Lecture 1

A Simple Vision System

Exciting times for computer vision

What is vision?

What does it mean, to see?

“to know what is where by looking”.

from Marr, 1982

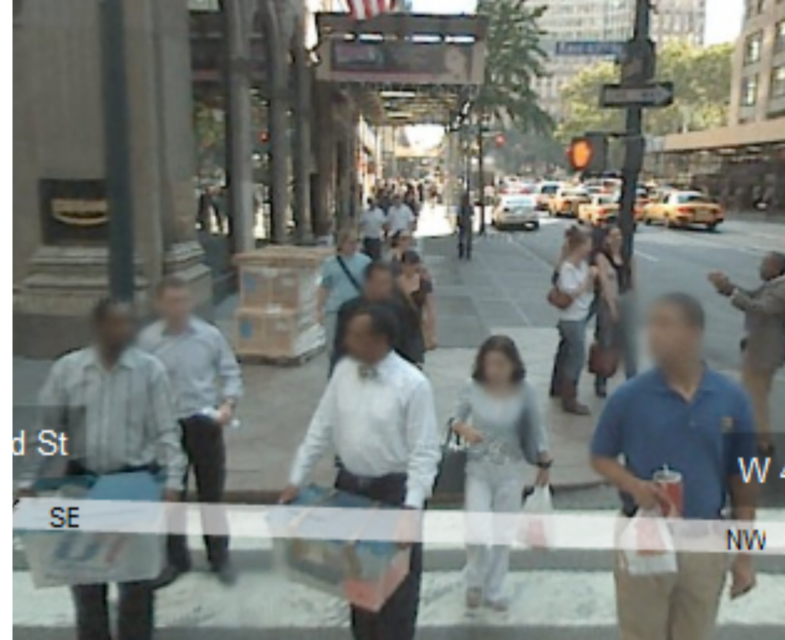
To discover from images what is present in the world, where things are, what actions are taking place.

A bit of history…

50 years ago…

30 years ago…

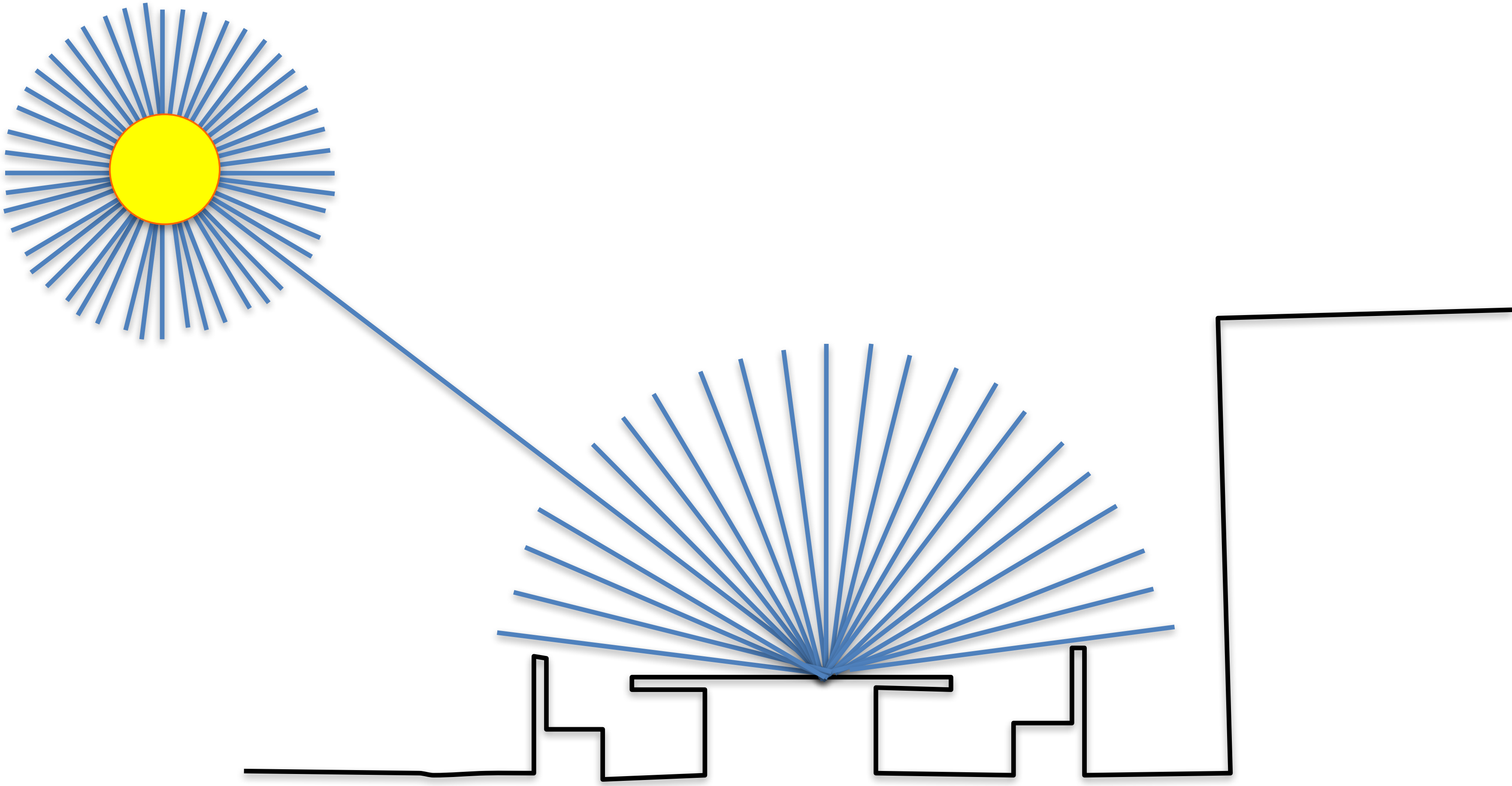
But 15 years ago…



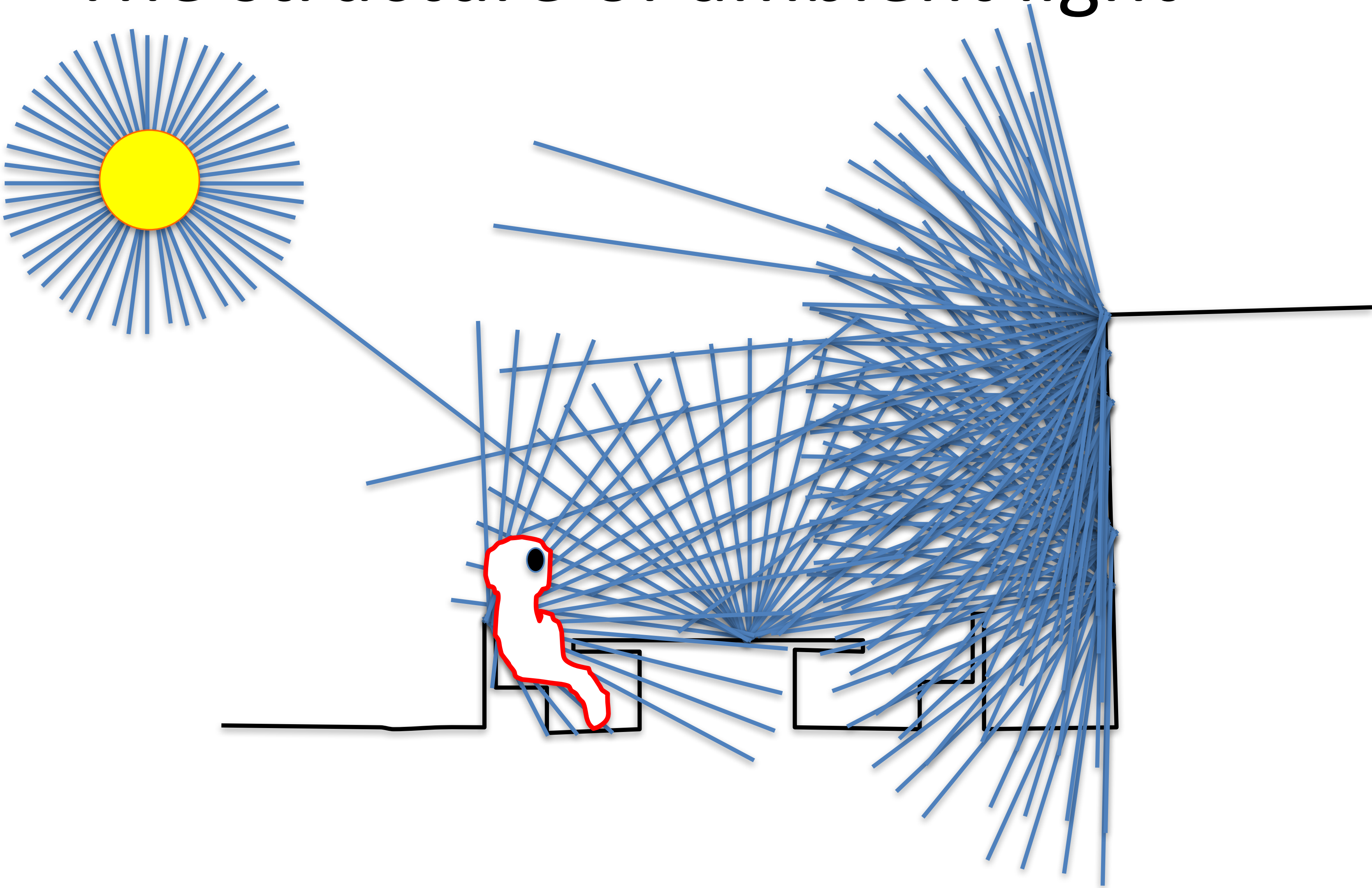
- The representation and matching of pictorial structures Fischler, Elschlager (1973).
- Face recognition using eigenfaces M. Turk and A. Pentland (1991).
- Human Face Detection in Visual Scenes - Rowley, Baluja, Kanade (1995)
- Graded Learning for Object Detection - Fleuret, Geman (1999)
- Robust Real-time Object Detection - Viola, Jones (2001)
- Feature Reduction and Hierarchy of Classifiers for Fast Object Detection in Video Images - Heisele, Serre, Mukherjee, Poggio (2001)
- ….

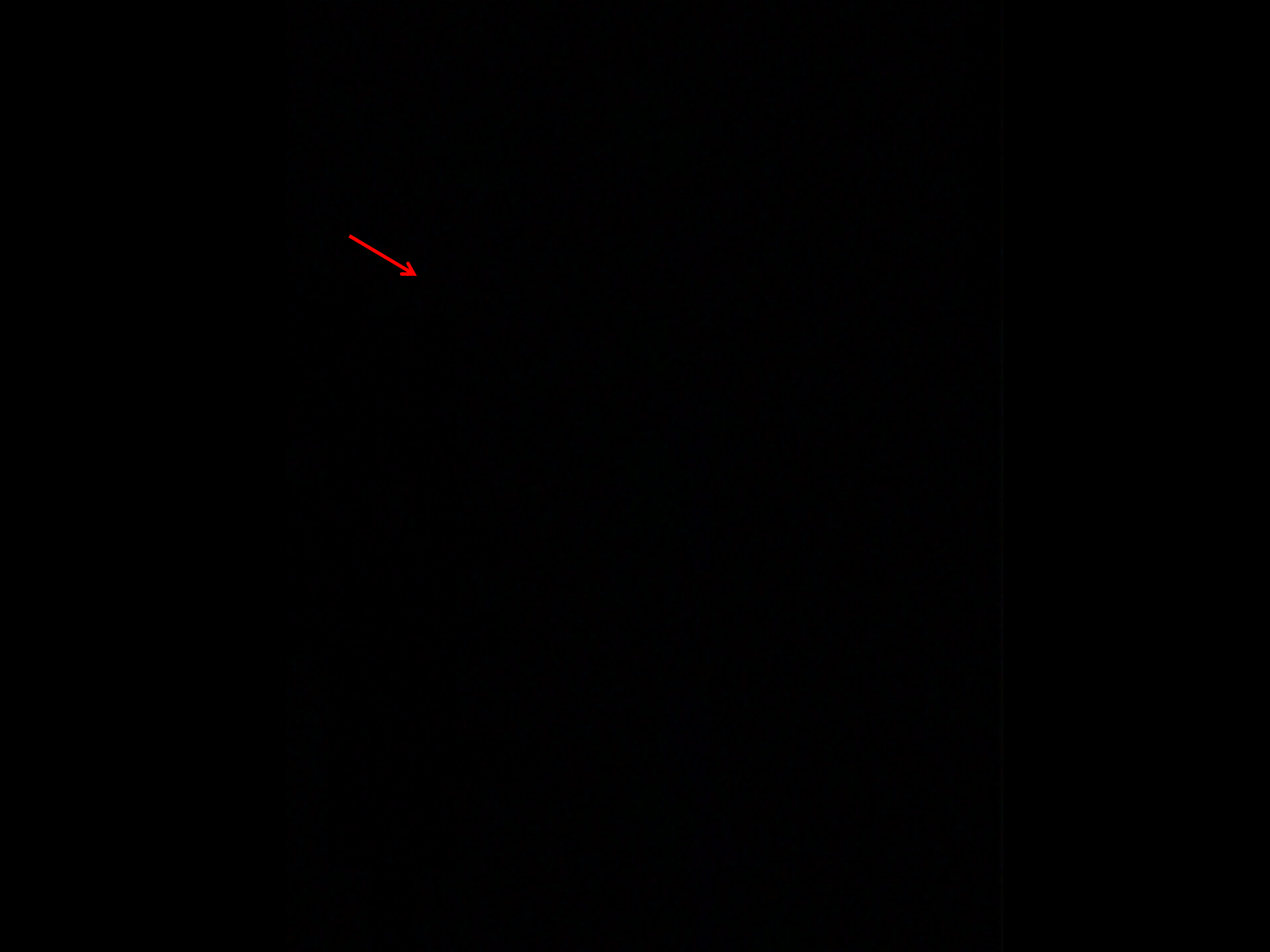
Why is vision hard?

The structure of ambient light



The structure of ambient light



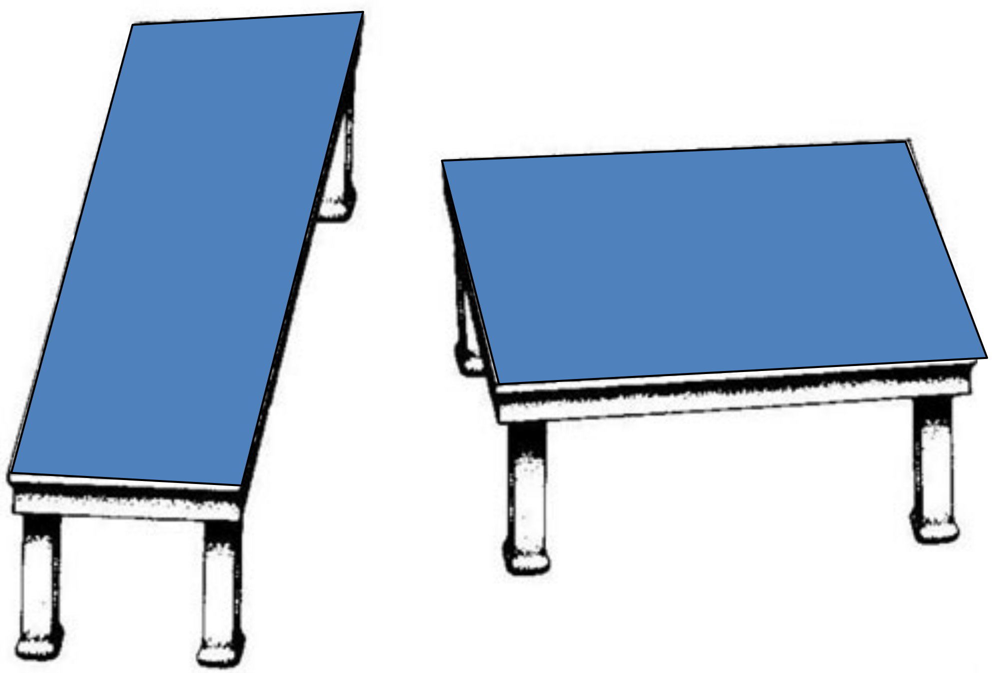


Why is vision hard?

