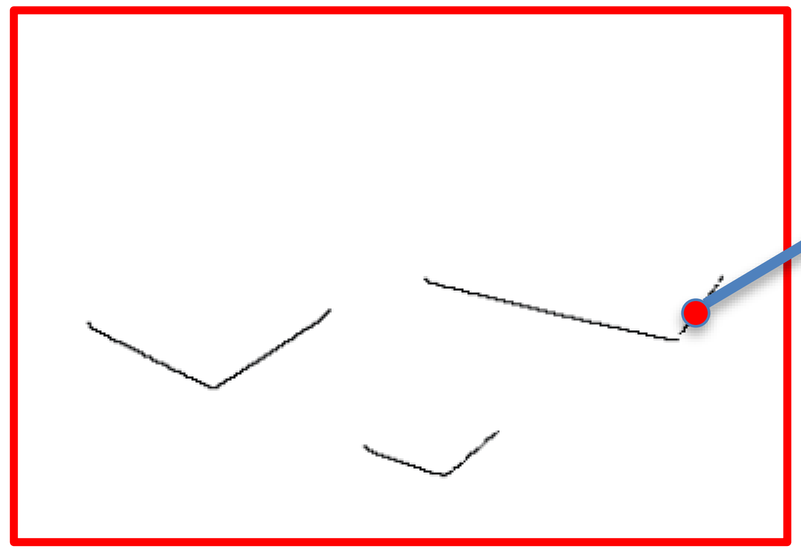
From edges to surface constraints

- Ground



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

- Contact edge

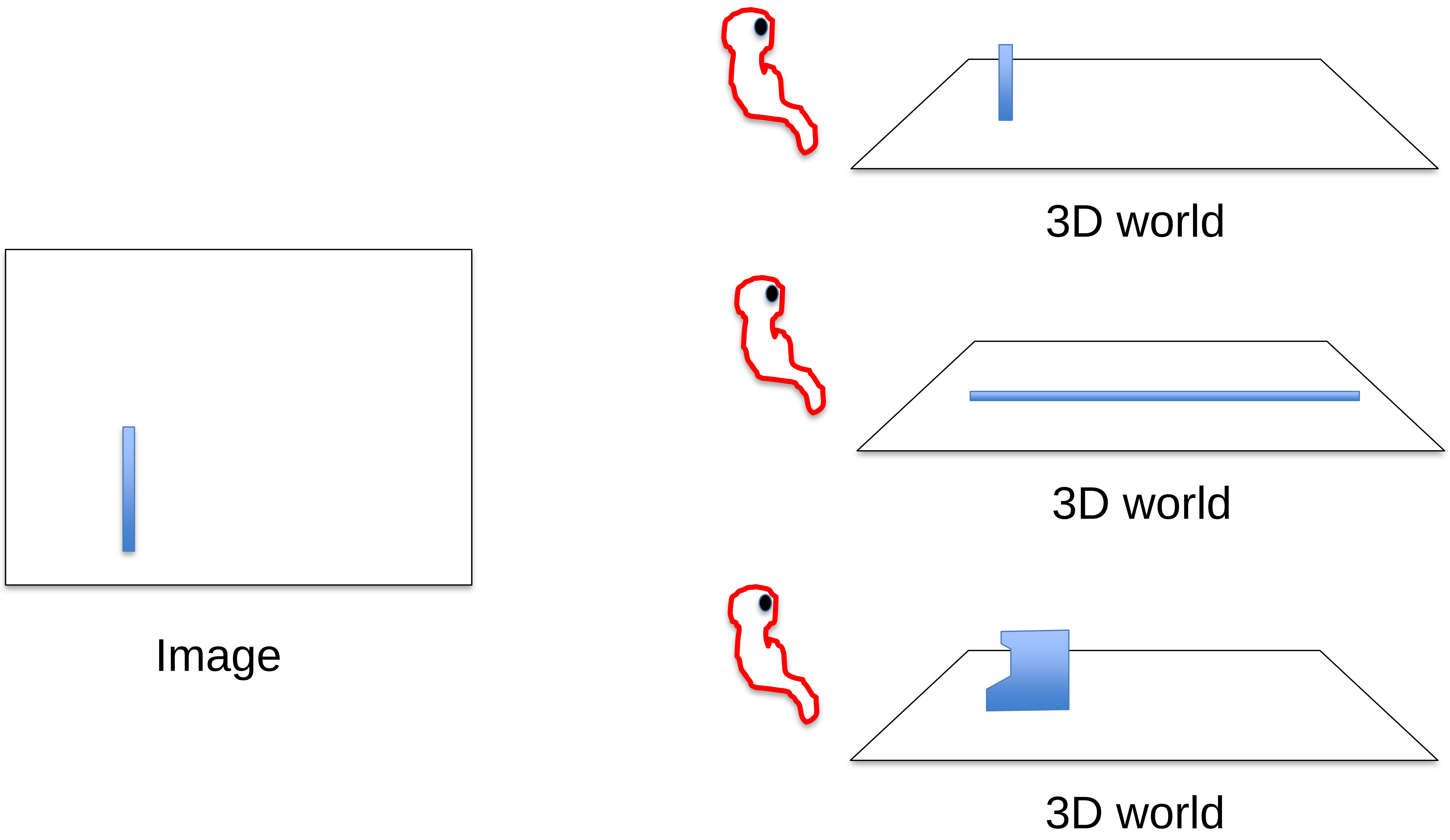


Y(x,y) = 0   if (x,y) belongs to foreground and is a contact edge

- What happens inside the objects?

… now things get a bit more complicated.

