

<http://groups.csail.mit.edu/vision/courses/6.869/>



MIT CSAIL

6.869: Advances in Computer Vision

MIT
COMPUTER
VISION

Spring 2011

Instructors: William T. Freeman Antonio Torralba
Office: 32-D476 Office: 32-D432
email: billf@mit.edu email: torralba@csail.mit.edu
Office hours: Wednesday 2:30-3:30pm Office hours: Monday 2:30-3:30pm

TA: Michael Rubinstein
Office: 32-D460
email: mrub@mit.edu
Office hours: Tuesday 4-5pm D4-south open area

Lectures: MW 1-2:30pm (36-144)
Units: 3-0-9 (Graduate H-level, Area II AI TQE)
Prerequisites: 6.041 or 6.042; 18.06
Course Website: <http://groups.csail.mit.edu/vision/courses/6.869>
Stellar: <https://stellar.mit.edu/S/course/6/sp11/6.869>

Assignments

- Problem sets (2/3)
 - Almost weekly
 - Graded as 1-3
 - Late policy
 - Collaboration policy
- Final project (1/3)
 - Project proposal
 - 5min class presentation
 - Report
- No exams or quizzes

Readings

Untitled3

Computer vision: Algorithms and Applications

<http://research.microsoft.com/en-us/um/people/szeliski/Book/>

Computer Vision: Algorithms and Applications

(c) [Richard Szeliski](#), Microsoft Research

Welcome to the repository for drafts of my computer vision textbook.

This book is largely based on the computer vision courses that I have co-taught at the University of Washington ([2008](#), [2005](#), [2001](#)) and Stanford (2003) with [Steve Seitz](#) and [David Fleet](#).

While I am working on the book, I would *love* to have people "test-drive" it in their computer vision courses (or their research) and [send me feedback](#).

The PDFs should be enabled for commenting directly in your viewer. Also, hyper-links to sections, equations, and references are enabled. To get back to where you were, use Alt-Left-Arrow in Acrobat.

This Web site is also a placeholder for the site that will accompany my computer vision textbook once it is published. Once I get further along with the project, I hope to publish supplemental course material here, such as figures and images from the book, slides sets, pointers to software, and a bibliography.

Latest draft

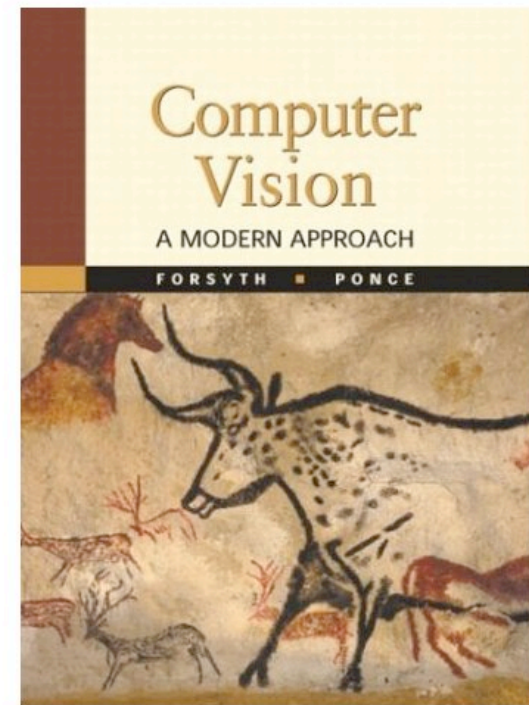

[February 2, 2009](#)

Older drafts

[January 20, 2009](#)

[January 12, 2009](#)

Done



nb group (nb.csail.mit.edu)

NB 2.0 mru@mit.edu - Mozilla Firefox

File Edit View History Bookmarks Tools Help

http://nb.csail.mit.edu/

NB 2.0 mru@mit.edu

Files Nav Home chapter1.pdf*

1/13 105%

Chapter 1

A Simple Vision System

Ya right...

1.1 Introduction

In 1966, Seymour Papert wrote a proposal for building a vision system as a summer project [4]. The abstract of the proposal starts stating a simple goal: "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 report then continues dividing all the tasks (most of which also are common parts of modern computer vision approaches) among a group of MIT students. This project was a reflection of the optimism existing on the early days of computer vision. However, the task proved to be harder than anybody expected.

The goal of this first lecture is to present several of the main topics that we will cover during this course.

Class is in the front

Staff Sign Myself

Why is vision hard?

It is difficult to say exactly what makes vision hard as we do not have a solution yet [3]. In this section we will mention two of the aspects that make vision hard: first is the structure of the input, and second is the structure of the output.

Discard Save

1.2 The structure of ambient light

From a light source a dense array of light rays emerges in all directions. But before these light rays reach the eye, the light interacts with the objects in the world. Lets consider a single ray emerging from the light source (we will use here a simple geometric interpretation of the structure of light). If the light ray is not directed towards the eye, this ray will strike some surface in the world and, as result, a new set of rays will be emitted in many directions. This process will continue producing more and more interactions and filling the space. In the middle of the space, the observer will sense only a subset of the light rays that will strike the eye. As a result of this complex pattern of interactions (reflections, absorptions, etc.) even a single ray will form a complex image in the eye of the observer. Figure 1.1 shows some pictures taken when illuminating a complex scene with a laser pointer which will illuminate the scene with a narrow beam of light (this is the best approximation to the image produced by a single light ray that I could do at home). The resulting images show the complexity of the interactions between the different surfaces. The sum of the contribution of all the light rays emitted by the light source will give rise to a naturally looking picture. Despite of the complexity of the structure of the ambient light (a term coined by J. J. Gibson [1]) with multiple reflections, shadows, specular surfaces (which provide incorrect disparity information

Notes*

9 Threads

sorted by: ans

chapter1.pdf (page 2)

Figure 1.1: Left) scene illuminated with a ceiling lamp. Right) the two images on the right have been obtained by illuminating the scene with a laser pointer. On each image, it is indicated (red arrow) the approximate direction of the light beam produced by pointer.

to our two eyes, our visual system has no problem in interacting with this scene, even if it is among the first things that we have to see just after waking up.

The pattern of light filling the space can be described by the function:

$$P(\theta, \Phi, \lambda, t, X, Y, Z)$$

where P is the light intensity of a ray passing by the world location (X, Y, Z) in the direction given by the angle (θ, Φ) , and with wavelength λ (we will study color in lecture 7) at an instant in time t . This function, called the plenoptic function by Adelson¹ and Bergen [2], contains all the information needed to describe the complete pattern of light rays that fills the space. The plenoptic function does not include information about the observer. The observer does not have access to the entire plenoptic function, only to a small slice of it and in lecture 11 we will describe how different mechanisms can produce images by sampling the ambient light in different ways.

For a given observer, most of the light rays are occluded. If it was not because of occlusion, vision would be a lot simpler. Occlusion is the best example of how hard vision can get. Many times, properly interpreting an image will require to understand what is occluded (e.g., we know that a person is not floating just because the legs are occluded behind a table). Unfortunately occlusions are common. In

chapter1.pdf (page 2)

to our two eyes, our visual system has no problem in interacting with this scene, even if it is among the first things that we have to see just after waking up.

The pattern of light filling the space can be described by the function:

$$P(\theta, \Phi, \lambda, t, X, Y, Z) \quad (1.1)$$

where P is the light intensity of a ray passing by the world location (X, Y, Z) in the direction given by the angle (θ, Φ) , and with wavelength λ (we will study color in lecture 7) at an instant in time t . This function, called the plenoptic function by Adelson¹ and Bergen [2], contains all the information needed to describe the complete pattern of light rays that fills the space. The plenoptic function does not include information about the observer. The observer does not have access to the entire plenoptic function, only to a small slice of it and in lecture 11 we will describe how different mechanisms can produce images by sampling the ambient light in different ways.

For a given observer, most of the light rays are occluded. If it was not because of occlusion, vision would be a lot simpler. Occlusion is the best example of how hard vision can get. Many times, properly interpreting an image will require to understand what is occluded (e.g., we know that a person is not floating just because the legs are occluded behind a table). Unfortunately occlusions are common. In

chapter1.pdf (page 2)

occlusion is the best example of how hard vision can get. Many times, properly interpreting an image will require to understand what is occluded (e.g., we know that a person is not floating just because the legs are occluded behind a table). Unfortunately occlusions are common. In fact, in the world, there is more occluded matter than visible to any given observer.

Although knowing the entire Plenoptic function can have many applications, fortunately, the goal of vision is not to recover this function.

1.2.2 Measuring light vs. measuring scene properties

If the goal of vision was to simply measure the light intensity coming from a particular direction of (like a photometer) then things will be clear. However, the goal of vision is to provide an interpretation of the world in front of the observer in terms of "meaningful" surfaces, objects, materials, etc., in order to extract all the different elements that compose the scene (anything that will be relevant to the observer). This problem is hard because most of the information is lost and the visual system needs to make a number of assumptions about the structure of the visual world in order to be able to recover the information from a small sample of the plenoptic function. It is also hard because of our understanding of what is relevant for the observer is also incomplete.

¹Adelson was going to call it the holoscopic function, but a well-known holographer told him that he would punch the nose if he called it that



MIT CSAIL

6.869: Advances in Computer Vision

MIT
COMPUTER
VISION

Lecture 1

A Simple Vision System

What is vision?

- What does it mean, to see? “to know what is where by looking”.
- How to discover from images what is present in the world, where things are, what actions are taking place.

The importance of images

Some images are more important than others

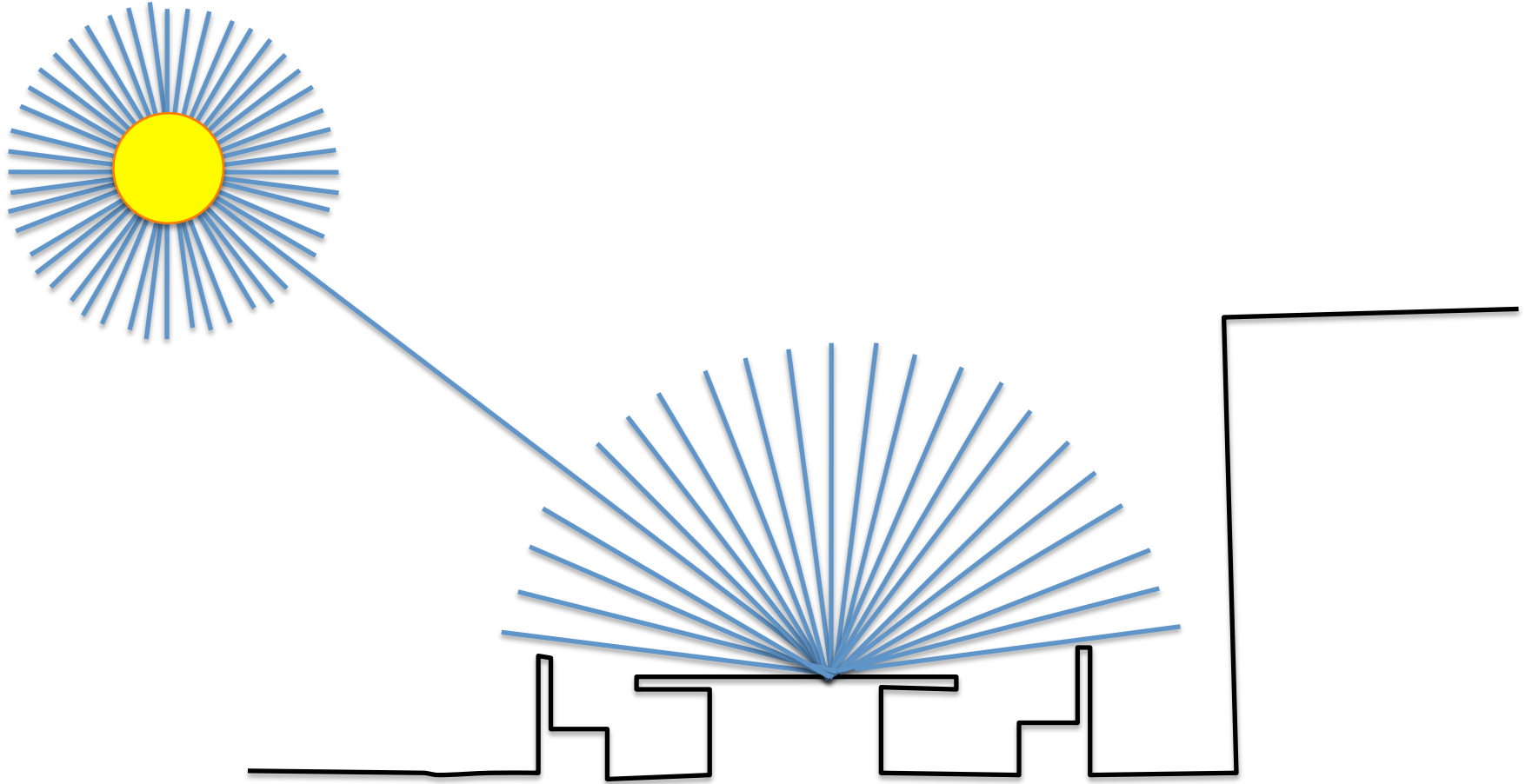


"Dora Maar au Chat"
Pablo Picasso, 1941

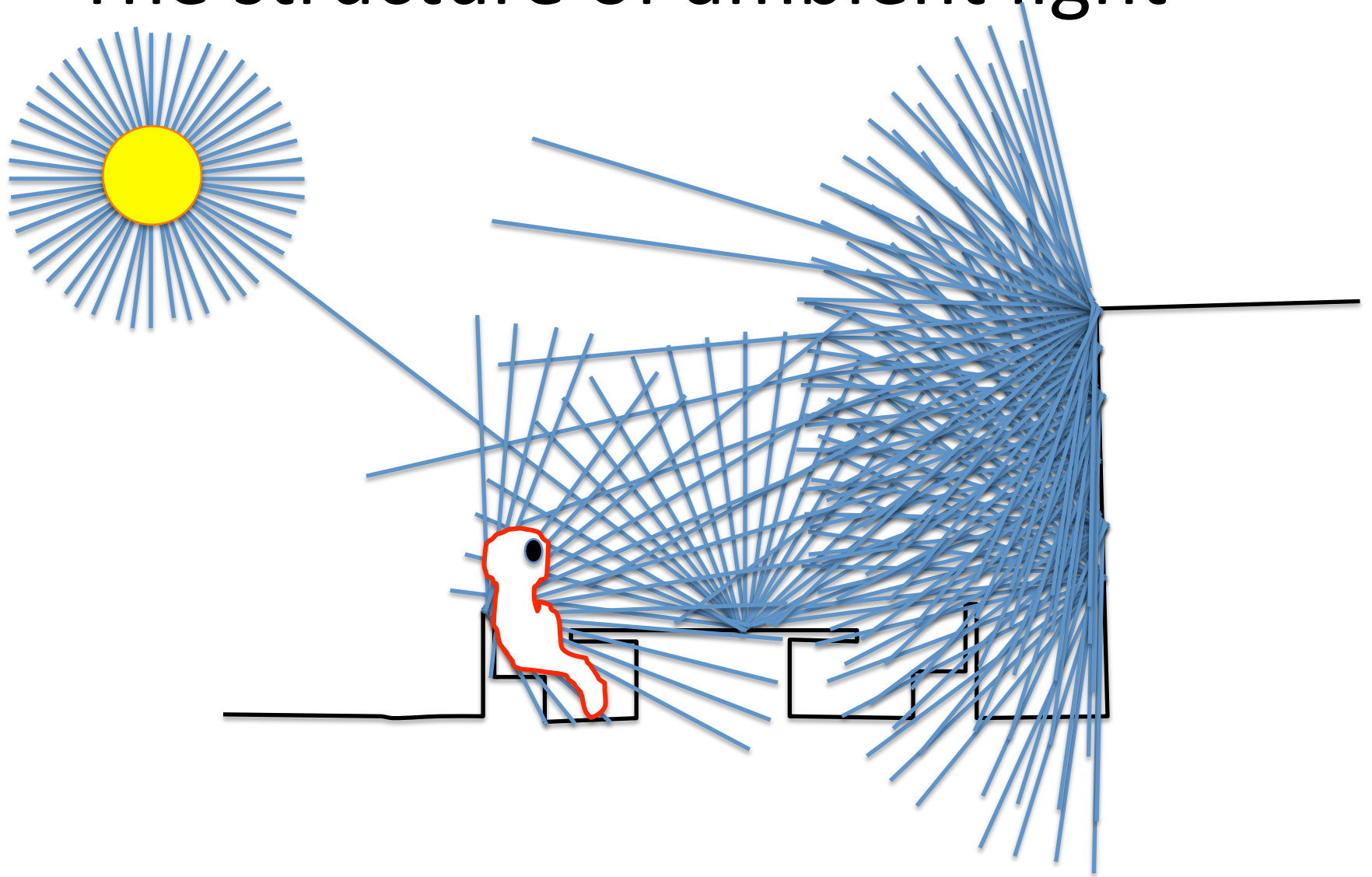
100 million \$

Why is vision hard?