Measuring light vs. measuring scene properties

We perceive two squares, one on top of each other.

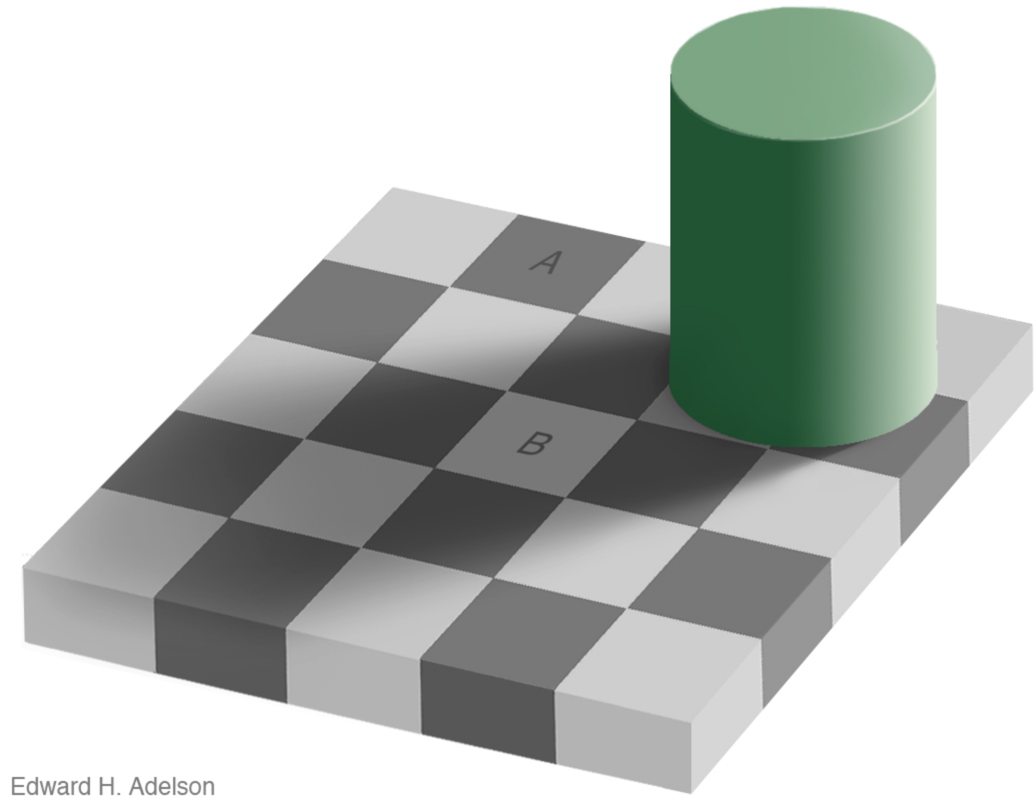
Measuring light vs. measuring scene properties



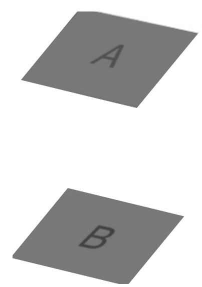
by Roger Shepard (”Turning the Tables”)

Depth processing is automatic, and we can not shut it down…

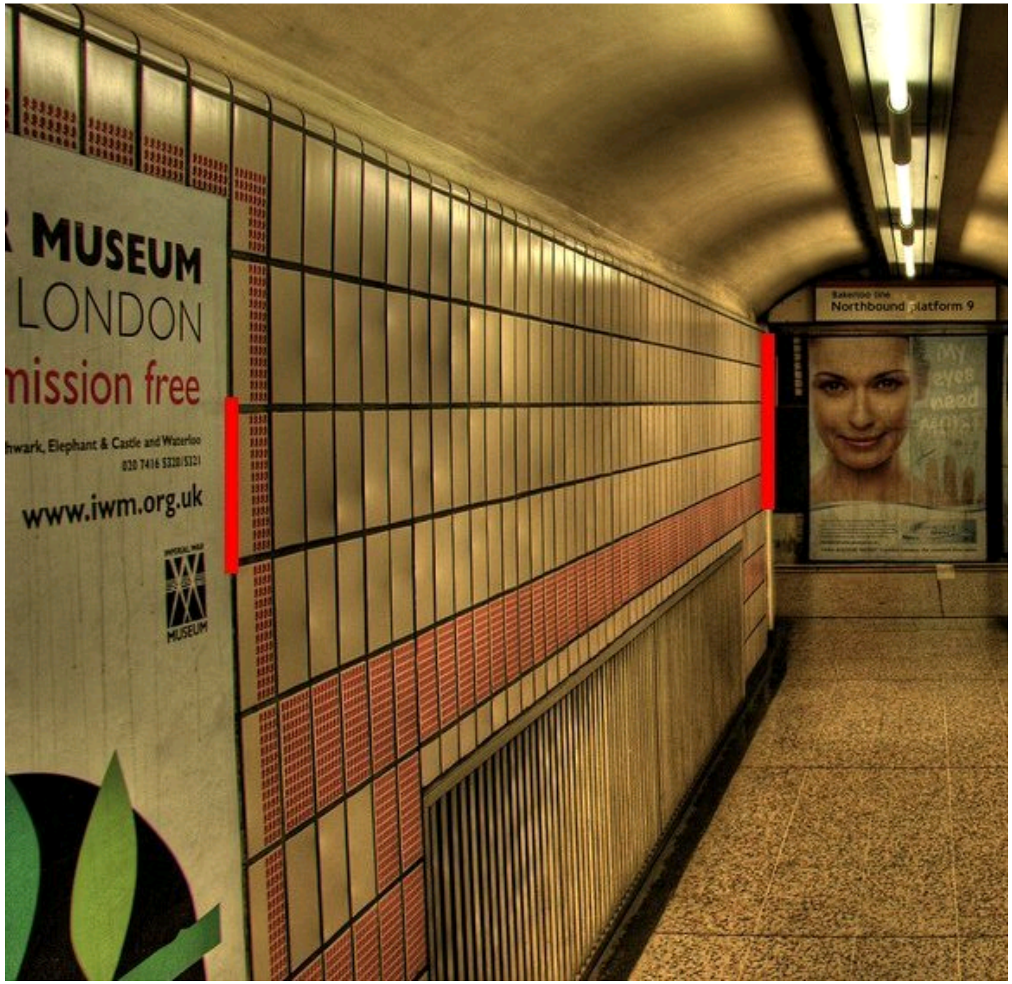
Measuring light vs. measuring scene properties



Measuring light vs. measuring scene properties

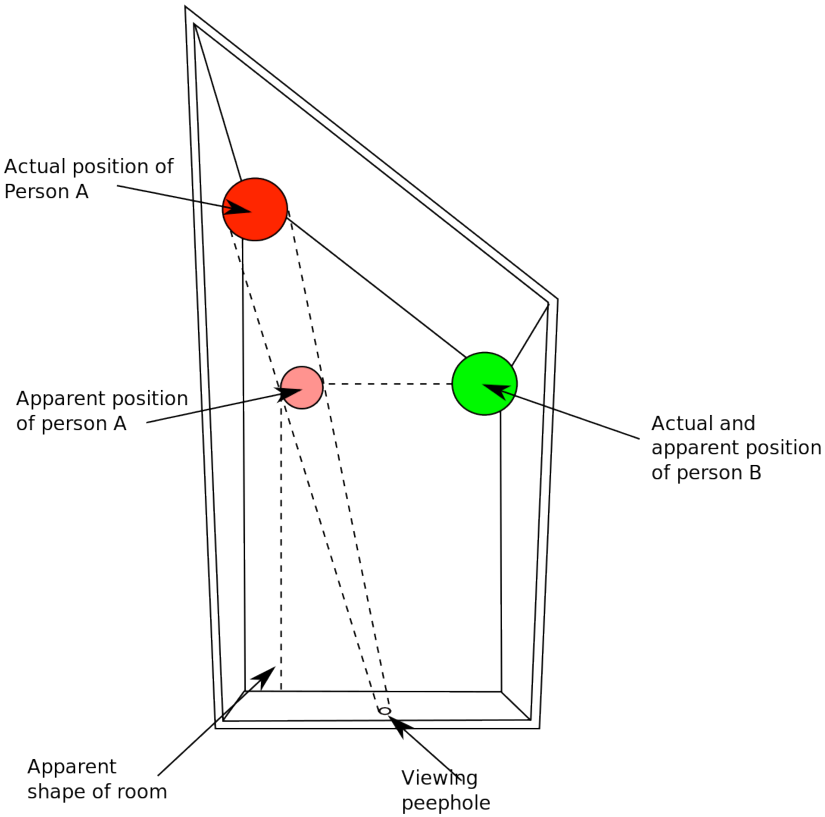


Measuring light vs. measuring scene properties



(c) 2006 Walt Anthony

Assumptions can be wrong



Ames room (1934)

Why is vision hard?

Some things have strong variations in appearance






Some things know that you have eyes

Brady, M. J., & Kersten, D. (2003). Bootstrapped learning of novel objects. J Vis, 3(6), 413-422

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".

Problem set 1
The “one week” vision project

The goal of the first problem set is
to solve vision

A Simple Visual System

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

A Simple World

A Simple World

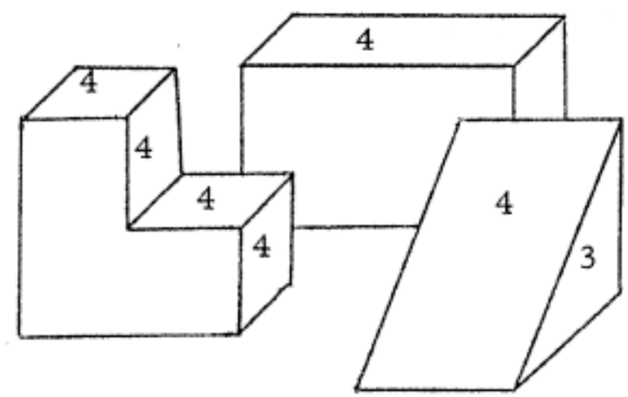
MACHINE PERCEPTION OF THREE-DIMENSIONAL SOLIDS

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.

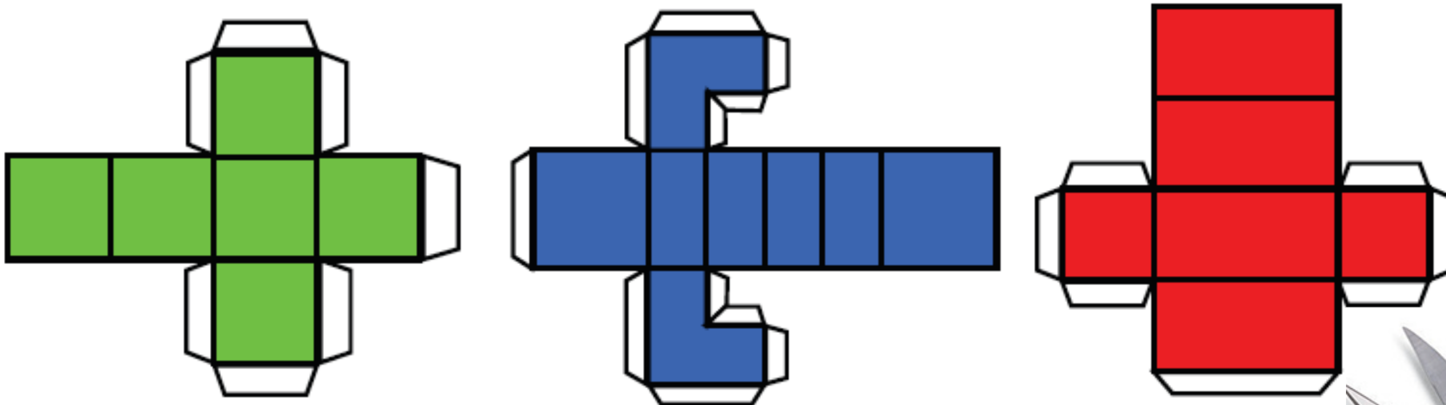
The problem of machine recognition of pictorial data has long been a challenging goal, but has seldom been attempted with anything more complex than alphabetic characters. Many people have felt that research on character recognition would be a first step, leading the way to a more general pattern recognition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

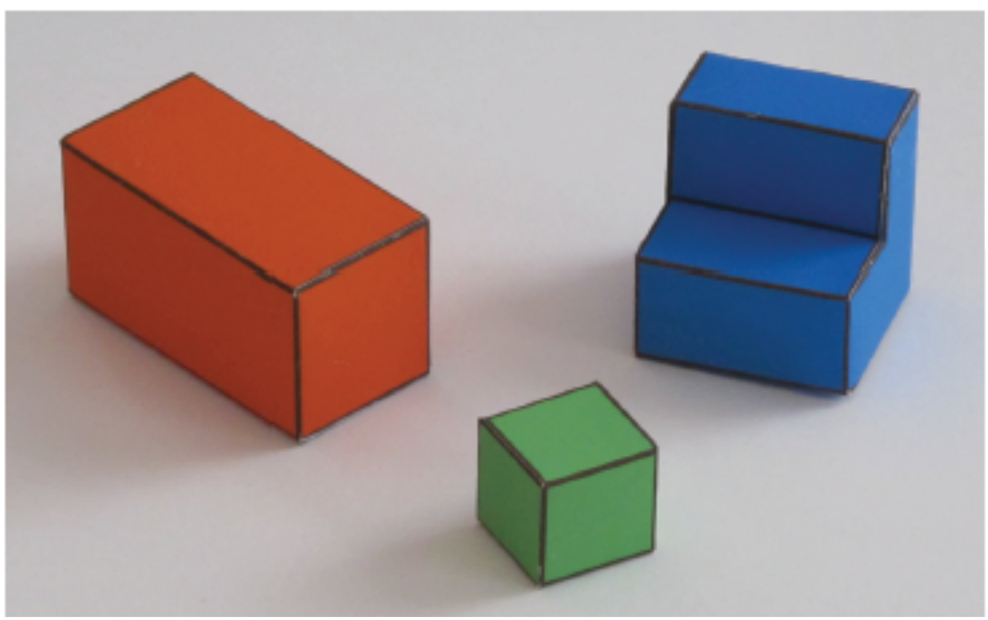


Complete Convex Polygons. The polygon selection procedure would select the numbered polygons as complete and convex. The number indicates the probable number of sides. A polygon is incomplete if one of its points is a collinear joint of another polygon.