Generic view assumption

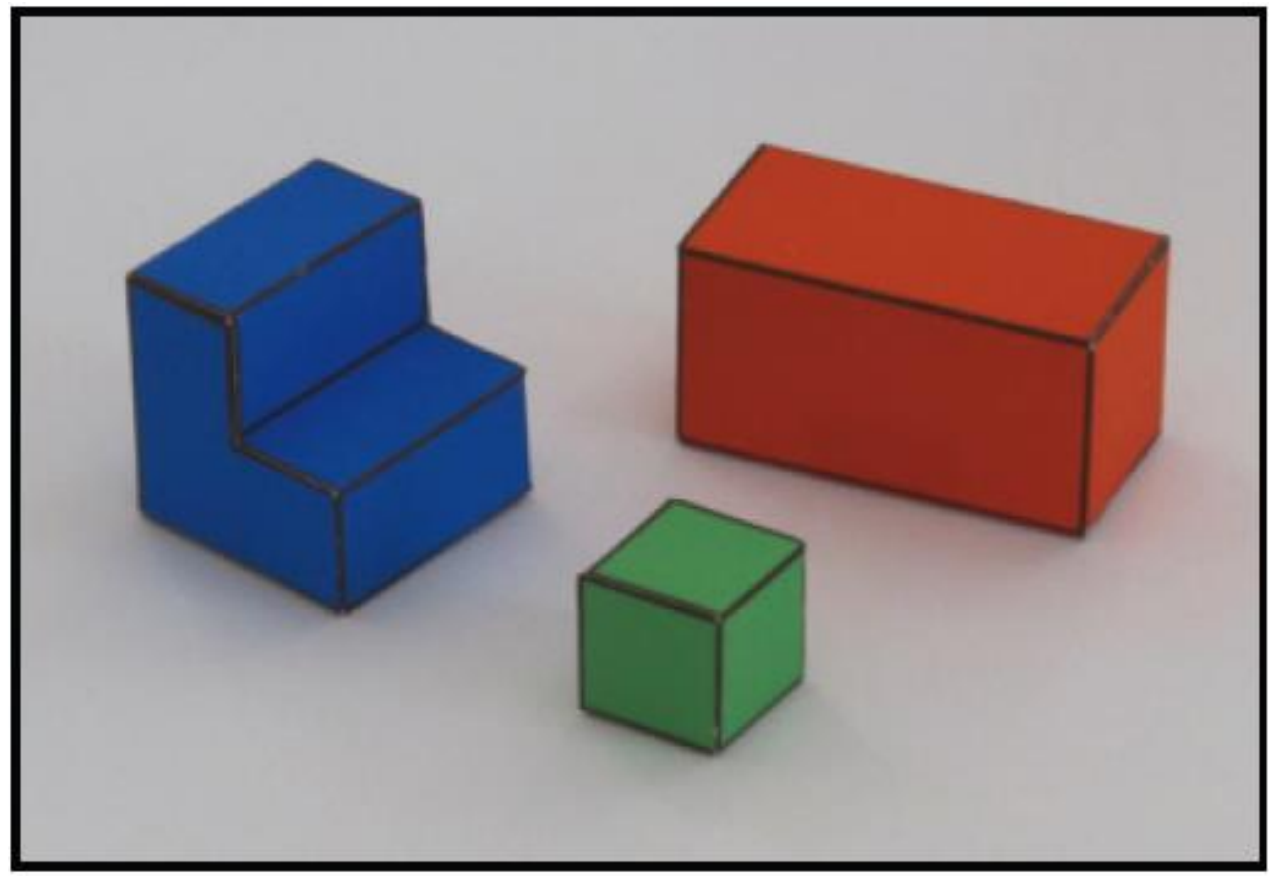


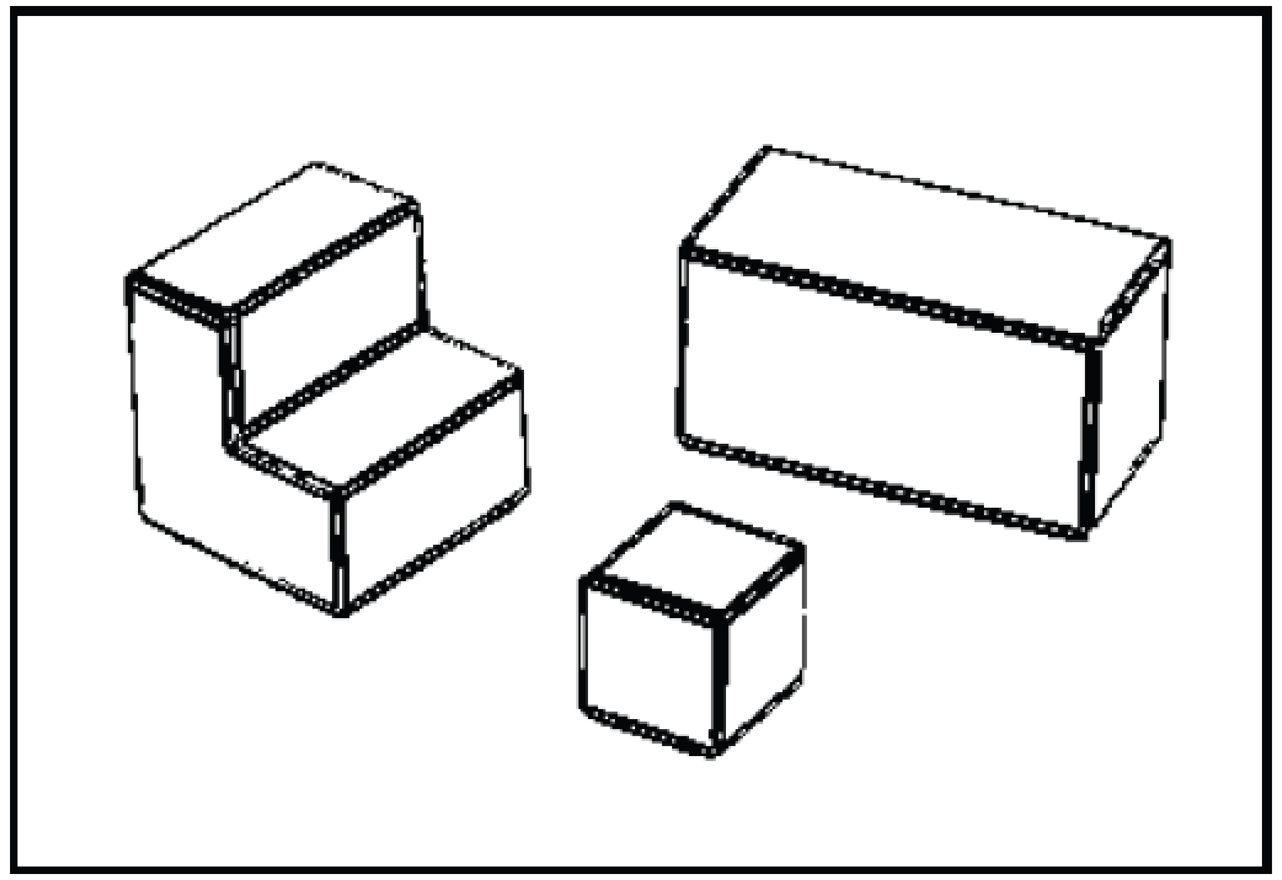
Generic view assumption: the observer should not assume that he has a special position in the world… The most generic interpretation is to see a vertical line as a vertical line in 3D.

Freeman, 93

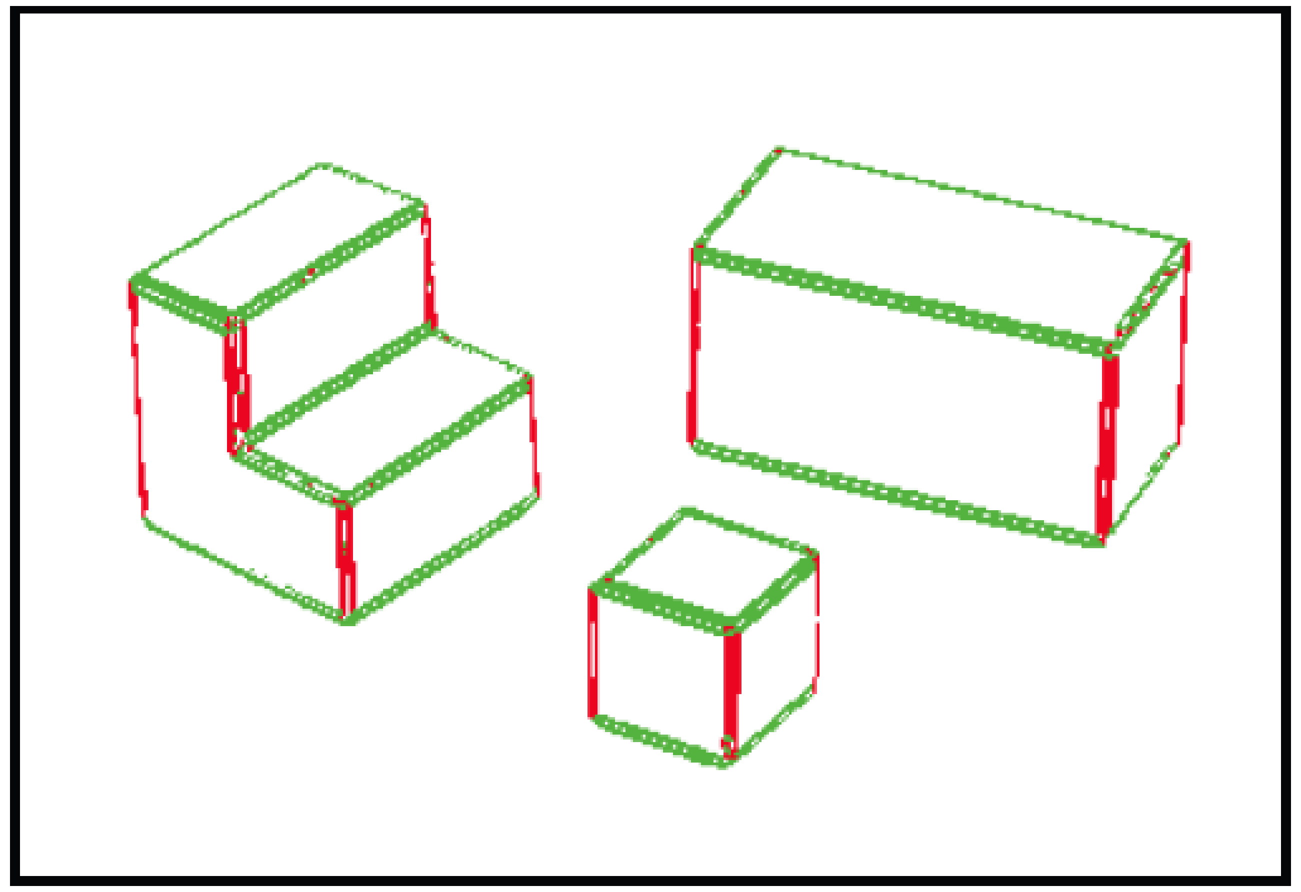
Non-accidental properties in the simple world

generic generic generic accidental generic generic generic





Using E(x,y)