The structure of ambient light

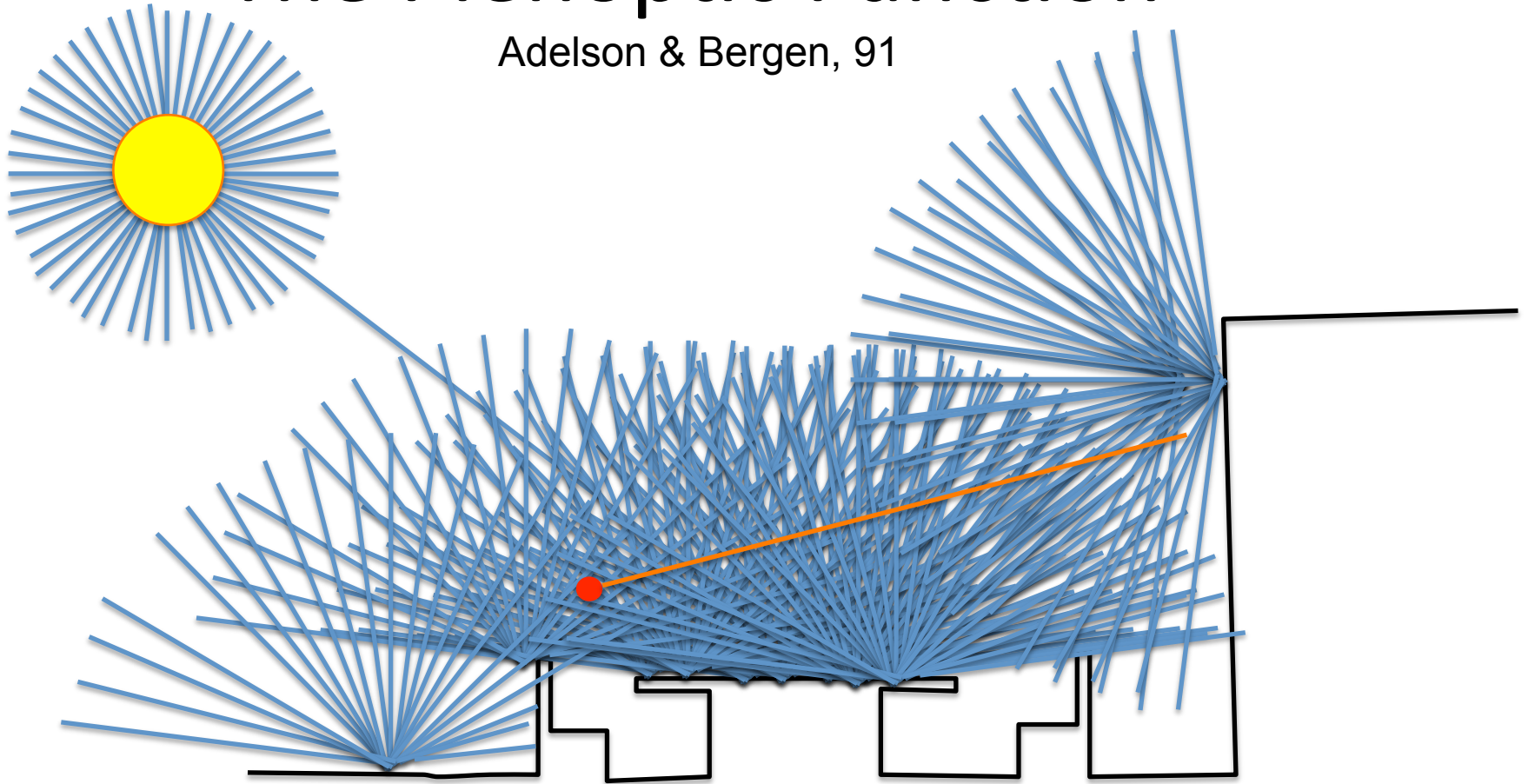


The structure of ambient light



The Plenoptic Function

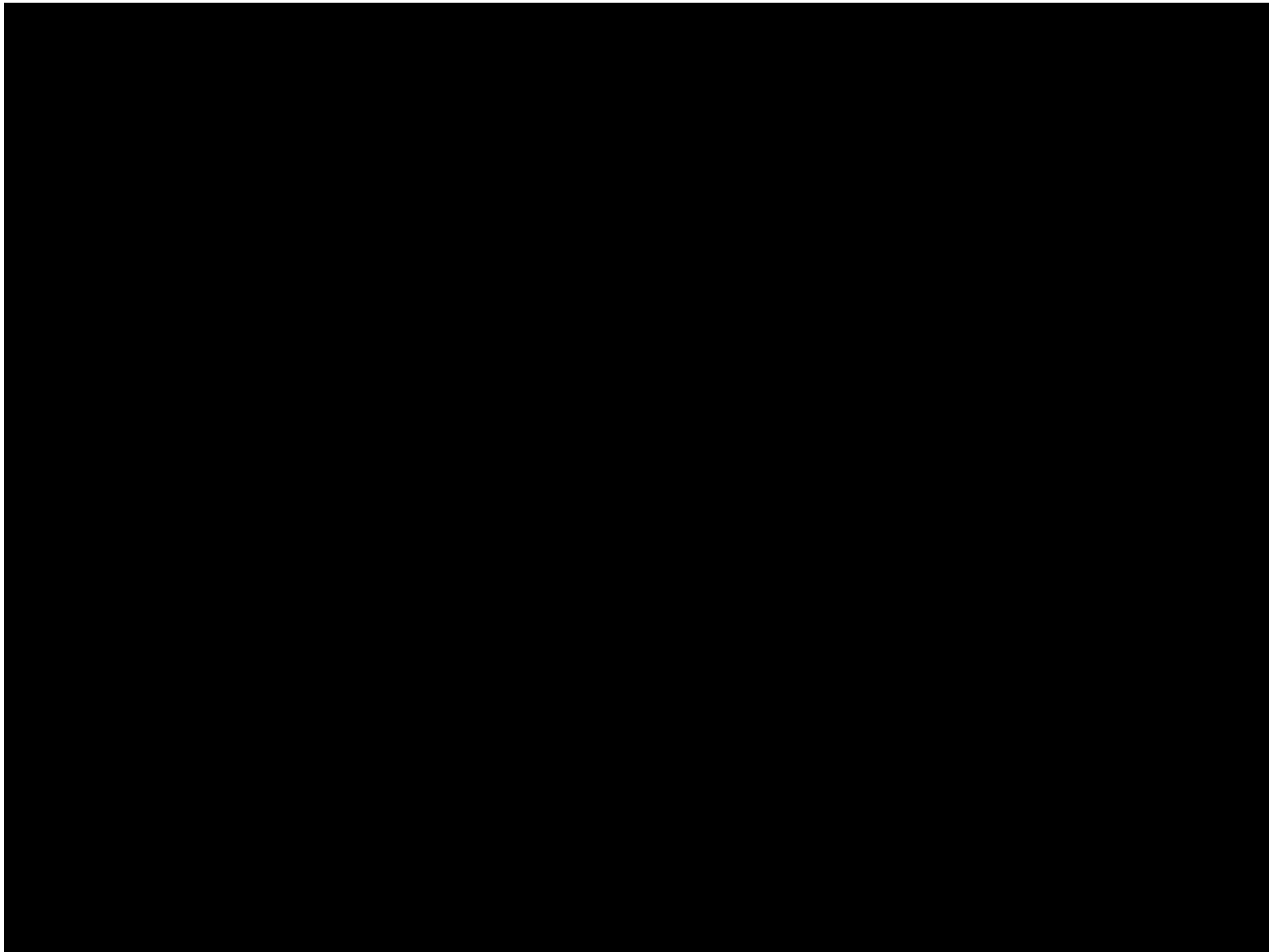
Adelson & Bergen, 91



The intensity P can be parameterized as:

$$P(\theta, \phi, \lambda, t, X, Y, Z)$$

“The complete set of all convergence points constitutes the permanent possibilities of vision.” Gibson









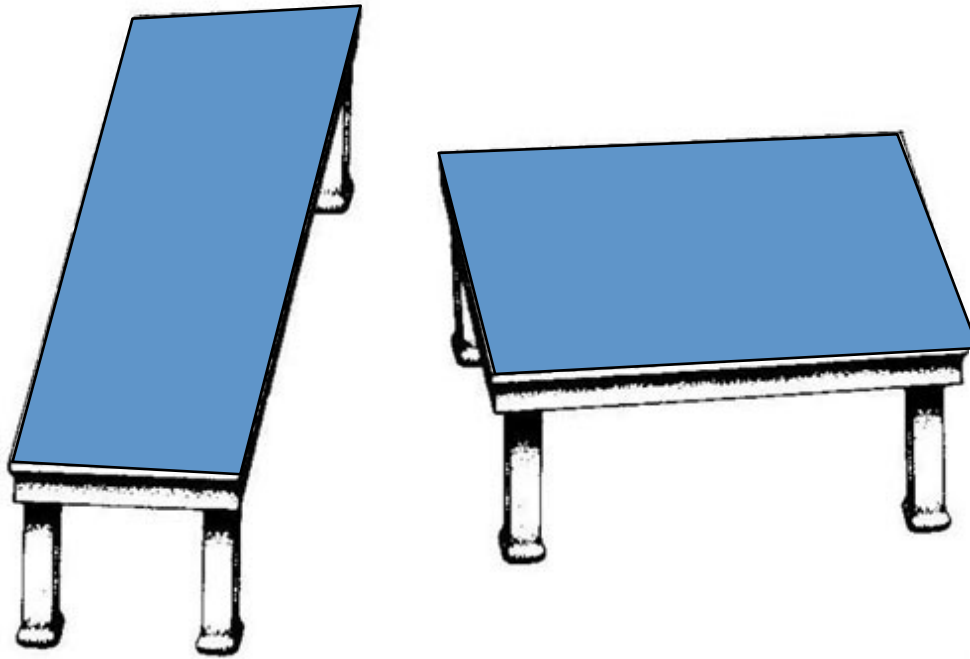
Why is vision hard?

Measuring light vs. measuring scene properties



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

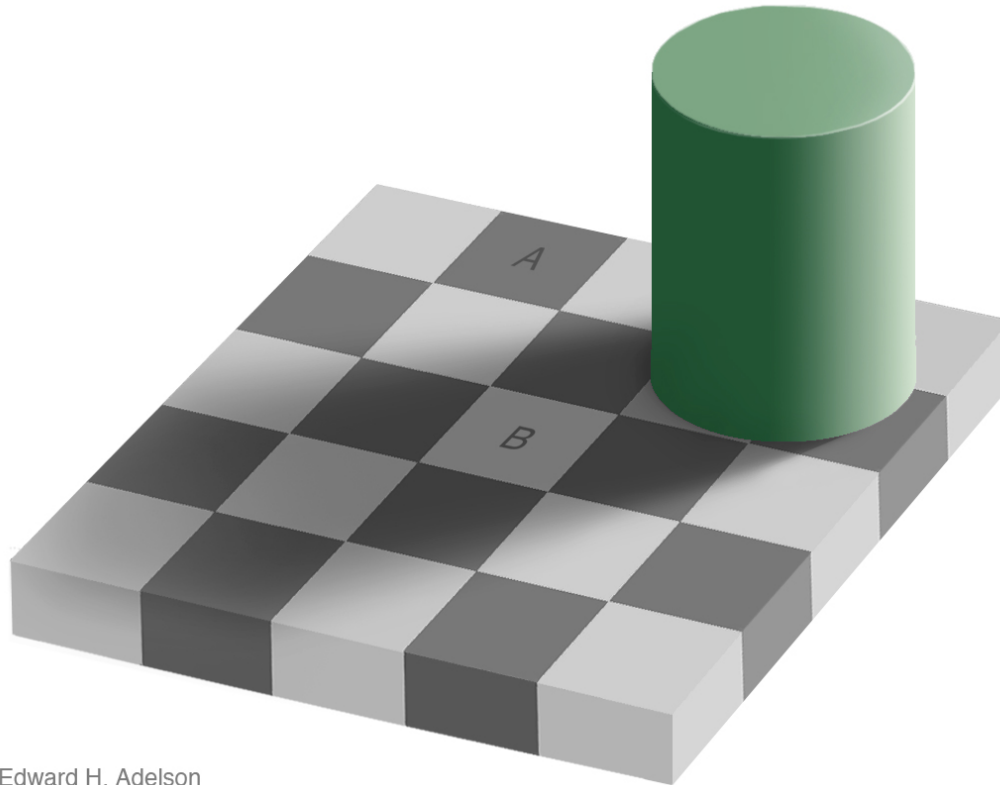
Measuring light vs. measuring scene properties



by Roger Shepard ("Turning the Tables")

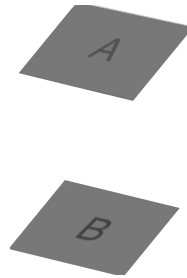
Depth processing is automatic, and we can not shut it down...