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

A Simple World

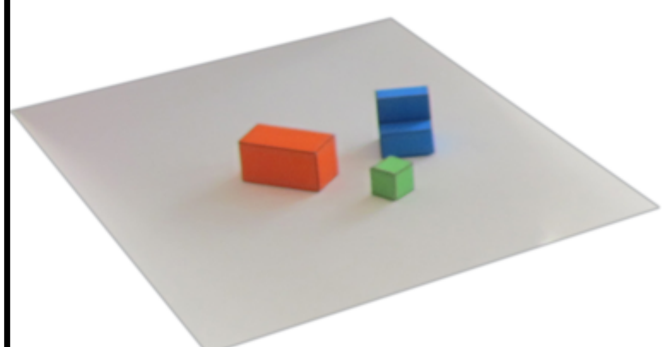




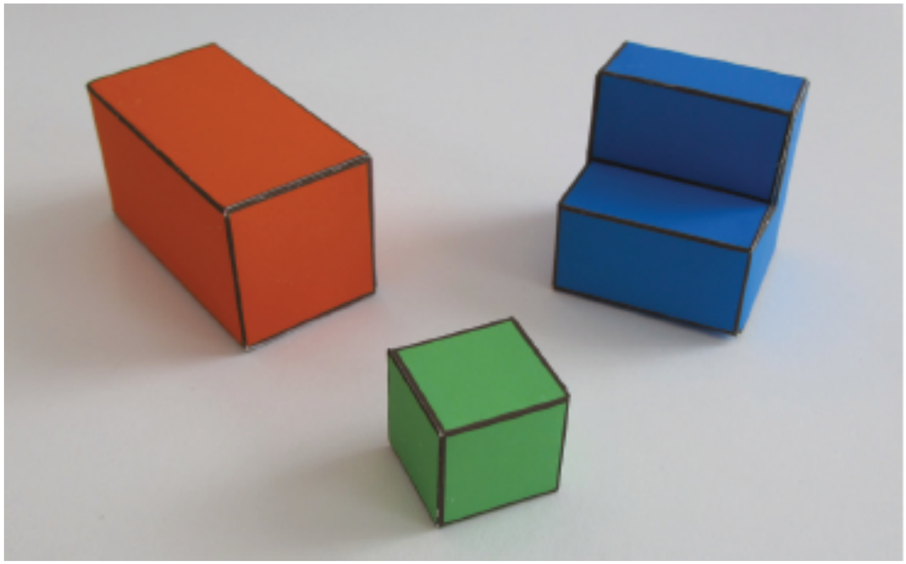
A simple image formation model

Simple world rules:

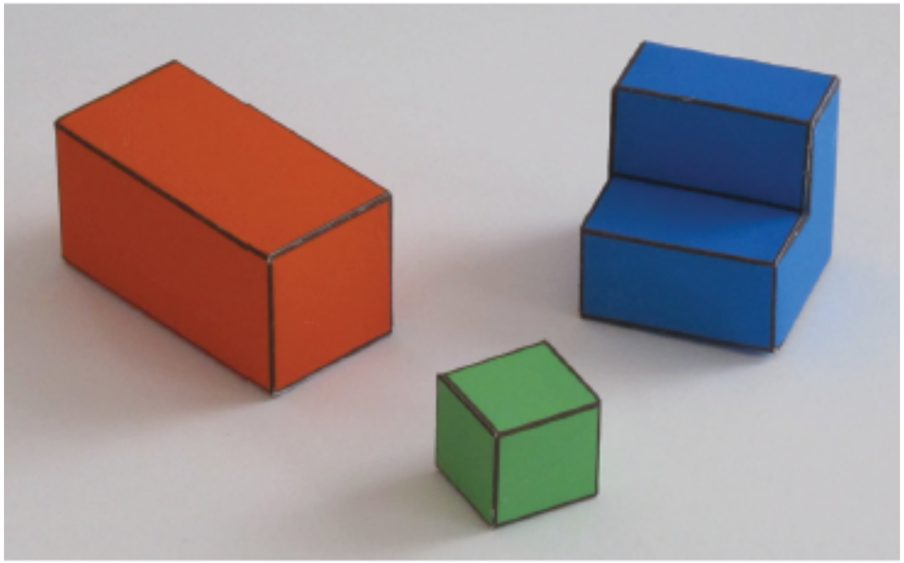
- Surfaces can be horizontal or vertical.
- Objects will be resting on a white horizontal ground plane



A simple image formation model

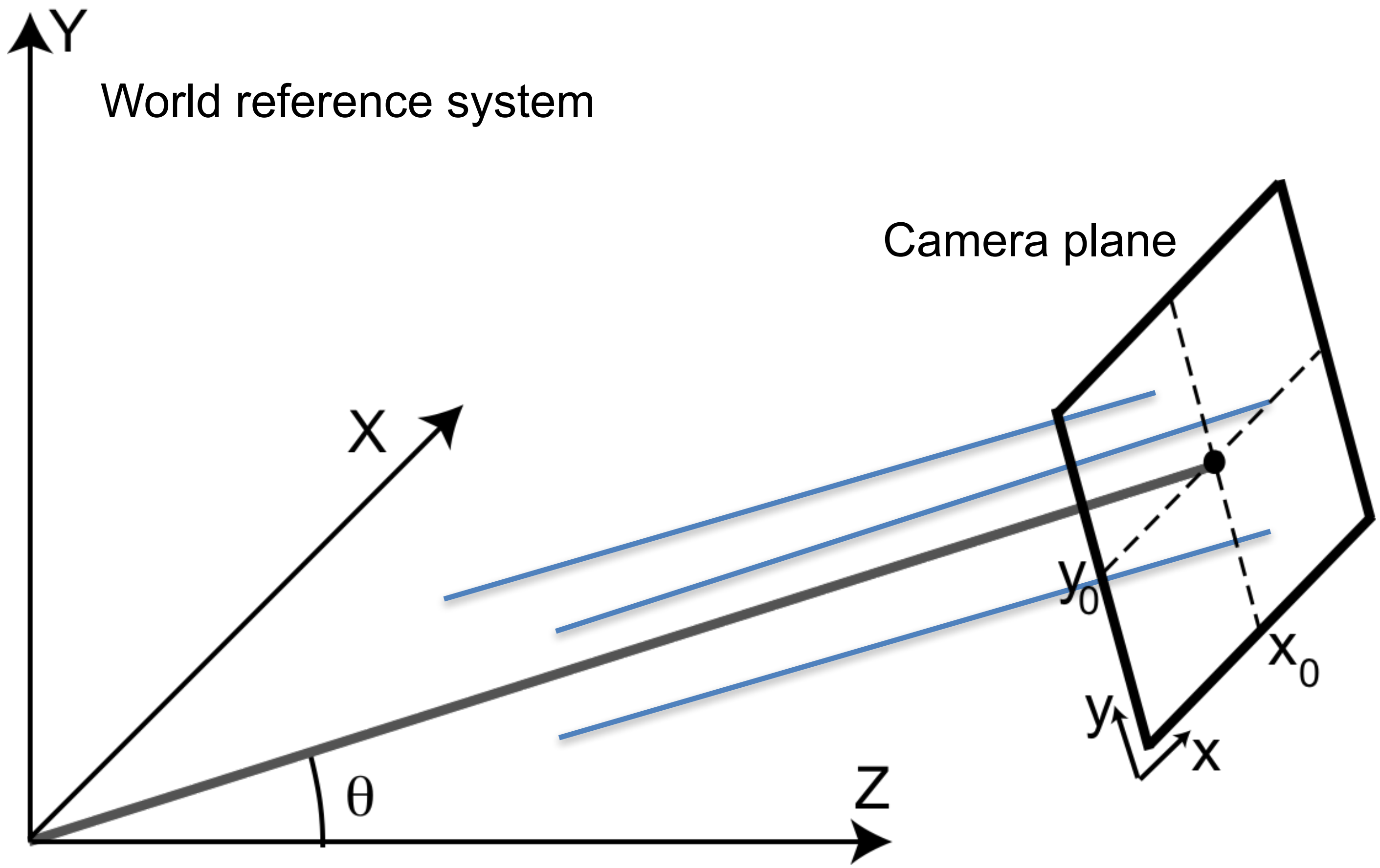


Perspective projection



Parallel (orthographic) projection

A simple image formation model



(right-handed reference system)

A simple image formation model

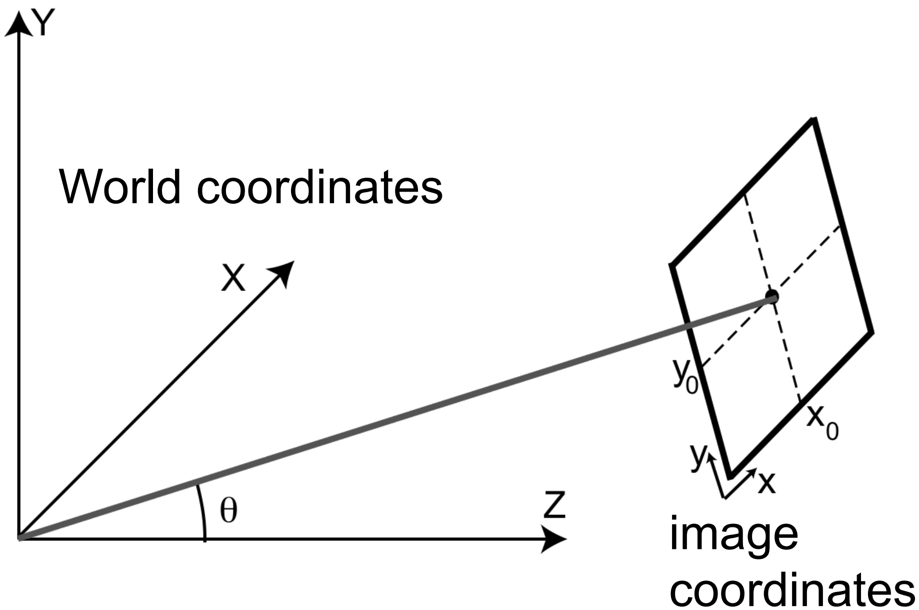
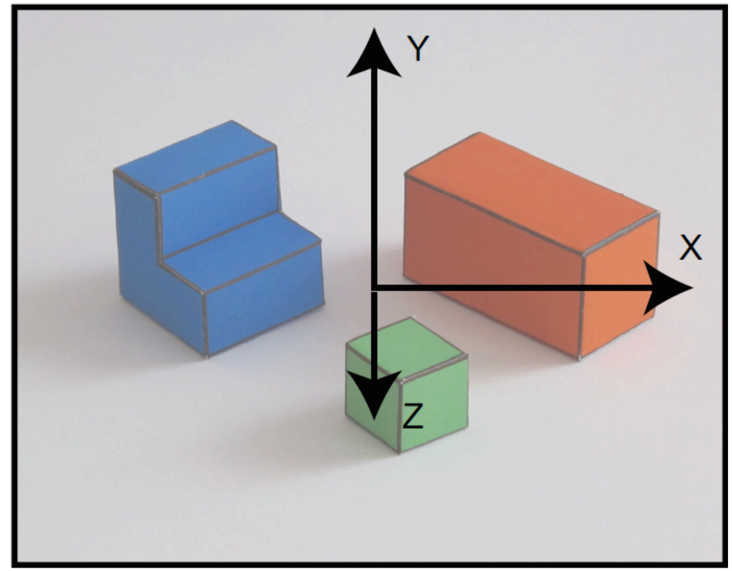


Image and projection of the world coordinate axes into the image plane



World coordinates

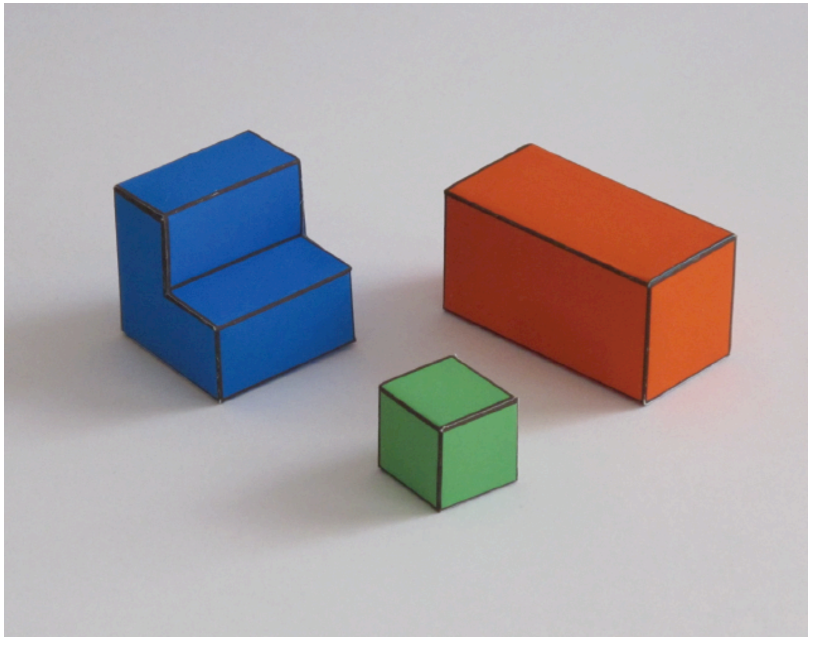
$$x = X + x_0$$

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

image coordinates

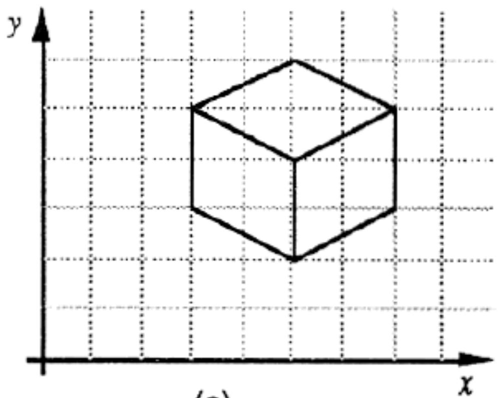
A simple goal

To recover the 3D structure of the world



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

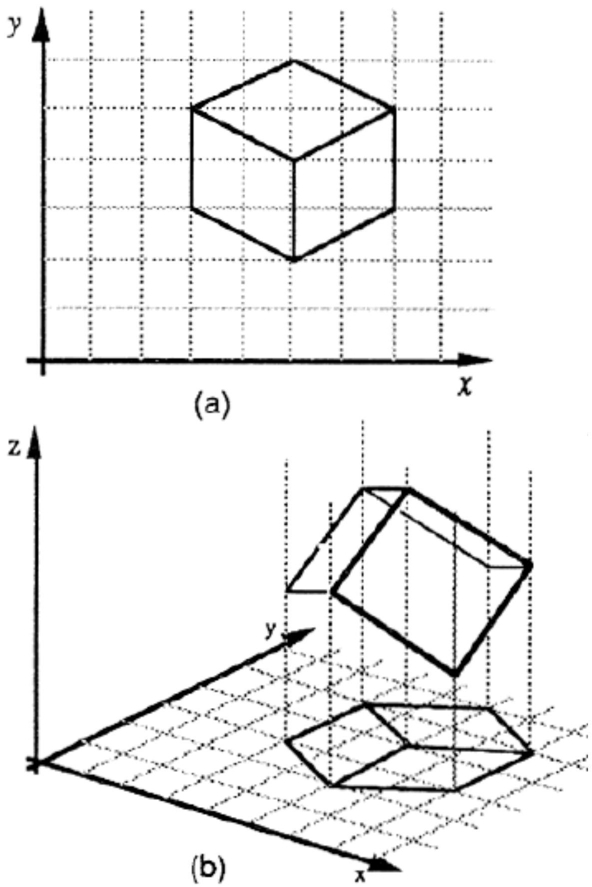
Why is this hard?