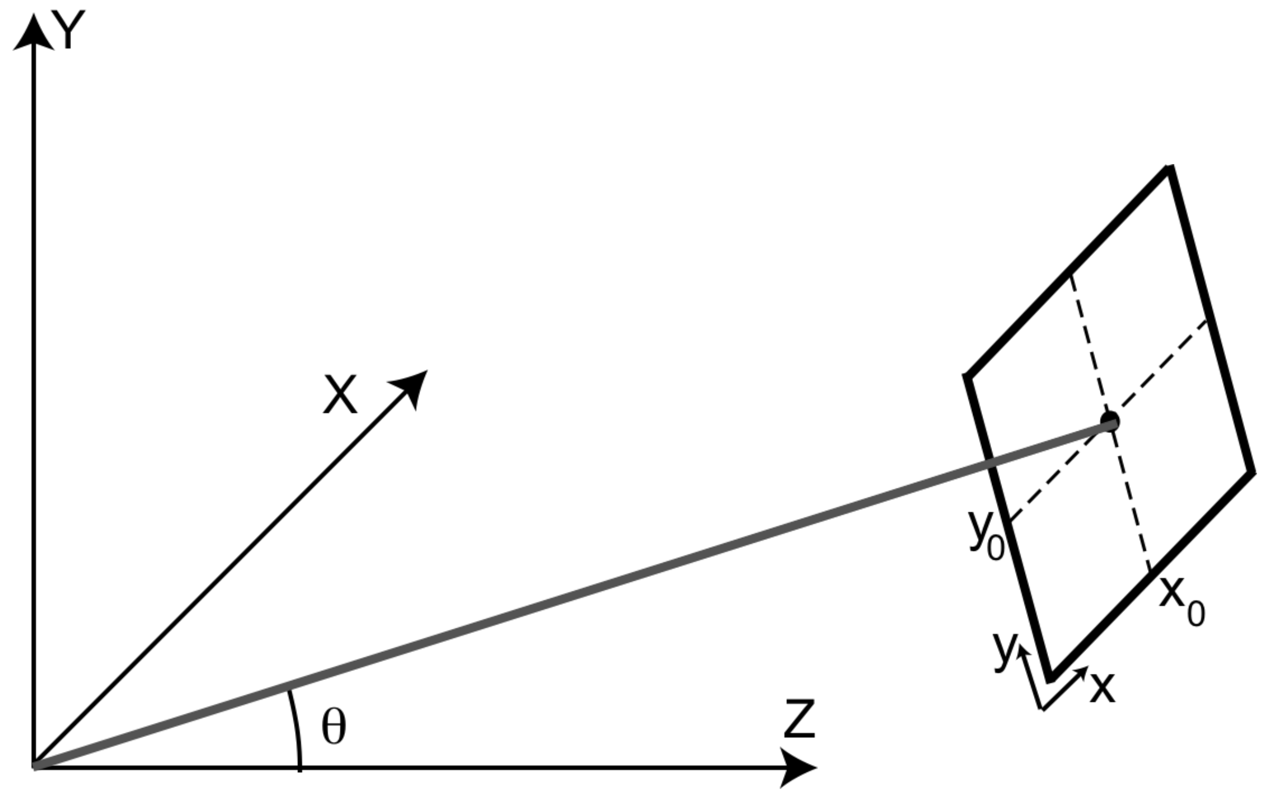


Using $\theta(x,y)$

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



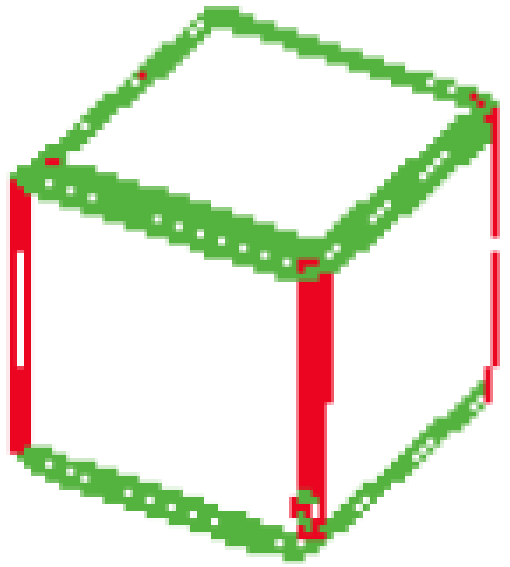
World coordinates

$x = X + x_0$

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

image coordinates

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

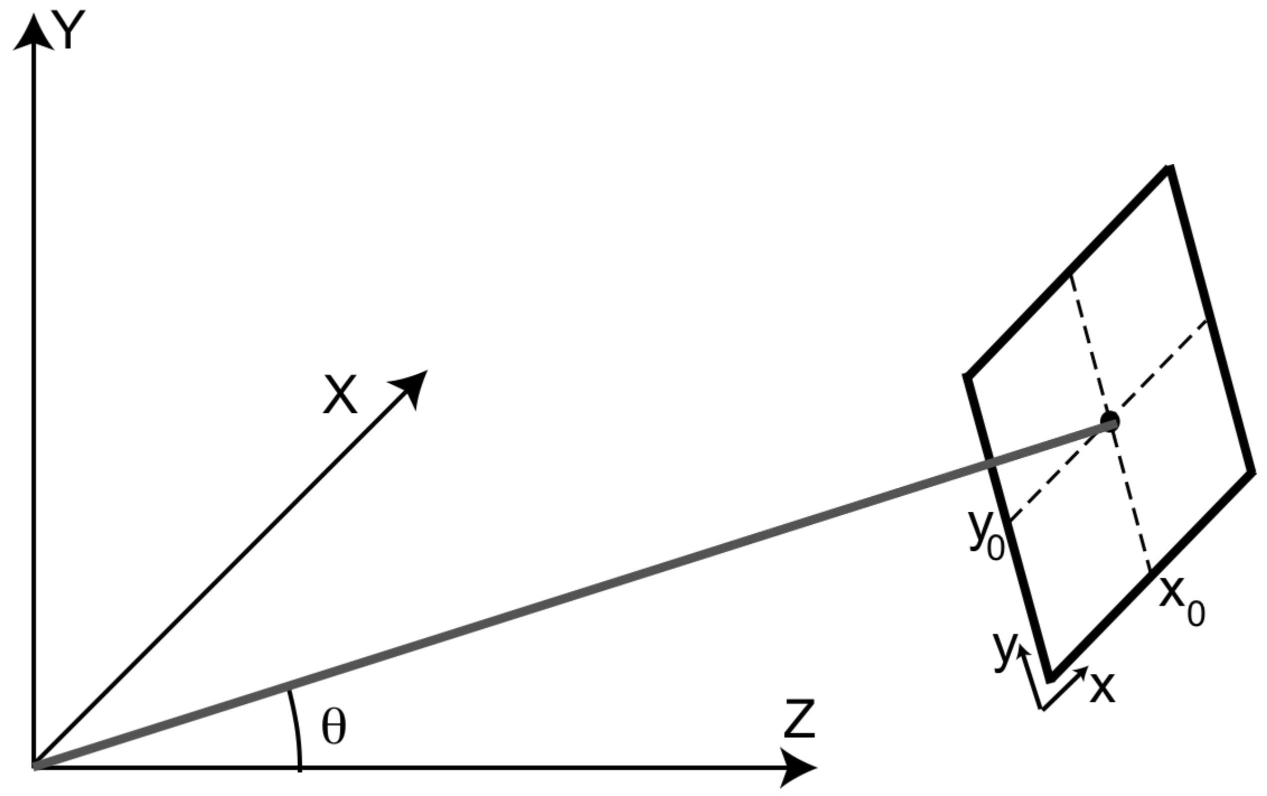


Z = constant along the edge

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

From edges to surface constraints

- Horizontal edges are 3D horizontal lines



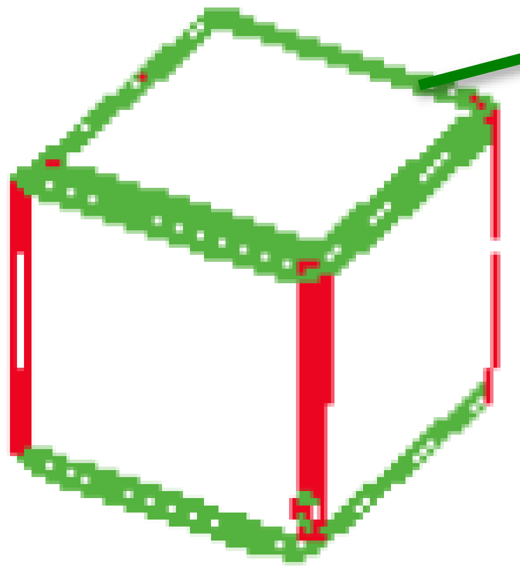
World coordinates

x = X + x_0

y = cos(θ) Y – sin(θ) 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?



