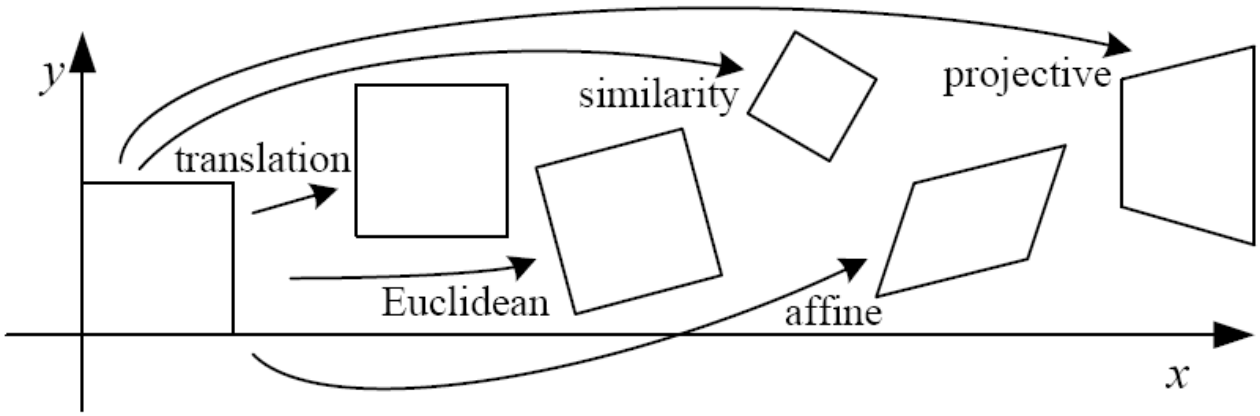
Stereo

Vinay P. Namboodiri

- Slide credit to James Hays, Robert T. Collins, Steve Seitz, Derek Hoiem

Review

2D image transformations (reference table)



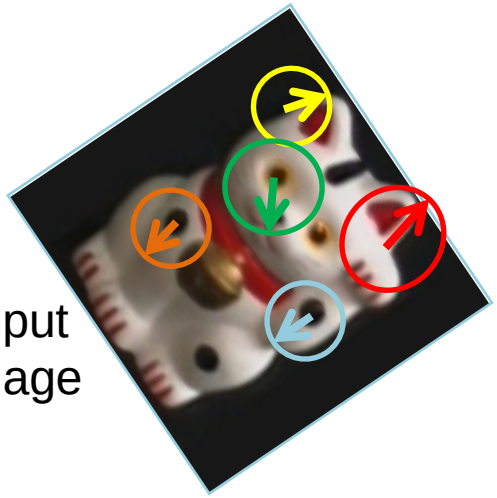
Name	Matrix	# D.O.F.	Preserves:	Icon
translation	$\left[\begin{array}{c\|c} \boldsymbol{I} & \boldsymbol{t} \end{array}\right]_{2\times3}$	2	orientation $+\cdots$	
rigid (Euclidean)	$\left[\begin{array}{c\|c} \boldsymbol{R} & \boldsymbol{t} \end{array}\right]_{2\times3}$	3	lengths $+\cdots$	
similarity	$\left[\begin{array}{c\|c} s\boldsymbol{R} & \boldsymbol{t} \end{array}\right]_{2\times3}$	4	angles $+\cdots$	
affine	$\left[\begin{array}{c} \boldsymbol{A} \end{array}\right]_{2\times3}$	6	parallelism $+\cdots$	
projective	$\left[\begin{array}{c} \tilde{\boldsymbol{H}} \end{array}\right]_{3\times3}$	8	straight lines	

Szeliski 2.1

Algorithm Summary

- Least Squares Fit
  - closed form solution
  - robust to noise
  - not robust to outliers
- Robust Least Squares
  - improves robustness to noise
  - requires iterative optimization
- Hough transform
  - robust to noise and outliers
  - can fit multiple models
  - only works for a few parameters (1-4 typically)
- RANSAC
  - robust to noise and outliers
  - works with a moderate number of parameters (e.g, 1-8)
- Iterative Closest Point (ICP)
  - For local alignment only: does not require initial correspondences

Finding the objects (overview)

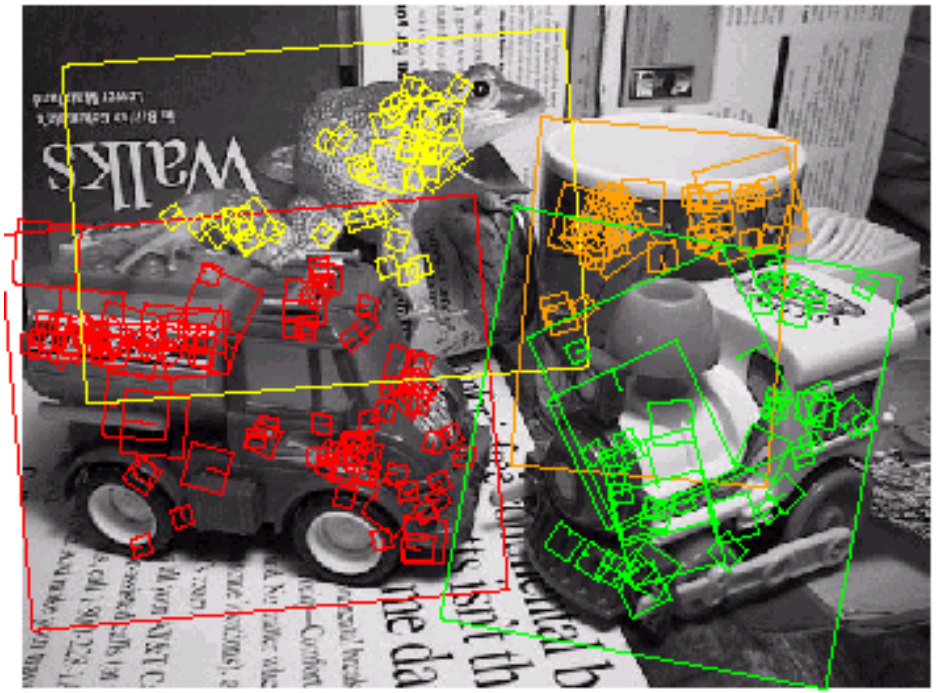


1. Match interest points from input image to database image
2. Matched points vote for rough position/orientation/scale of object
3. Find position/orientation/scales that have at least three votes
4. Compute affine registration and matches using iterative least squares with outlier check
5. Report object if there are at least T matched points

Finding the objects (SIFT, Lowe 2004)

1. Match interest points from input image to database image
2. Get location/scale/orientation using Hough voting
   - In training, each point has known position/scale/orientation wrt whole object
   - Matched points vote for the position, scale, and orientation of the entire object
   - Bins for x, y, scale, orientation
     - Wide bins (0.25 object length in position, 2x scale, 30 degrees orientation)
     - Vote for two closest bin centers in each direction (16 votes total)
3. Geometric verification
   - For each bin with at least 3 keypoints
   - Iterate between least squares fit and checking for inliers and outliers
4. Report object if > T inliers (T is typically 3, can be computed to match some probabilistic threshold)

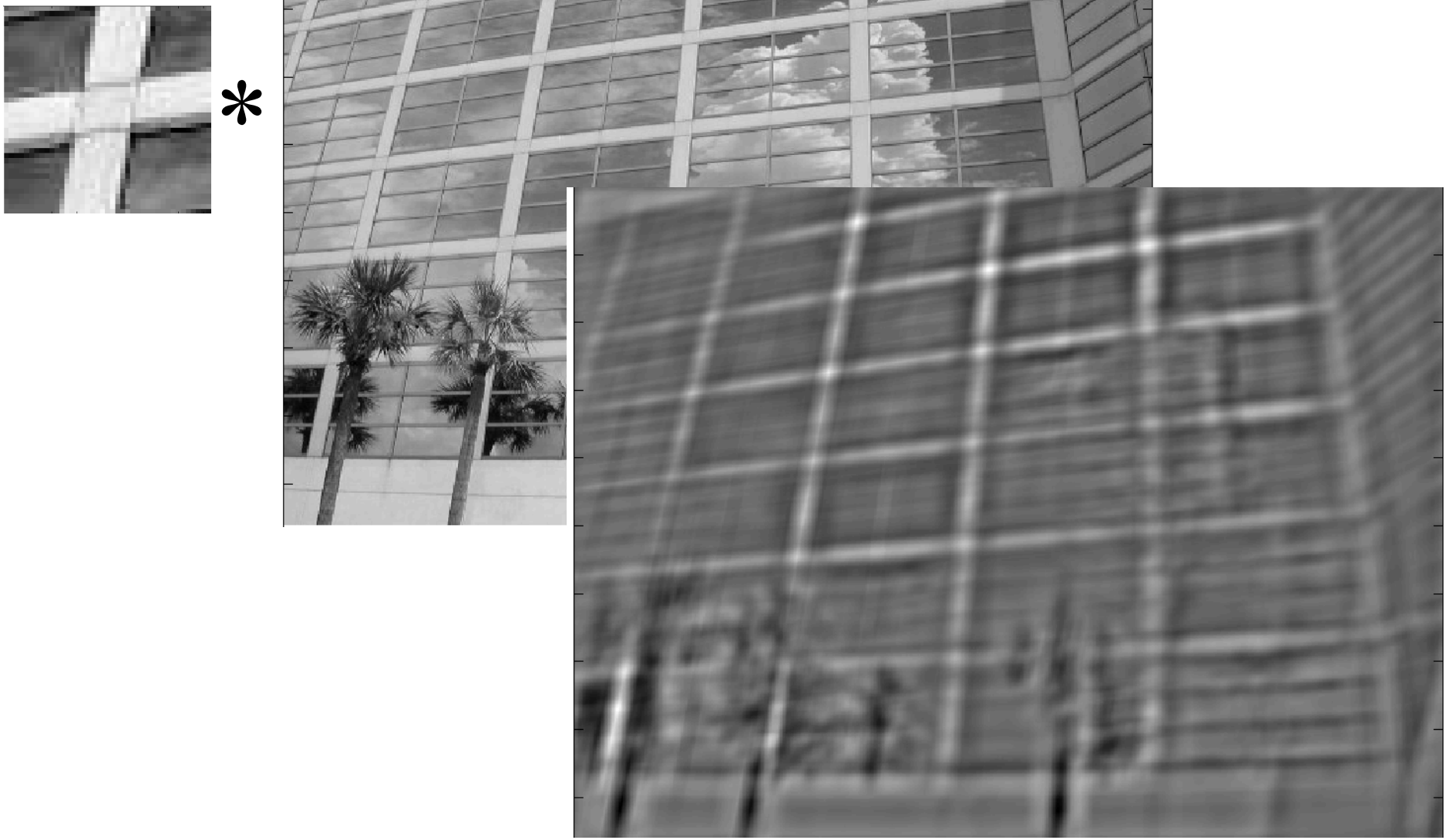
Examples of recognized objects



Dense Correspondence

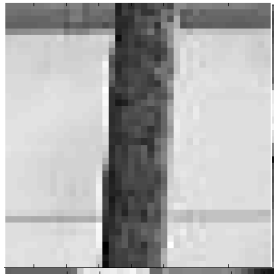
Robert Collins
CSE486, Penn State

Template Matching

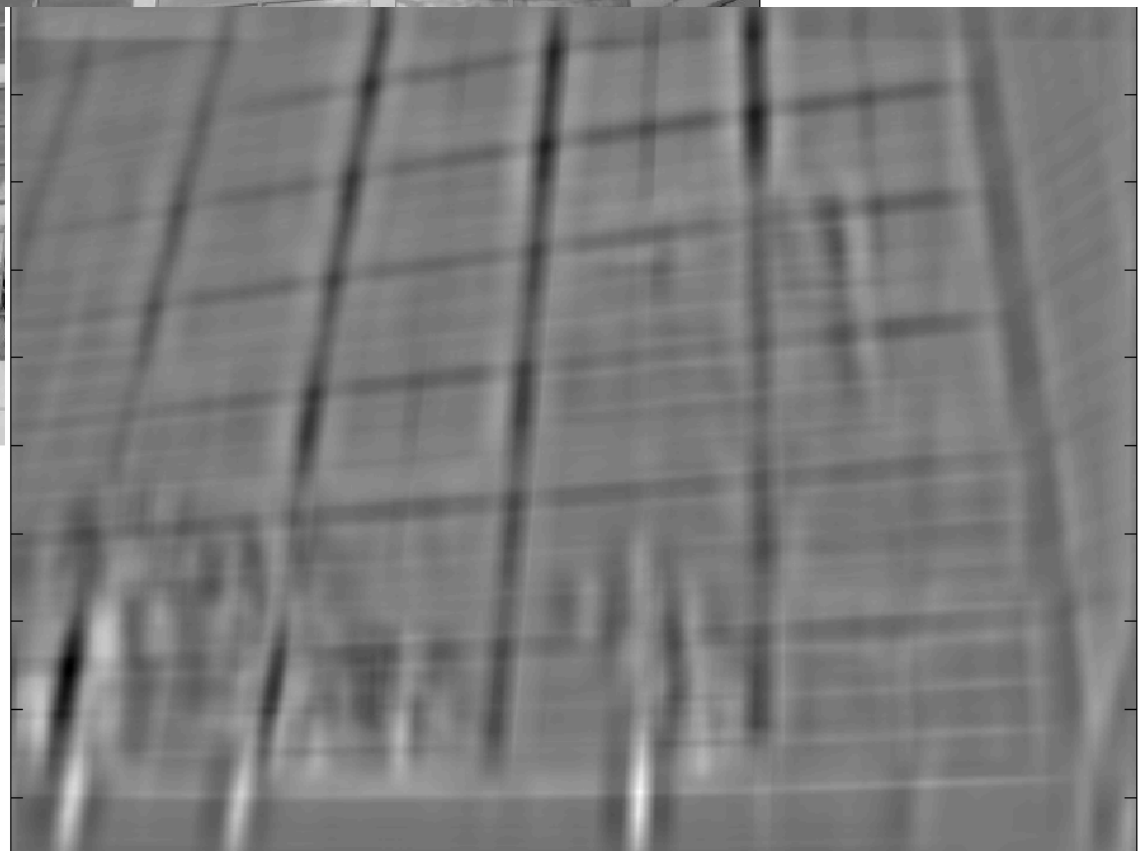


Robert Collins
CSE486, Penn State

Template Matching



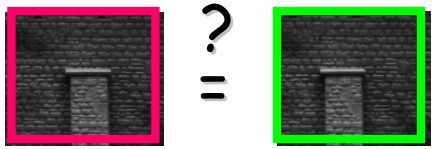
*



Correspondence Problem

- Two classes of algorithms:
  - Correlation-based algorithms
    - Produce a DENSE set of correspondences
  - Feature-based algorithms
    - Produce a SPARSE set of correspondences

Comparing Windows:



f g

Some possible measures:

$$\max_{[i,j]\in R} |f(i,j) - g(i,j)|$$

$$\sum_{[i,j]\in R} |f(i,j) - g(i,j)|$$

$$SSD = \sum_{[i,j]\in R} (f(i,j) - g(i,j))^2$$

$$C_{fg} = \sum_{[i,j]\in R} f(i,j)g(i,j)$$

Most popular

Example

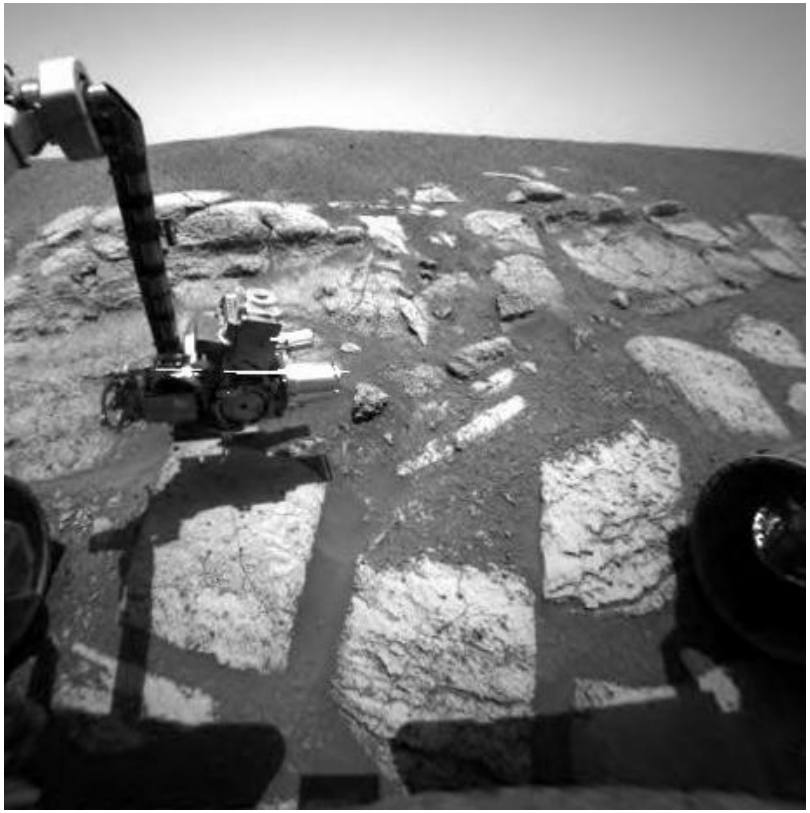


Image 1

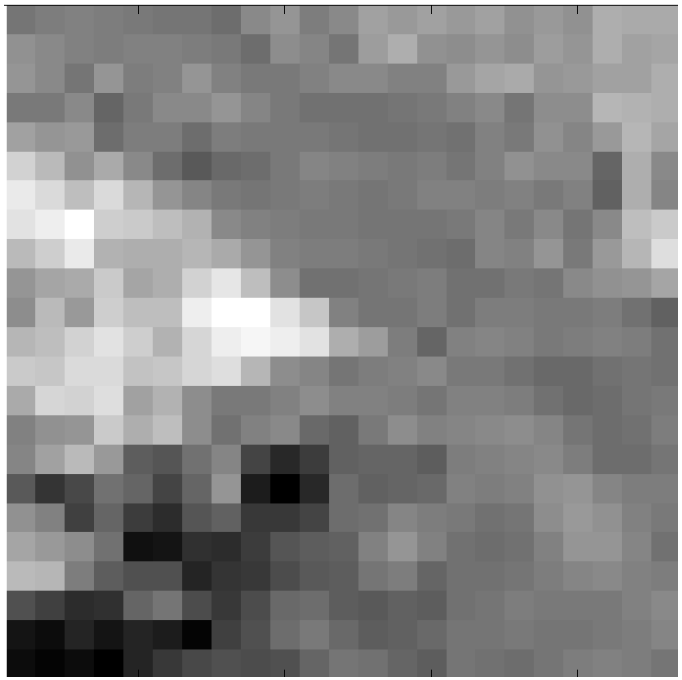
Image 2

Note: this is a stereo pair from the NASA mars rover. The rover is exploring the “El Capitan” formation.