(a)

Sinha & Adelson 93

Why is this hard?



Sinha & Adelson 93

Why is this hard?

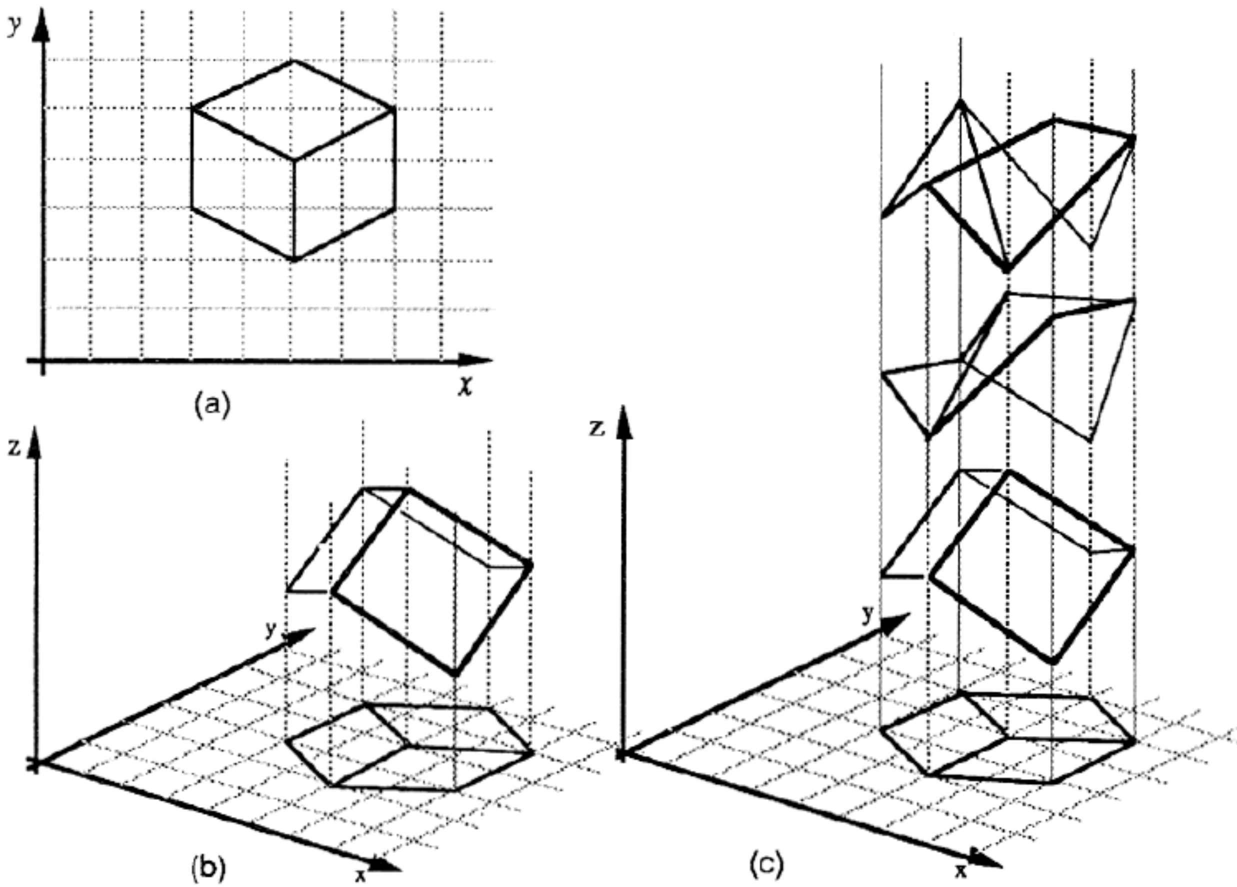
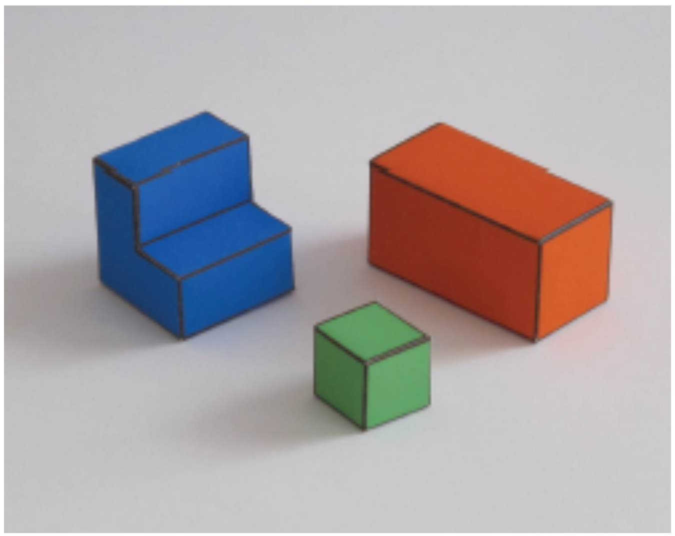
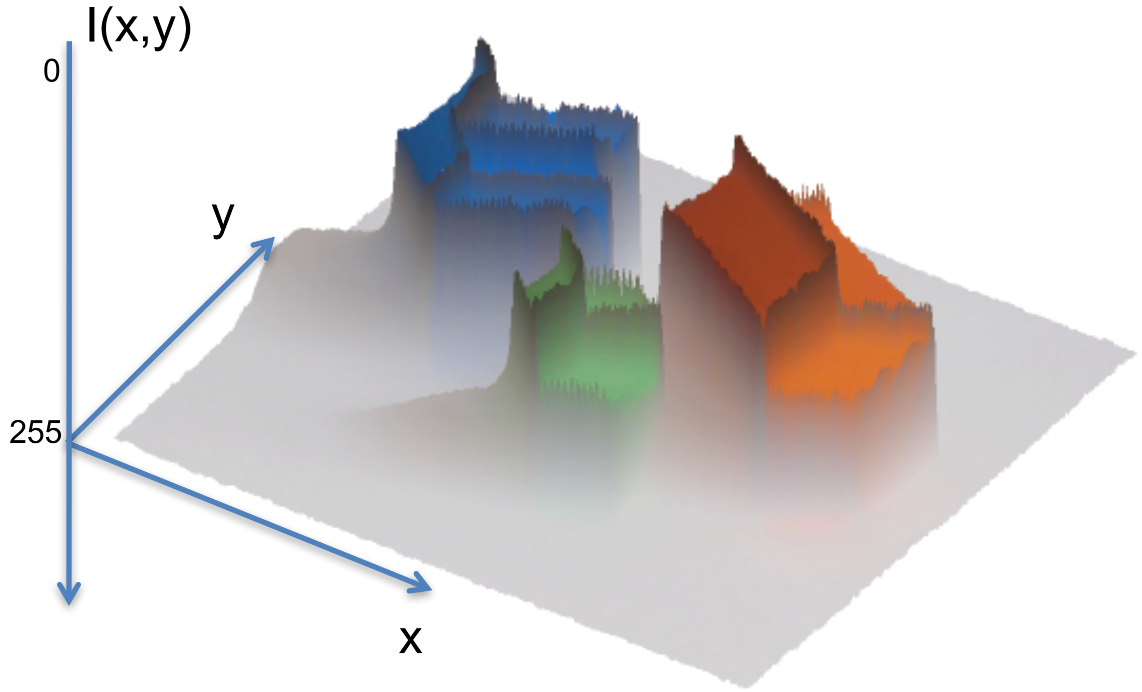


Figure 1. (a) A line drawing provides information only about the x, y coordinates of points lying along the object contours. (b) The human visual system is usually able to reconstruct an object in three dimensions given only a single 2D projection (c) Any planar line-drawing is geometrically consistent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

Sinha & Adelson 93

A simple visual system
The input image



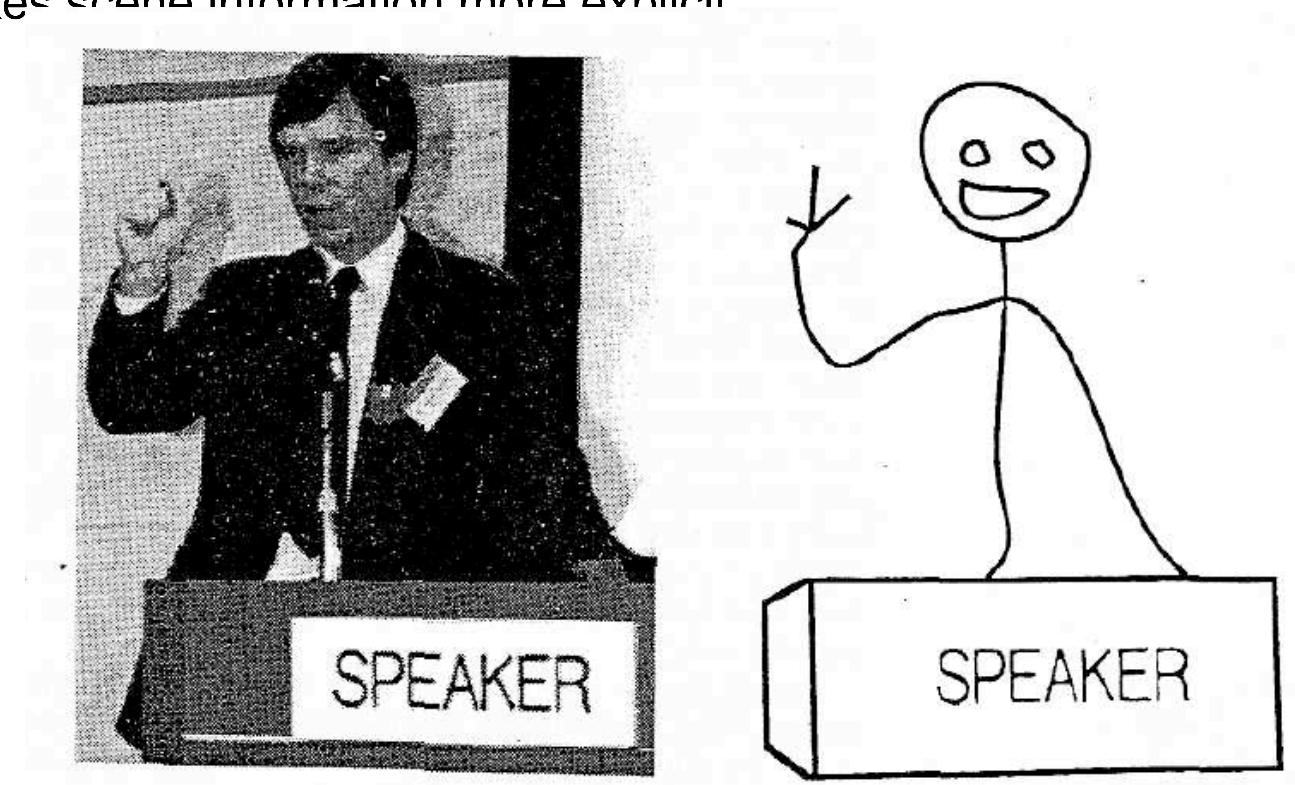


• Proposition 1. The primary task of early vision is to deliver a small set of useful measurements about each observable location.

• Proposition 2. The elemental operations of early vision involve the measurement of local change along various directions.

Adelson, Bergen. 91

• Goal: to transform the image into other representations (rather than pixel values) that makes scene information more explicit

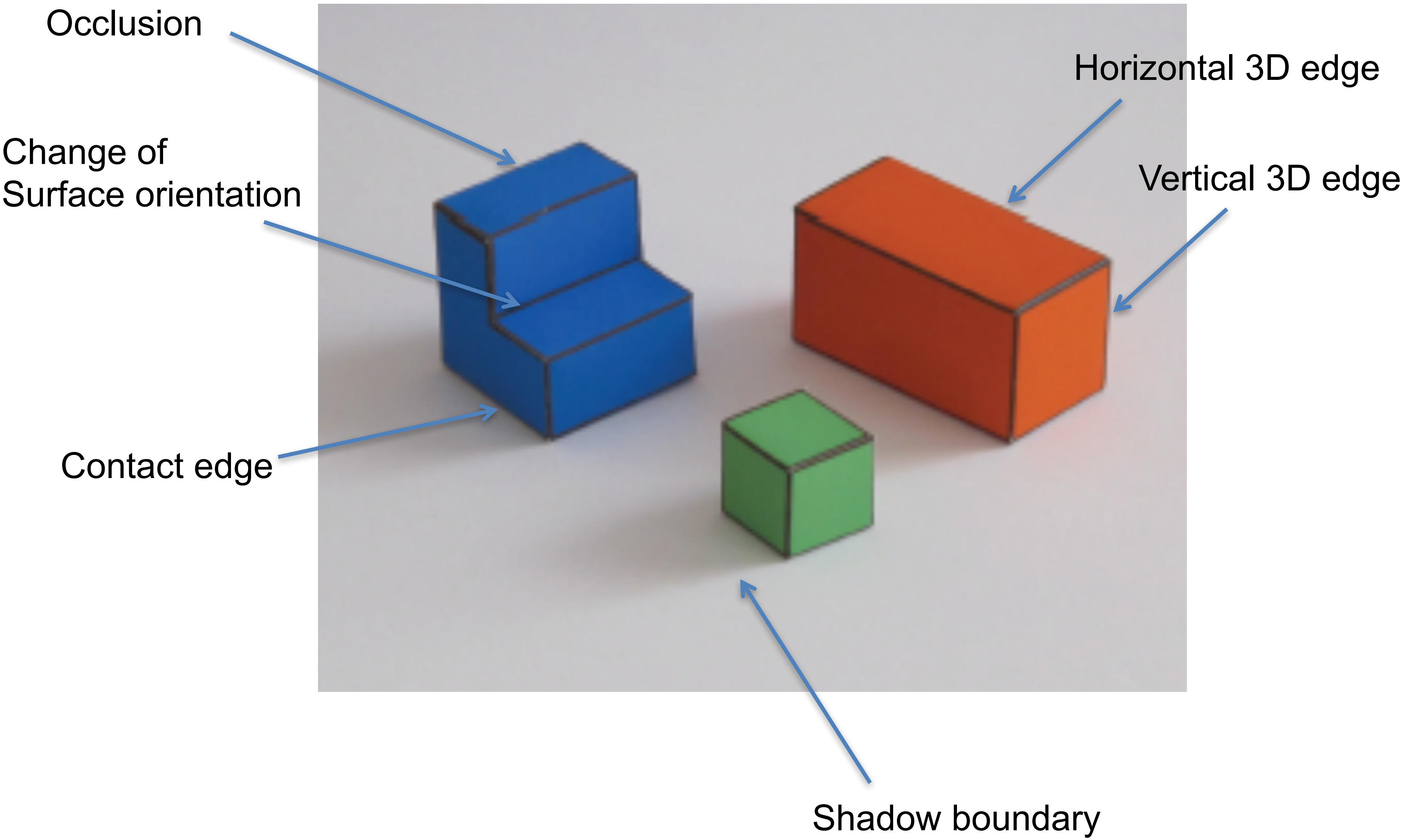


What we think we see

What we really see

Cavanagh, Perception 95

Edges


Occlusion
Change of
Surface orientation
Contact edge
Horizontal 3D edge
Vertical 3D edge
Shadow boundary