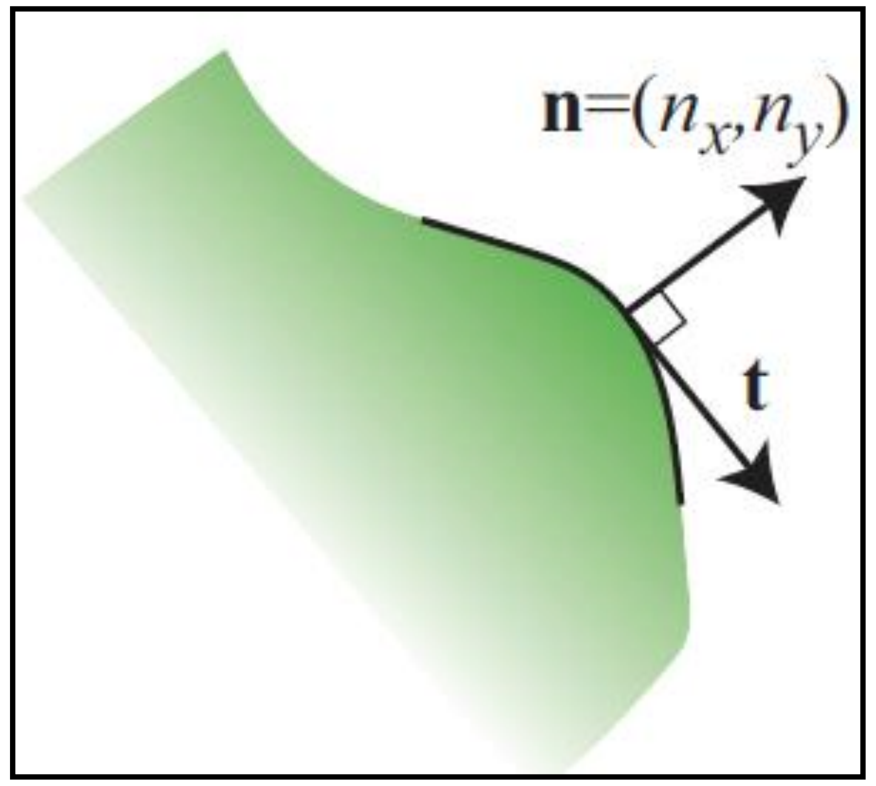
Y = constant along the edge

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

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

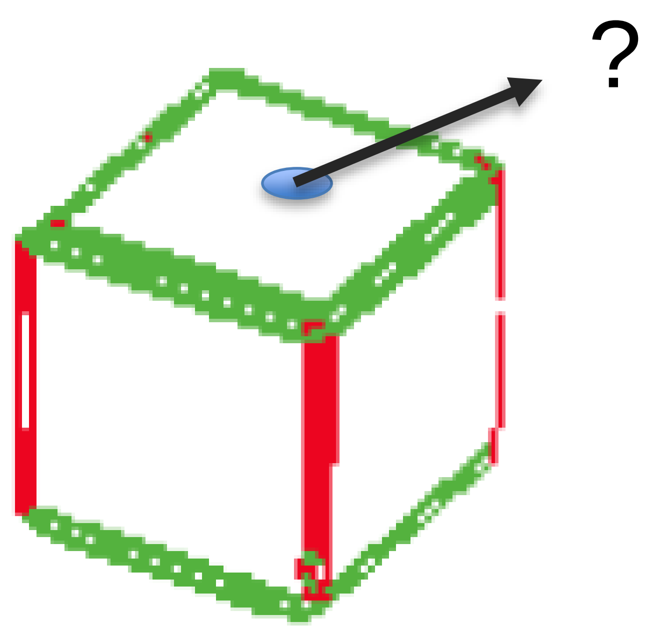
$$\mathbf{t} = (-n_y, n_x)$$

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



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

$$\boxed{\partial Y/\partial y \quad = \quad 1/\cos(\theta)}$$

if (x,y) belongs to a vertical edge

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

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

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

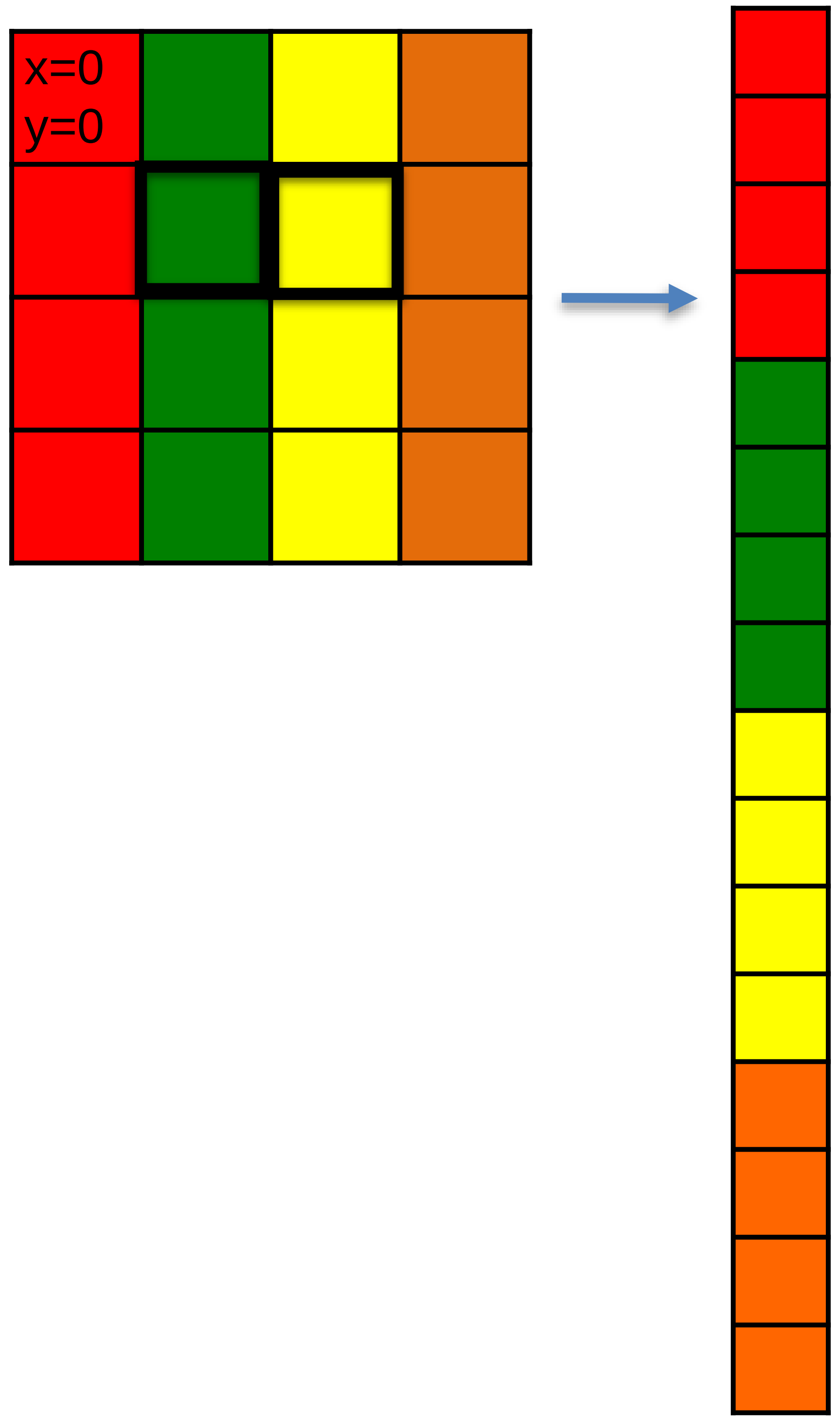
(it is symmetric, and it averages horizontal derivatives over 3 vertical locations)

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

Discrete approximation

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

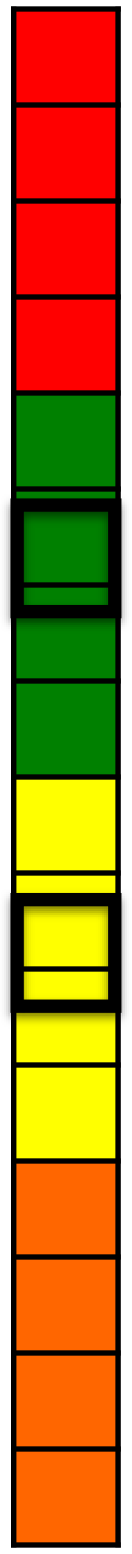
Y(x,y)