Measuring light vs. measuring scene properties

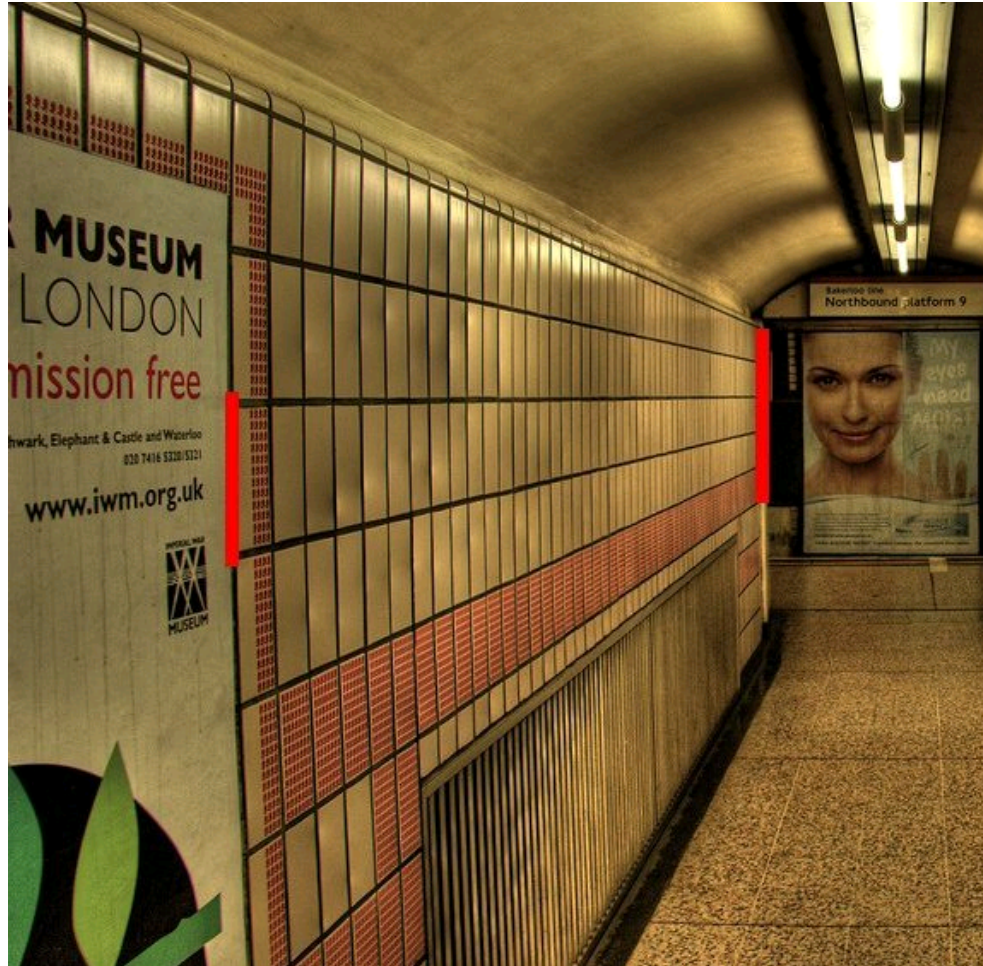


Edward H. Adelson

Measuring light vs. measuring scene properties

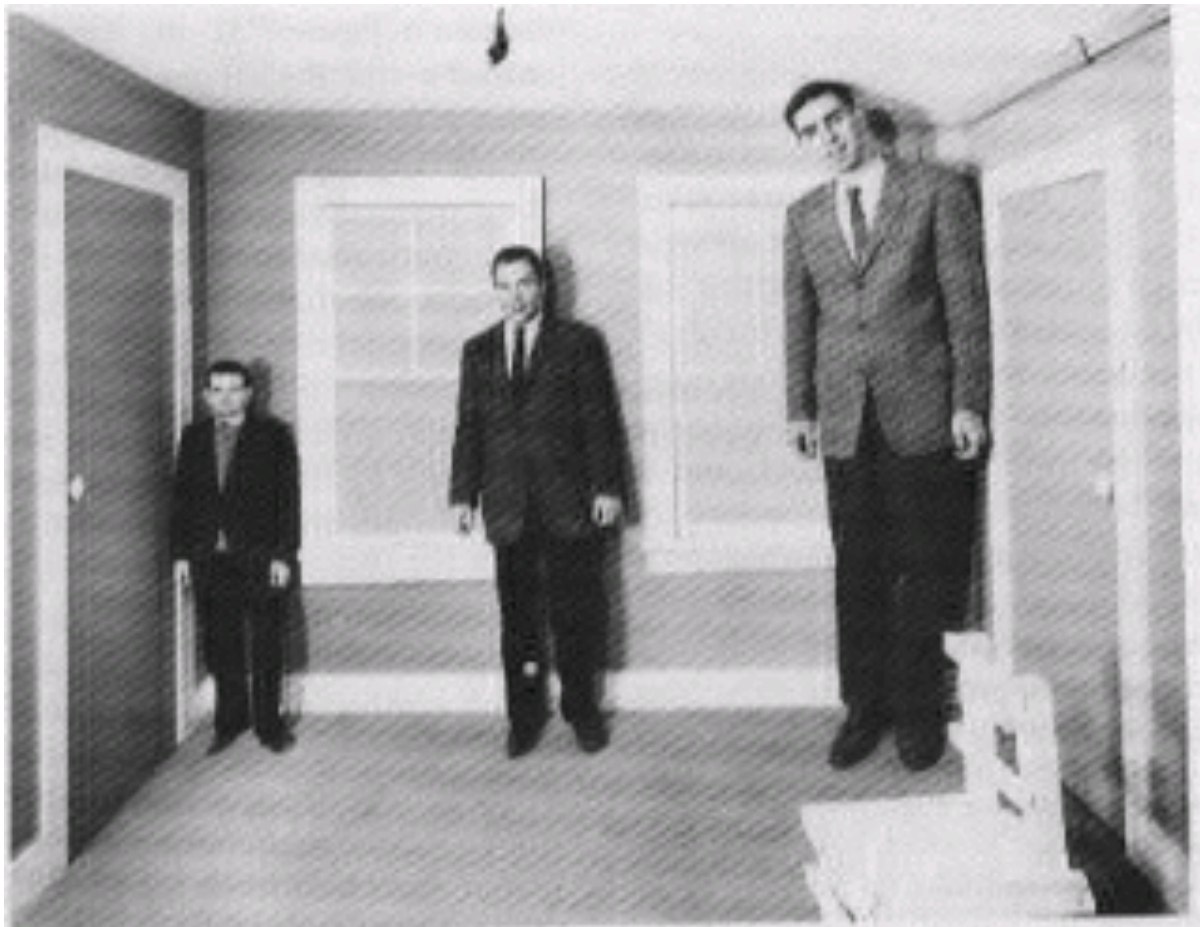


Measuring light vs. measuring scene properties



(c) 2006 Walt Anthony

Assumptions can be wrong



Ames room



By Aude Oliva

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
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".

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

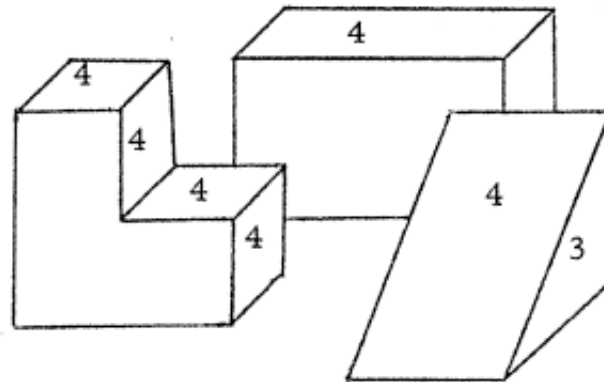
MACHINE PERCEPTION OF THREE-DIMENSIONAL SOLIDS

by

LAWRENCE GILMAN ROBERTS

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

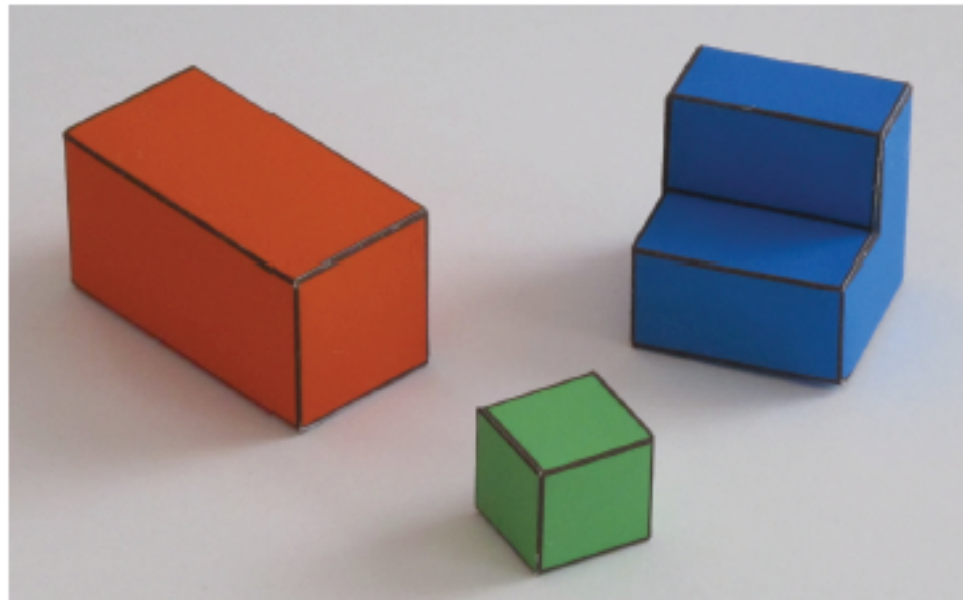
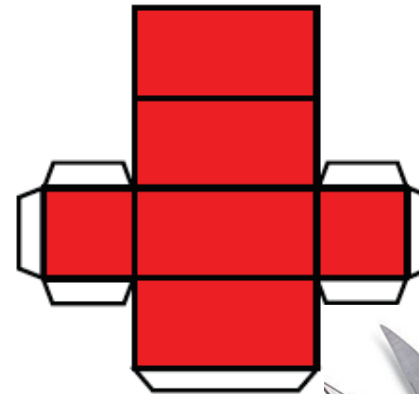
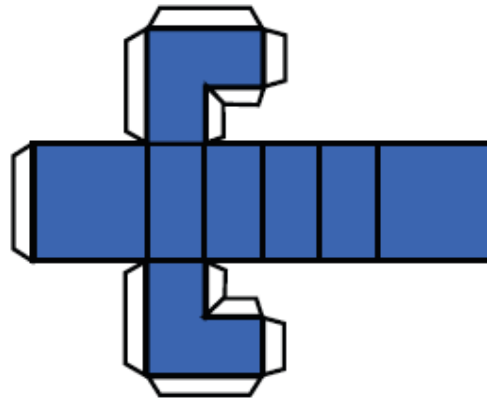
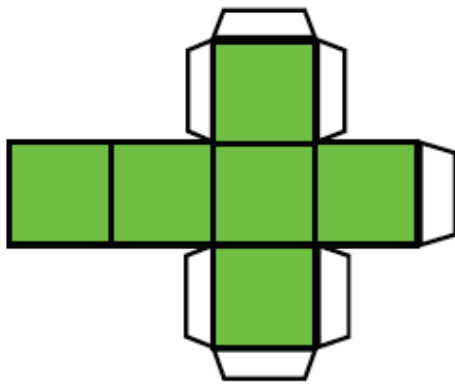
The problem of machine recognition of pictorial data has long been a challenging goal, but has seldom been attempted with anything more complex than alphabetic characters. Many people have felt that research on character recognition would be a first step, leading the way to a more general pattern recognition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

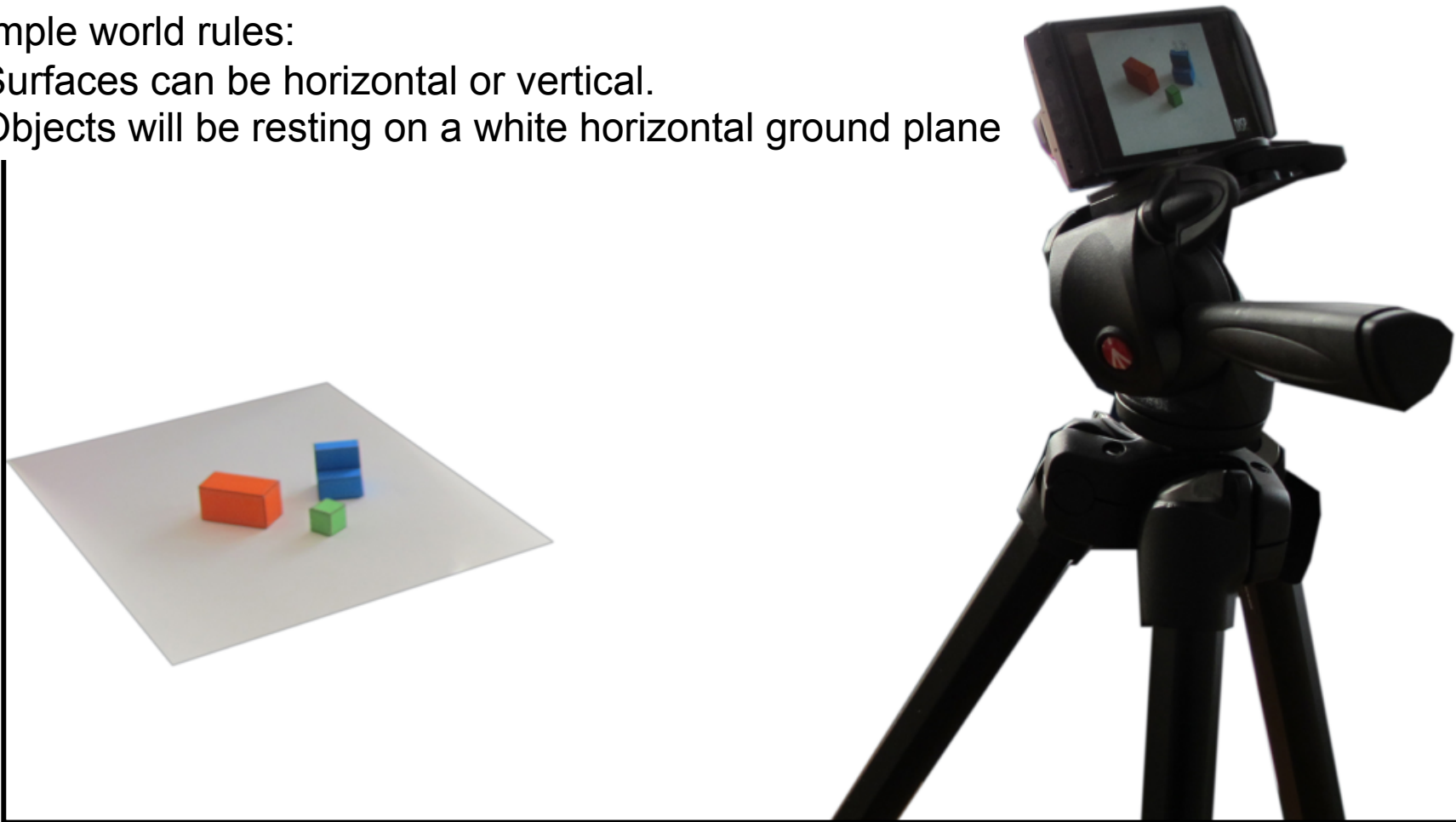
A Simple World



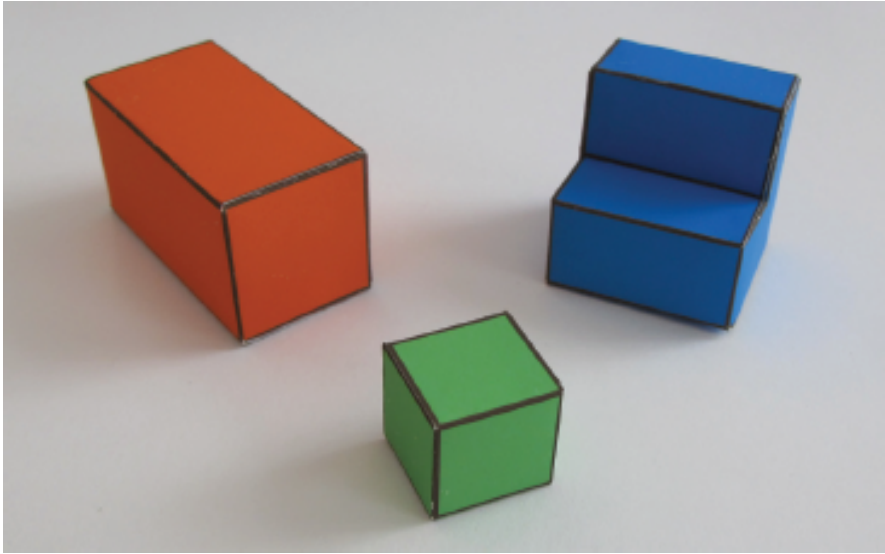
A simple image formation model

Simple world rules:

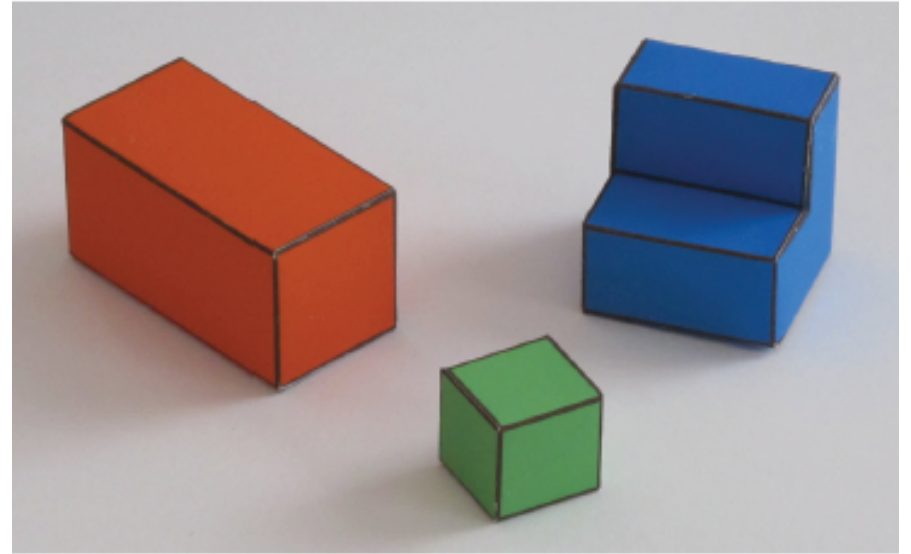
- Surfaces can be horizontal or vertical.
- Objects will be resting on a white horizontal ground plane



A simple image formation model

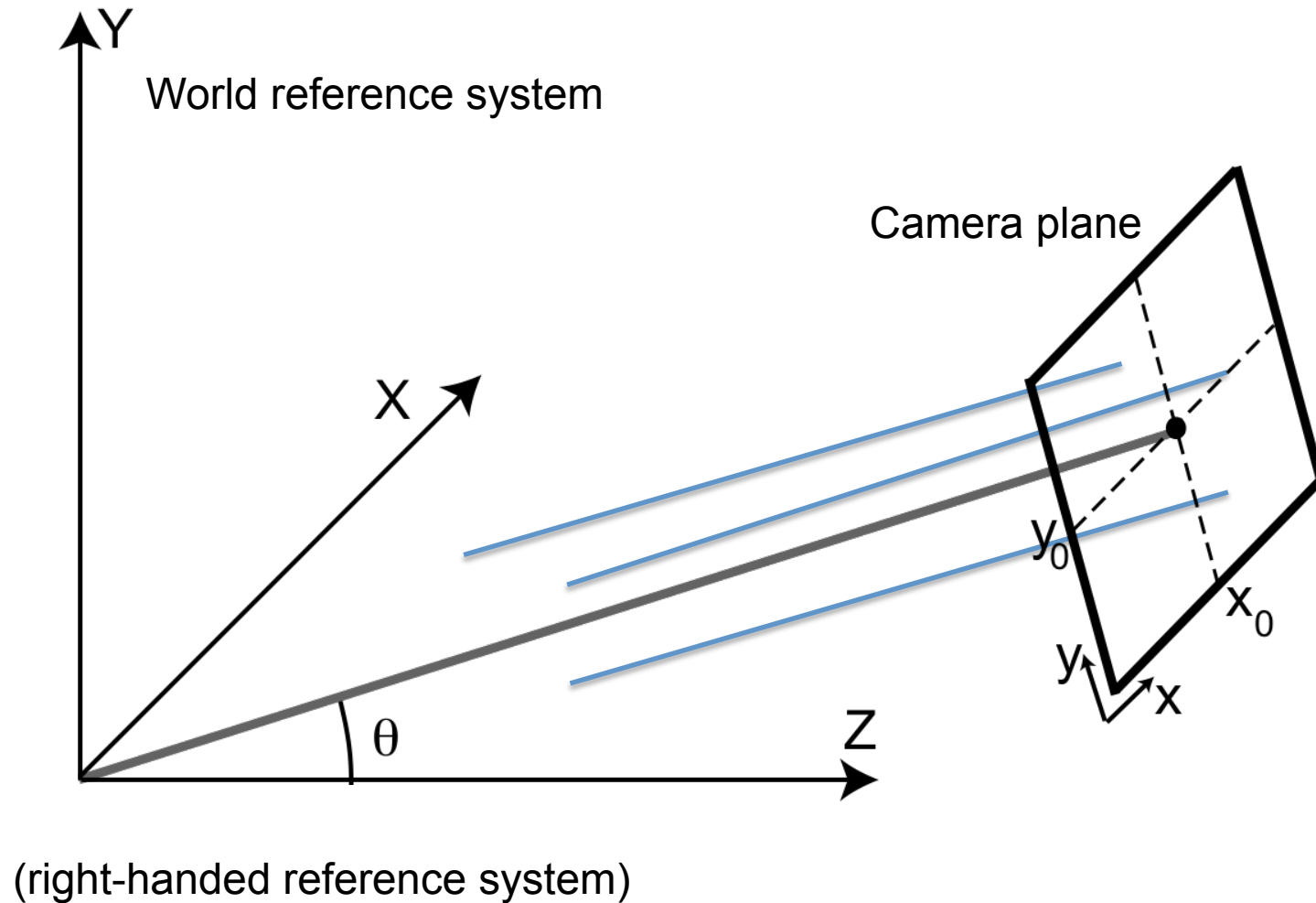


Perspective projection



Parallel (orthographic) projection

A simple image formation model



A simple image formation model

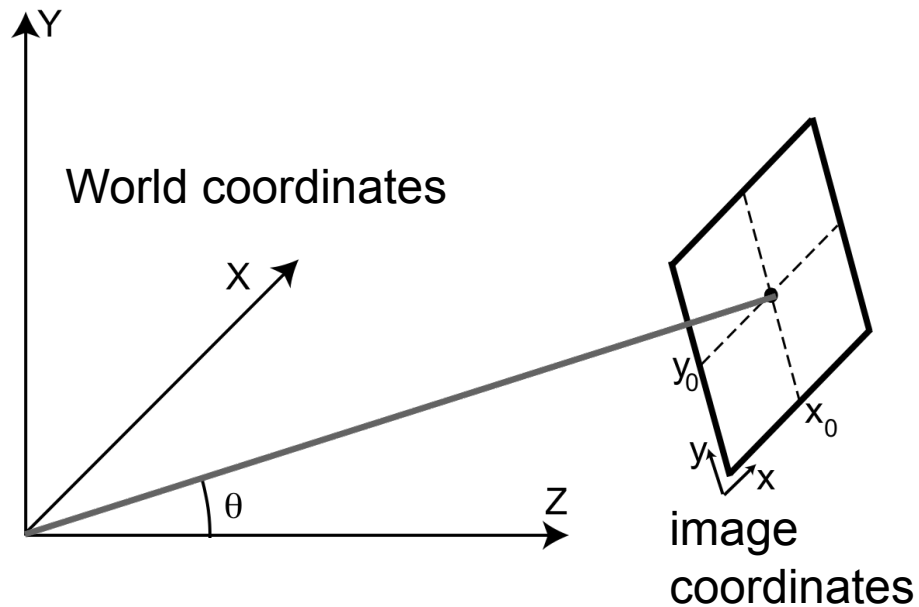
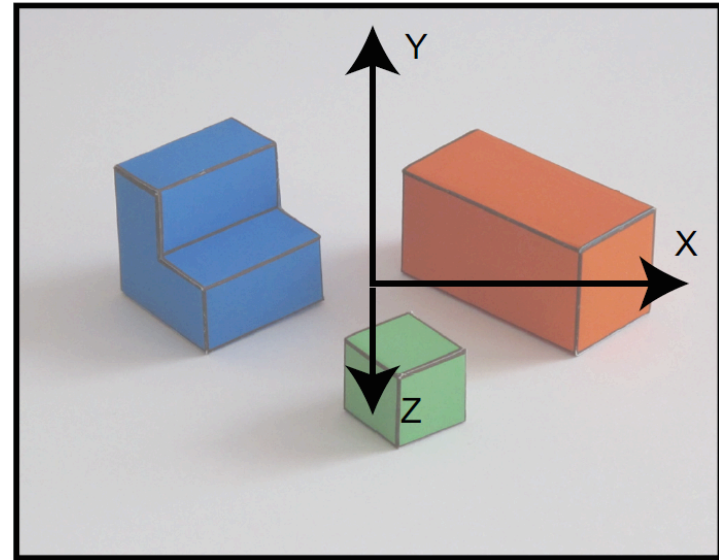


Image and projection of the world coordinate axes into the image plane



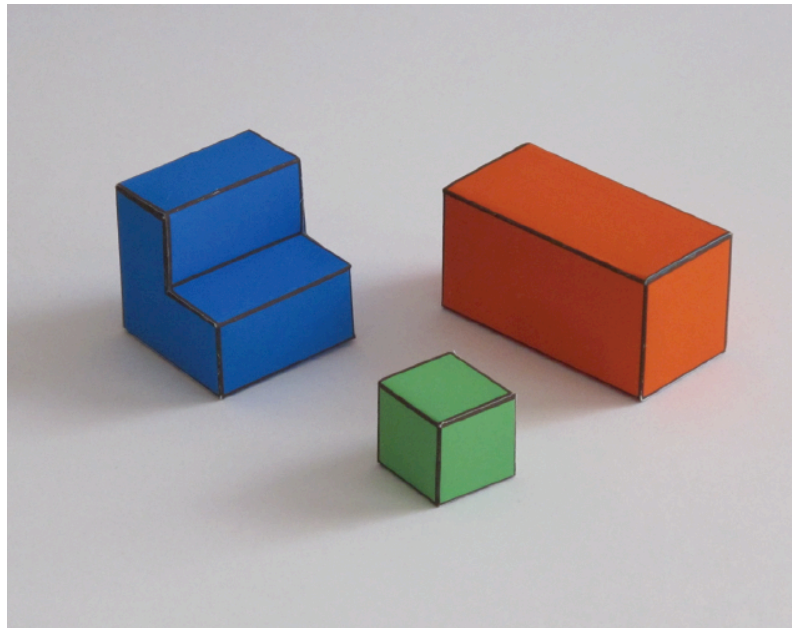
World coordinates

$$\begin{aligned} x &= X + x_0 \\ y &= \cos(\theta) Y - \sin(\theta) Z + y_0 \end{aligned}$$

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?

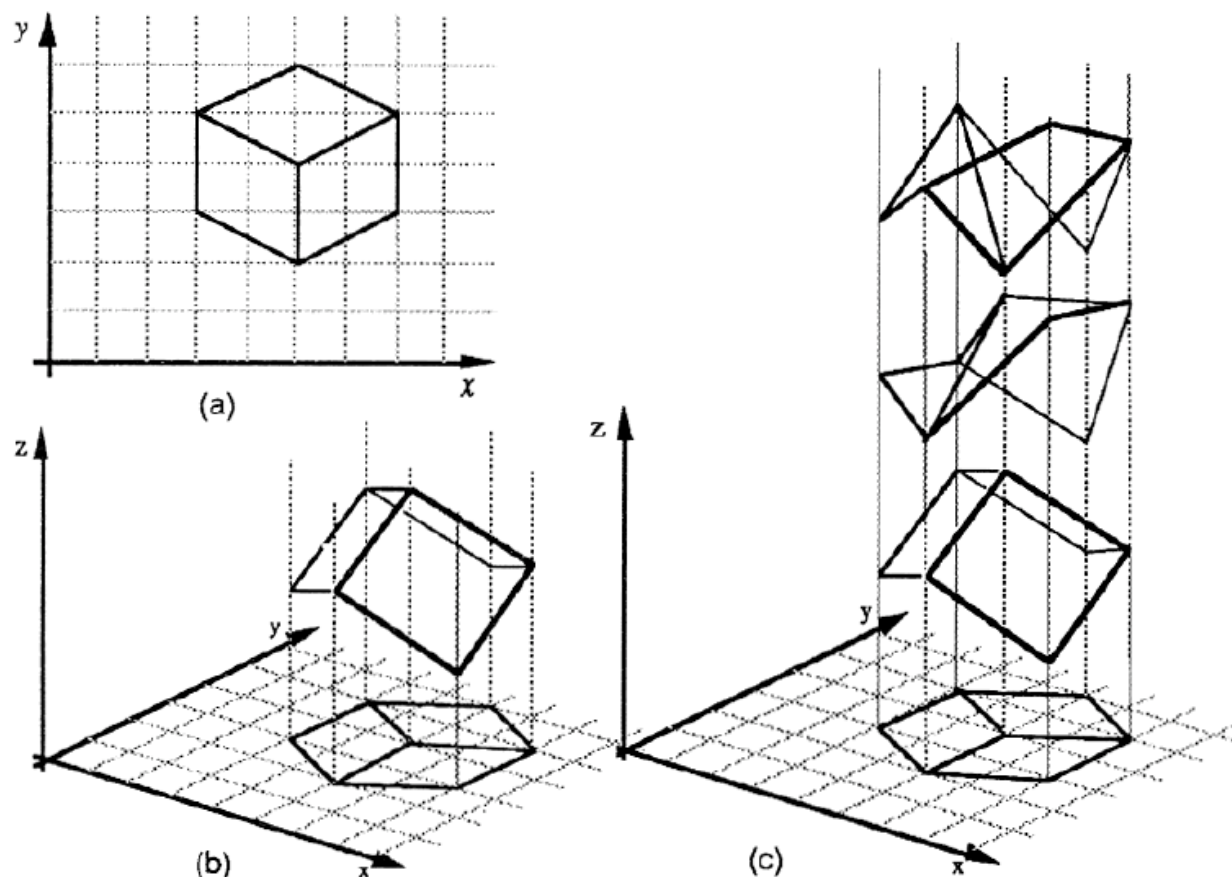
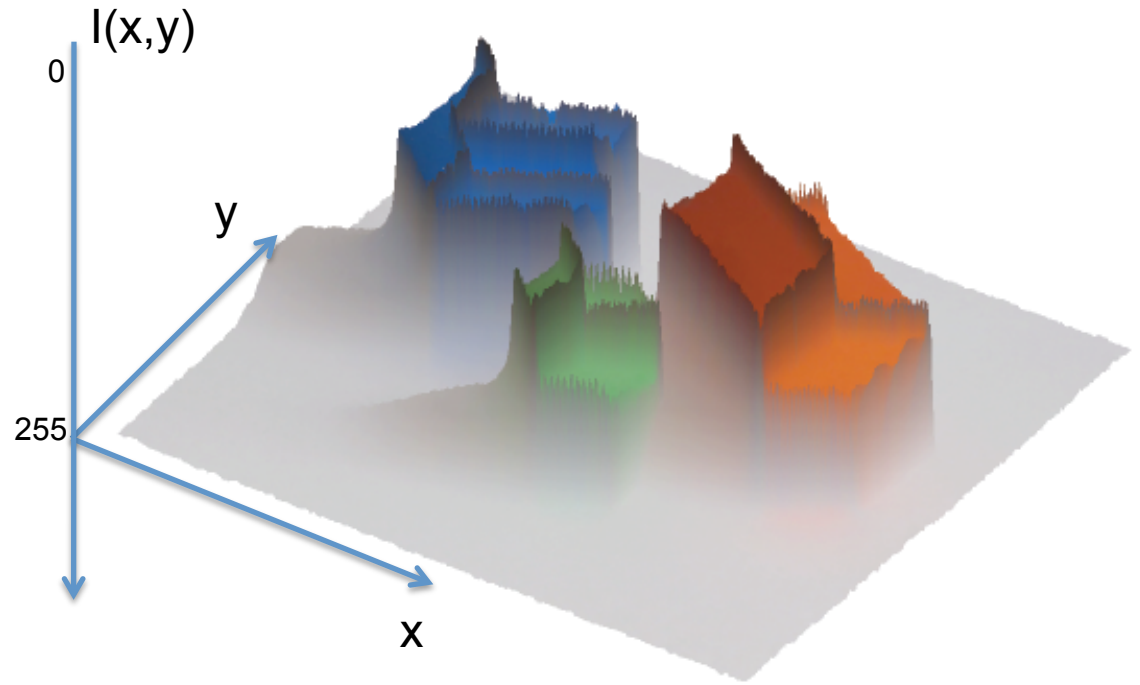
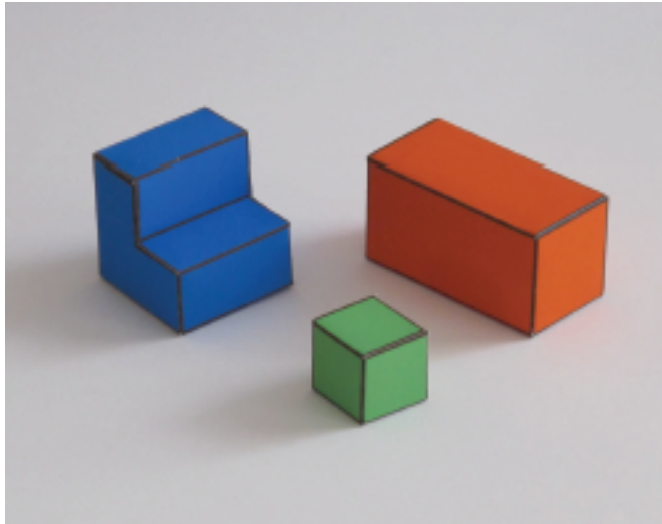