Robert Collins
CSE486, Penn State

Example

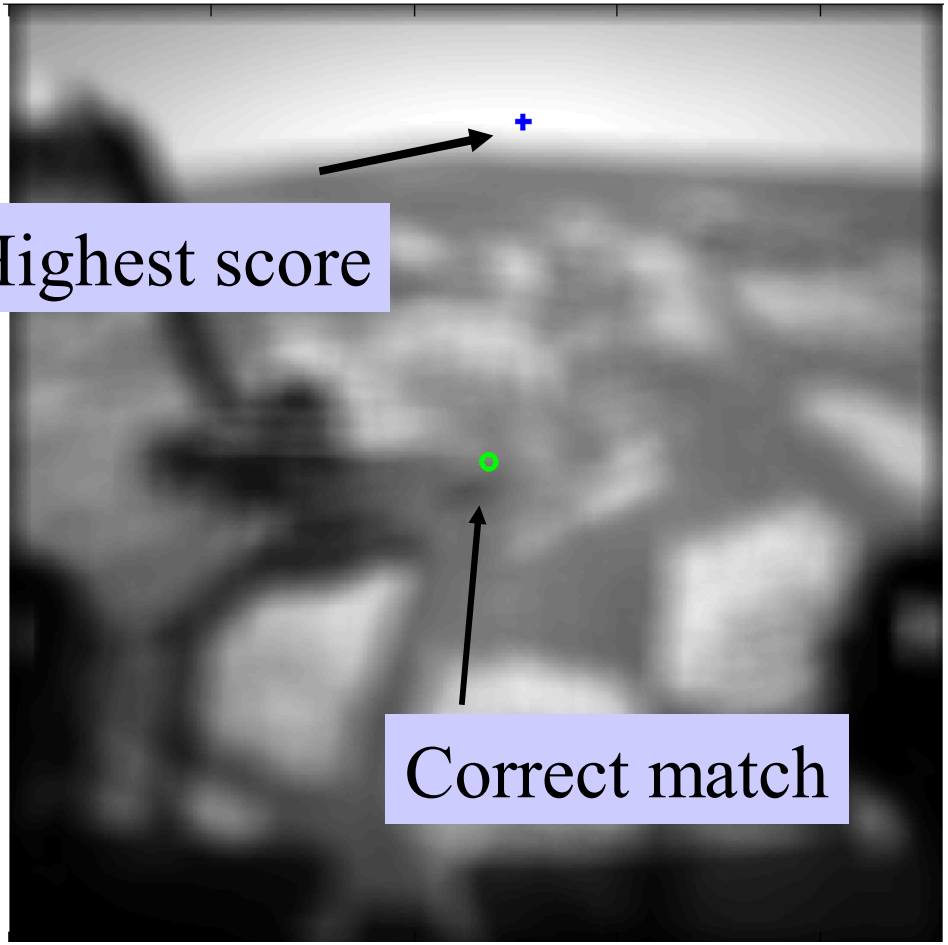
Image 1



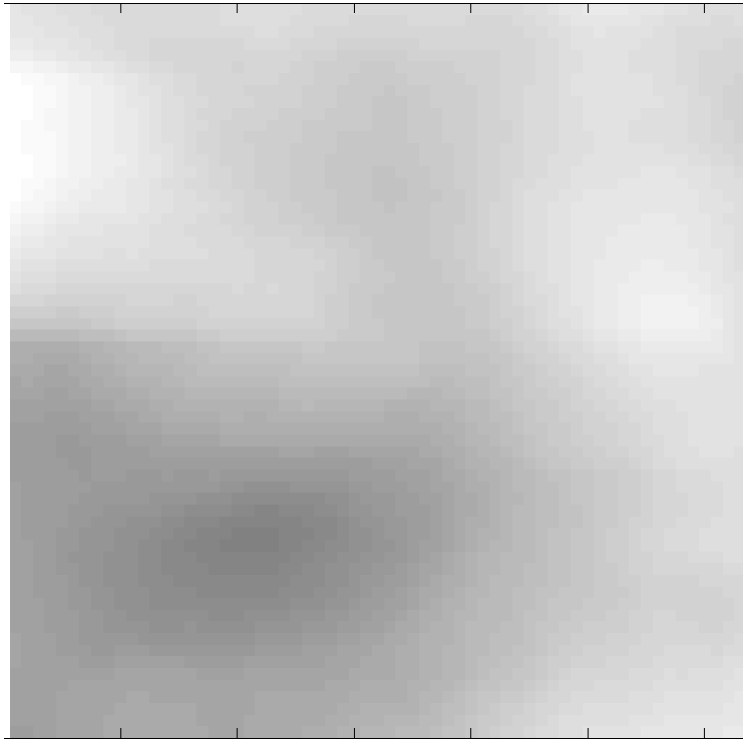
Template
(image patch)

Example: Raw Cross-correlation

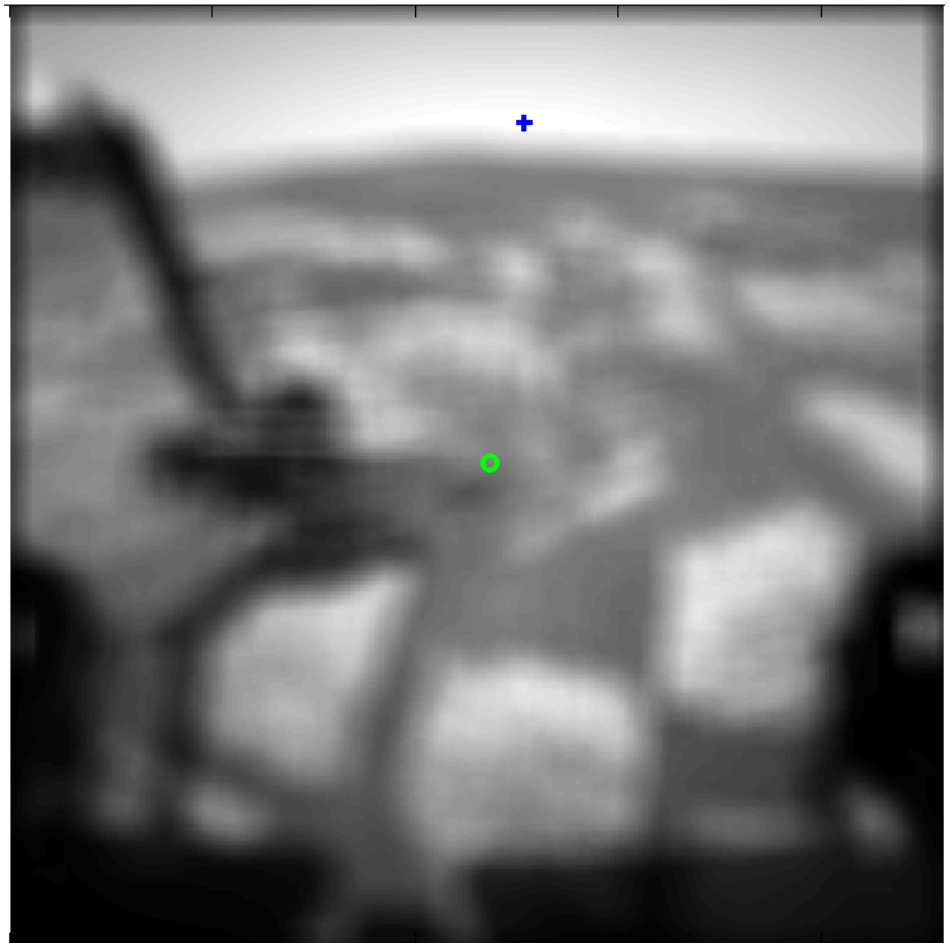
score = imfilter(image2,tmpl)



Score around correct match



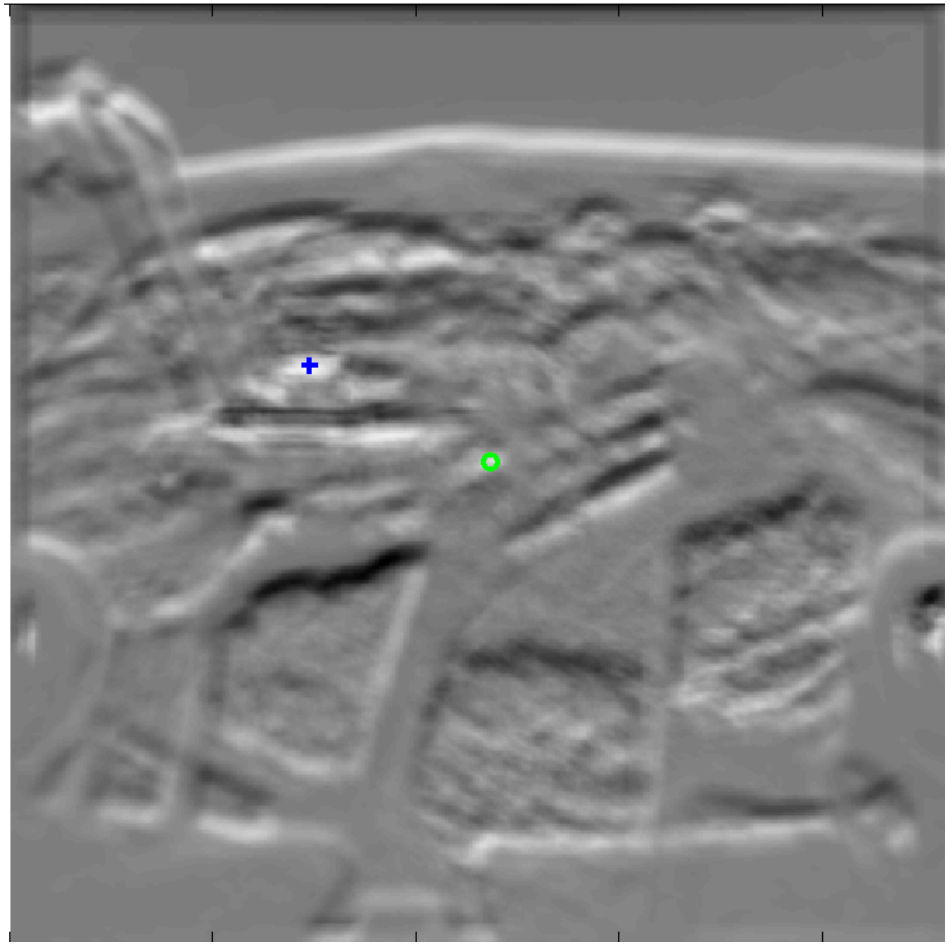
Example: Cross-correlation

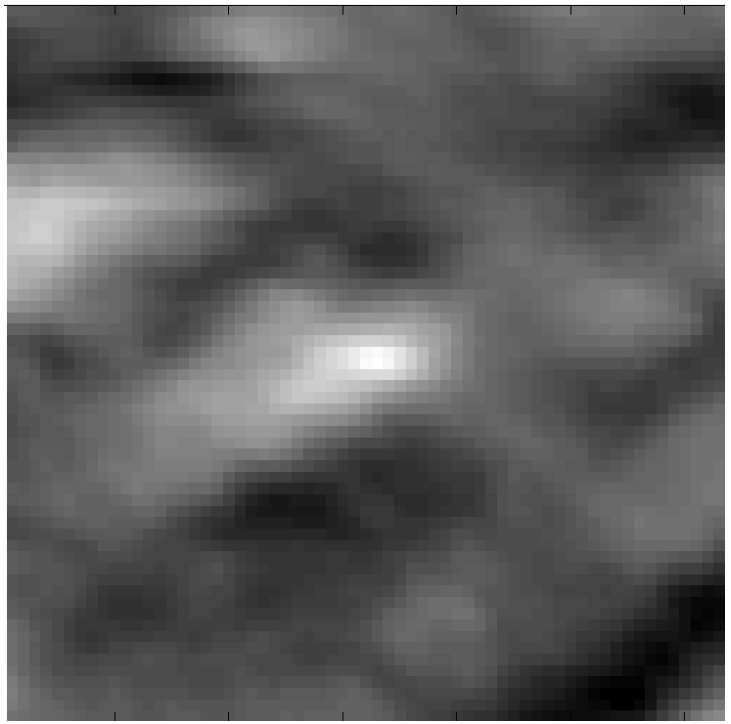


Note that score image looks a lot like a blurry version of image 2.

This clues us in to the problem with straight correlation with an image template.

Correlation, zero-mean template





Better! But highest score is still not the correct match.

Note: highest score IS best within local neighborhood of correct match.

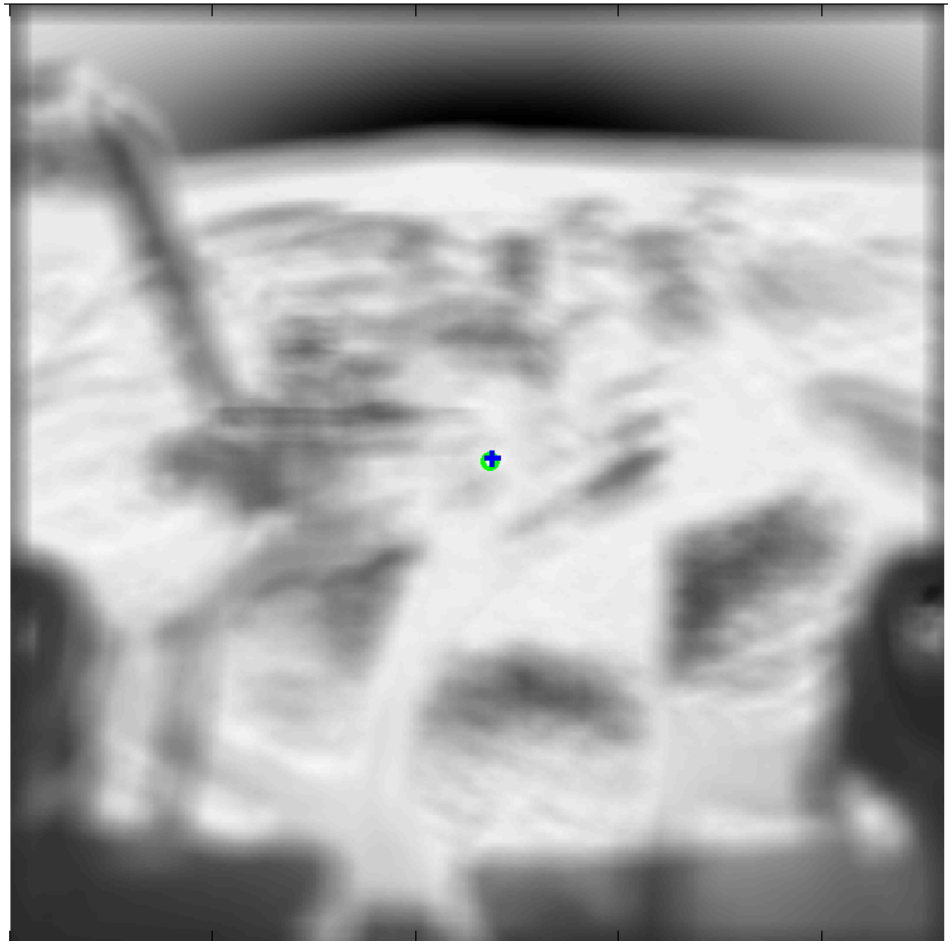
Robert Collins
CSE486, Penn State

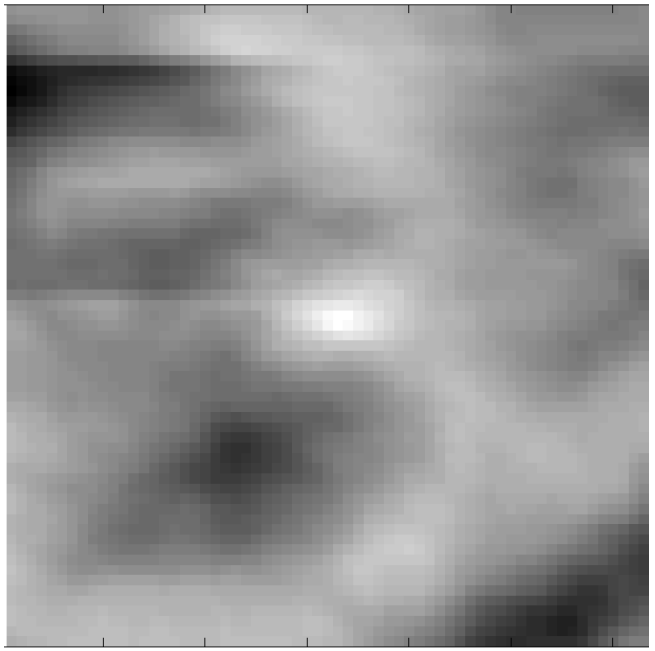
"SSD" or "block matching" (Sum of Squared Differences)

$$\sum_{[i,j] \in R} (f(i,j) - g(i,j))^2$$

1) The most popular matching score.
2) We used it when deriving Harris corners
3) T&V claim it works better than cross-correlation