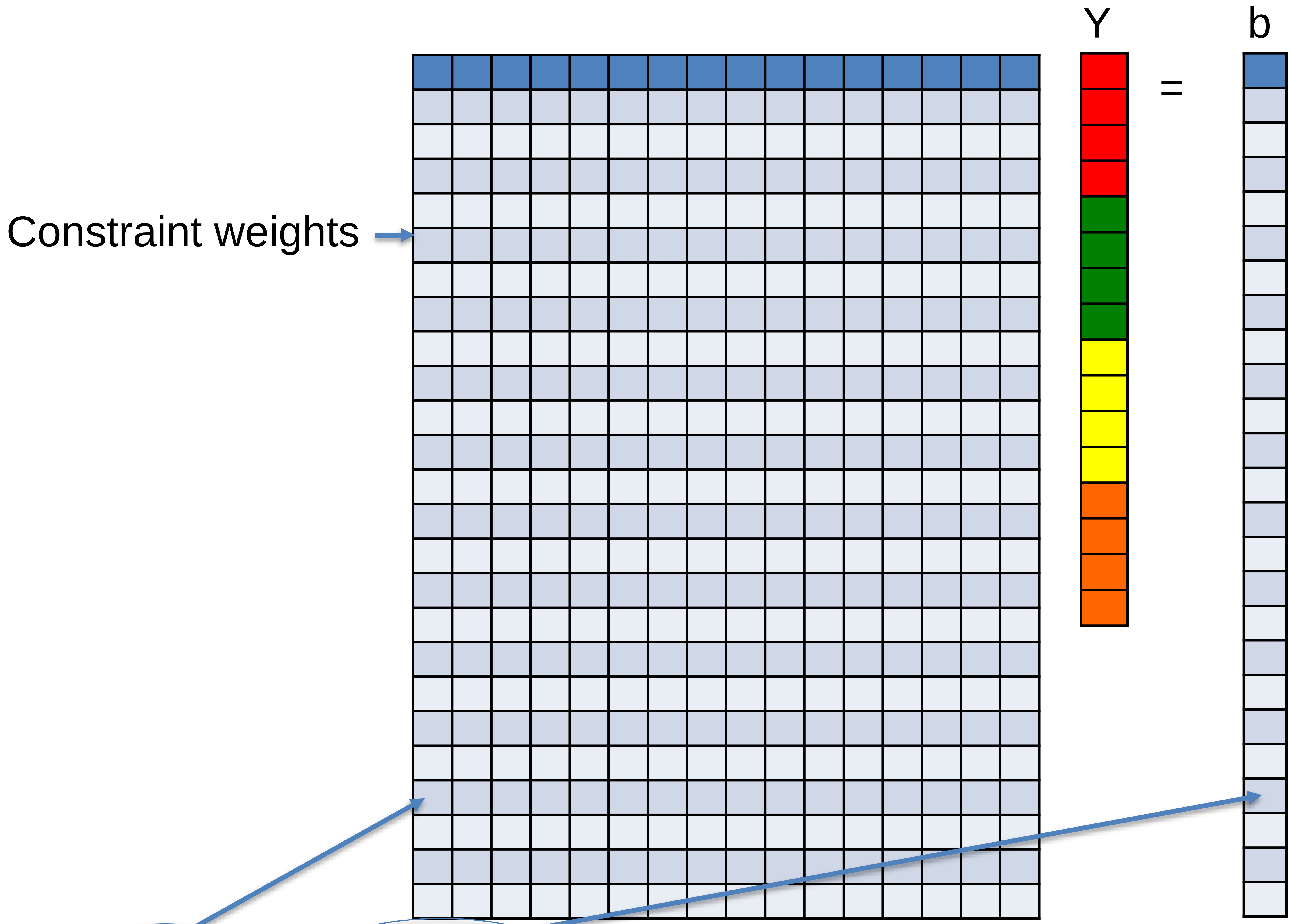
x=2, y=1

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

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



A simple inference scheme



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

$$A\,Y = b$$

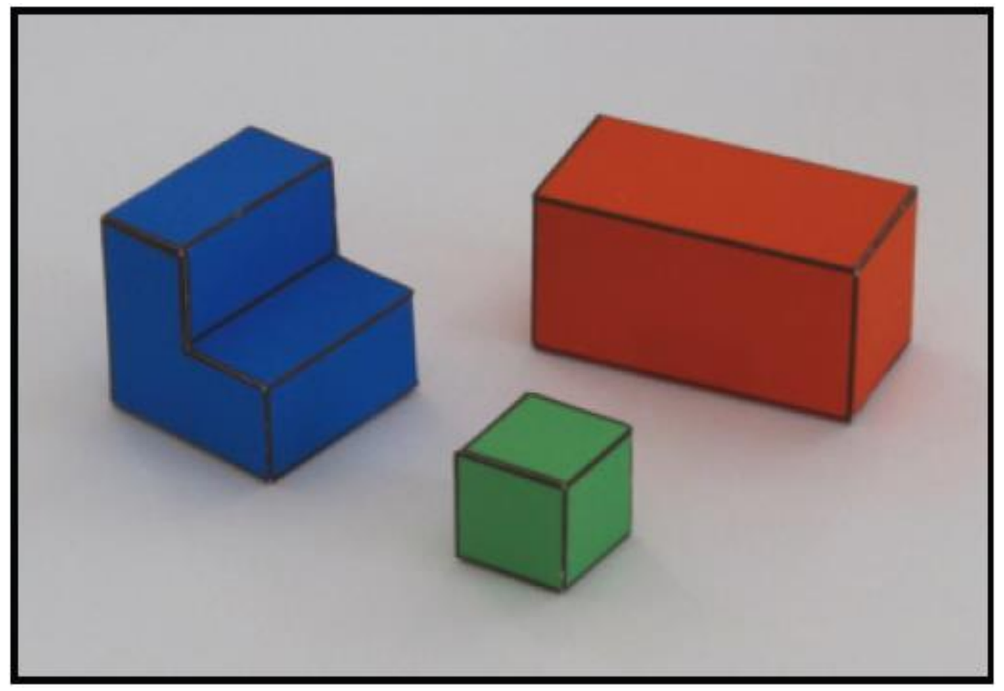
$$Y = (A^T A)^{-1} A^T b$$

Results

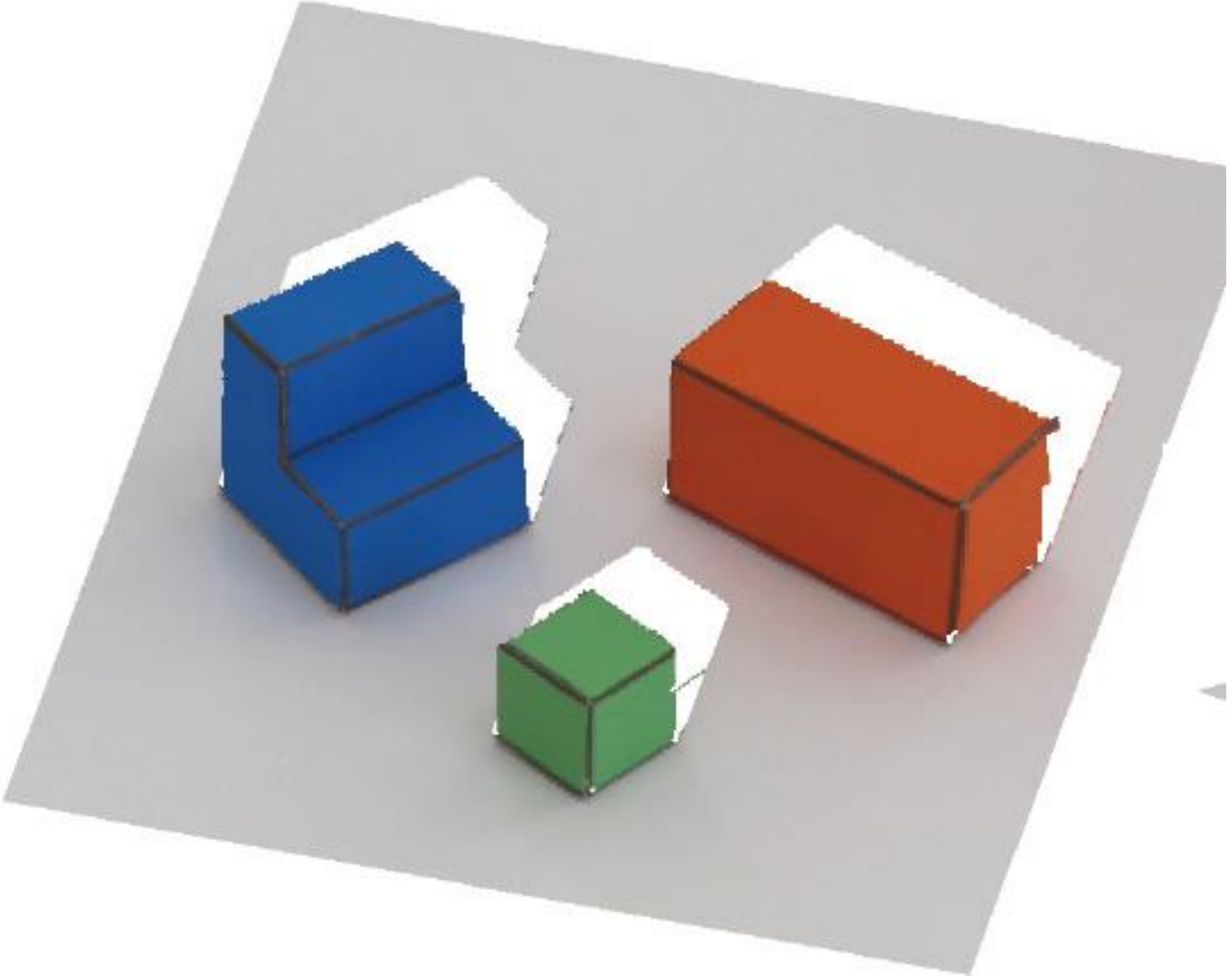
Input
Representation 1
Edge normals
Figure/ground
Representation 2
Edge strength
3D orientation
Contact edges
Depth discontinuities
Linear system
Output
X
Y
Z