Finding edges in the image

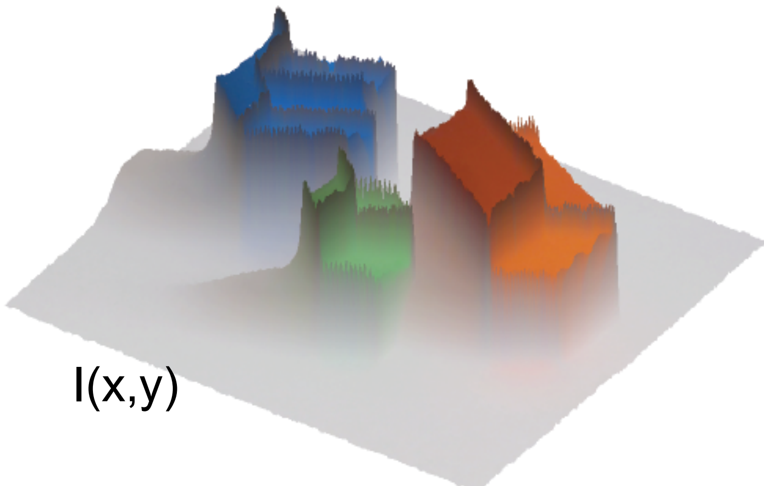


Image gradient:

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

Approximation image derivative:

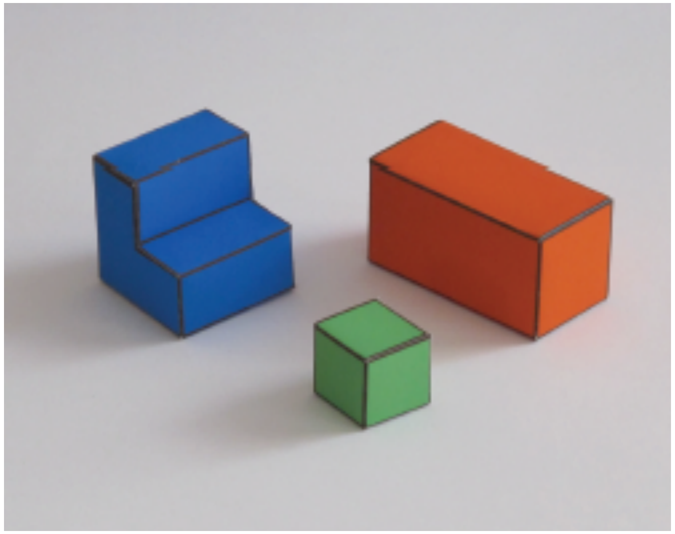
$$\frac{\partial \mathbf{I}}{\partial x} \simeq \mathbf{I}(x,y) - \mathbf{I}(x-1,y)$$

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

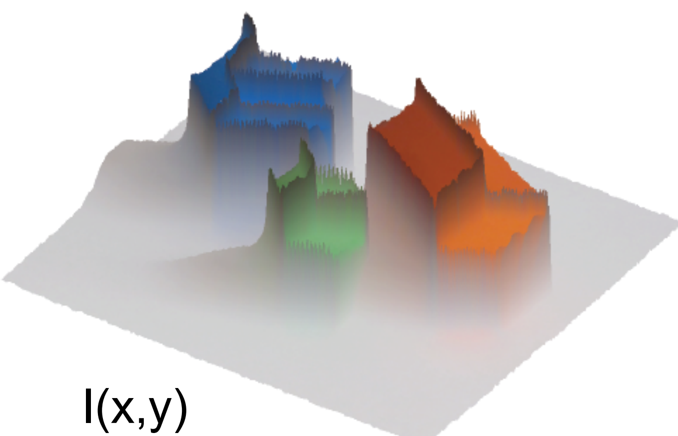
Edge orientation: $\theta(x,y) = \angle \nabla \mathbf{I} = \arctan \frac{\partial \mathbf{I}/\partial y}{\partial \mathbf{I}/\partial x}$

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

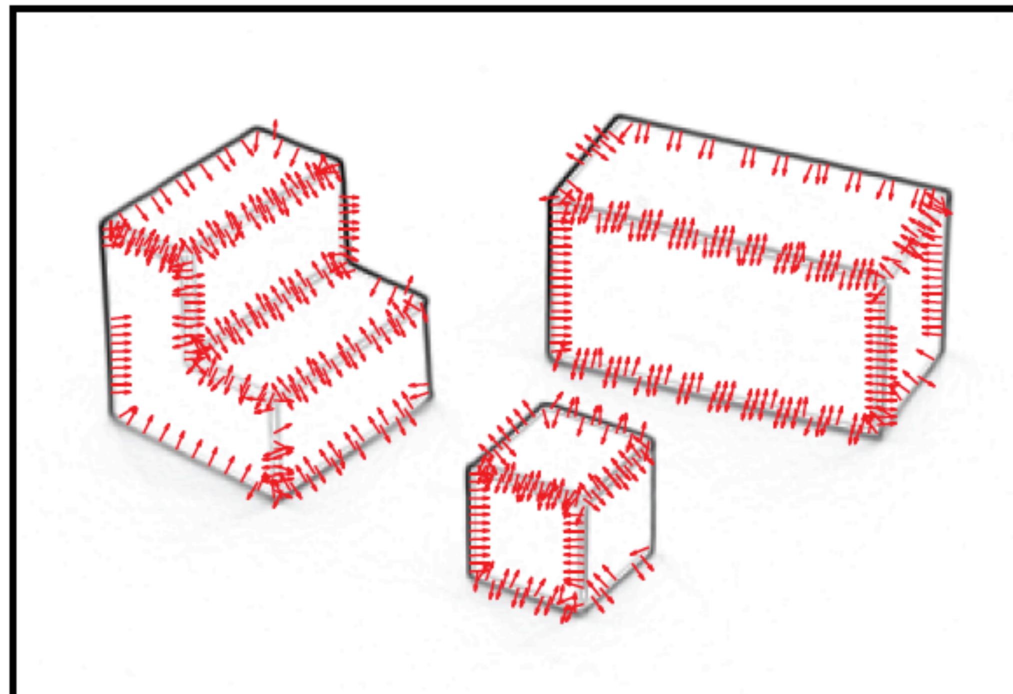
Finding edges in the image



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



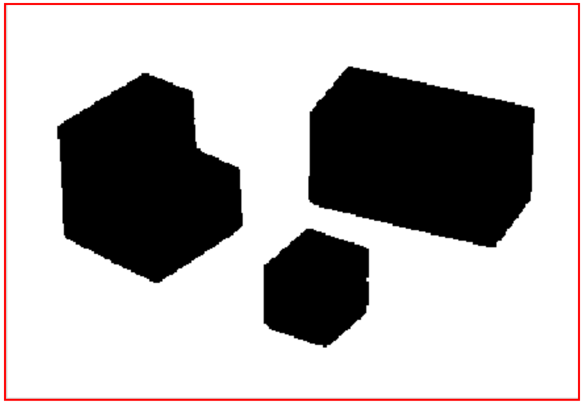
I(x,y)



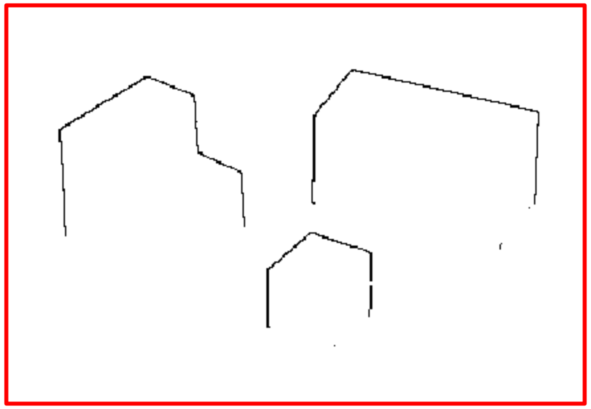
E(x,y) and *n*(x,y)

Edge classification

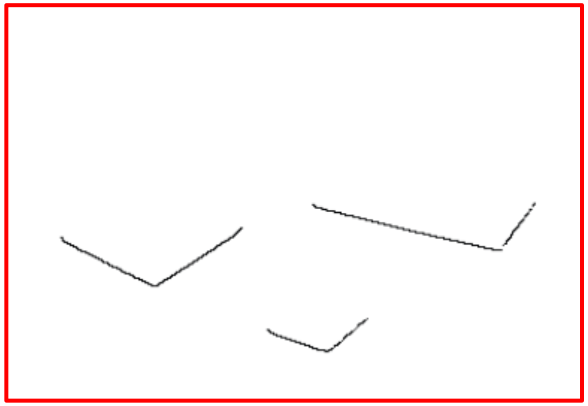
- Figure/ground segmentation
  - Using the fact that objects have color



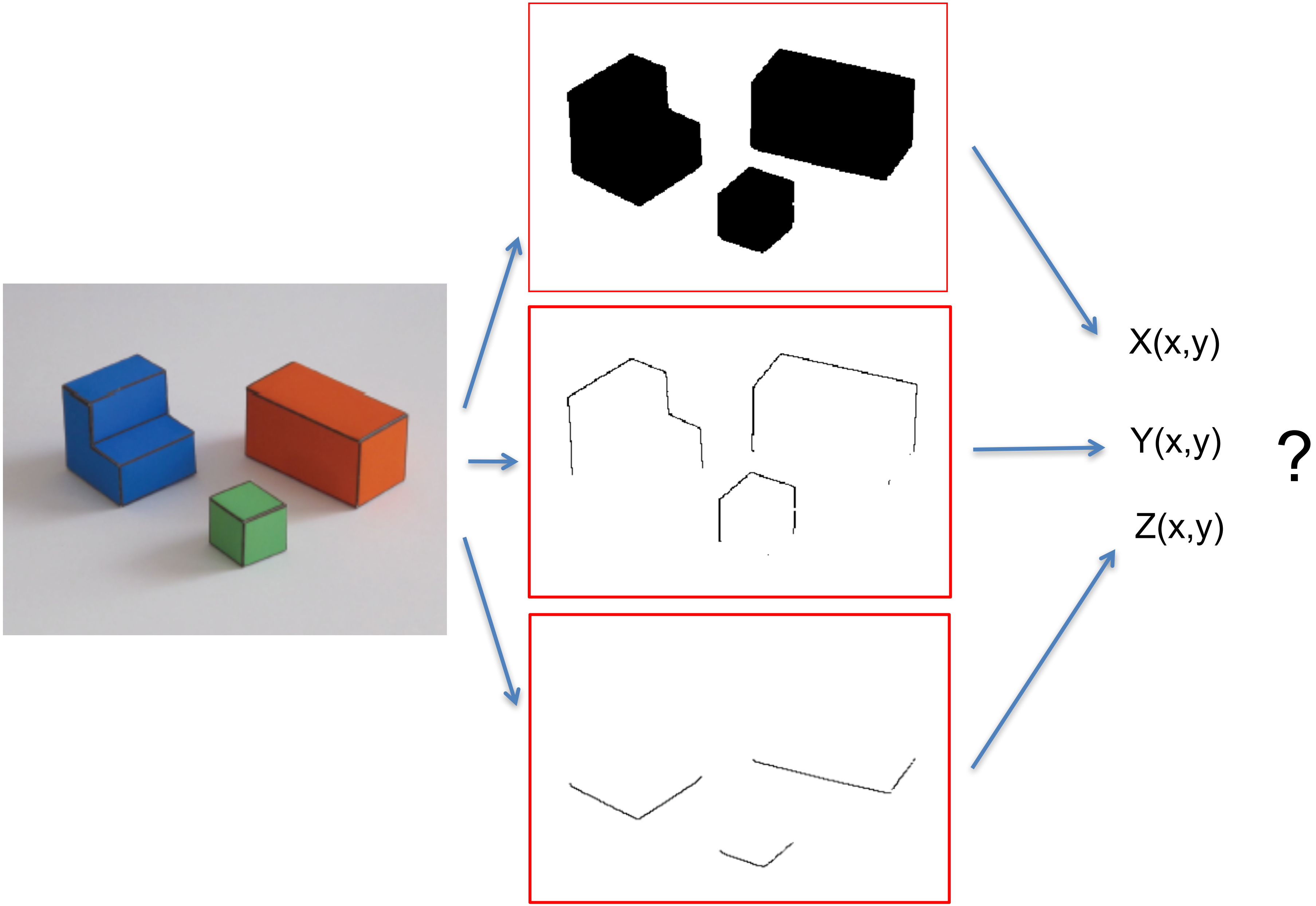
- Occlusion edges
  - Occlusion edges are owned by the foreground



- Contact edges

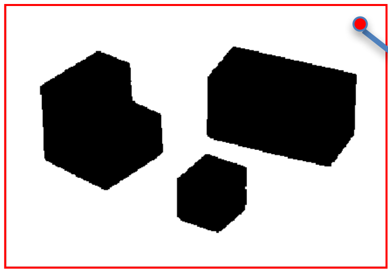


From edges to surface constraints



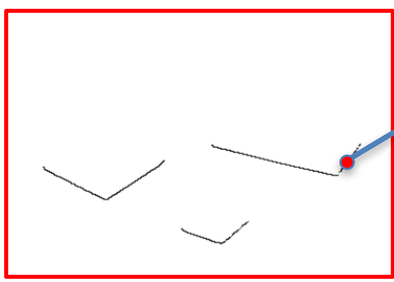
From edges to surface constraints

- Ground



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

- Contact edge

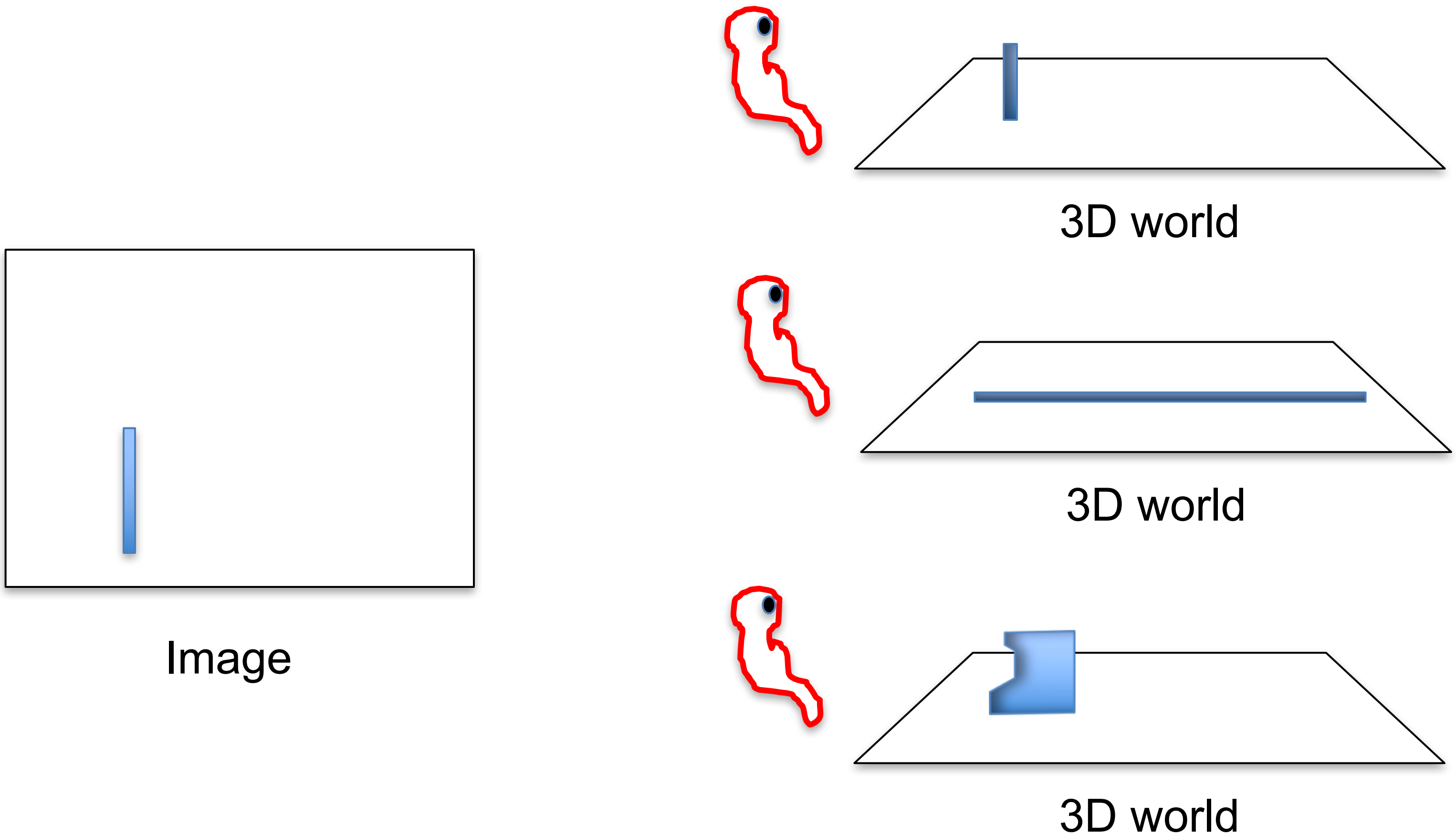


Y(x,y) = 0   if (x,y) belongs to foreground and is a contact edge

- What happens inside the objects?