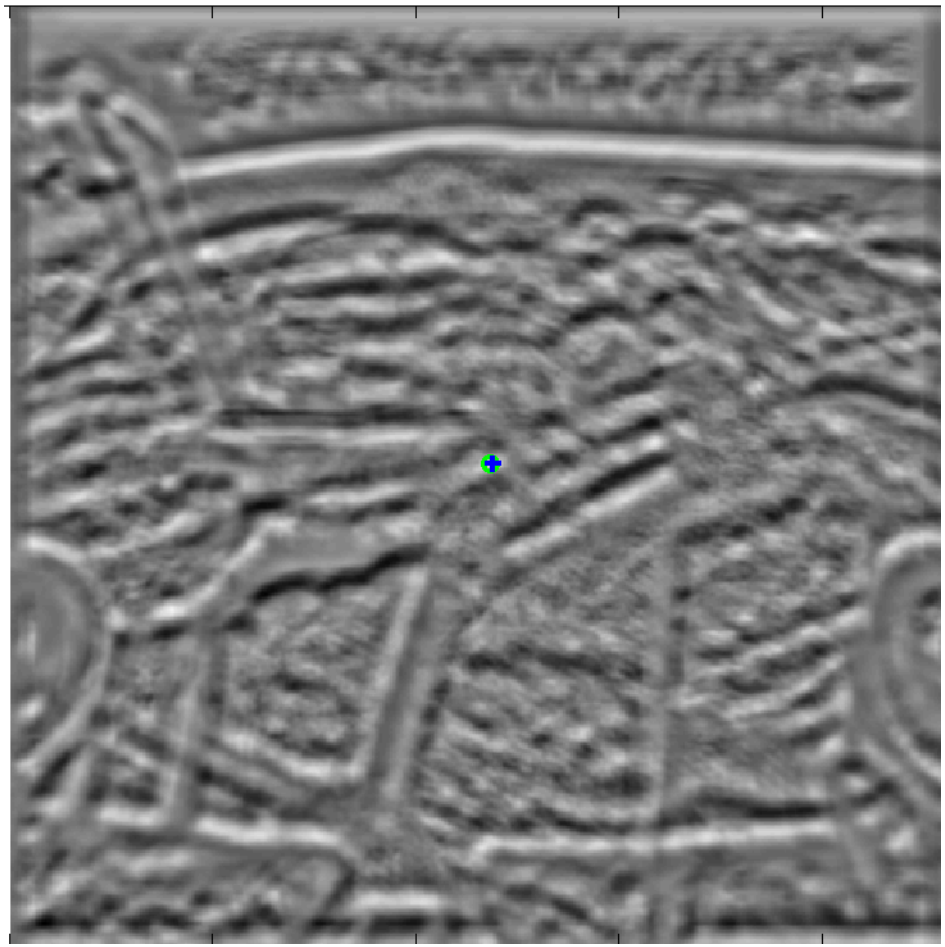
SSD

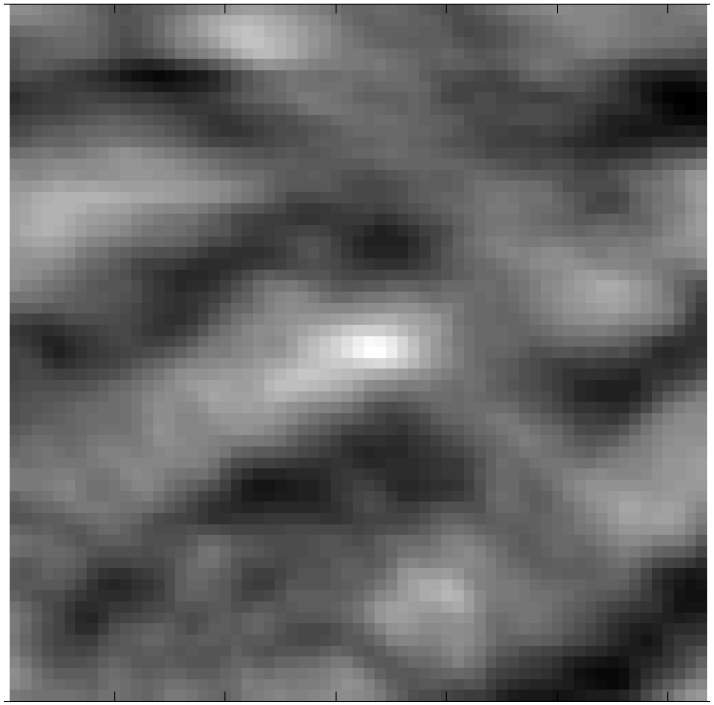




Best match (highest score) in image coincides with correct match in this case!

Normalized Cross Correlation

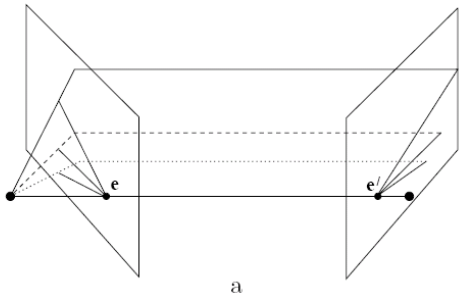




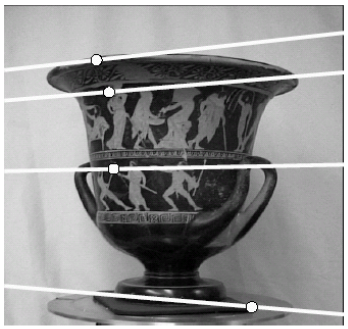
Highest score also coincides with correct match.
Also, looks like less chances of getting a wrong match.

Stereo

Multiple views



a



Hartley and Zisserman

stereo vision
structure from motion
optical flow

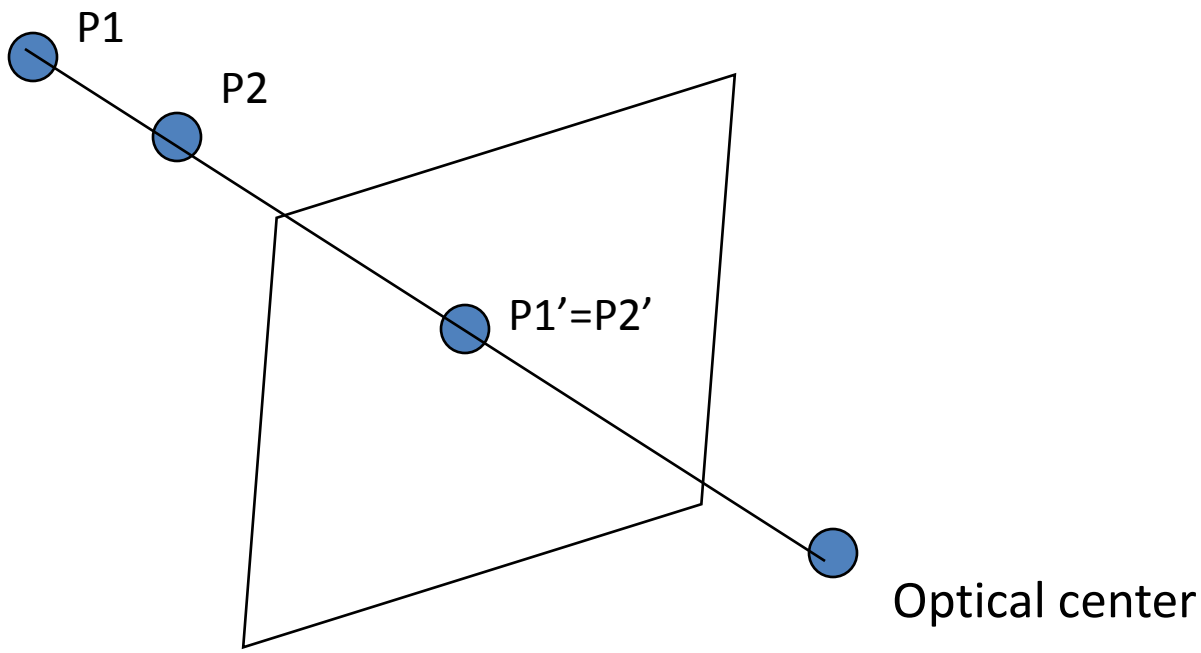
Why multiple views?

- Structure and depth are inherently ambiguous from single views.

Images from Lana Lazebnik

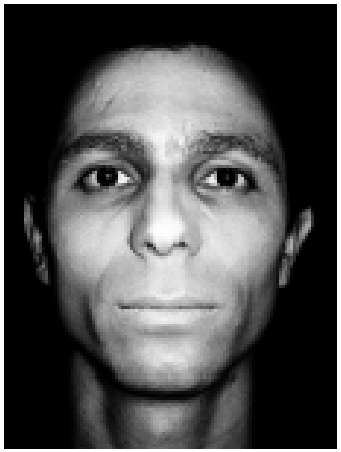
Why multiple views?

- Structure and depth are inherently ambiguous from single views.

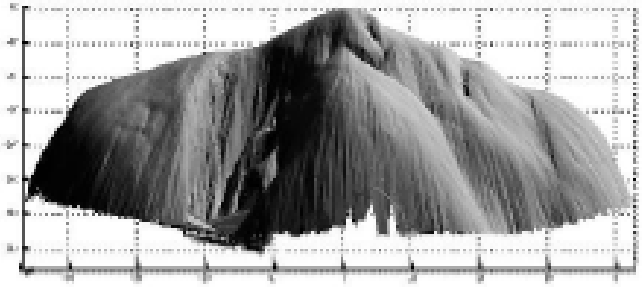


- What cues help us to perceive 3d shape and depth?

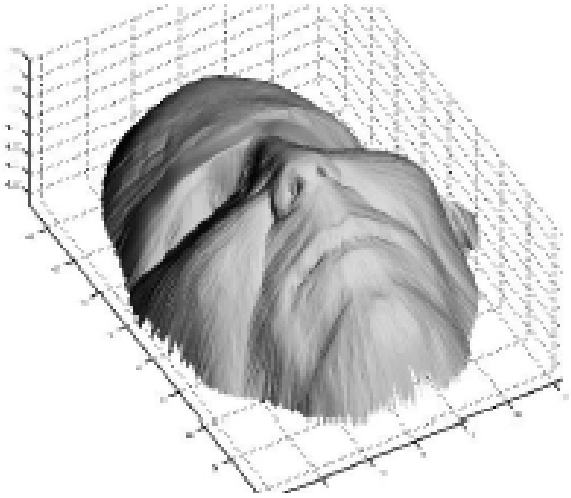
Shading



a)



b)

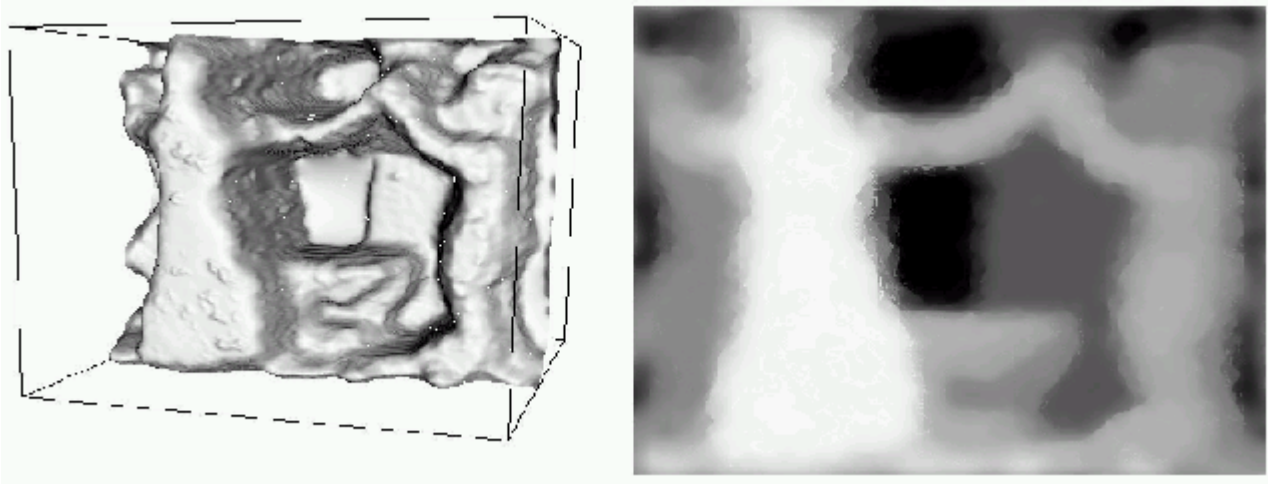


c)

[Figure from Prados & Faugeras 2006]

Focus/defocus

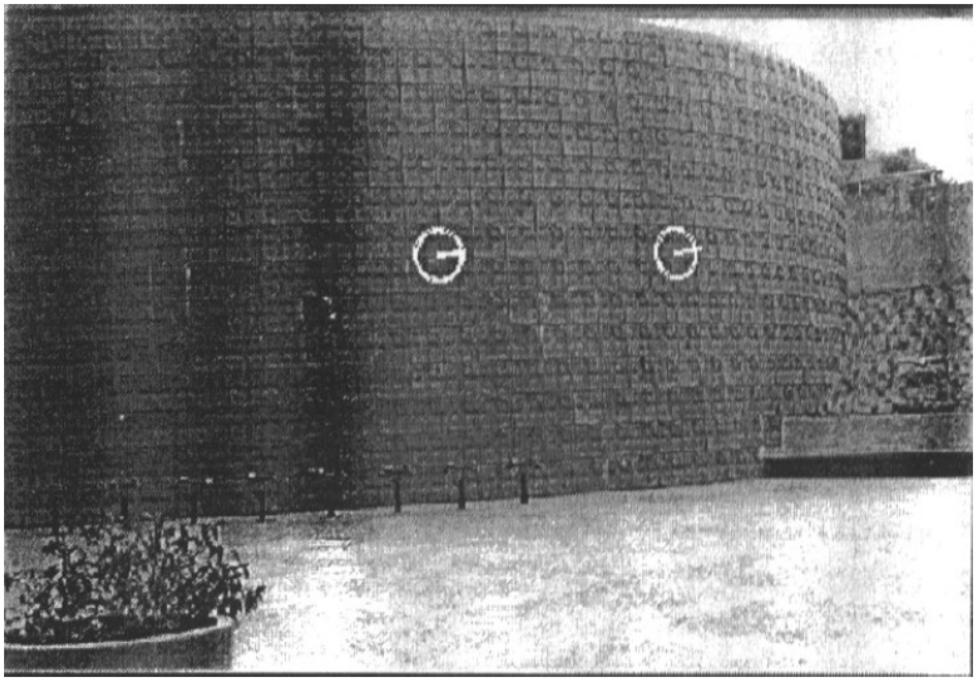
Images from same point of view, different camera parameters