Figure 1. (a) A line drawing provides information only about the x, y coordinates of points lying along the object contours. (b) The human visual system is usually able to reconstruct an object in three dimensions given only a single 2D projection (c) Any planar line-drawing is geometrically consistent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

A simple visual system

The input image



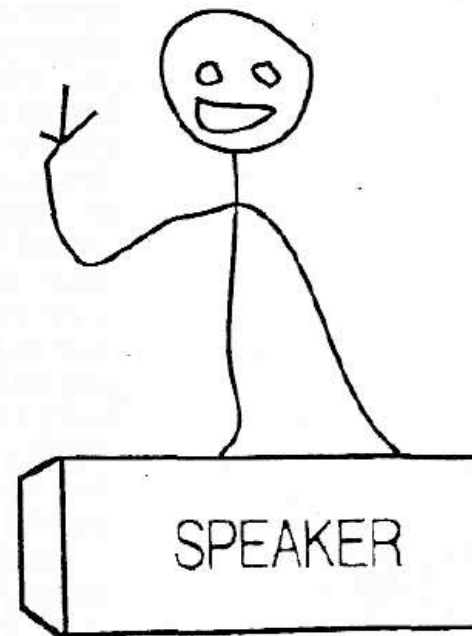
- Proposition 1. The primary task of early vision is to deliver a small set of useful measurements about each observable location in the plenoptic function.
- Proposition 2. The elemental operations of early vision involve the measurement of local change along various directions within the plenoptic function.

Adelson, Bergen. 91

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



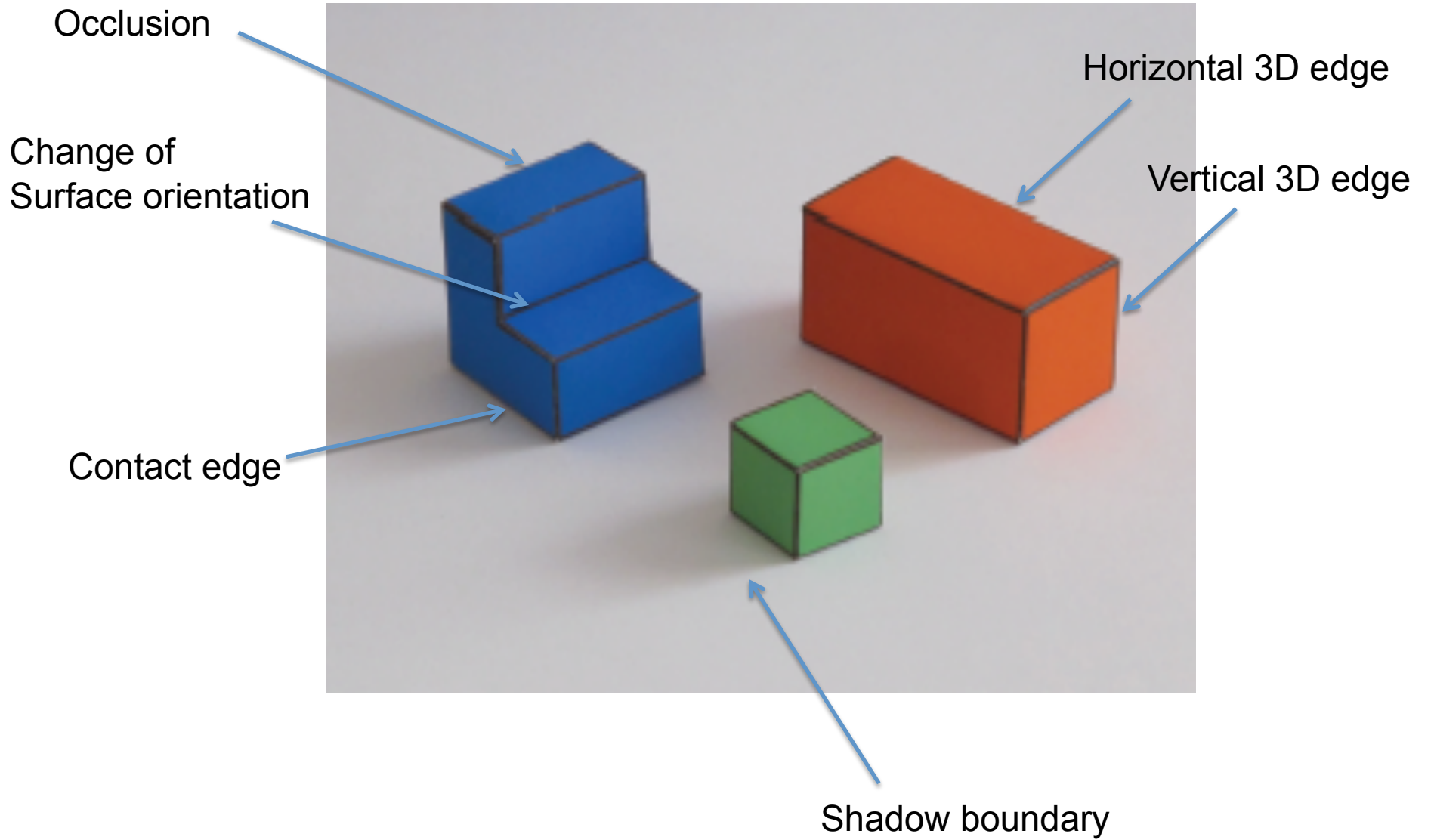
What we think we see



What we really see

Cavanagh, Perception 95

Edges



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

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

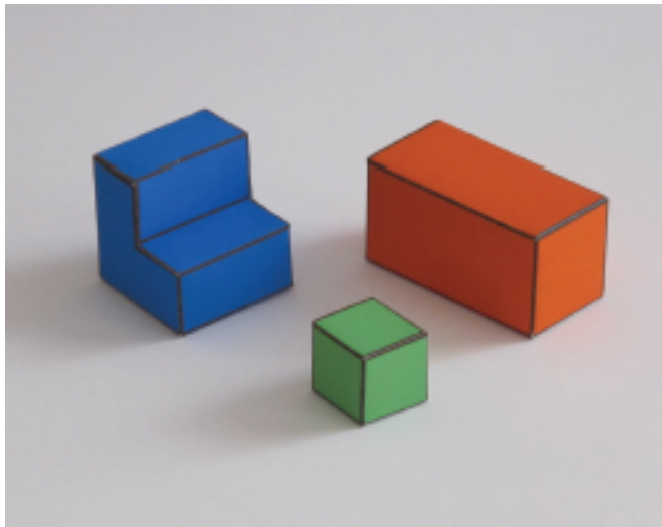
Edge orientation:

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

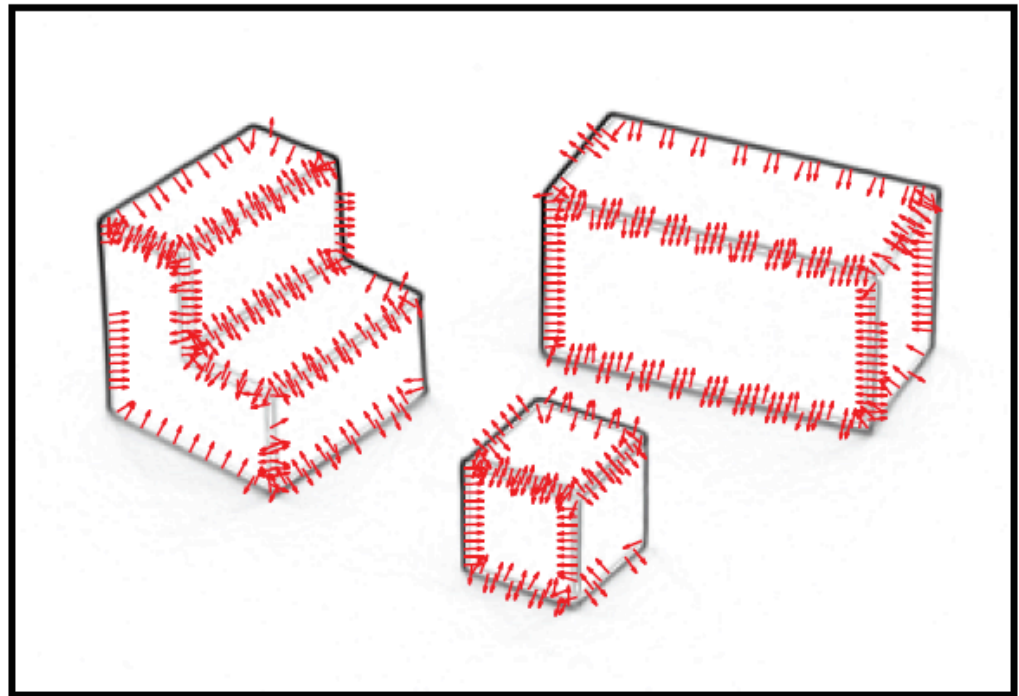
Edge normal:

$$\mathbf{n} = \frac{\nabla \mathbf{I}}{|\nabla \mathbf{I}|}$$

Finding edges in the image

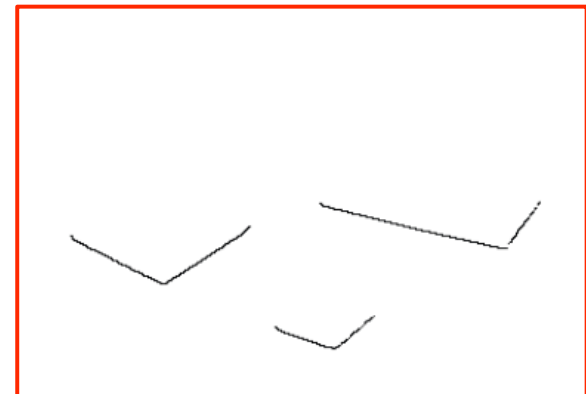
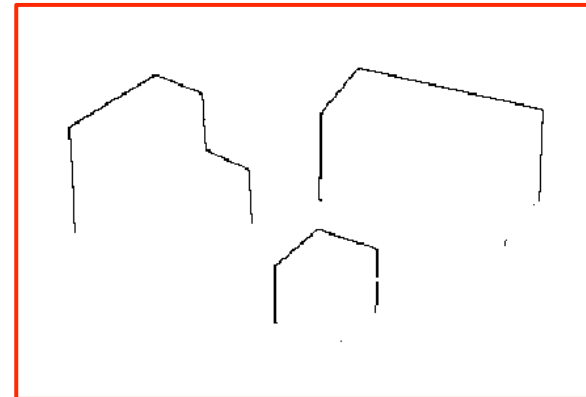
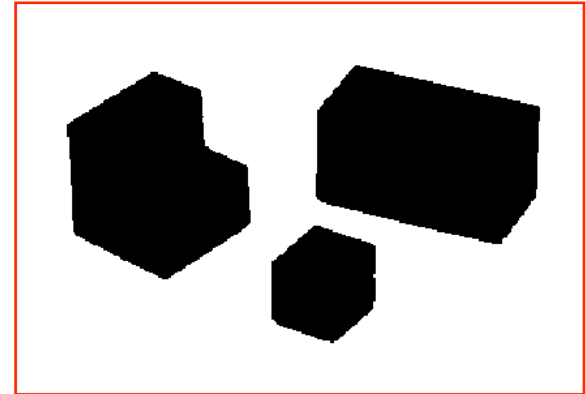


$$\nabla I = \left(\frac{\partial I}{\partial x}, \frac{\partial I}{\partial y} \right) \quad \mathbf{n} = \frac{\nabla I}{|\nabla I|}$$



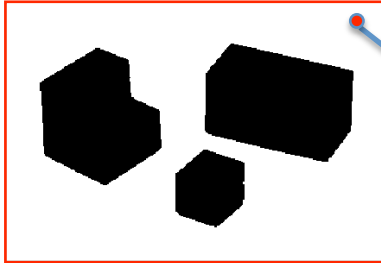
Edge classification

- Figure/ground segmentation
- Occlusion edges
 - Occlusion edges are owned by the foreground
- Contact edges