… now things get a bit more complicated.

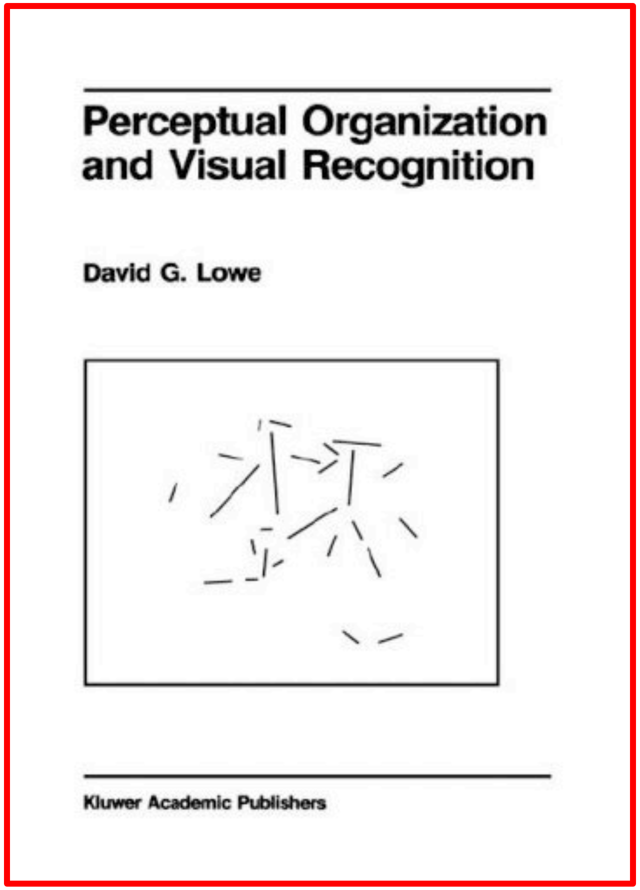
Generic view assumption



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

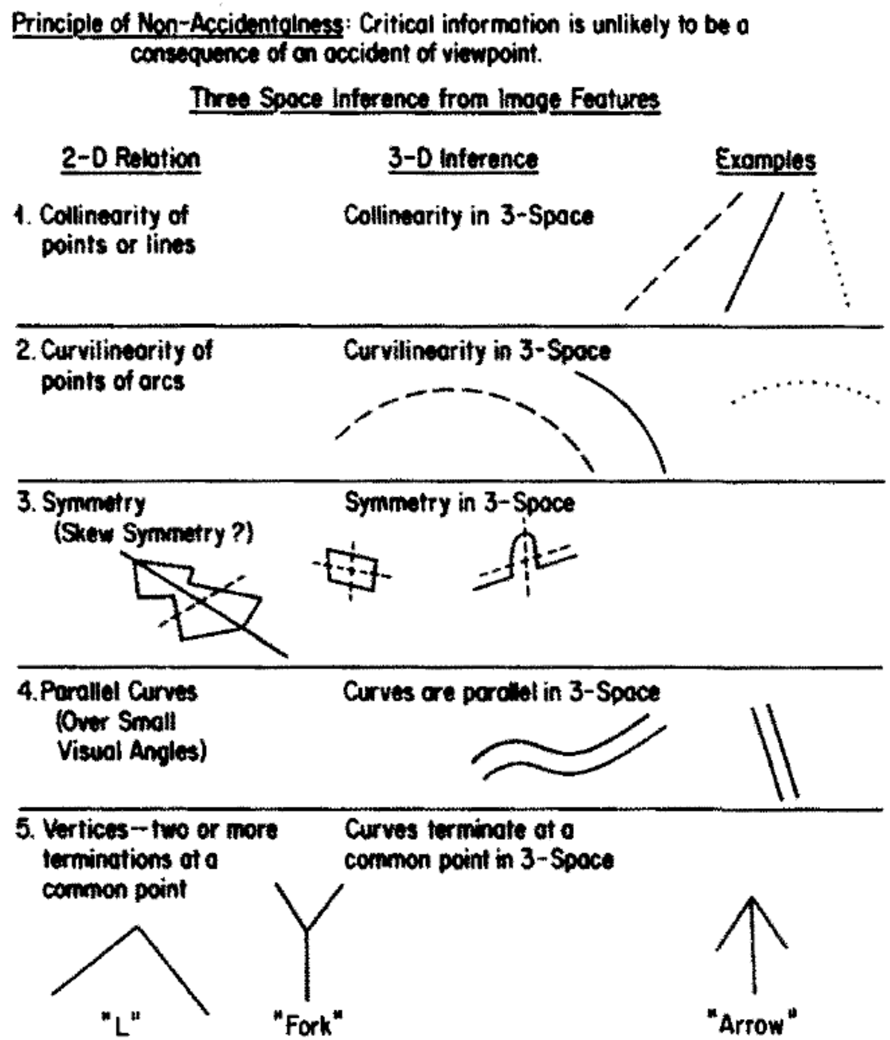
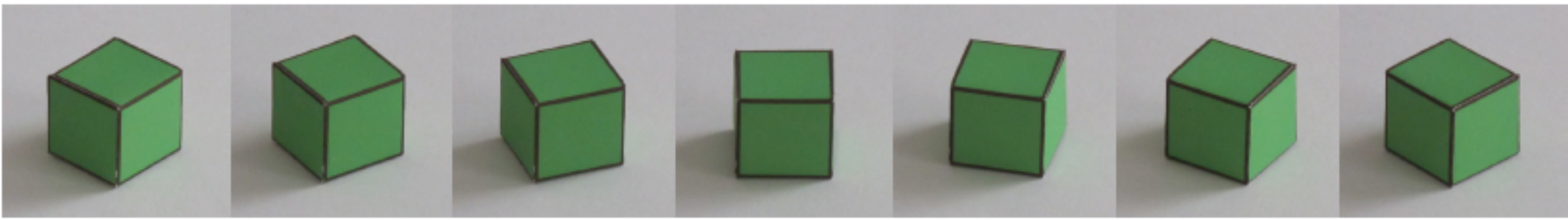


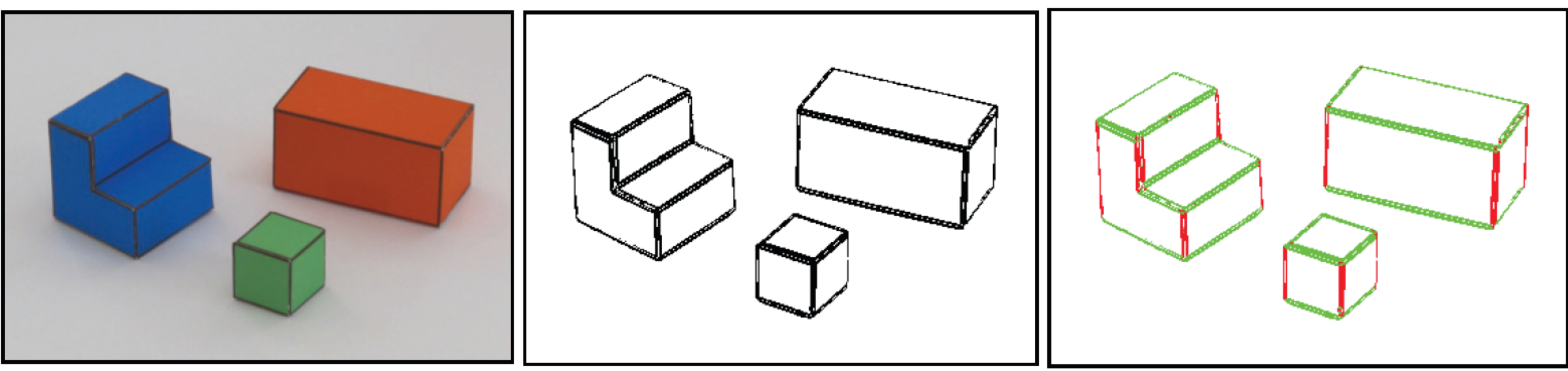
Figure 4. Five nonaccidental relations. (From Figure 5.2, *Perceptual organization and visual recognition* [p. 77] by David Lowe. Unpublished doctorial dissertation, Stanford University. Adapted by permission.)

Biederman_RBC_1987

Non-accidental properties in the simple world



generic generic generic accidental generic generic generic

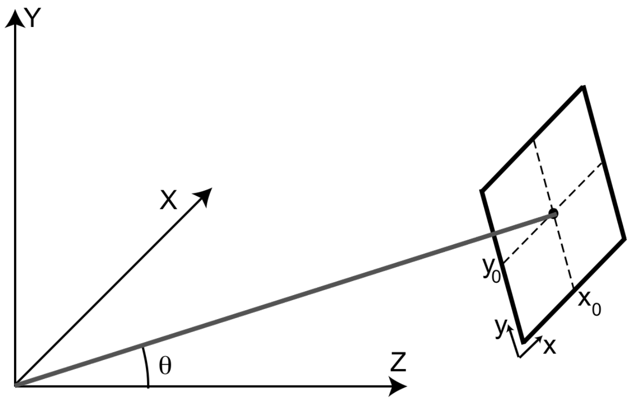


Using E(x,y) Using $\theta(x,y)$

From edges to surface constraints

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

- Vertical edges



World coordinates

$$x = X + x_0$$

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

image coordinates

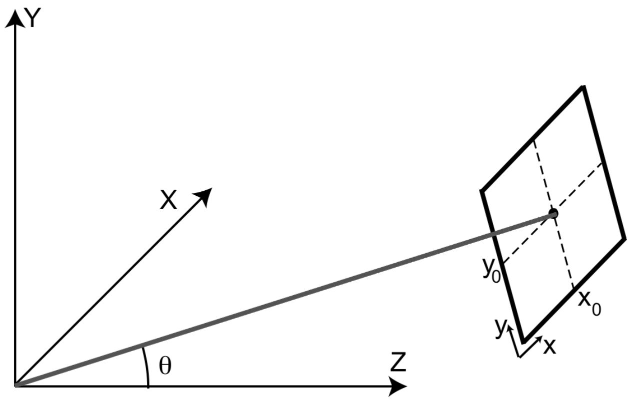
Given the image, what can we say about X, Y and Z in the pixels that belong to a vertical edge?

Z = constant along the edge

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

From edges to surface constraints

- Horizontal edges



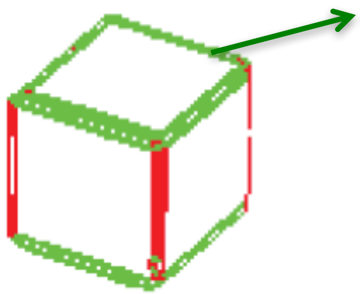
World coordinates

$x = X + x_0$

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

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

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

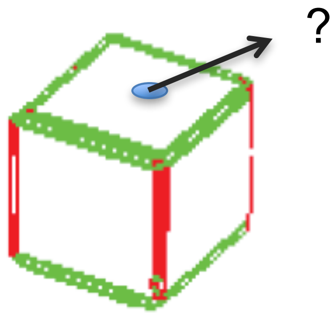
Where **t** is the vector parallel to the edge

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

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

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

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

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

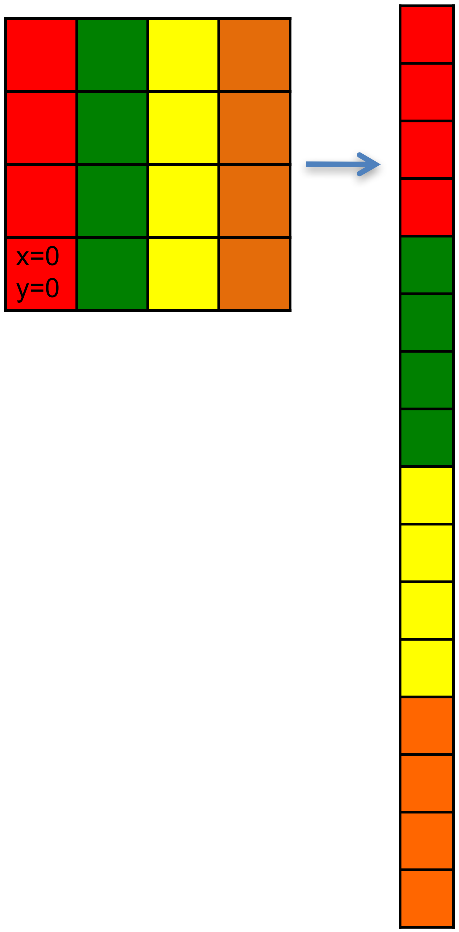
(it is symmetric, and it averages horizontal derivatives over 3 vertical locations)

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

Discrete approximation

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

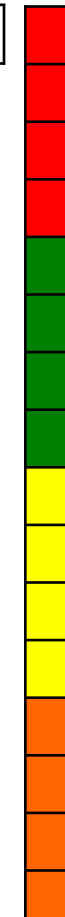
Y(x,y)