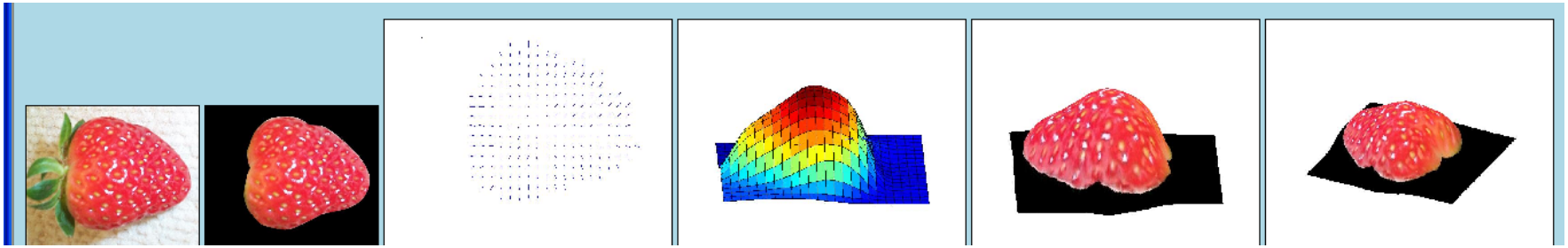


3d shape / depth estimates

[figs from H. Jin and P. Favaro, 2002]

Texture





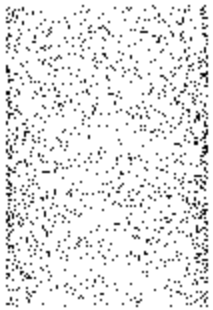
[From A.M. Loh. The recovery of 3-D structure using visual texture patterns. PhD thesis]

Perspective effects

Image credit: S. Seitz

Motion



Figures from L. Zhang

http://www.brainconnection.com/teasers/?main=illusion/motion-shape

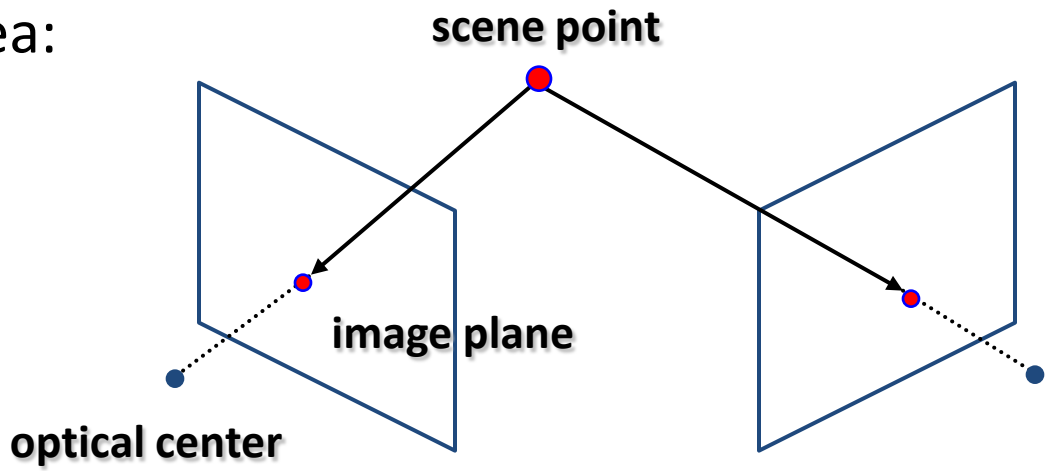
Occlusion

Rene Magritt'e famous painting *Le Blanc-Seing* (literal translation: "The Blank Signature") roughly translates as "free hand" or "free rein".

Estimating scene shape

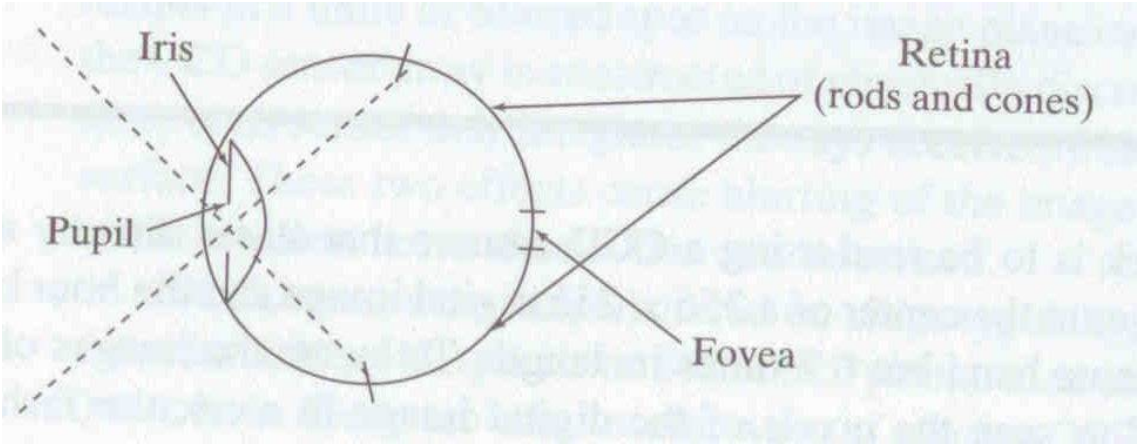
- "Shape from X": Shading, Texture, Focus, Motion…
- **Stereo**:
  - shape from "motion" between two views
  - infer 3d shape of scene from two (multiple) images from different viewpoints

Main idea:



Human eye

Rough analogy with human visual system:



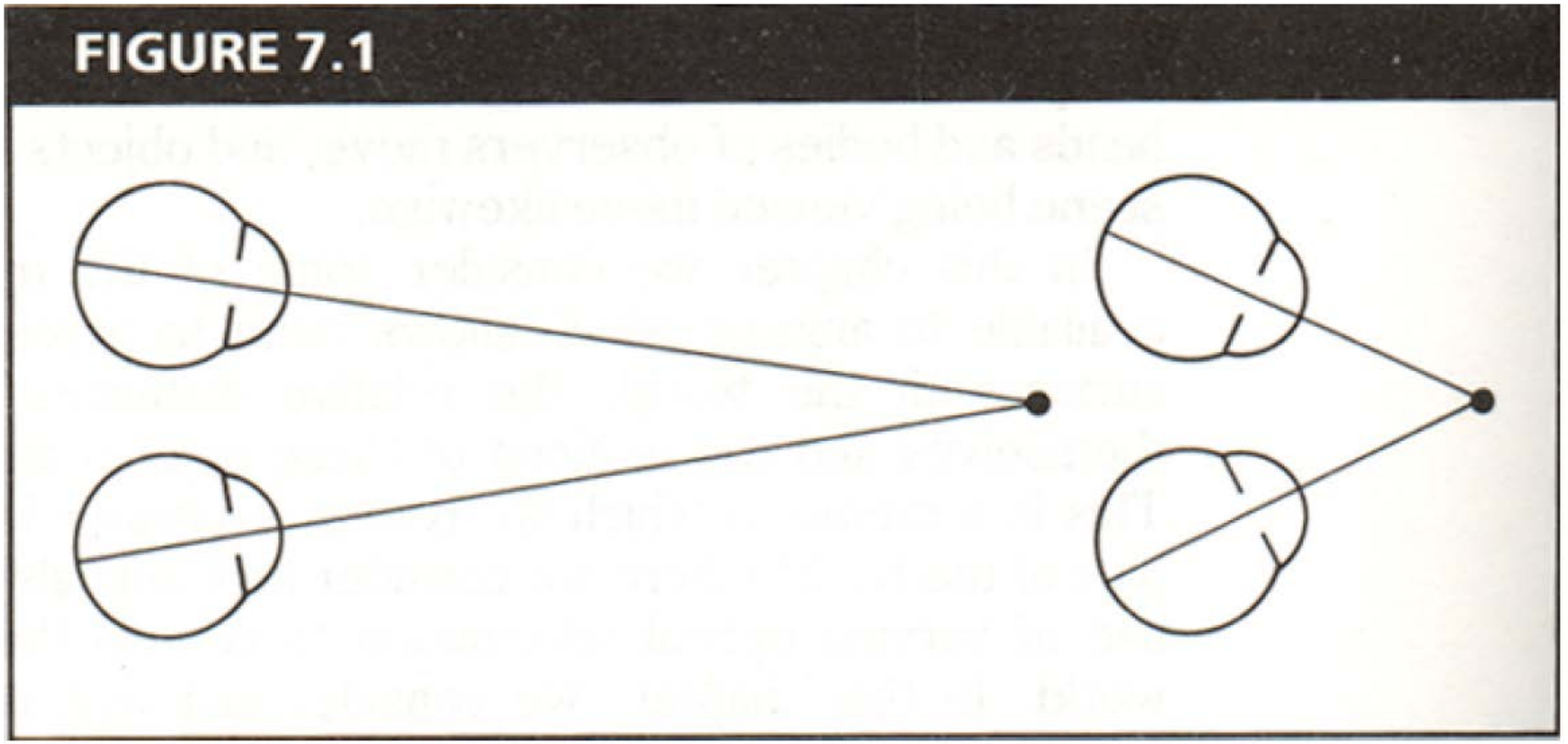
Pupil/Iris – control amount of light passing through lens

Retina - contains sensor cells, where image is formed

Fovea – highest concentration of cones

Fig from Shapiro and Stockman

Human stereopsis: disparity



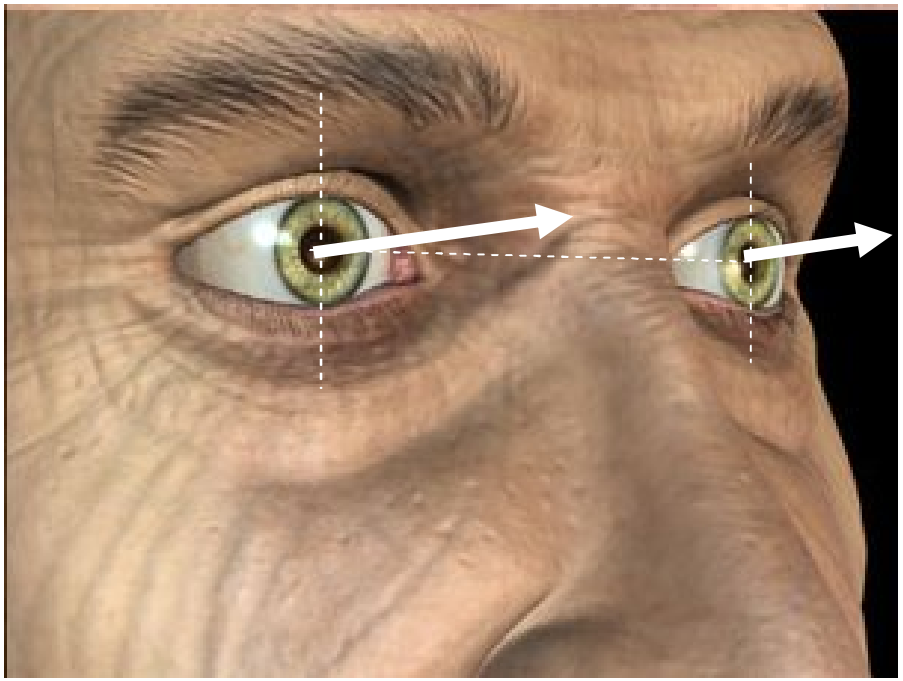
From Bruce and Green, Visual Perception, Physiology, Psychology and Ecology

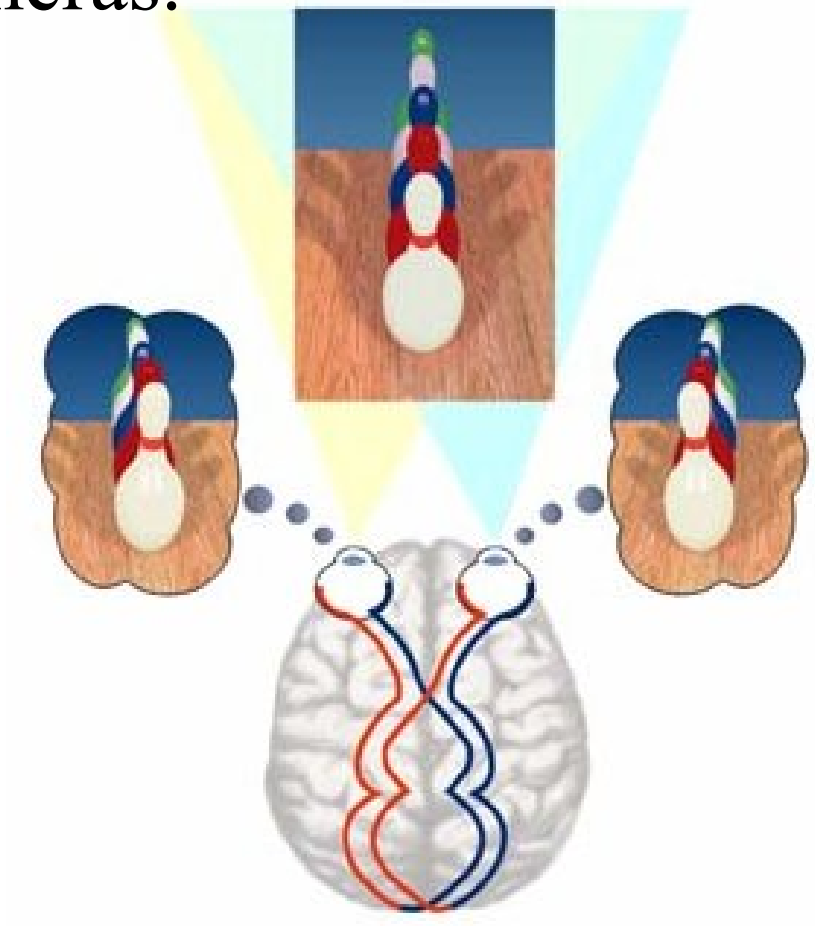
Human eyes **fixate** on point in space – rotate so that corresponding images form in centers of fovea.

Robert Collins
CSE486, Penn State

Stereo Vision

Inferring depth from images taken at the same time by two or more cameras.