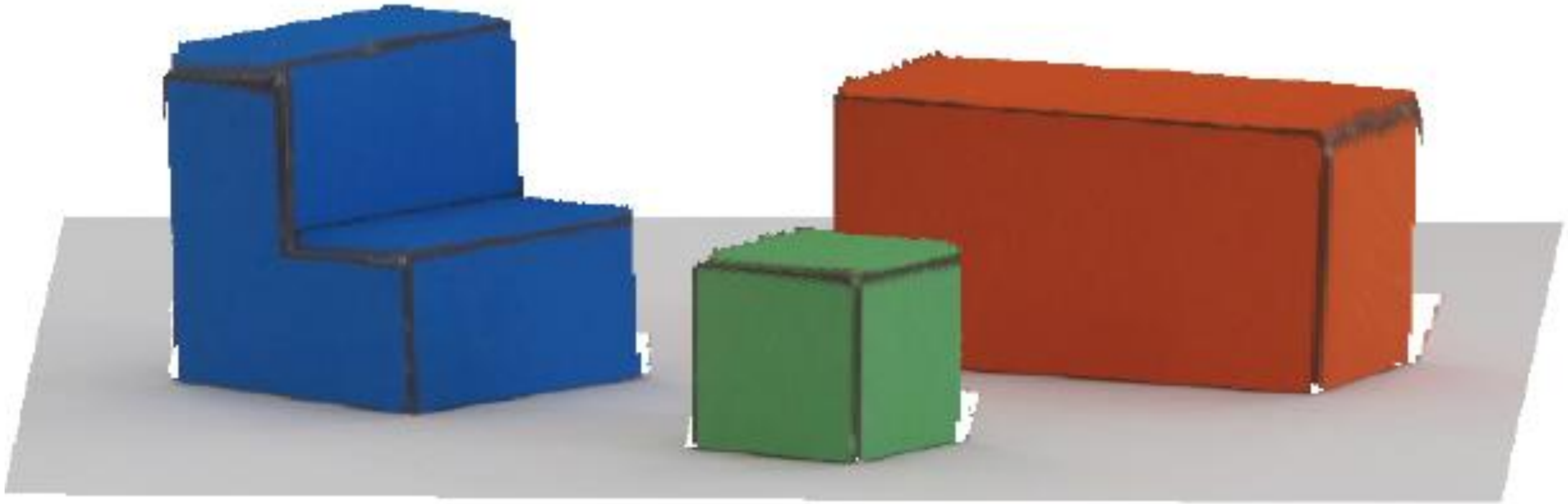
Changing view point

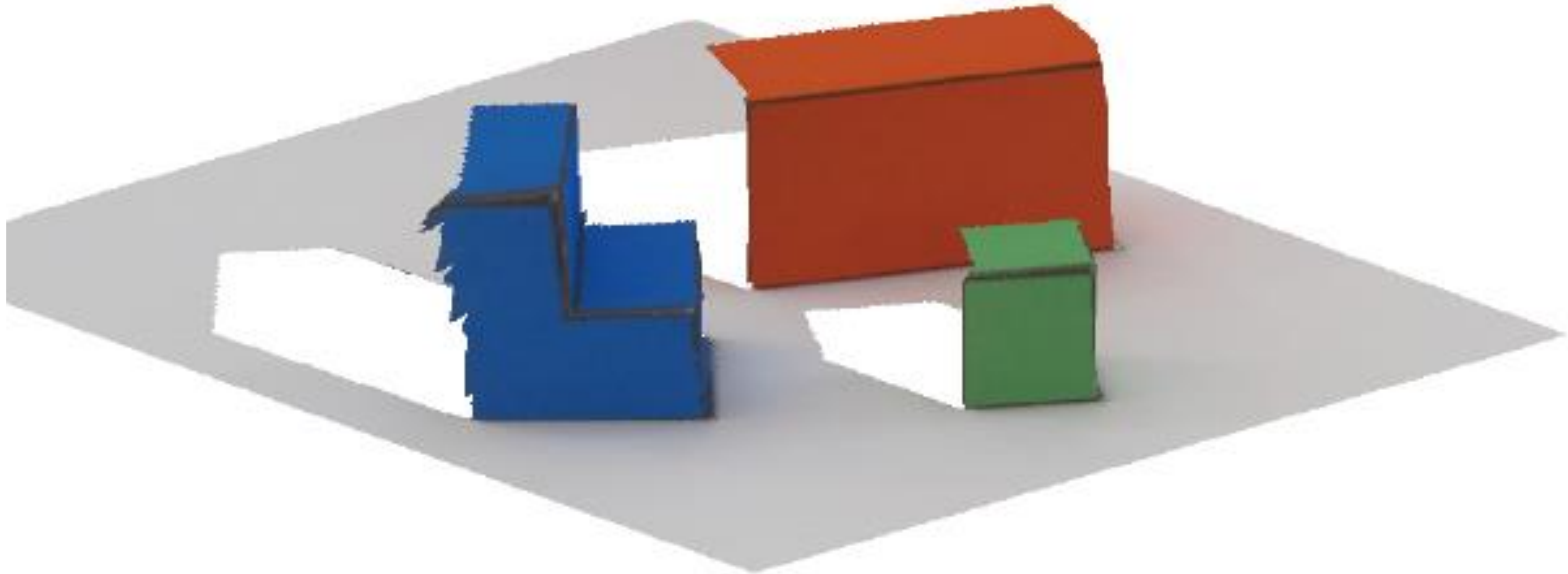
Input



New view points:

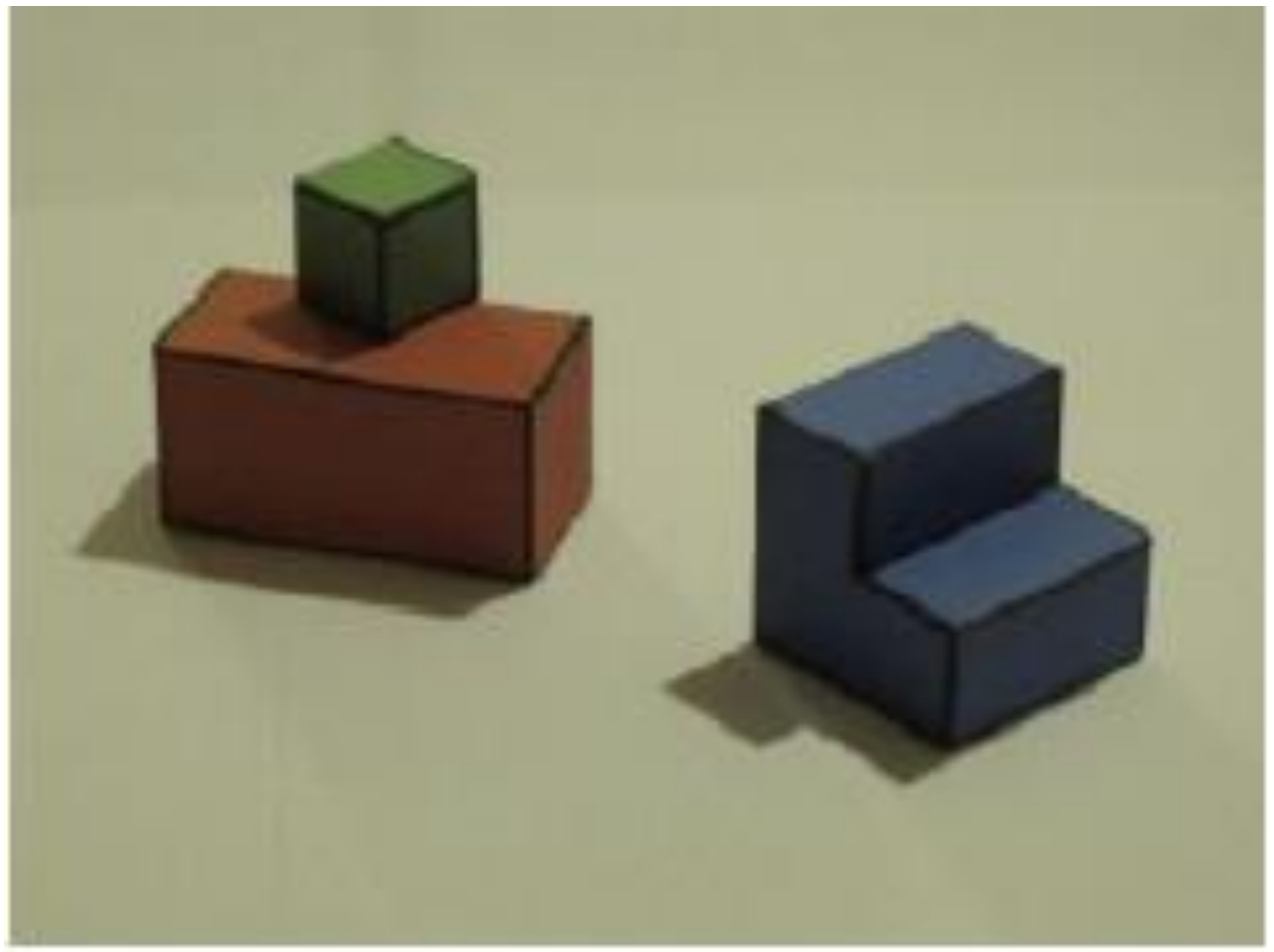




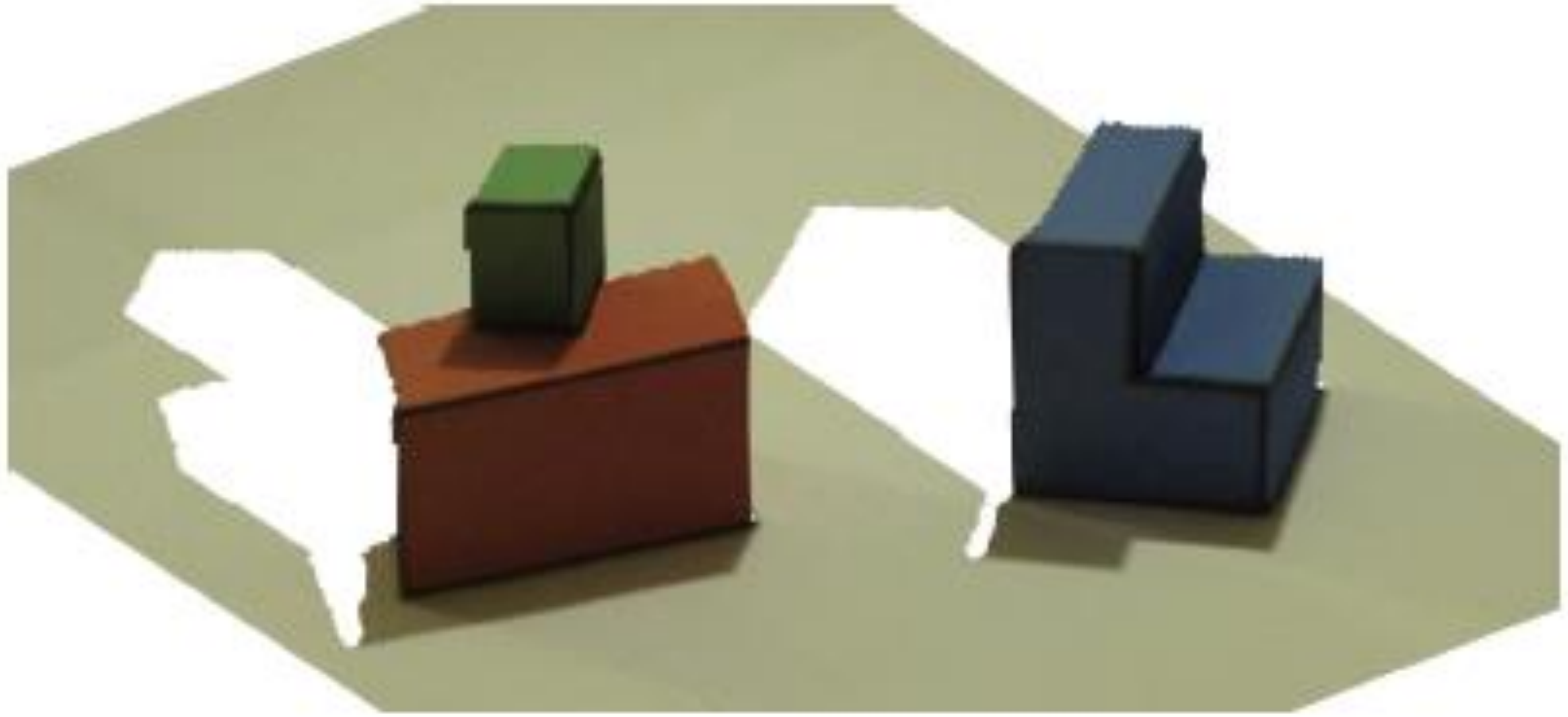


Generalization

Input

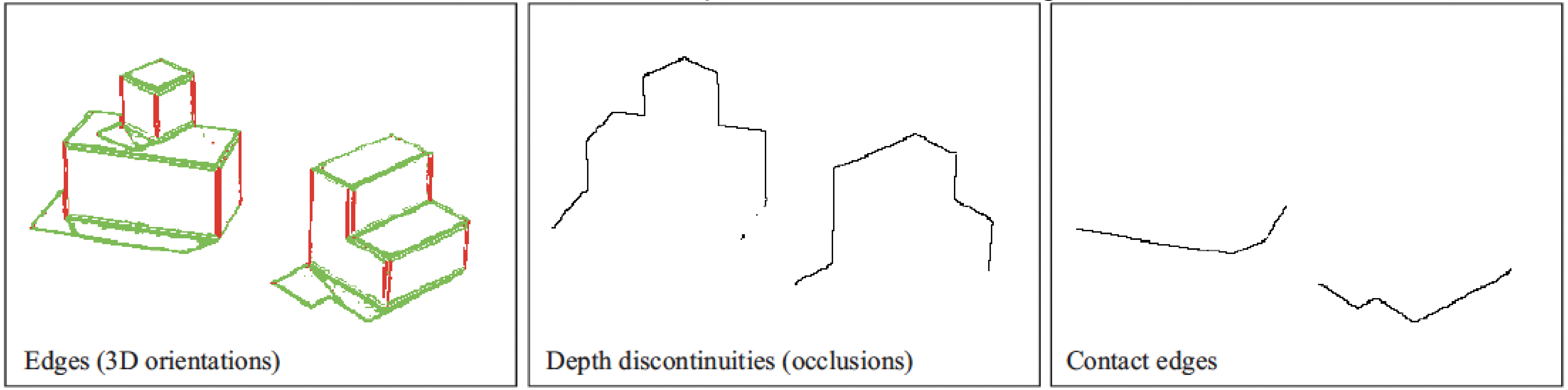


New view point:



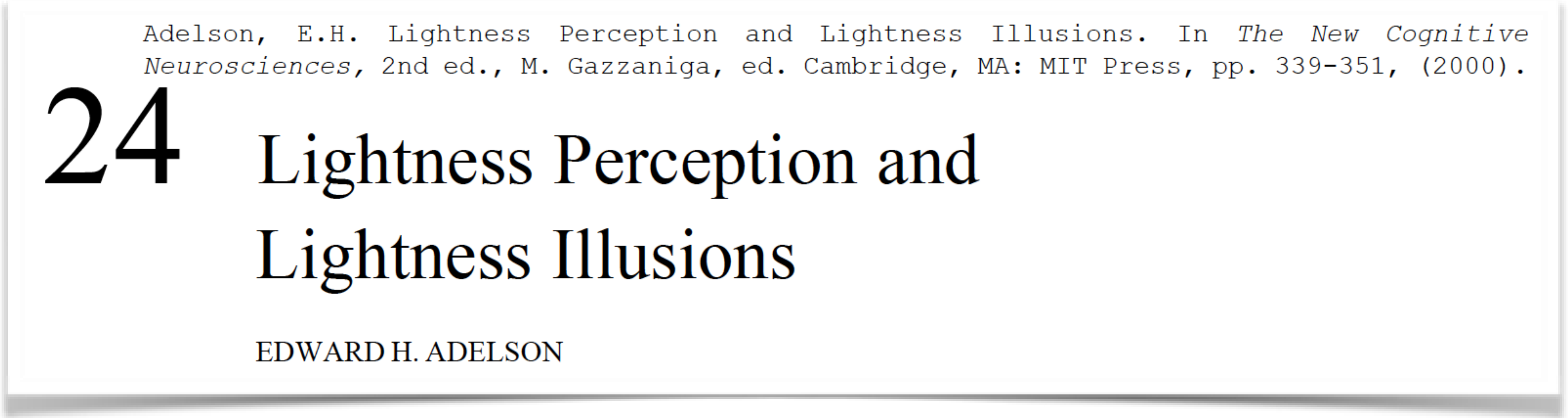
It seems to work!

… but the representation is wrong!



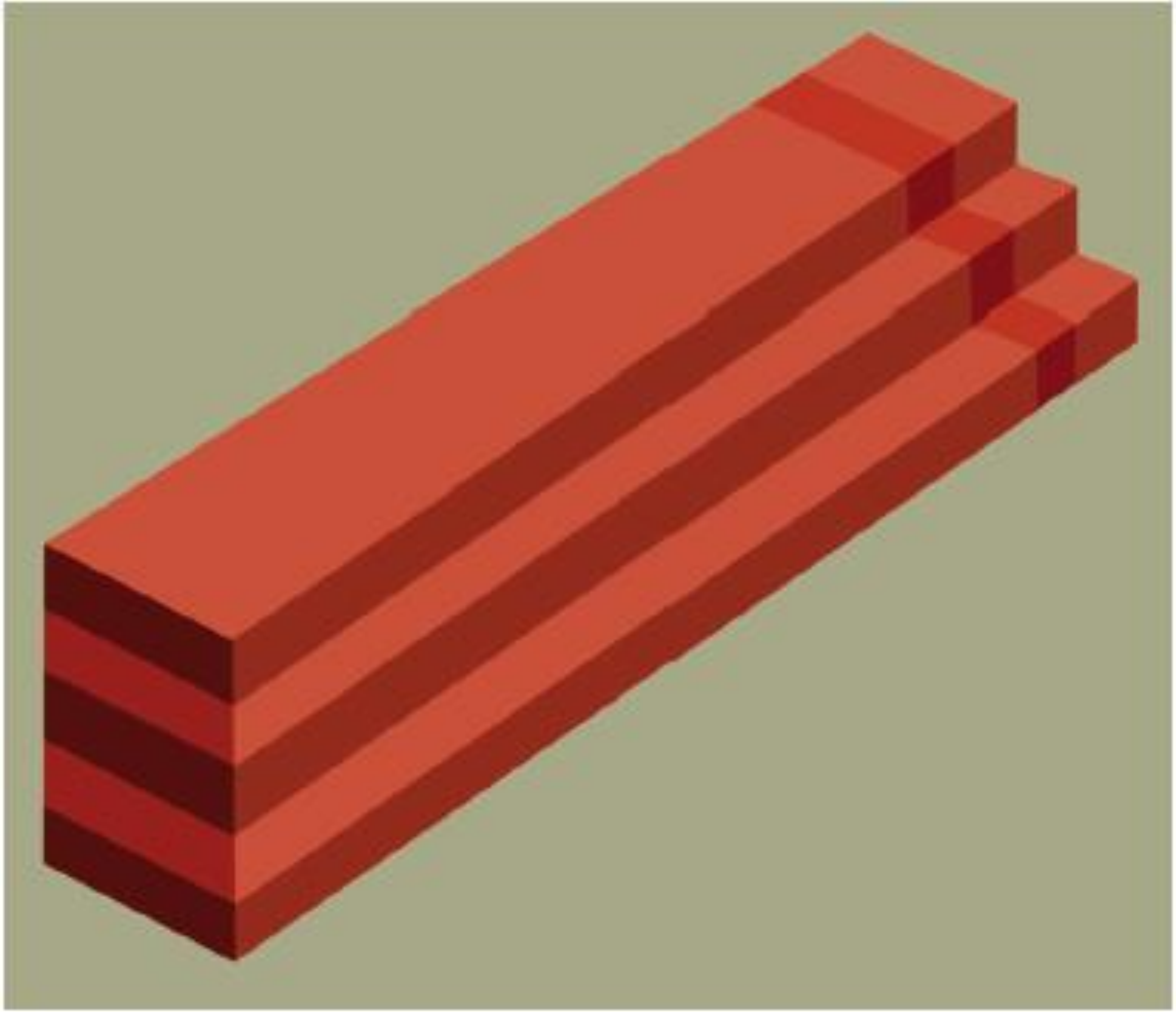
Generalization
2nd test

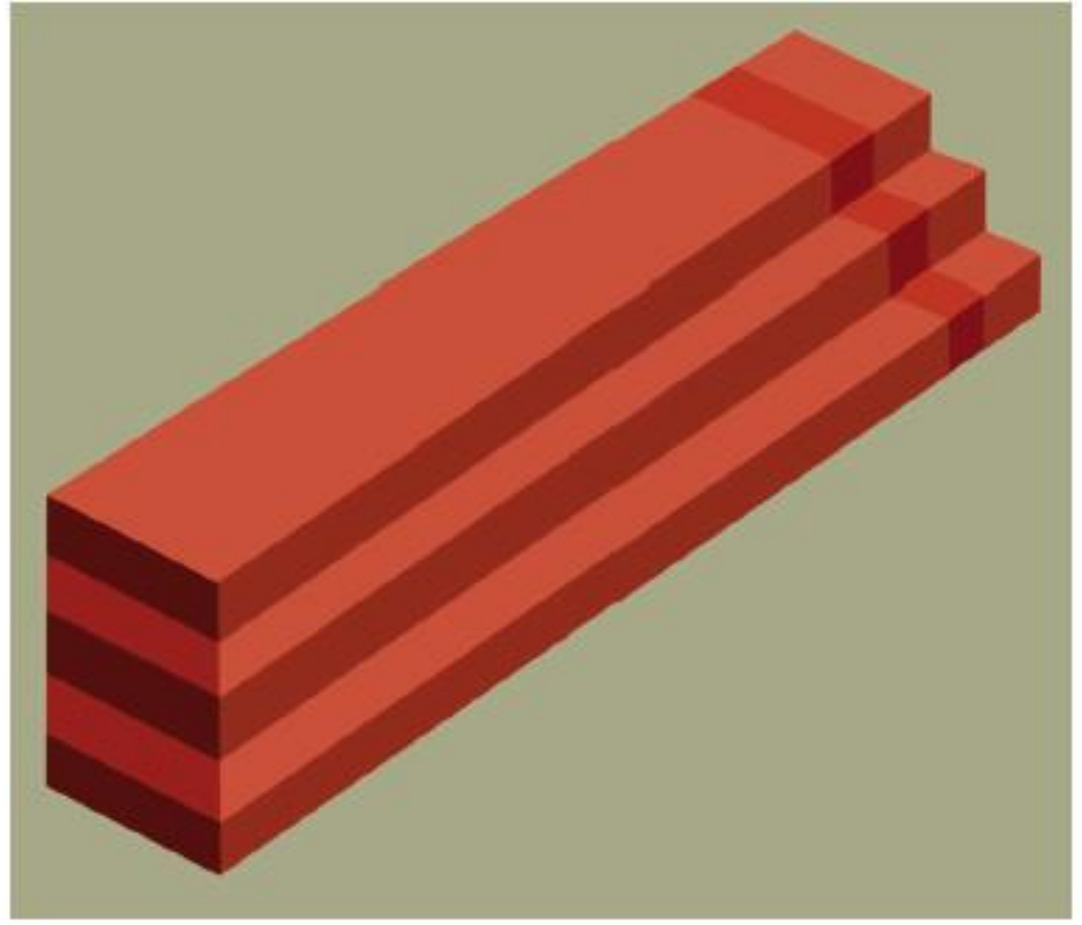
Impossible steps


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