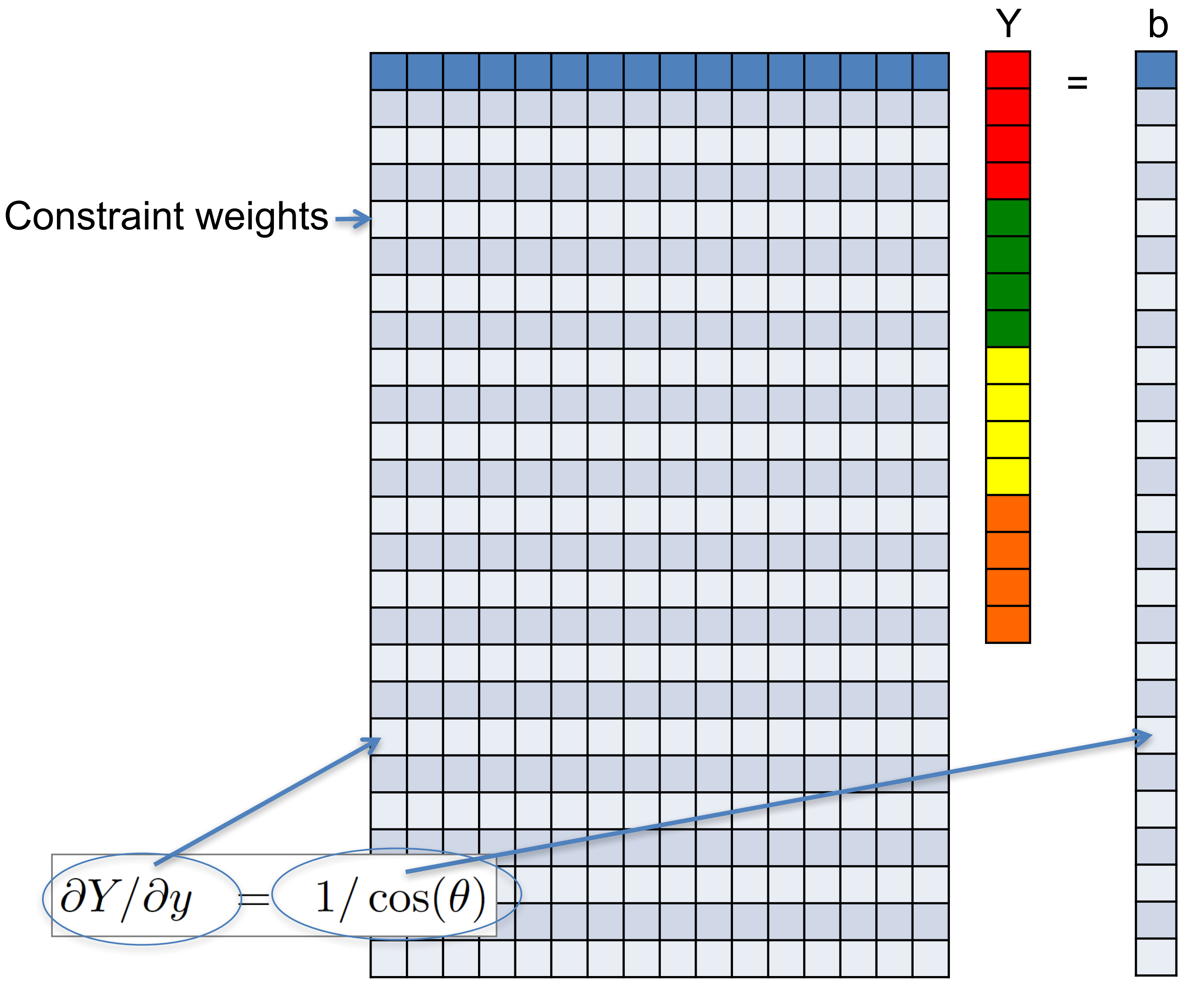
x=2, y=2

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

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



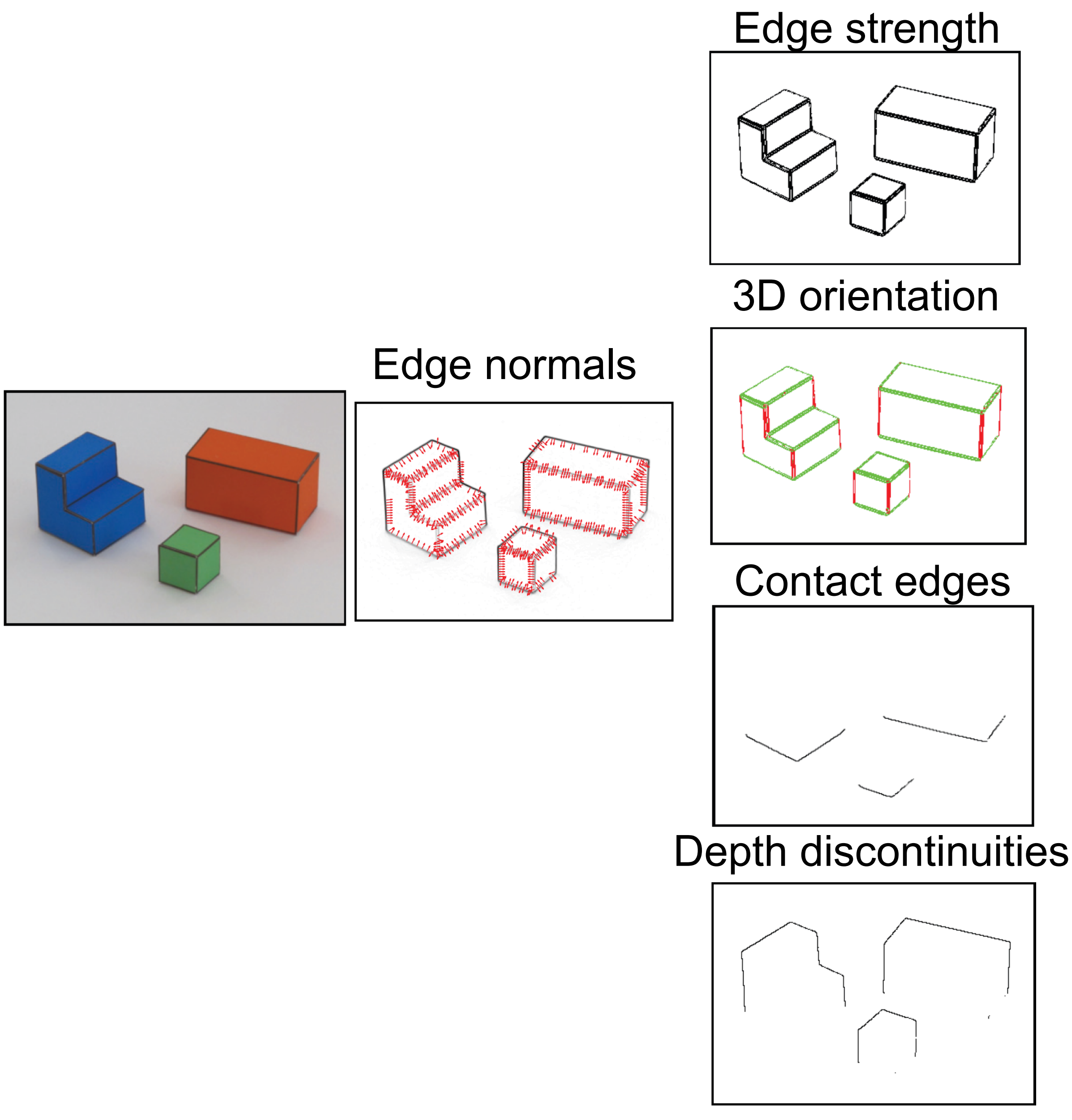
A simple inference scheme



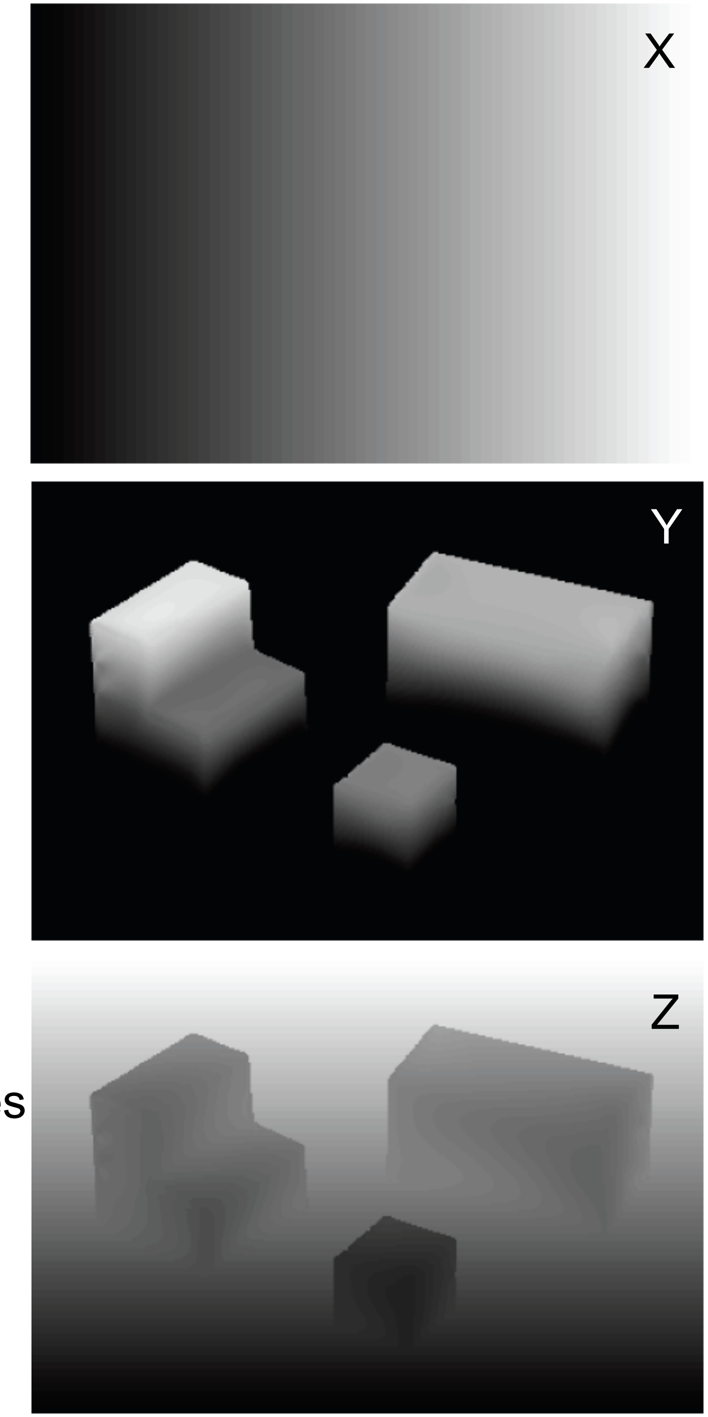
A Y = b

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

Results



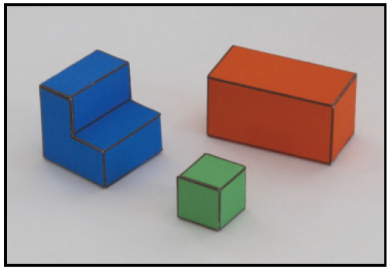
Edge strength
Edge normals
3D orientation
Contact edges
Depth discontinuities



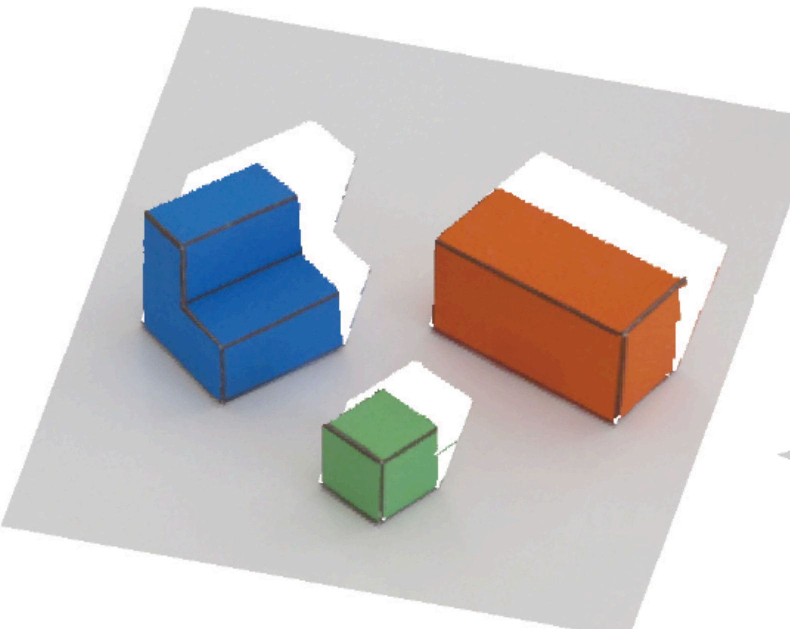
X
Y
Z

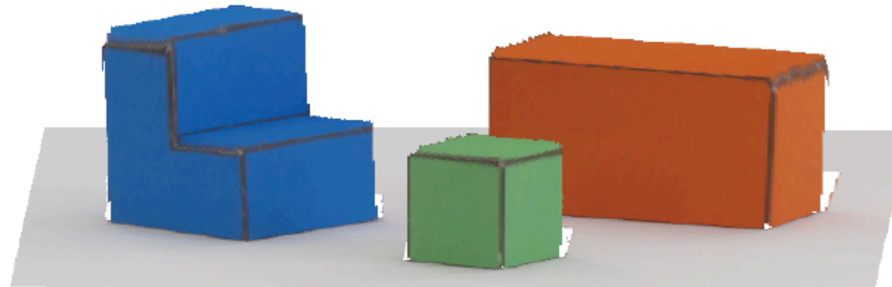
Changing view point

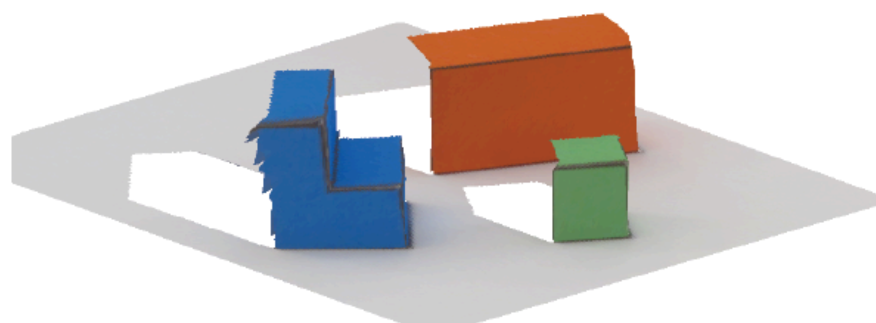
Input



New view points:







Violations of simple world assumptions

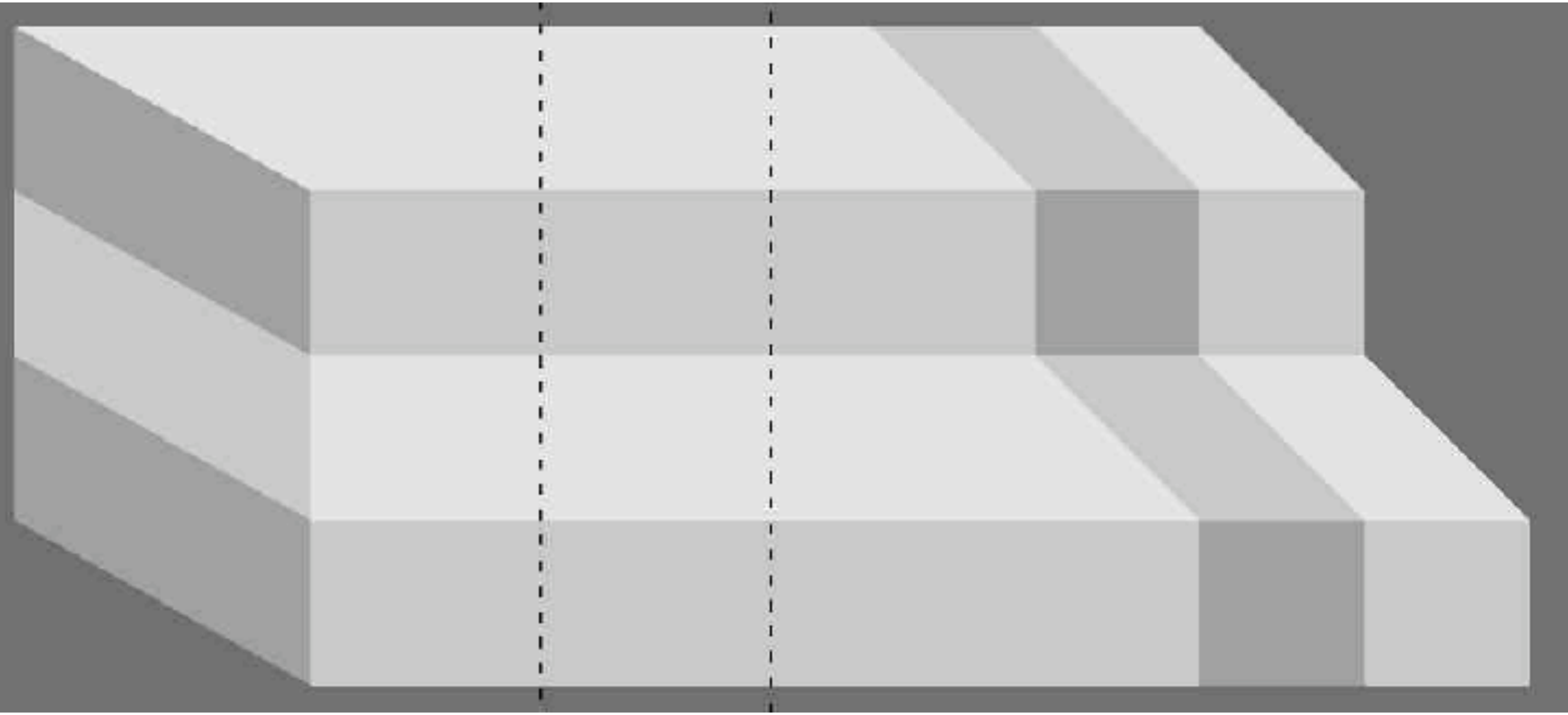


FIGURE 24.9 The impossible steps. On the left, the horizontal stripes appear to be due to paint; on the right, they appear to be due to shading.

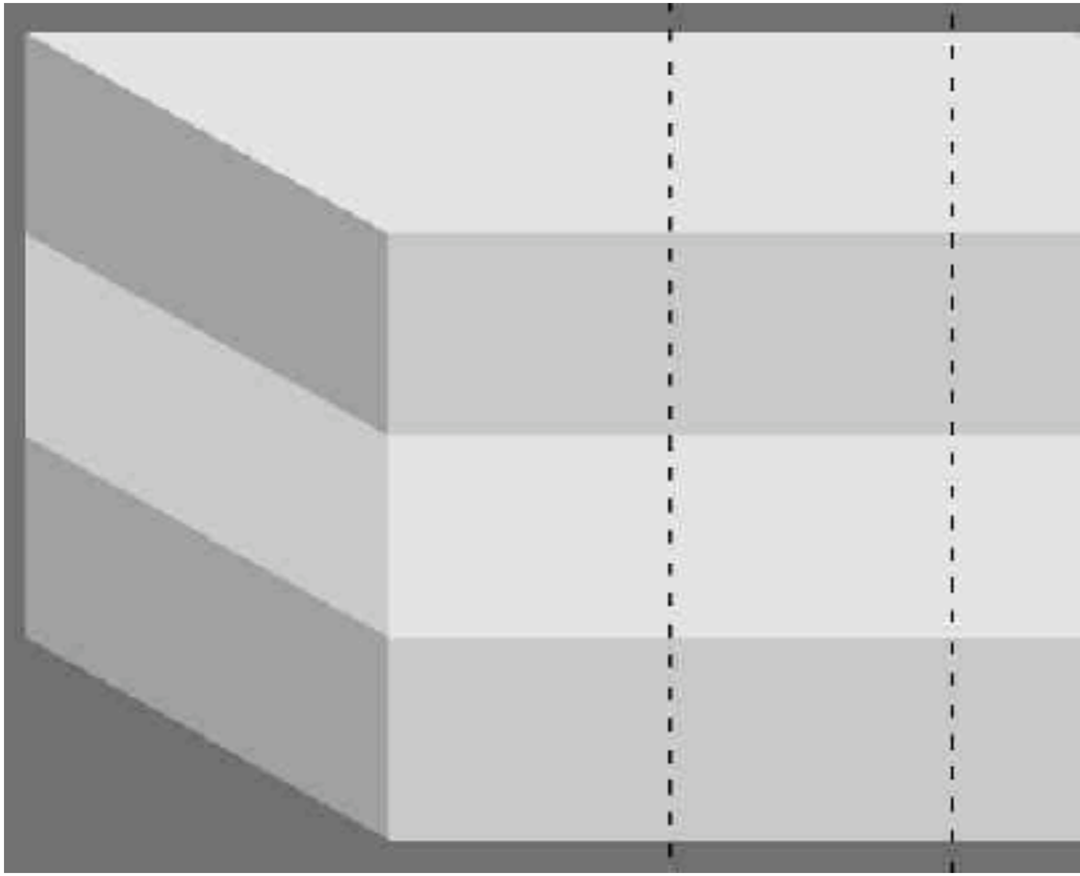
Adelson, E.H. Lightness Perception and Lightness Illusions. In *The New Cognitive Neurosciences,* 2nd ed., M. Gazzaniga, ed. Cambridge, MA: MIT Press, pp. 339-351, (2000).

24 Lightness Perception and Lightness Illusions

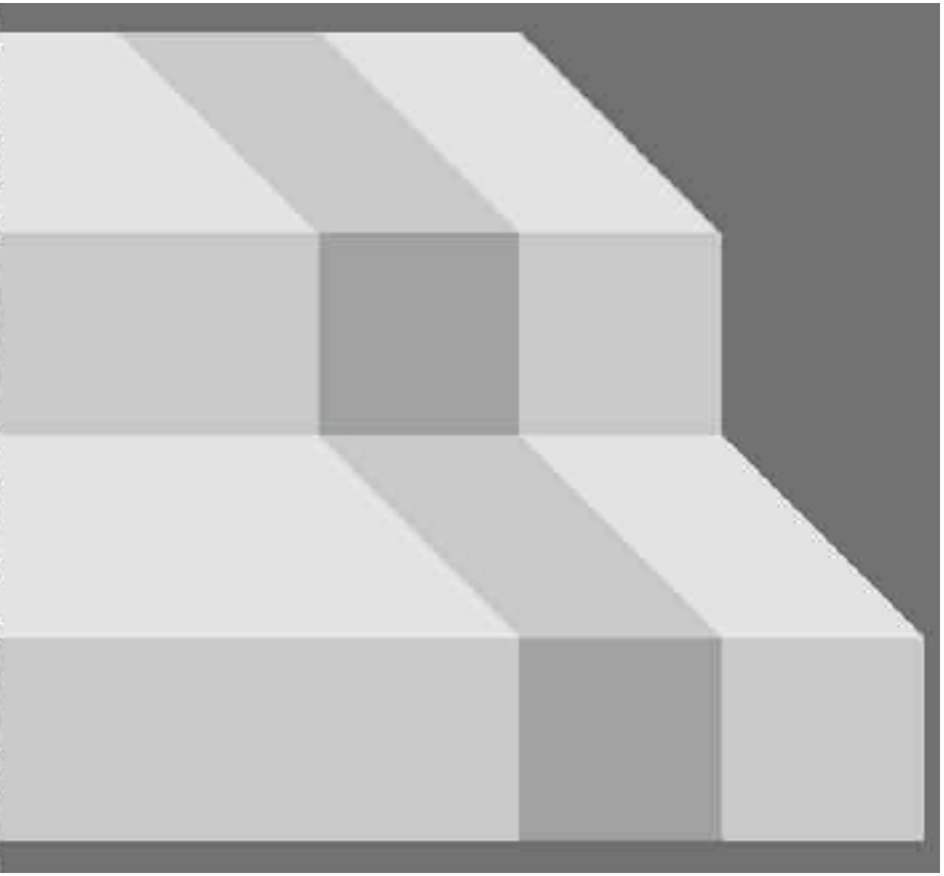
EDWARD H. ADELSON

Violations of simple world assumptions

Shading is due to painted stripes

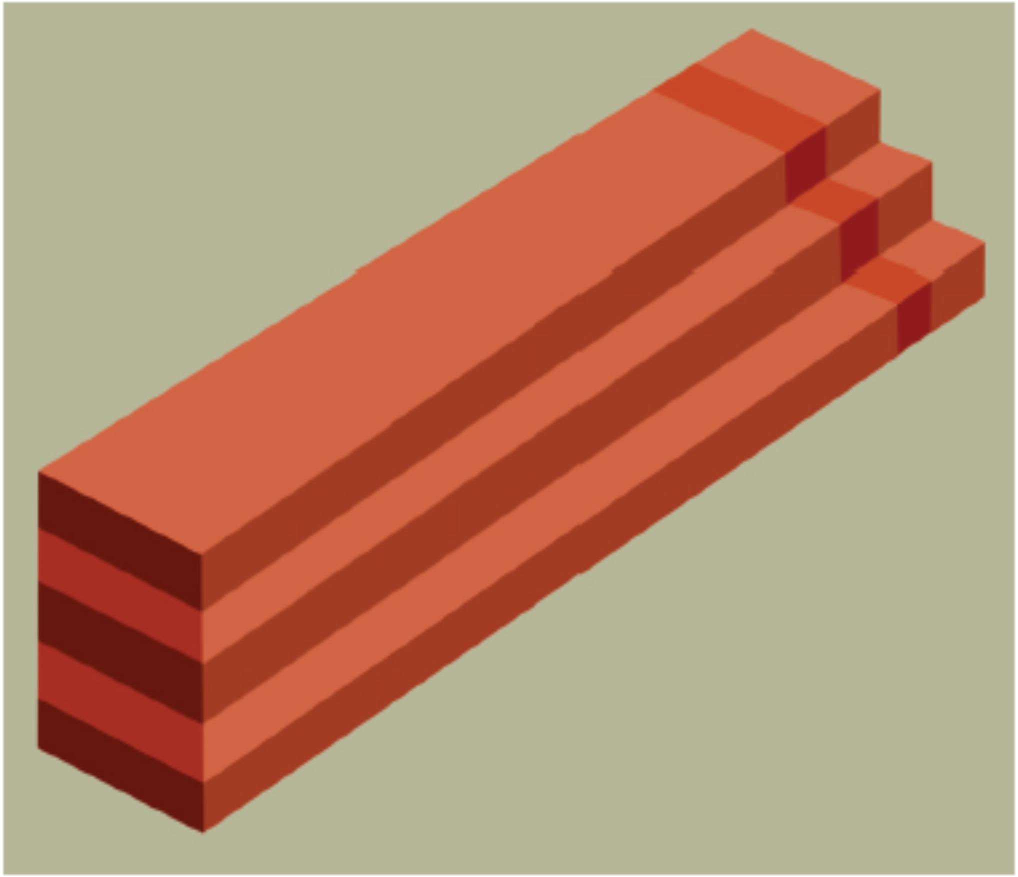


Violations of simple world assumptions

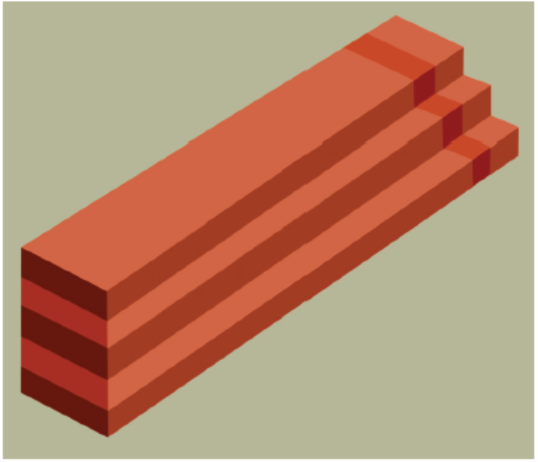


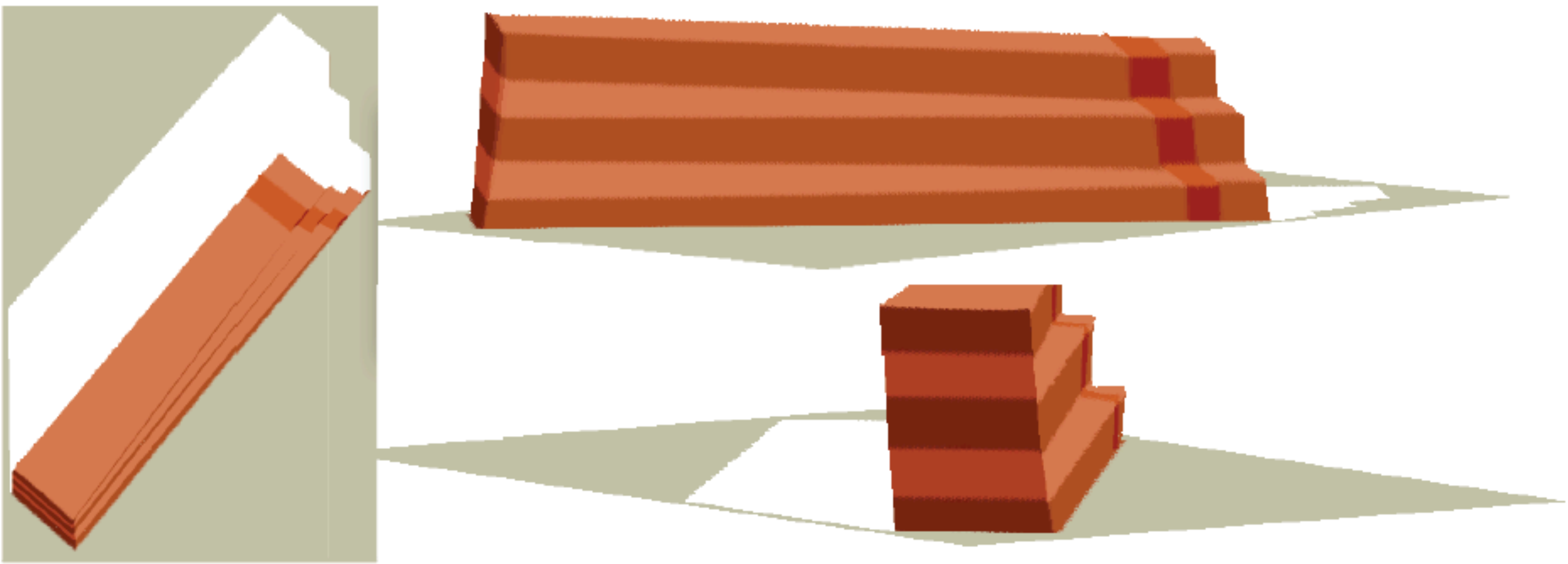
Shading is due to illumination

Impossible steps



Impossible steps





Problem set 1
The “one week” vision project