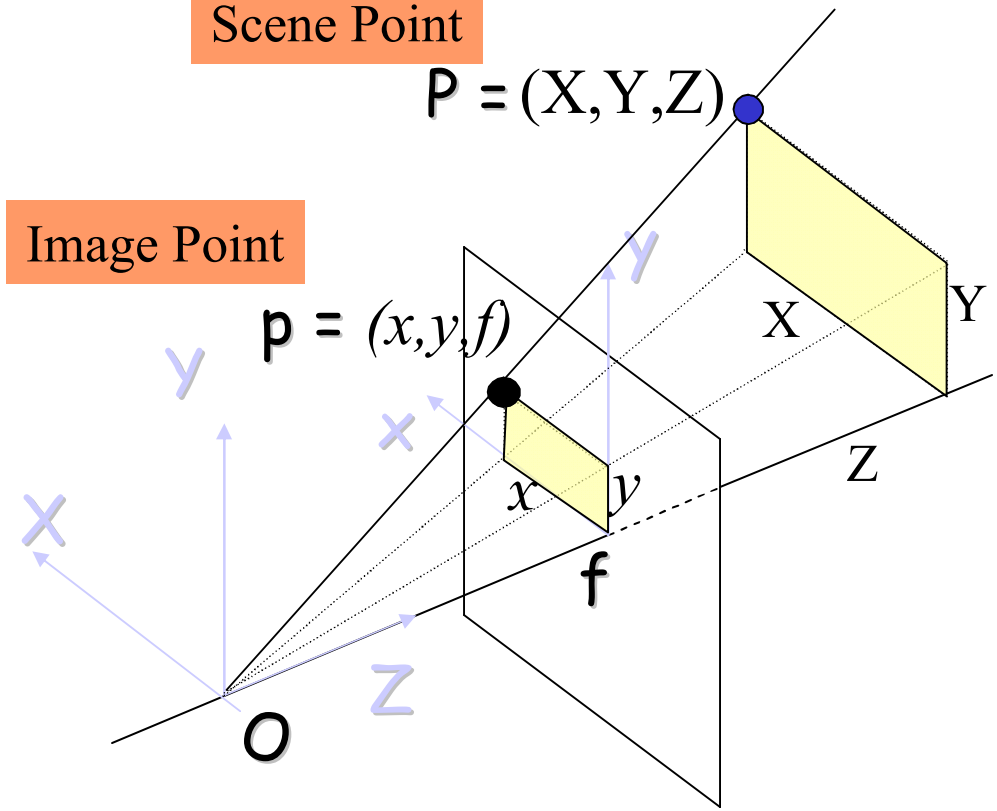




Robert Collins
CSE486, Penn State

Basic Perspective Projection

Scene Point



Perspective Projection Eqns

$$x = f\frac{X}{Z}$$

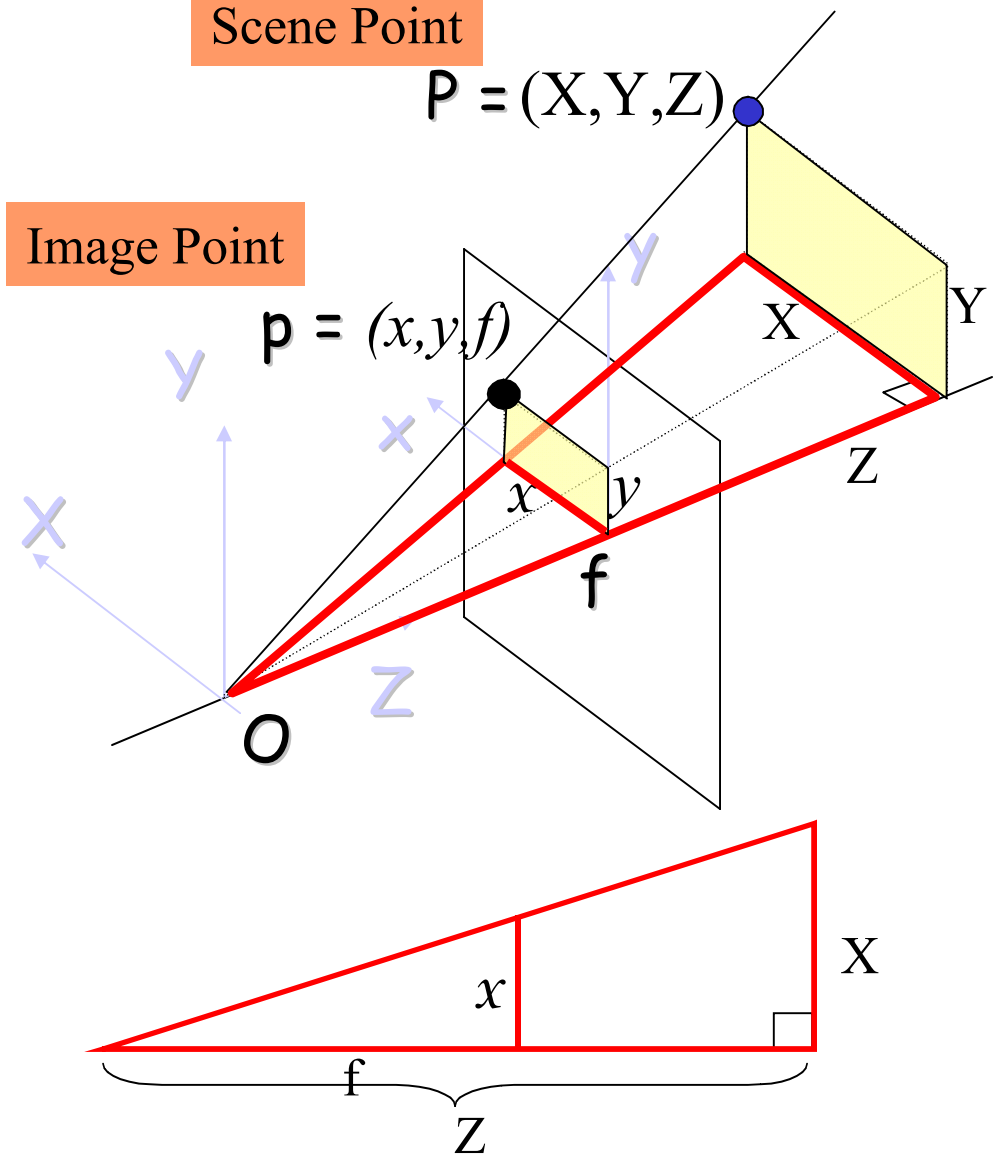
$$y = f\frac{Y}{Z}$$

Robert Collins
CSE486, Penn State

Basic Perspective Projection

Scene Point

Image Point



Perspective Projection Eqns

$$x = f\frac{X}{Z}$$

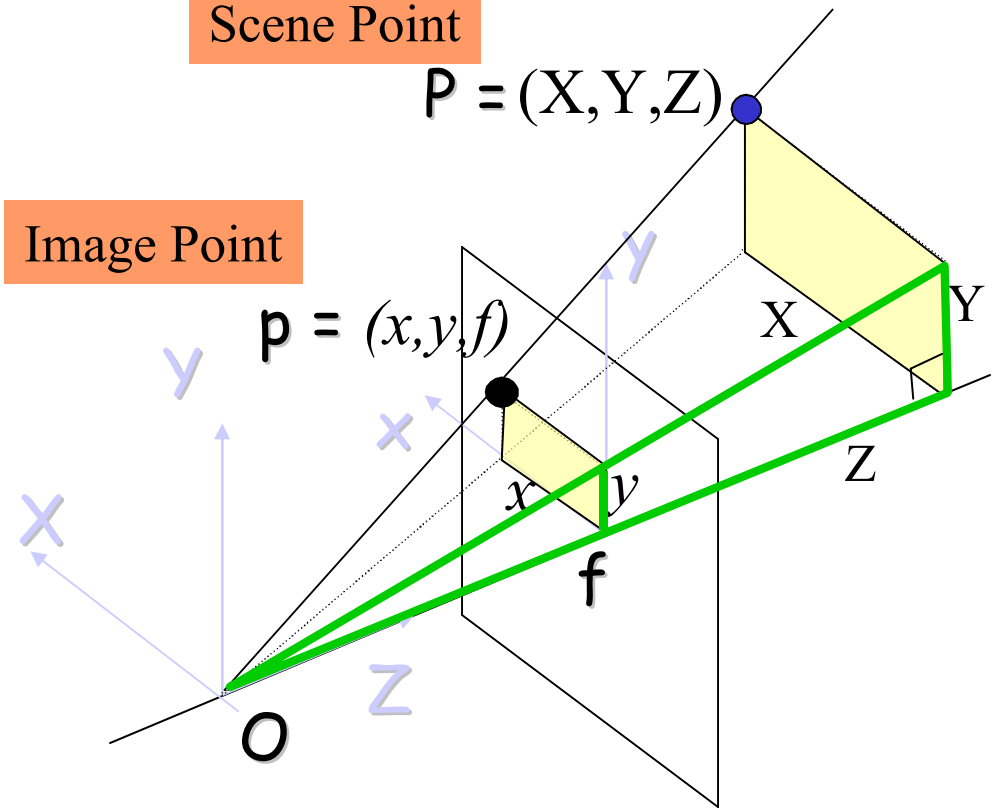
$$y = f\frac{Y}{Z}$$

Robert Collins
CSE486, Penn State

Basic Perspective Projection

Scene Point

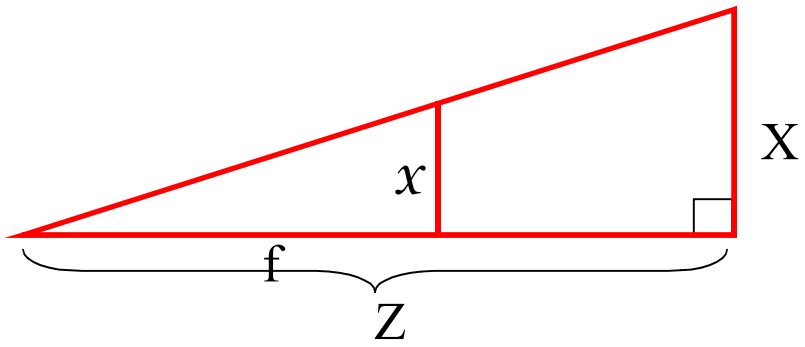
Image Point

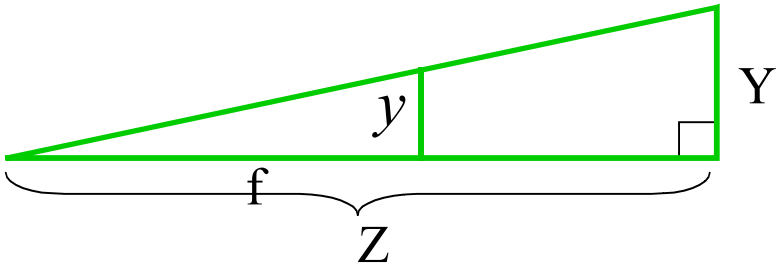


Perspective Projection Eqns

$$x = f\frac{X}{Z}$$

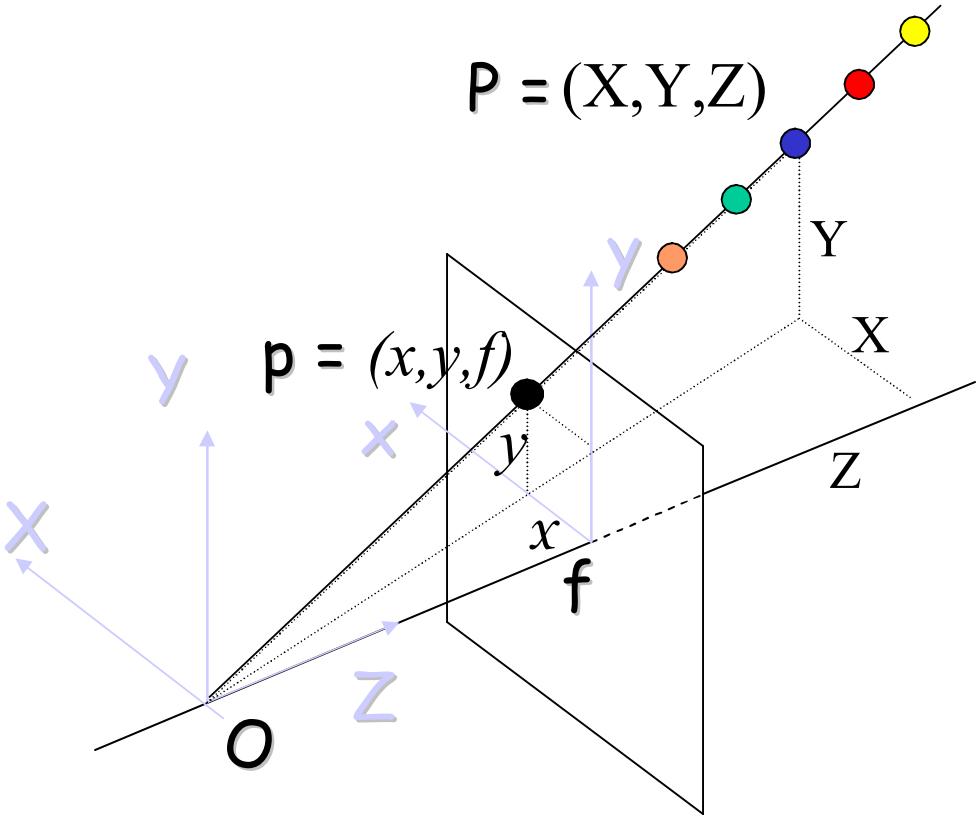
$$y = f\frac{Y}{Z}$$





Robert Collins
CSE486, Penn State

Why Stereo Vision?



$$x = f\frac{X}{Z} = f\frac{kX}{kZ}$$

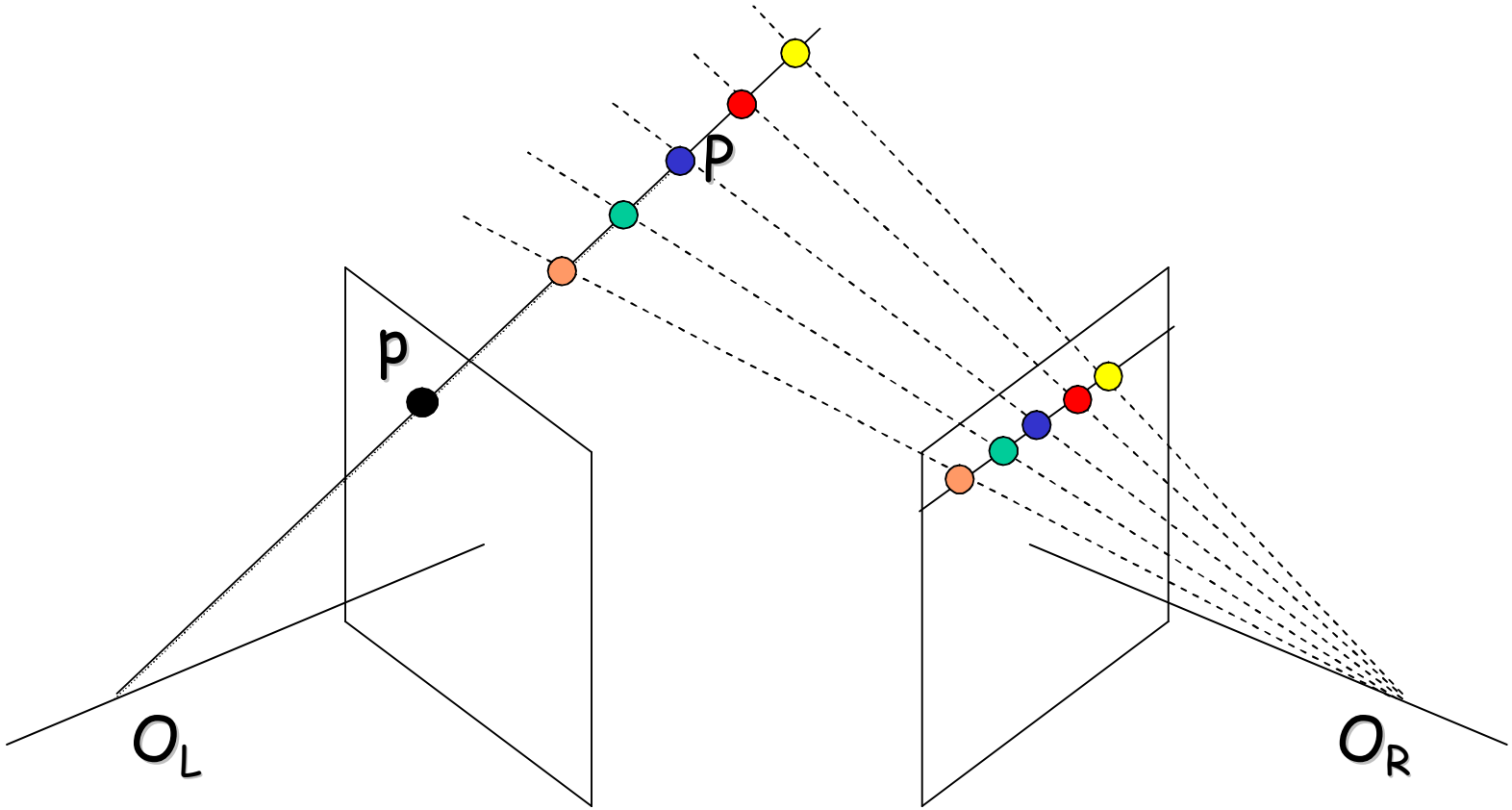
$$y = f\frac{Y}{Z} = f\frac{kY}{kZ}$$

Fundamental Ambiguity:

Any point on the ray OP has image p

Robert Collins
CSE486, Penn State

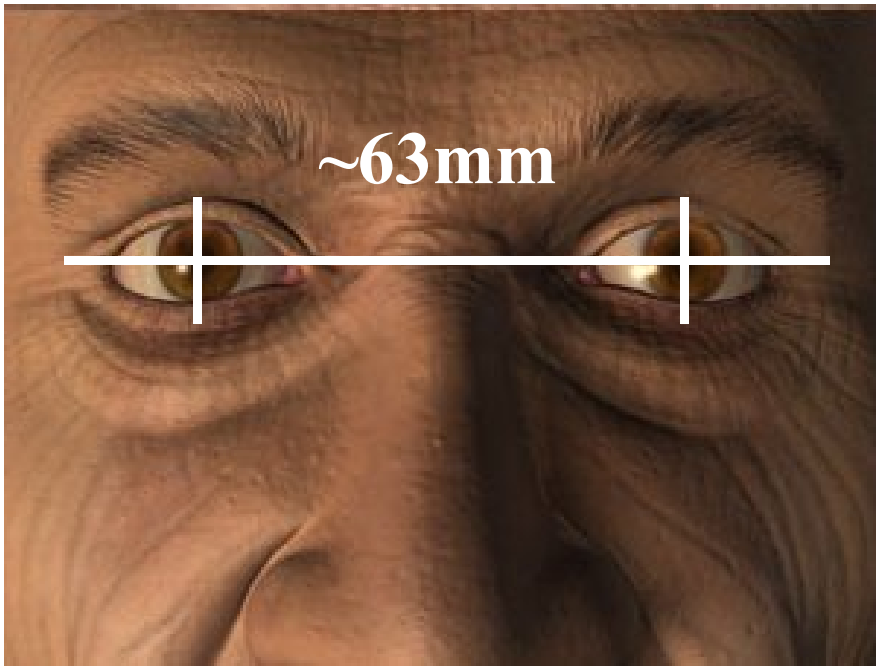
Why Stereo Vision?



A second camera can resolve the ambiguity, enabling measurement of depth via triangulation.

Robert Collins
CSE486, Penn State

Why Stereo Vision?



Your two eyes form a stereo system

The right and left eyes see the world from slightly shifted vantage points.