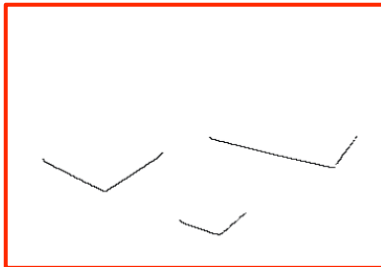
From edges to surface constraints

- Ground



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

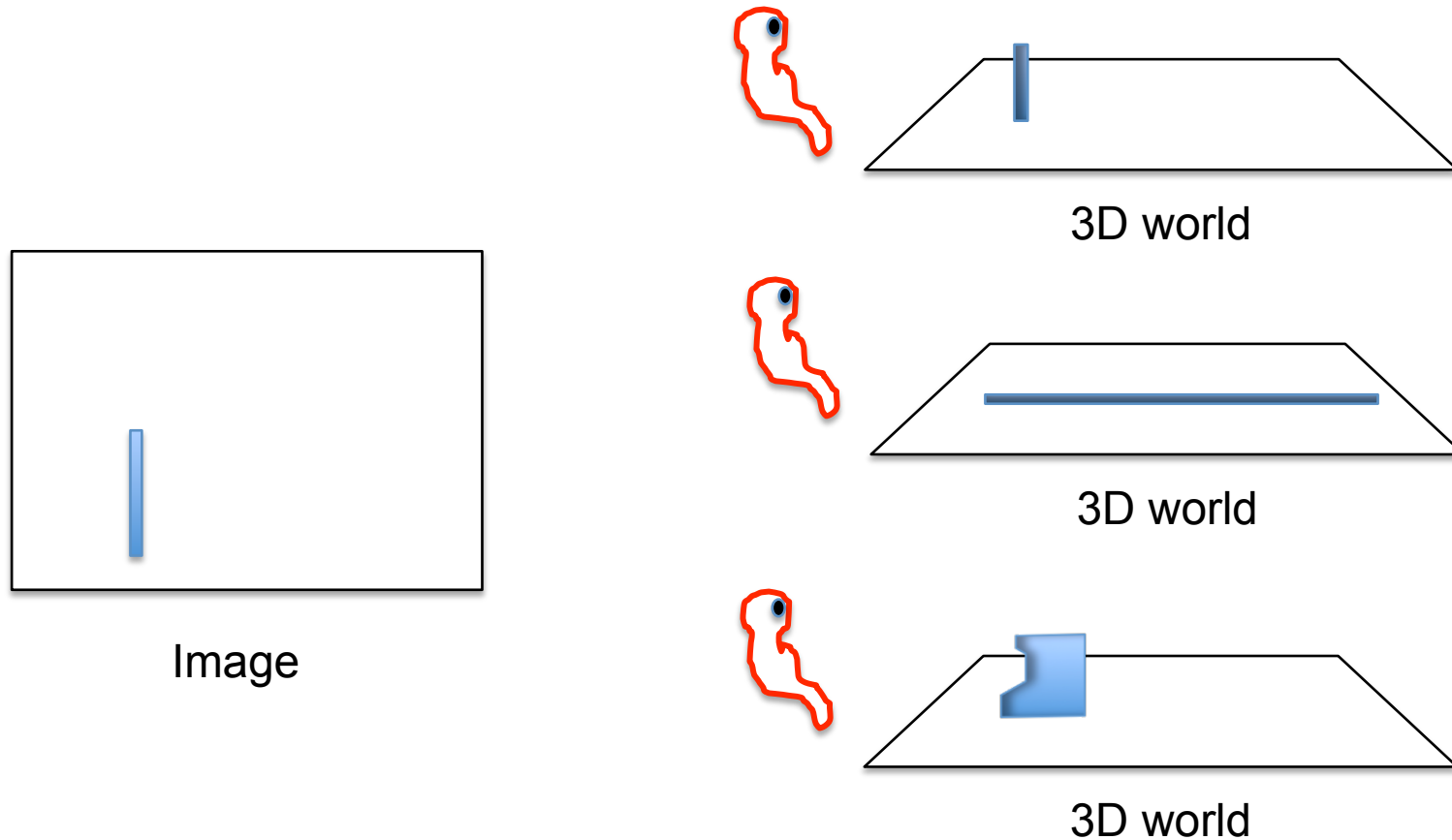
- Contact edge



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

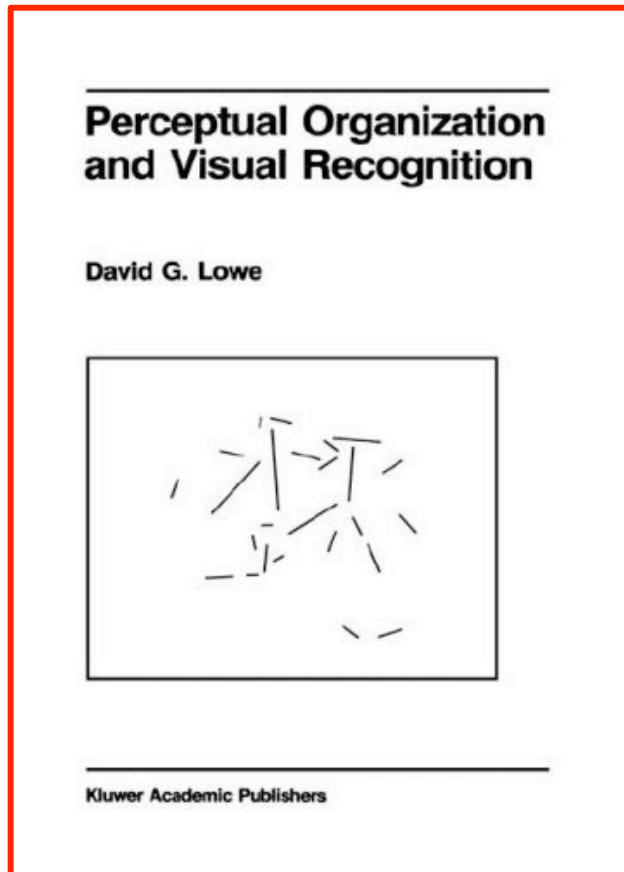
Next, things get a bit more complicated.

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.

Non-accidental properties



D. Lowe, 1985

Principle of Non-Accidentalness: Critical information is unlikely to be a consequence of an accident of viewpoint.

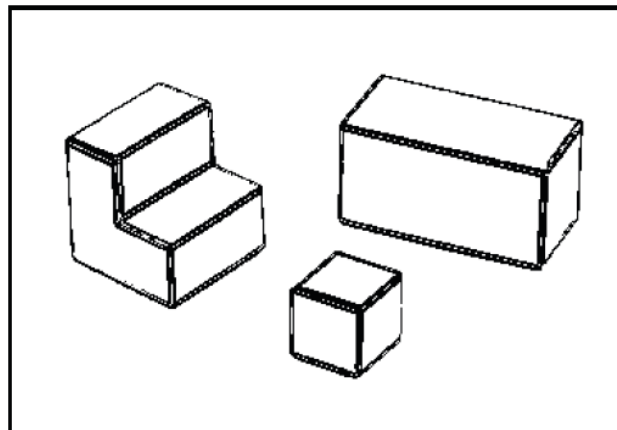
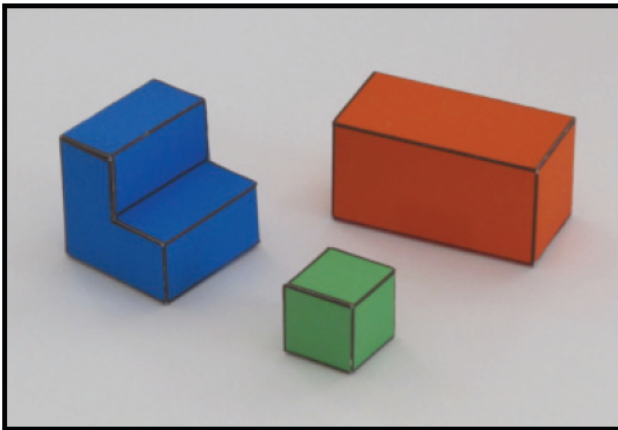
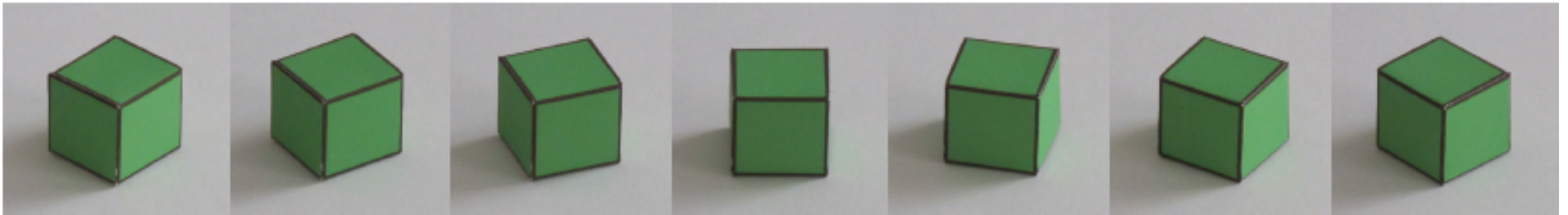
Three Space Inference from Image Features

<u>2-D Relation</u>	<u>3-D Inference</u>	<u>Examples</u>
1. Collinearity of points or lines	Collinearity in 3-Space	
2. Curvilinearity of points or arcs	Curvilinearity in 3-Space	
3. Symmetry (Skew Symmetry?)	Symmetry in 3-Space	
4. Parallel Curves (Over Small Visual Angles)	Curves are parallel in 3-Space	
5. Vertices—two or more terminations at a common point	Curves terminate at a common point in 3-Space	

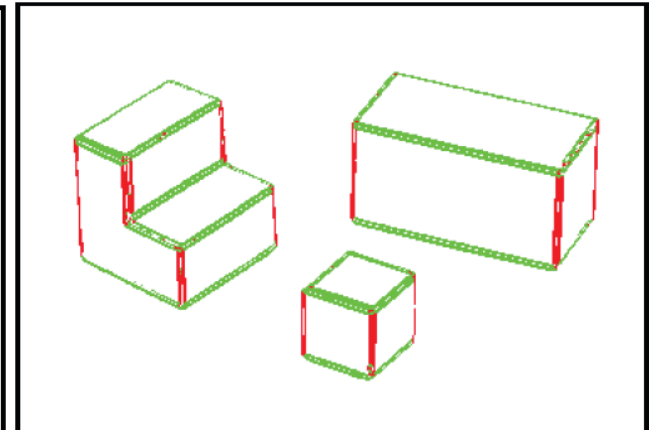
Figure 4. Five nonaccidental relations. (From Figure 5.2, *Perceptual organization and visual recognition* [p. 77] by David Lowe. Unpublished doctoral dissertation, Stanford University. Adapted by permission.)

Biederman_RBC_1987

Non-accidental properties in the simple world



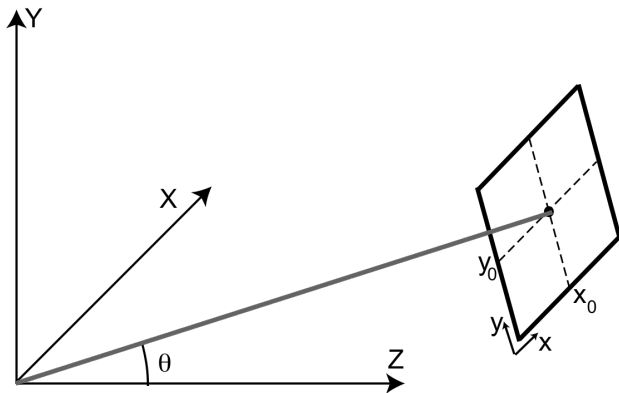
Using $E(x,y)$



Using $\theta(x,y)$

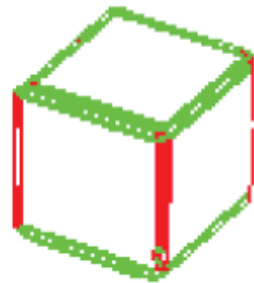
From edges to surface constraints

- Vertical edges



$$x = X + x_0$$

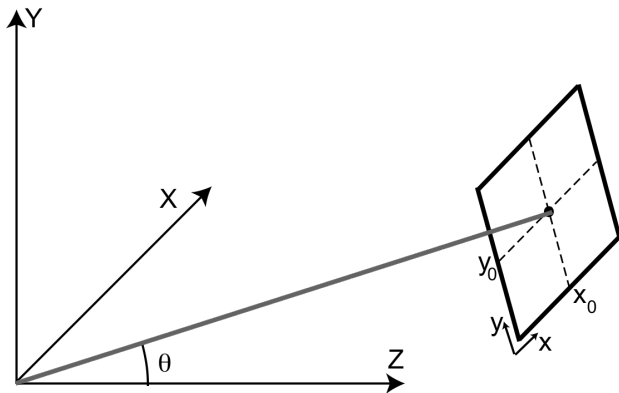
$$y = \cos(\theta) Y - \sin(\theta) Z + y_0$$



$$\rightarrow \begin{cases} Z = \text{constant along the edge} \\ \partial Y / \partial y = 1 / \cos(\theta) \end{cases}$$

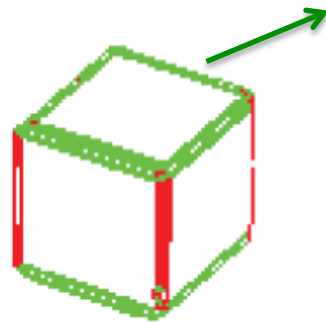
From edges to surface constraints

- Horizontal edges



$$x = X + x_0$$

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



$$\begin{cases} Y = \text{constant along the edge} \\ \partial Y / \partial \mathbf{t} = 0 \end{cases}$$

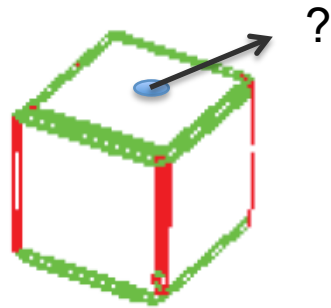
Where \mathbf{t} is the vector parallel to the edge

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

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

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 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