Robert Collins
CSE486, Penn State

Do-it-Yourself Parallax Demo

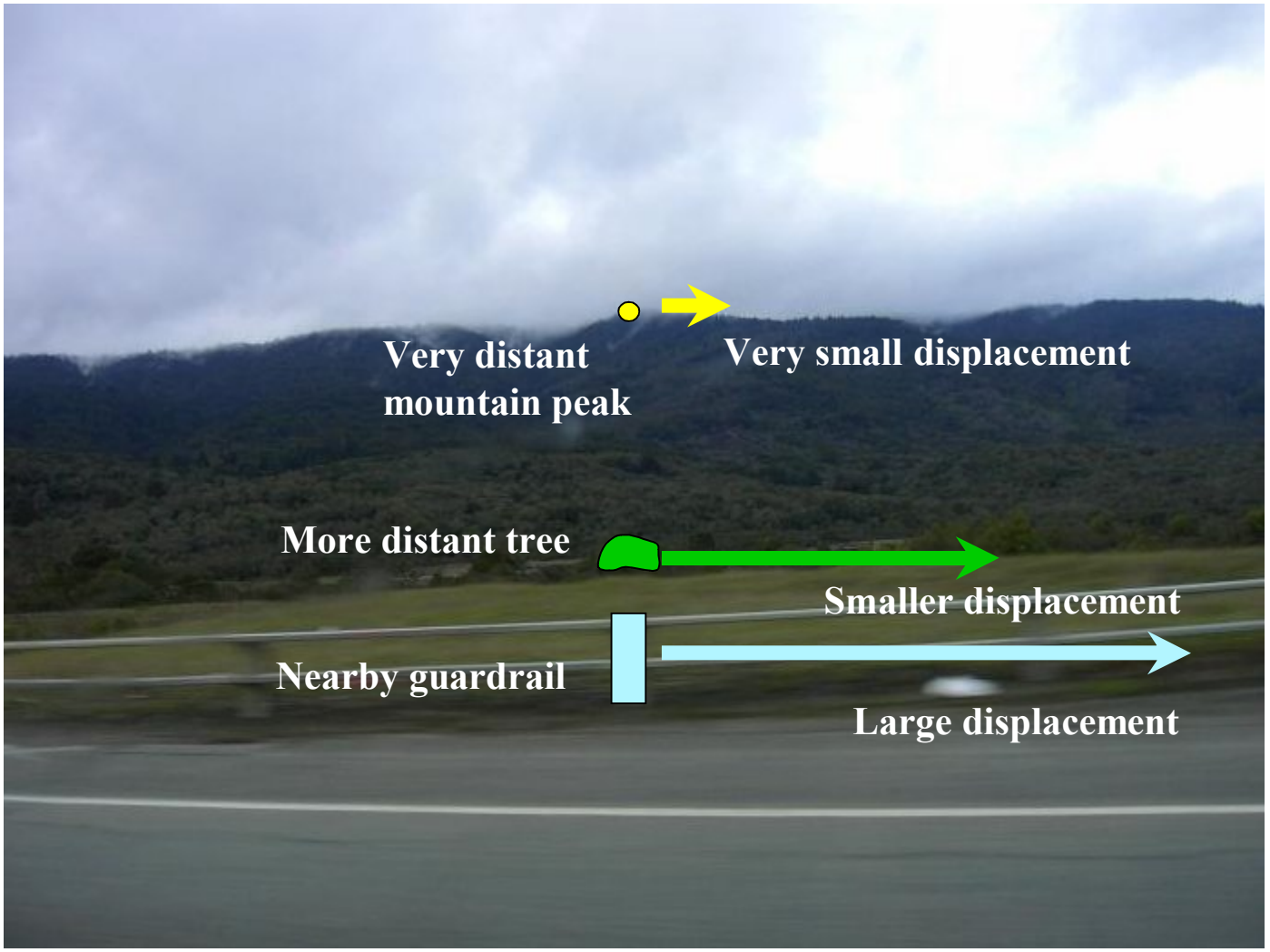
Show:

- Points at different depths displace differently
- Nearby points displace more than far ones

Robert Collins
CSE486, Penn State

A Hitchhiker's Guide to Parallax

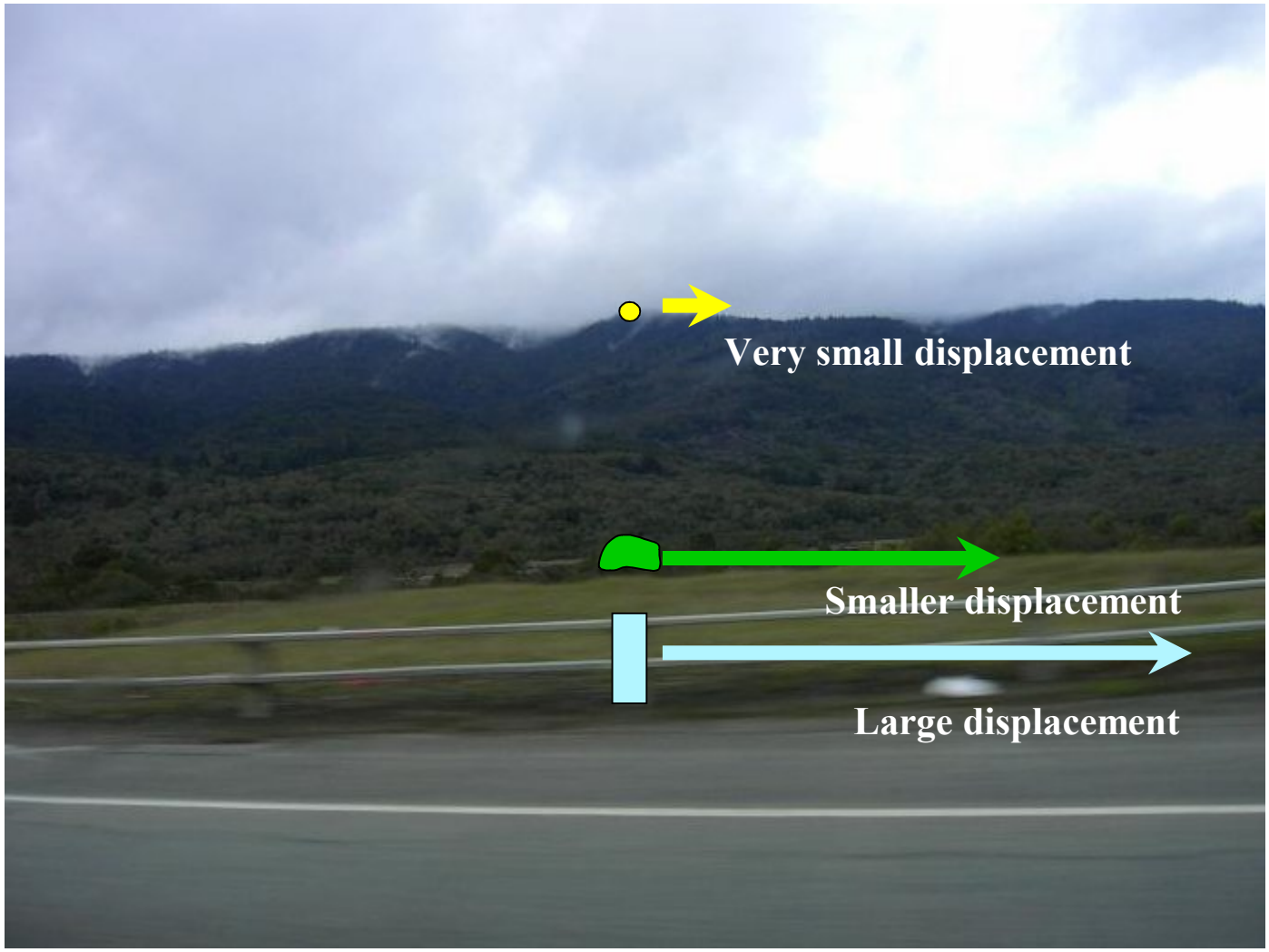
Parallax = apparent motion of scene features located at different distances



Robert Collins
CSE486, Penn State

General Idea of Stereo

Infer distance to scene points by measuring parallax.



INFER

Far

Midrange

Close

Robert Collins
CSE486, Penn State

Anaglyphs

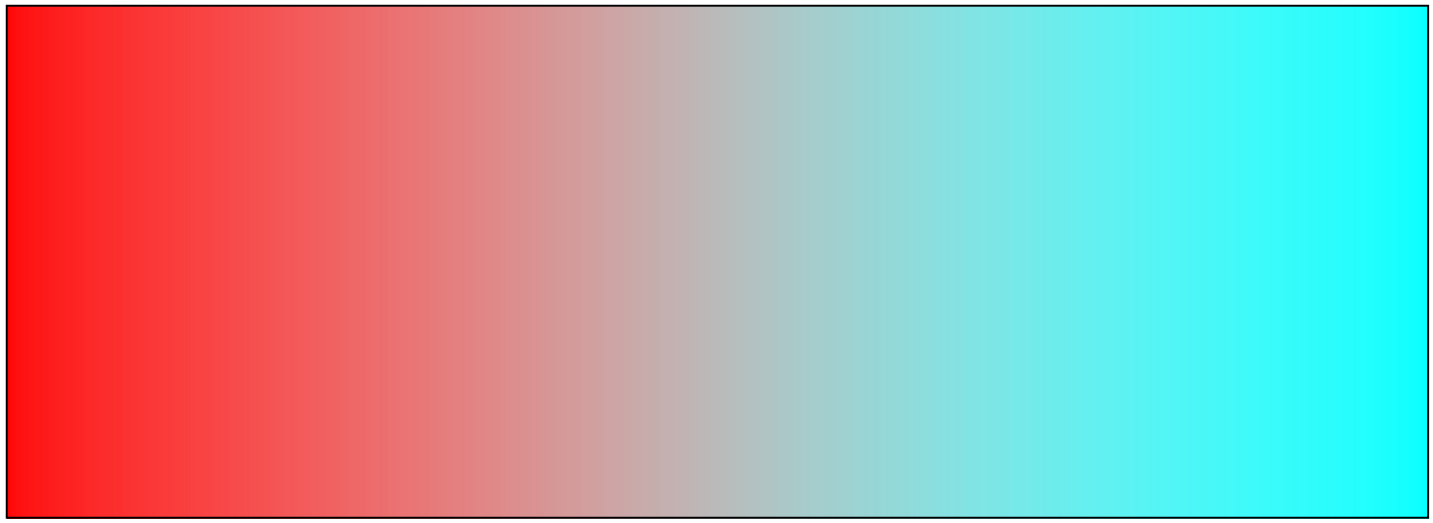
Anaglyphs are a way of encoding parallax in a single picture. Two slightly different perspectives of the same subject are superimposed on each other in contrasting colors, producing a three-dimensional effect when viewed through two correspondingly colored filters

Put red filter over left eye

Robert Collins
CSE486, Penn State

Robert Collins
CSE486, Penn State

How Anaglyphs Work



Close right eye, then close left. What do you observe?

Red filter selectively passes red color, and similarly for cyan filter and cyan color.

Robert Collins
CSE486, Penn State

Making an Anaglyph

Take a greyscale stereo pair.

Copy the left image to the red channel of a new image (the anaglyph image)

Copy the right image to the green and blue channels of the anaglyph image (note: green+blue = cyan)

Now when you view with red-cyan glasses, the left eye sees only the left image, and the right eye sees only the right image. The brain fuses to form 3D.

Robert Collins
CSE486, Penn State

Stereo Pyschophysics

How does stereo depth perception work?

In particular, at what "level" in the visual system does it occur at?

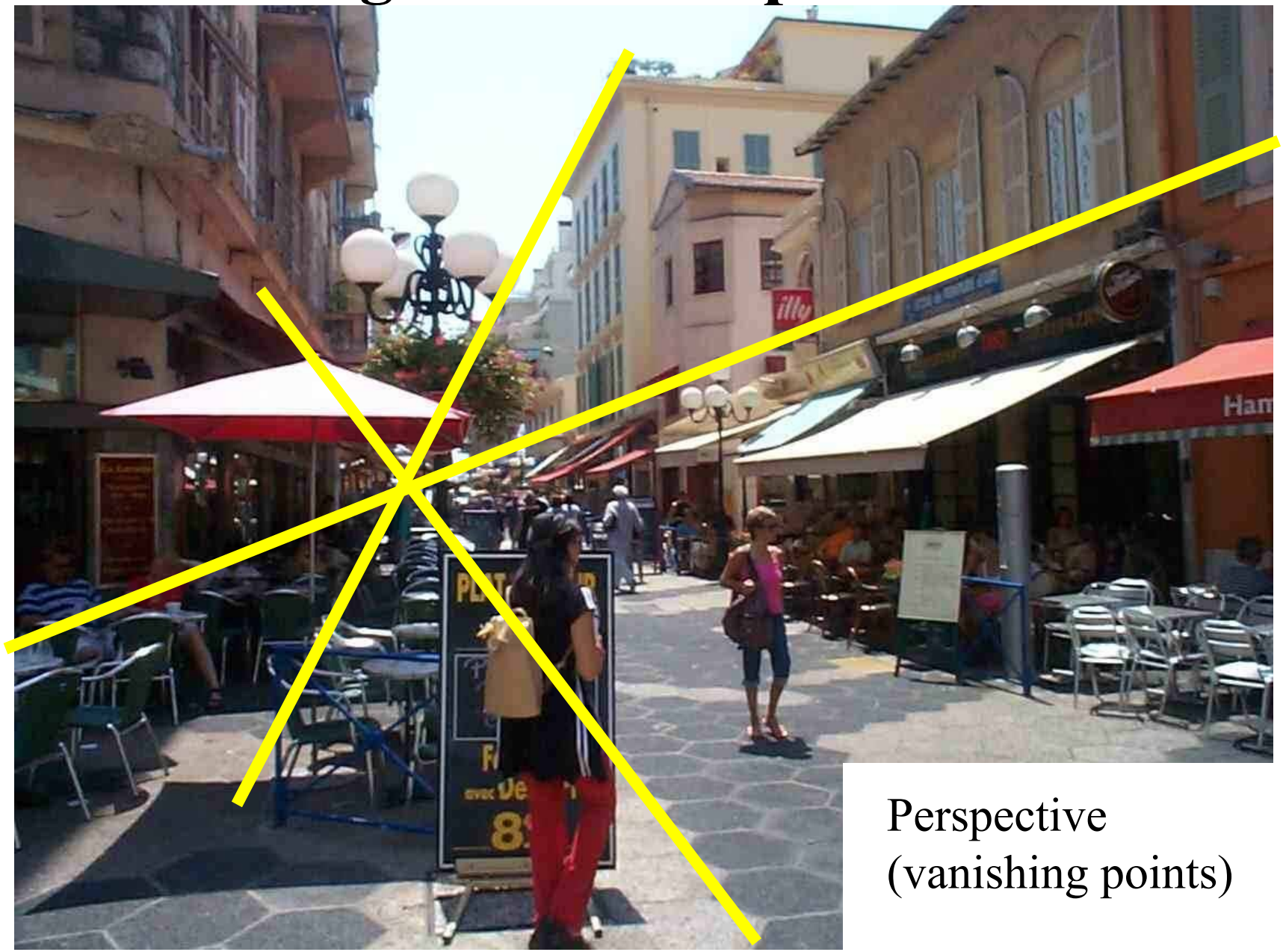
An early debate: do we infer depth from higher-level information like perspective and contours, or does it occur at a much lower level?

"The basis of this three-dimensional perception was hotly debated between Wheatstone and fellow physicist Sir David Brewster. (Though it may seem odd for physicists to concern themselves with the physiology of optics, this was felt to be a natural extension of the study of the physics of optics.) Brewster opined that perspective was the source of the apprehension of an object's shape. Wheatstone insisted that the images in the each eye had identifiable landmarks that were combined to assign depth to the landmarks."

-- Ralph M. Siegel Choices: The Science of Bela Julesz

Robert Collins
CSE486, Penn State

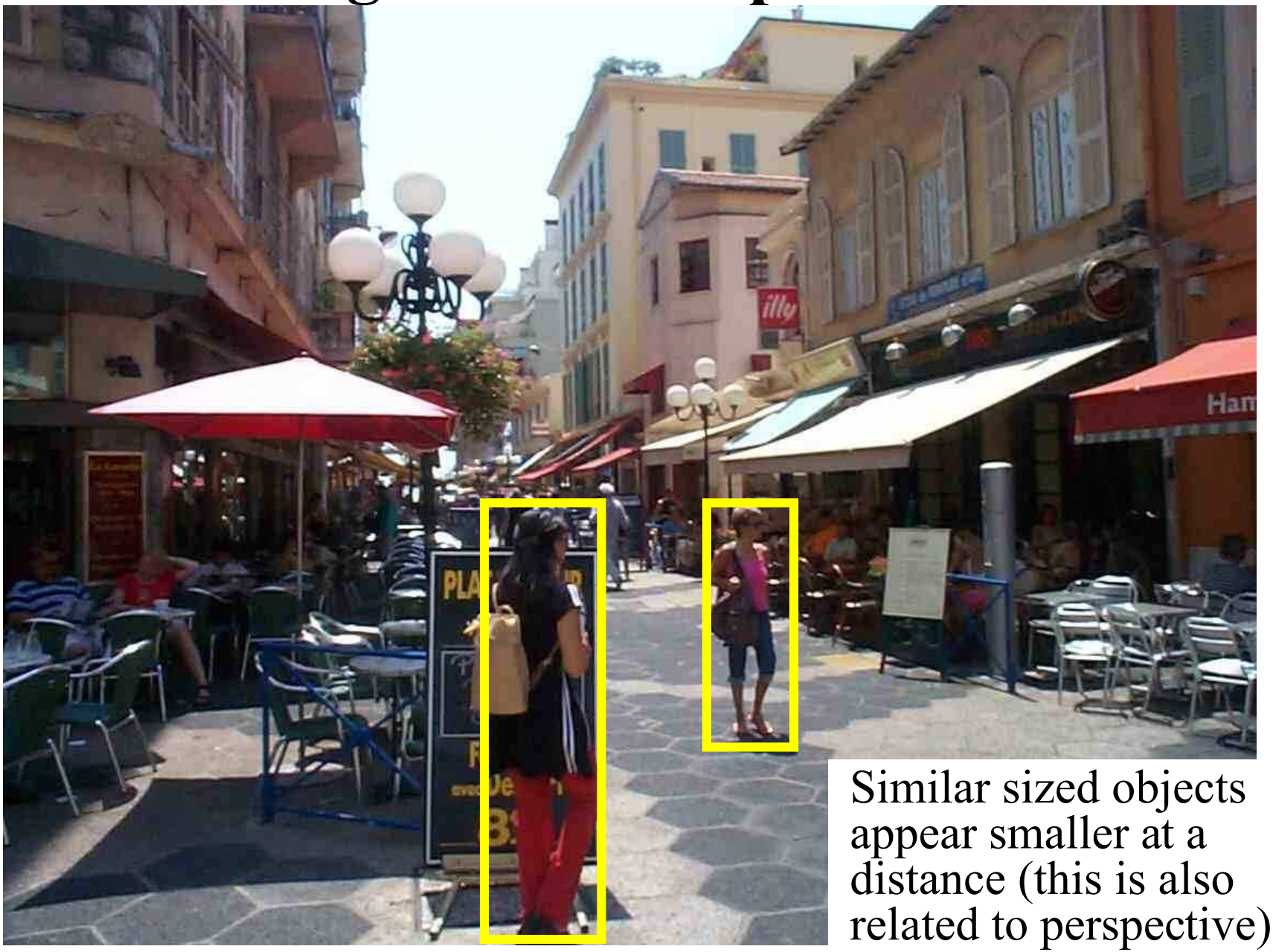
Higher-level Depth Cues



Perspective
(vanishing points)

Robert Collins
CSE486, Penn State

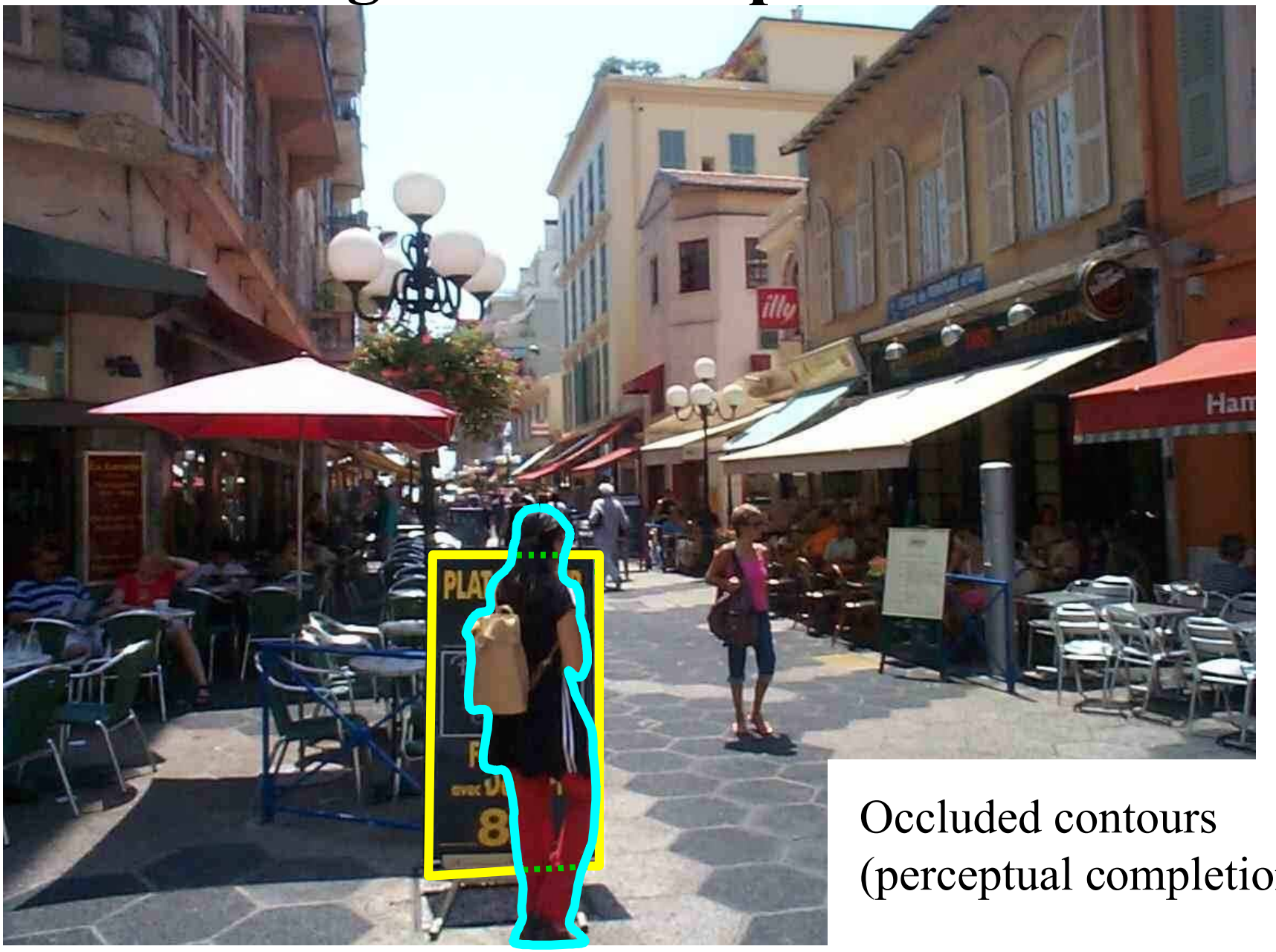
Higher-level Depth Cues



Similar sized objects appear smaller at a distance (this is also related to perspective)

Robert Collins
CSE486, Penn State

Higher-level Depth Cues



Occluded contours
(perceptual completion)

Robert Collins
CSE486, Penn State

Stereo Pyschophysics

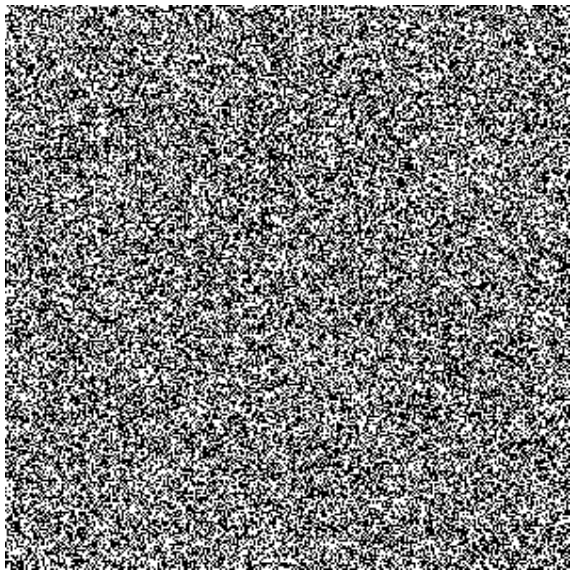
Obviously perspective and contours are important, (particularly for monocular depth perception), but are they necessary for binocular stereo depth perception?

Bela Julesz answered this question in 1960 with his experiments with random dot stereograms.

"In 1960, Bela's experiment with what eventually became known as Julesz random dot stereograms unambiguously demonstrated that stereoscopic depth could be computed in the absence of any identifiable objects, in the absence of any perspective, in the absence of any cues available to either eye alone." **-- Ralph M. Siegel** Choices: The Science of Bela Julesz

Julesz Random-Dot Experiment

Generate a random dot pattern using a computer

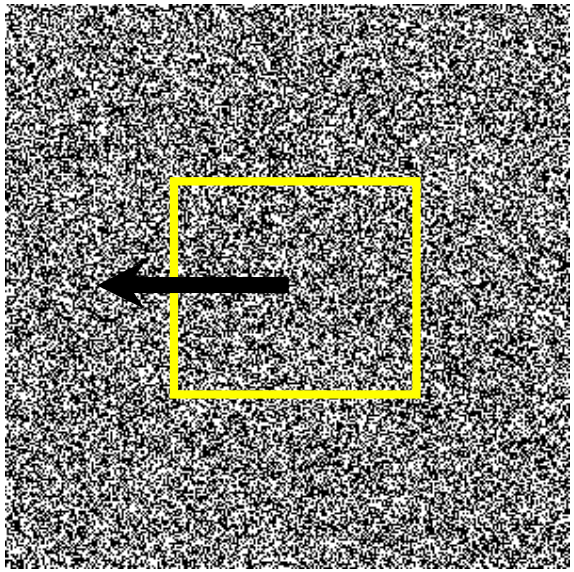


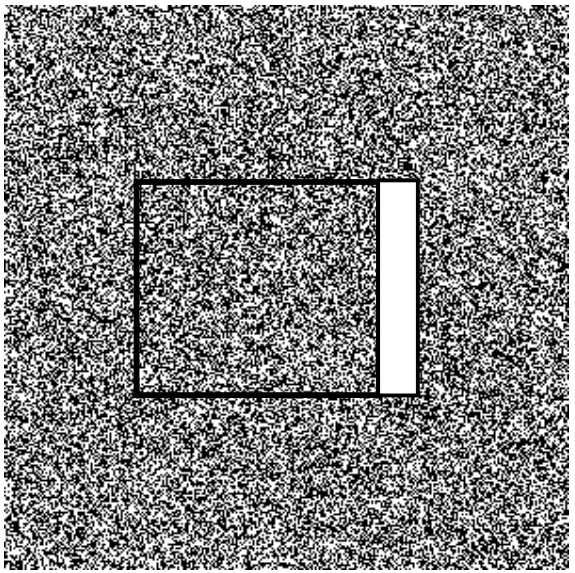
e.g. im = roicolor(rand(300,300), 0.5, 1);

By definition, this is just “noise”, so there are obviously no monocular depth cues here.

Robert Collins
CSE486, Penn State

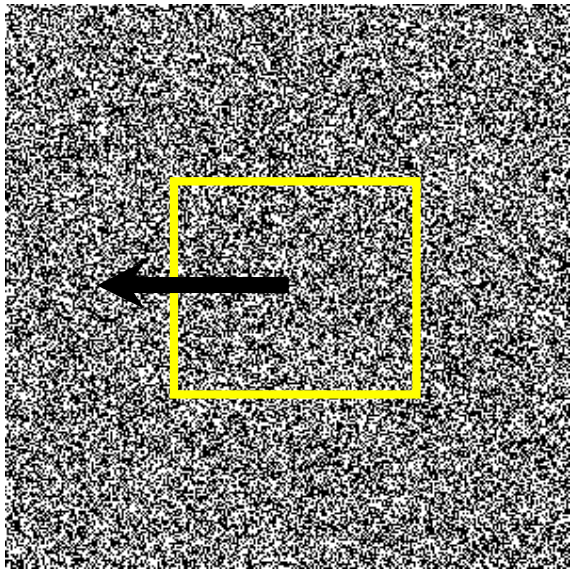
Julesz Random-Dot Experiment

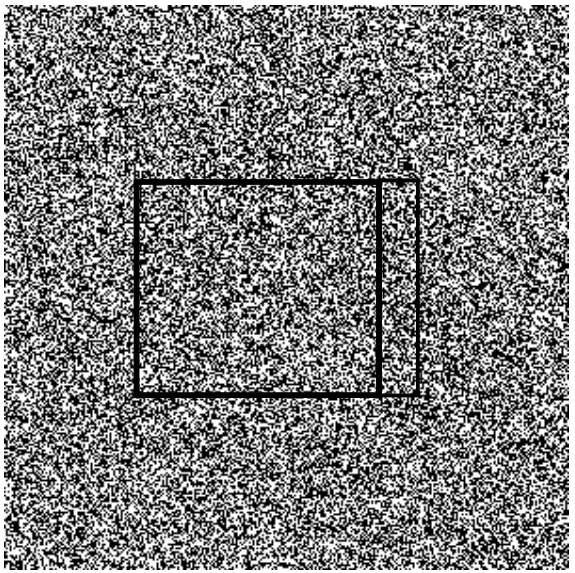




Clip out a square region and shift it to the left

Julesz Random-Dot Experiment



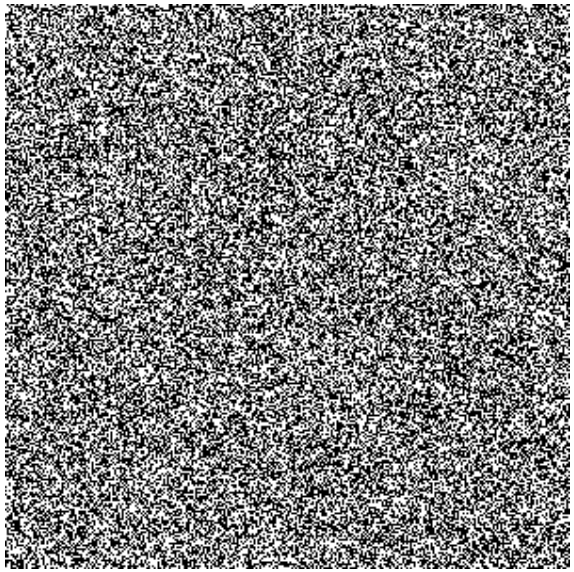


Clip out a square region and shift it to the left

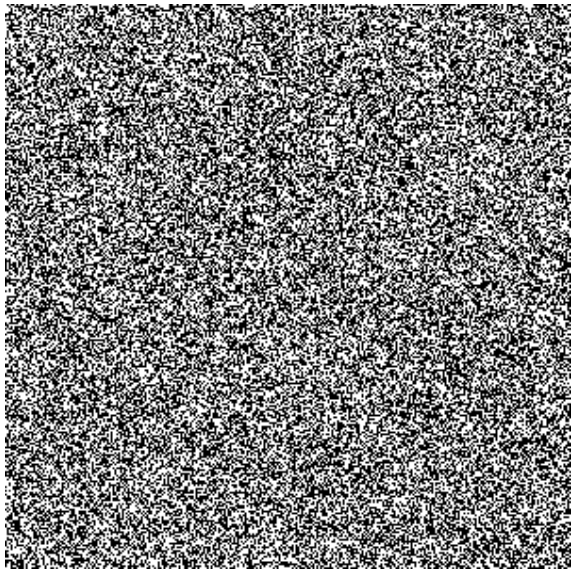
Fill in the “hole” left behind with more random dots.

Robert Collins
CSE486, Penn State

Julesz Random-Dot Experiment



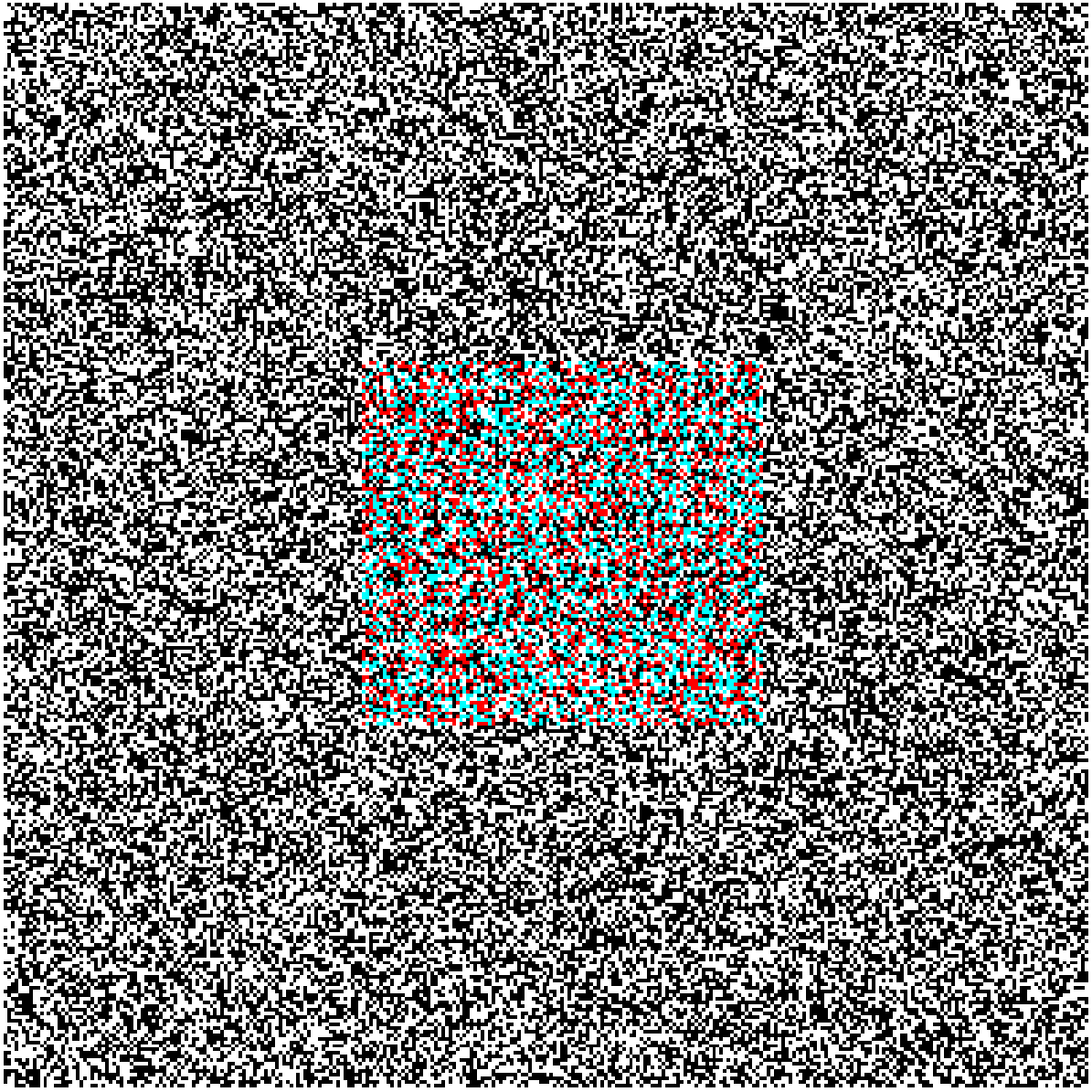
Original dot image



Dot image with shifted square

Now view as a stereo pair.
Julesz used a special viewer, but we will display as an anaglyph (get your glasses!)