$$\partial Y / \partial t = 0$$

if (x,y) belongs to an horizontal edge

$$\partial^2 Y / \partial x^2 = 0$$

$$\partial^2 Y / \partial y^2 = 0$$

$$\partial^2 Y / \partial y \partial x = 0$$

if (x,y) is not an edge

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
----	---

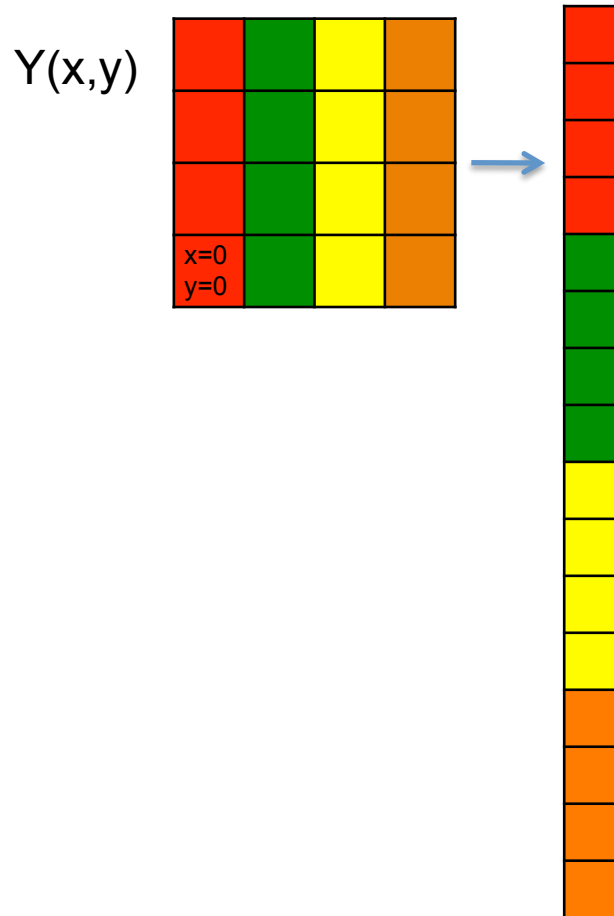
A slightly better approximation

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

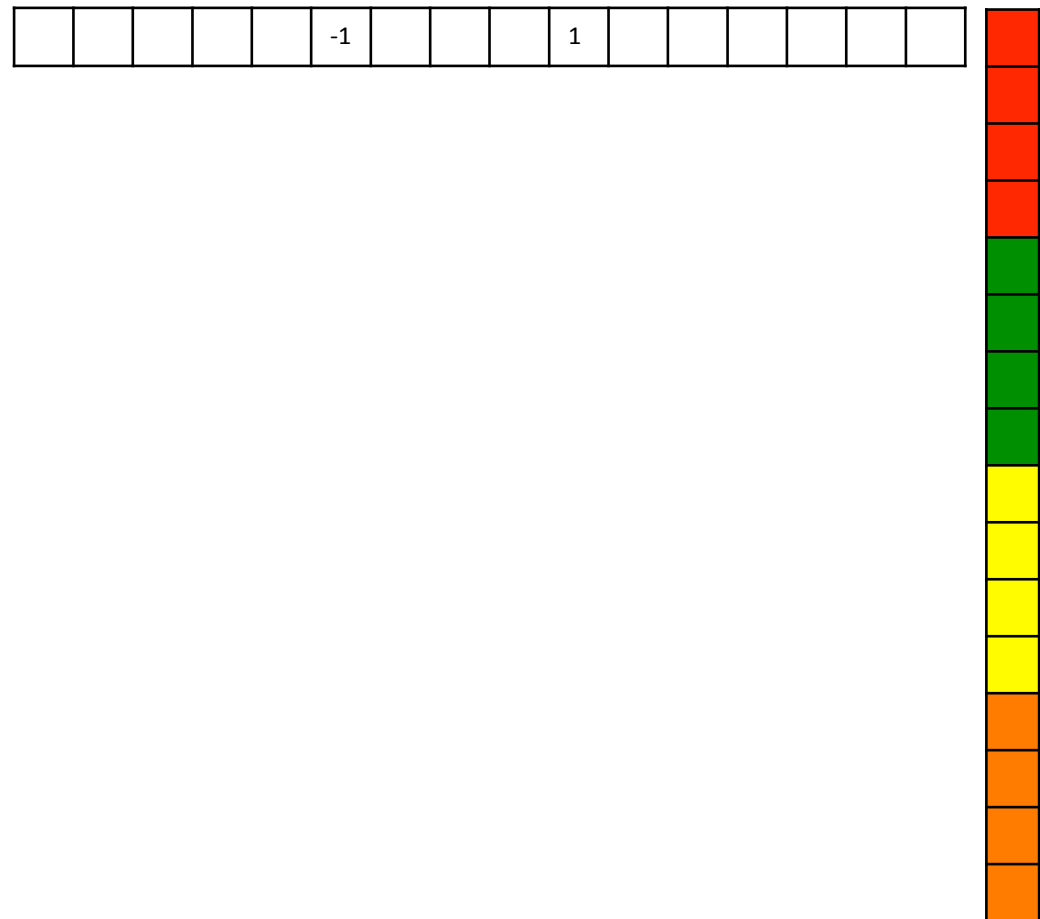
-1	0	1
-2	0	2
-1	0	1

Discrete approximation

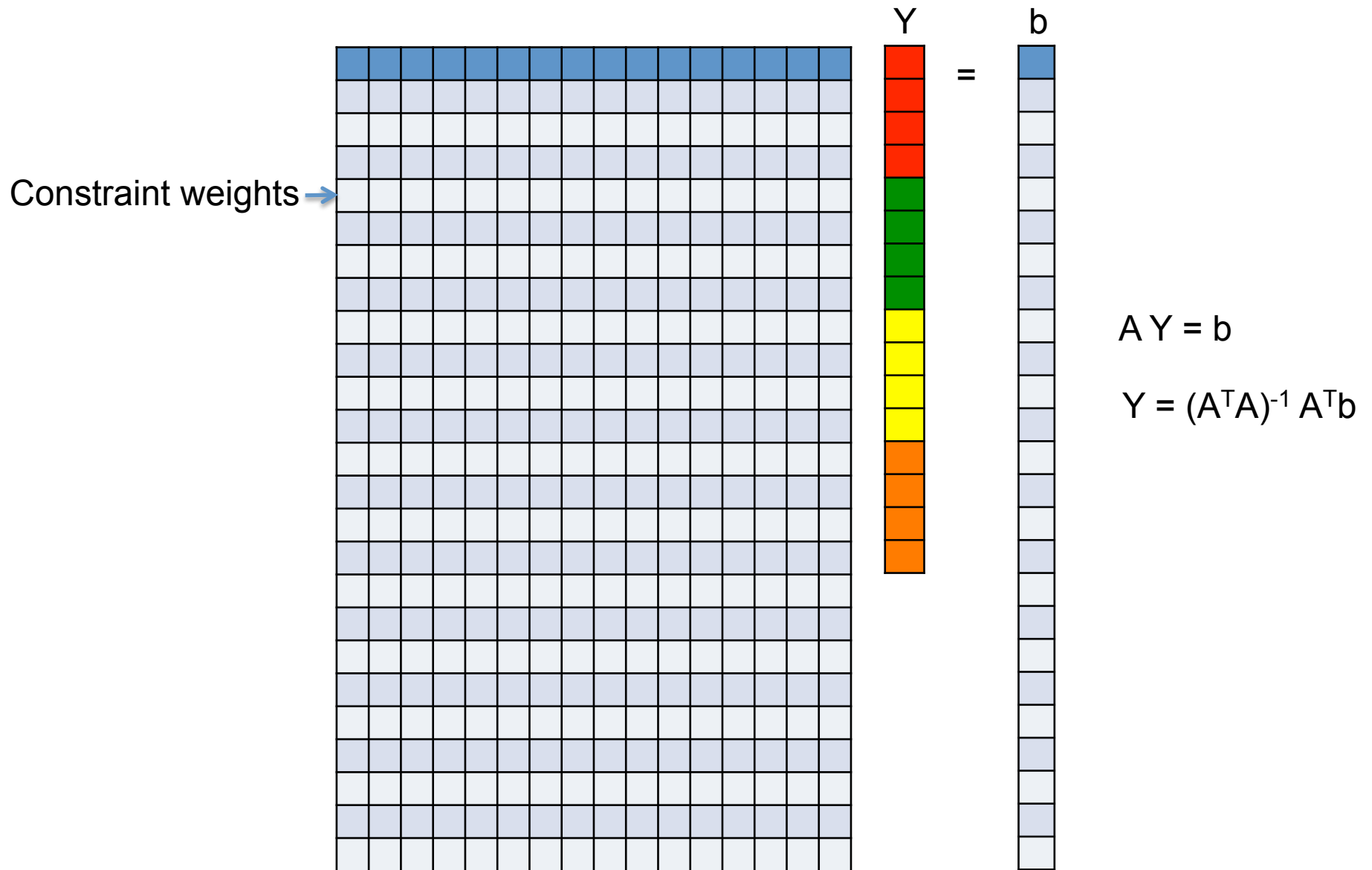
Transform the “image” $Y(x,y)$ into a column vector:



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

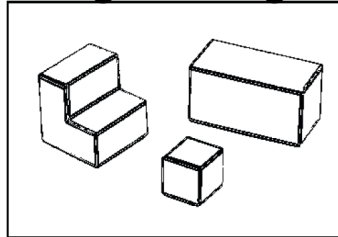


A simple inference scheme

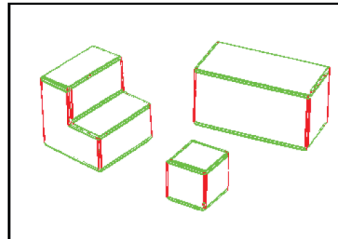


Results

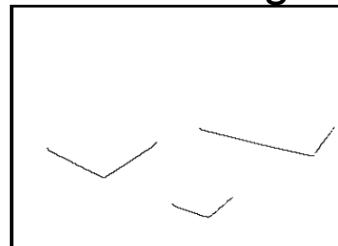
Edge strength



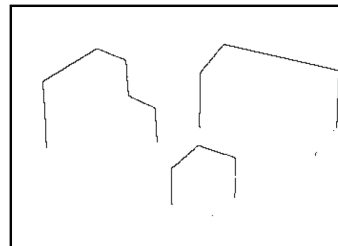
3D orientation



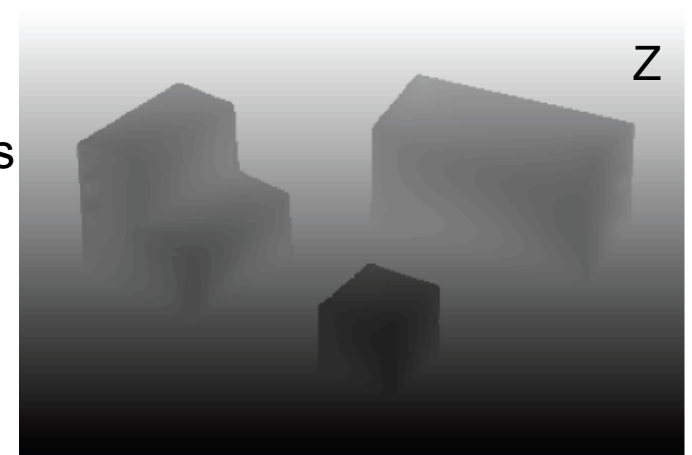
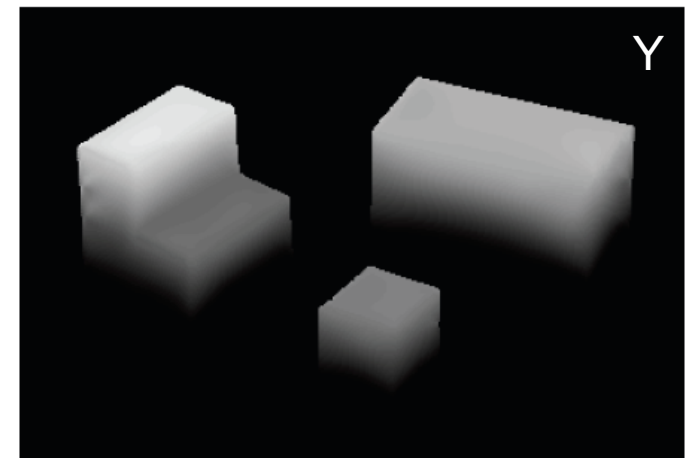
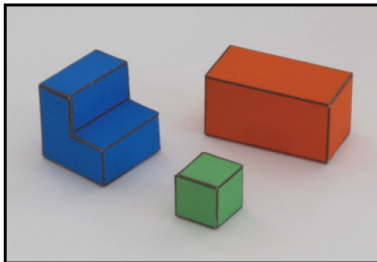
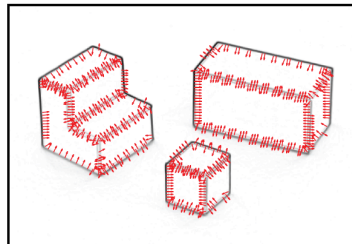
Contact edges



Depth discontinuities

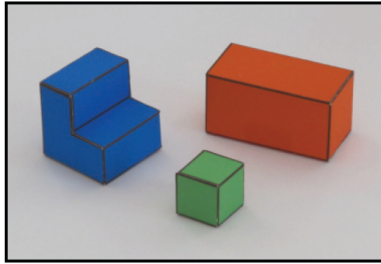


Edge normals

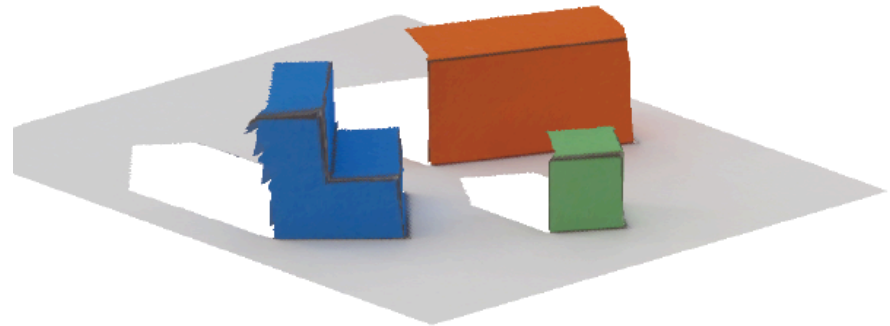
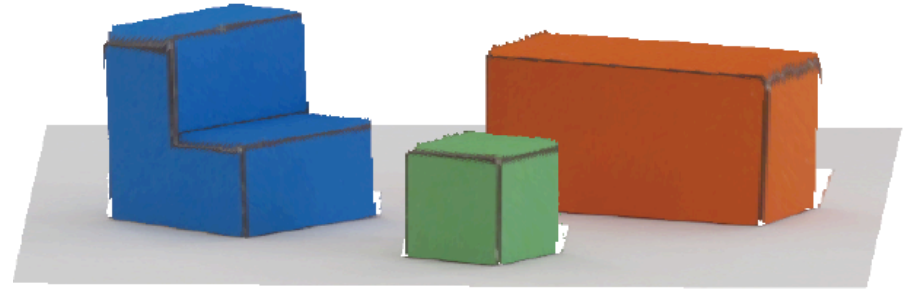
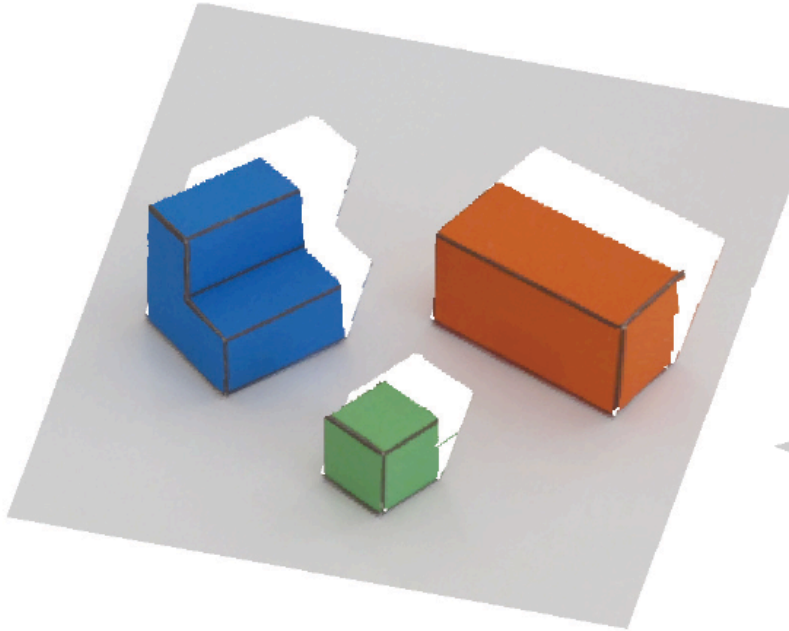


Changing view point

Input

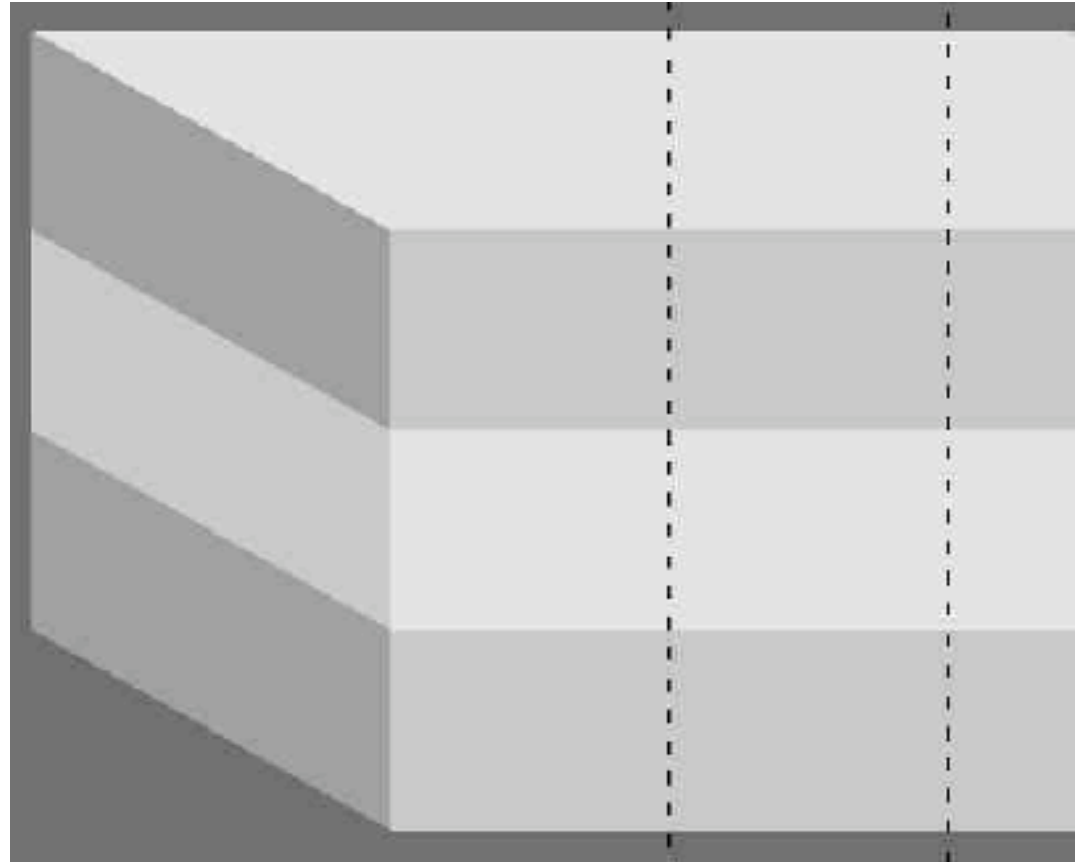


New view points:

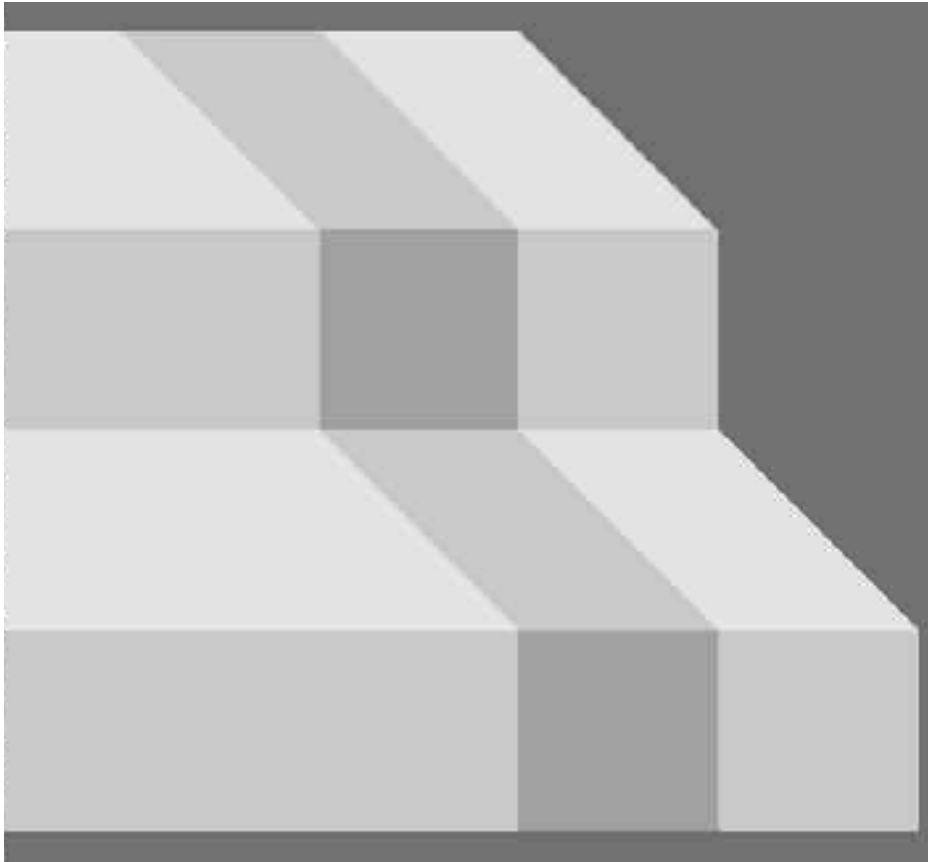


Violations of simple world assumptions

Shading is due to painted stripes



Violations of simple world assumptions



Shading is due to illumination

Violations of simple world assumptions

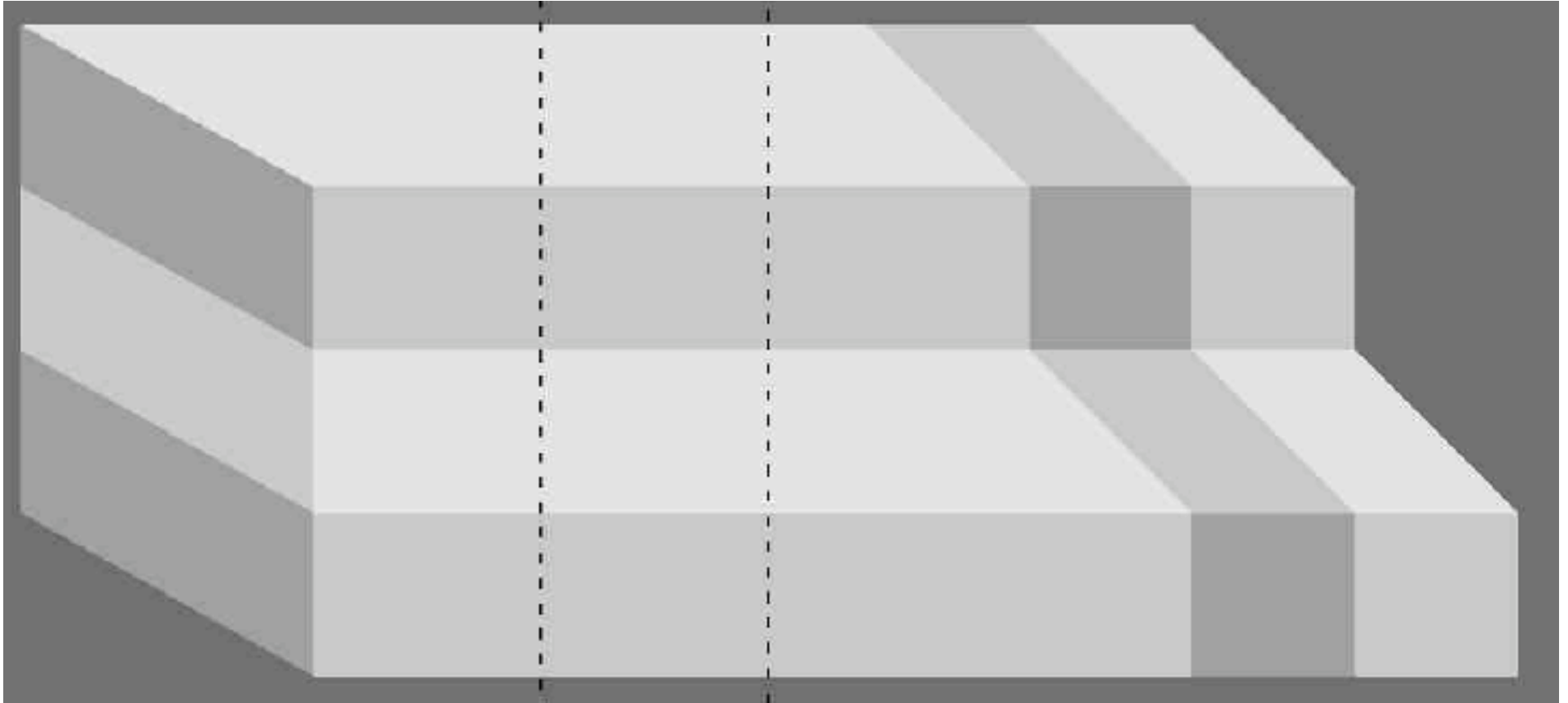


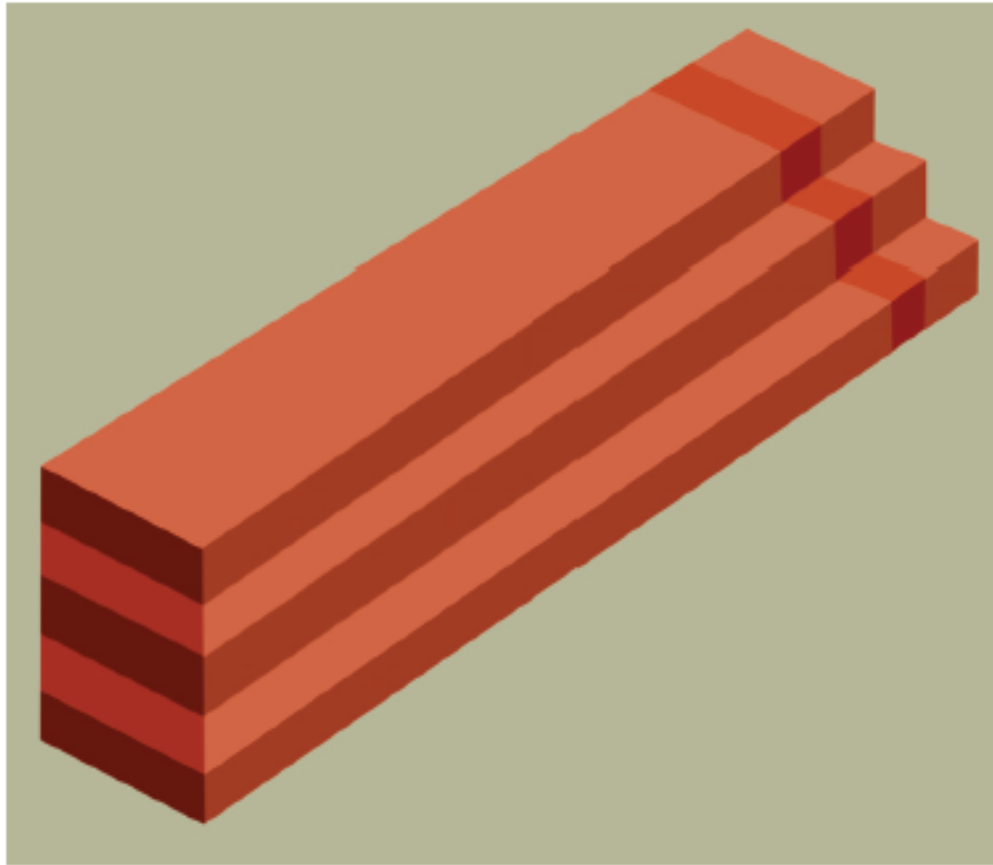
FIGURE 24.9 The impossible steps. On the left, the horizontal stripes appear to be due to paint; on the right, they appear to be due to shading.

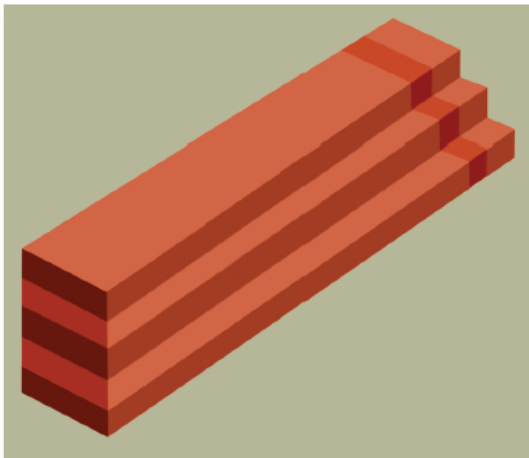
Adelson, E.H. Lightness Perception and Lightness Illusions. In *The New Cognitive Neurosciences*, 2nd ed., M. Gazzaniga, ed. Cambridge, MA: MIT Press, pp. 339-351, (2000).

24 Lightness Perception and Lightness Illusions

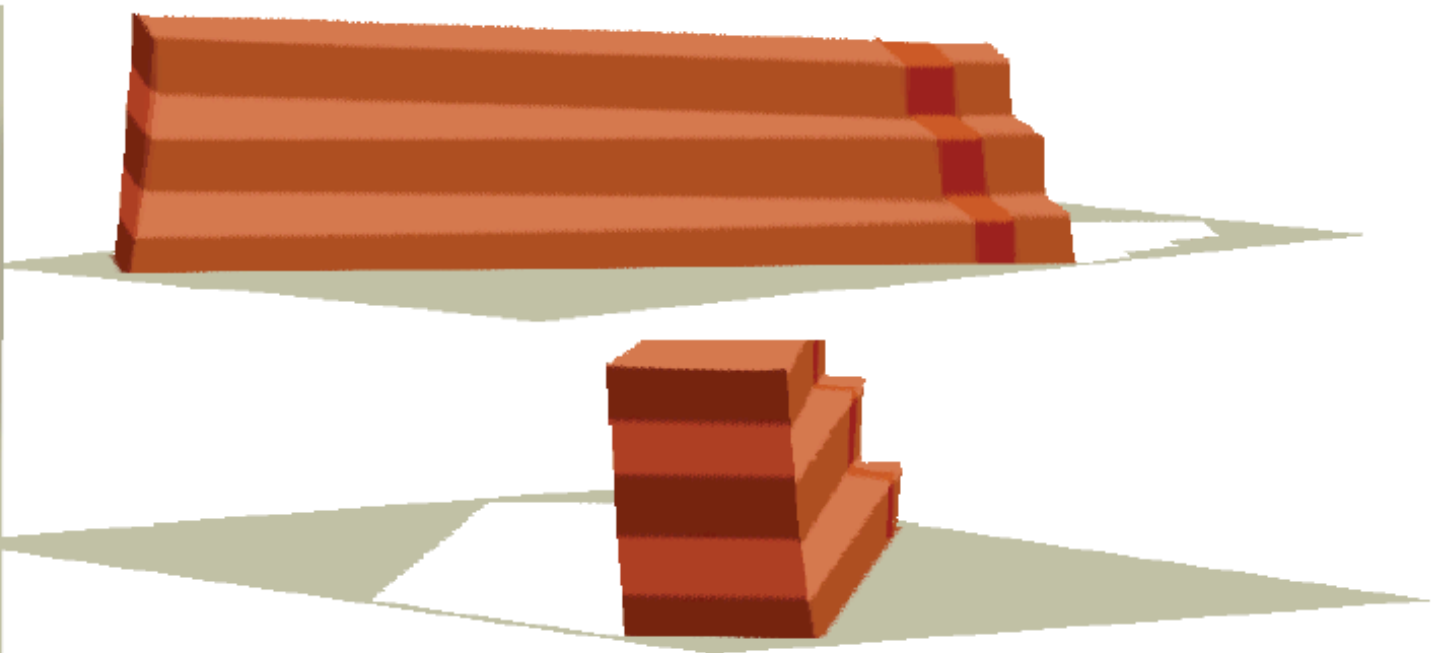
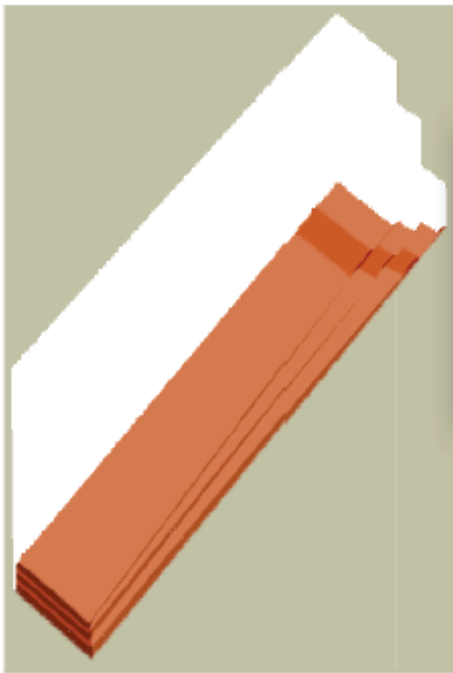
EDWARD H. ADELSON

Impossible steps





Impossible steps



What is missing?

- Recognition