Robert Collins
CSE486, Penn State



Robert Collins
CSE486, Penn State

Make Your Own

%make an image with random dots
```
im = roicolor(rand(300,300),.5,1);
```
%second image starts as a copy of that
```
im2 = im;
```
%shift a square of pixels to the right
```
im2(100:200,110:210) = im(100:200,100:200);
```
%fill in the "hole" with more random dots
```
im2(100:200,100:110) = roicolor(rand(101,11),.5,1);
```

%encode image2 in red channel of a color image
```
ana = 255*im2;
```
%encode image1 in blue and green channels
```
ana(:,:,2) = 255*im;
ana(:,:,3) = 255*im;
```
%take a look (remember to wear your red/cyan glasses!)
```
image(uint8(ana))
```

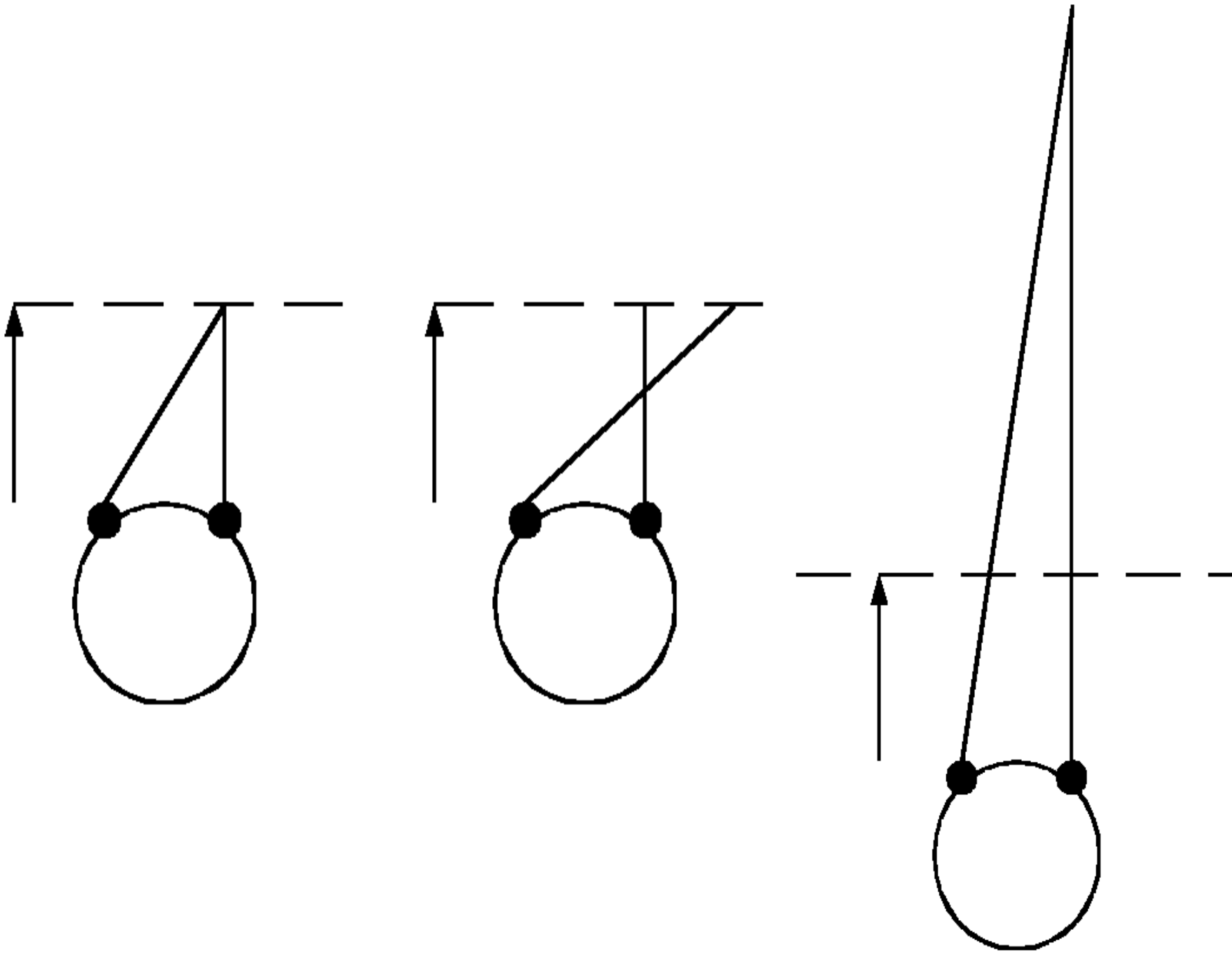
Try this: what happens when you shift the square to the left instead of to the right?

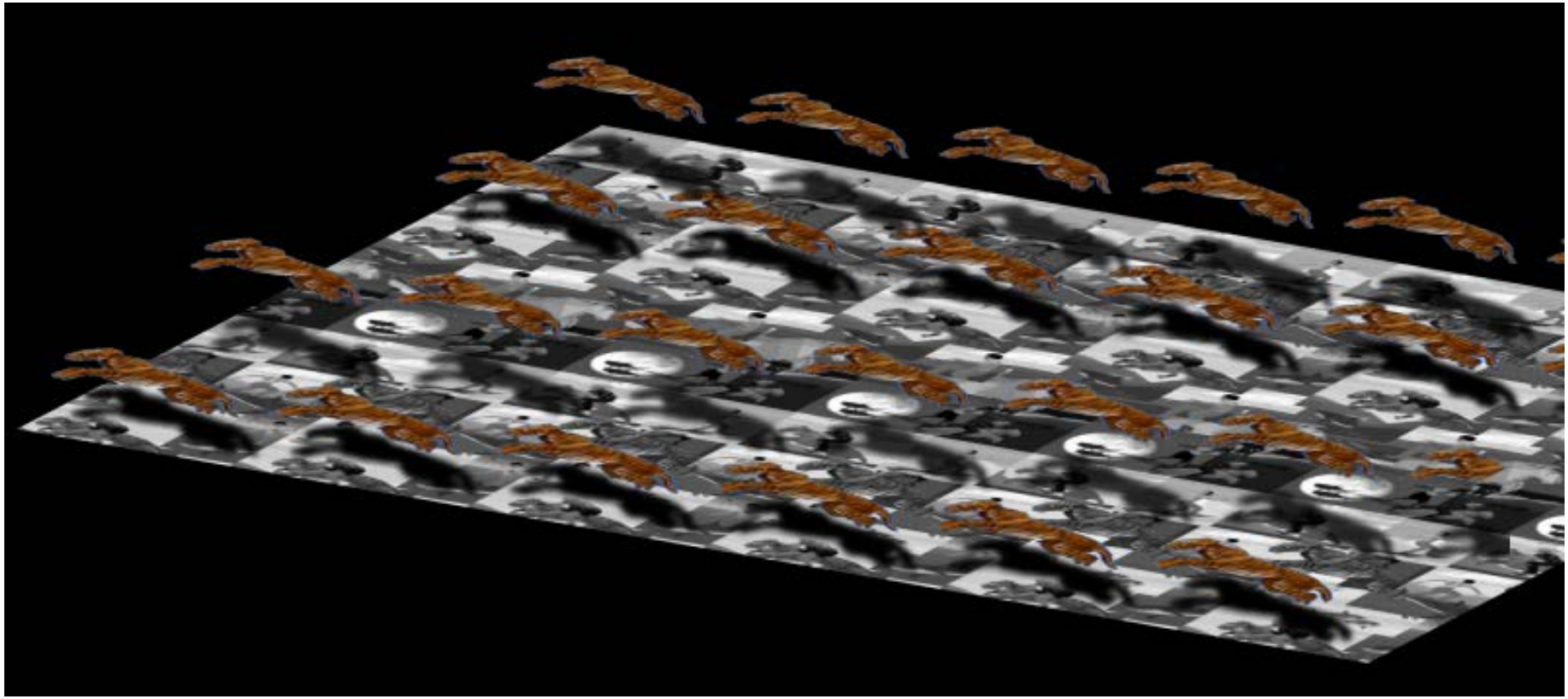
Robert Collins
CSE486, Penn State

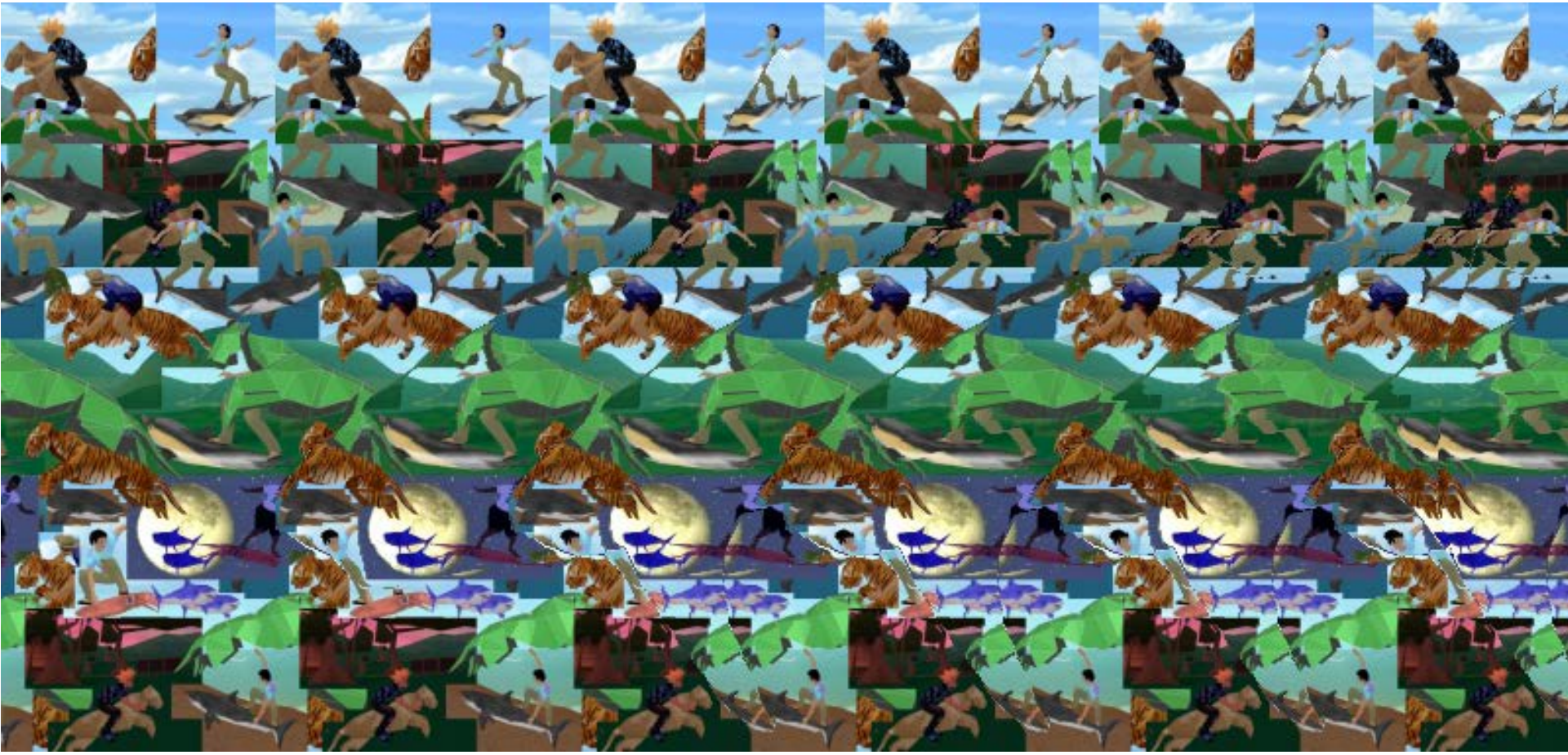
Stereograms

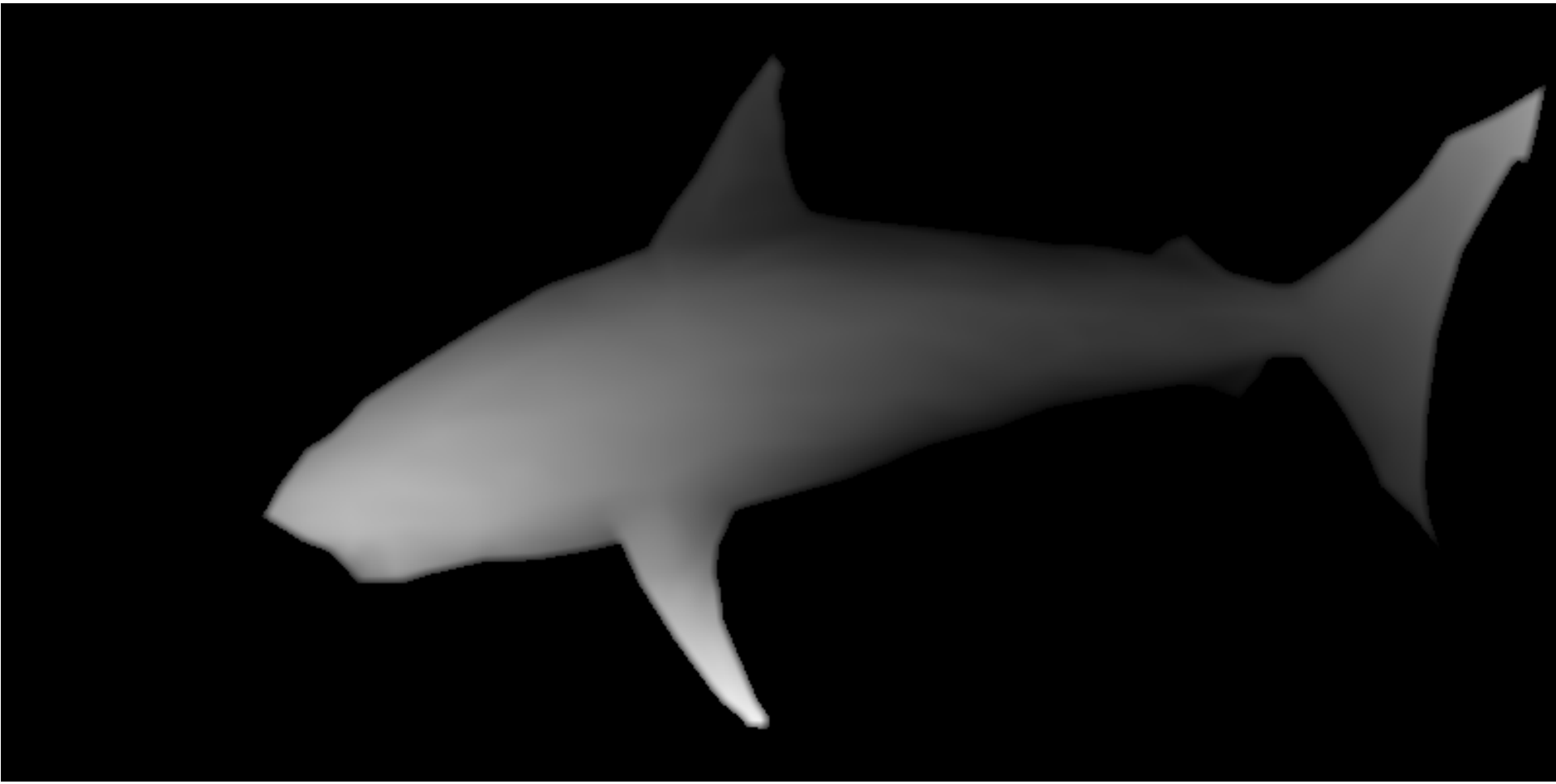
Another method of encoding parallax in a single image. Subtle shifts of repeated texture encode disparity of depths in a scene (a technique made famous under the “Magic Eye” brand name).

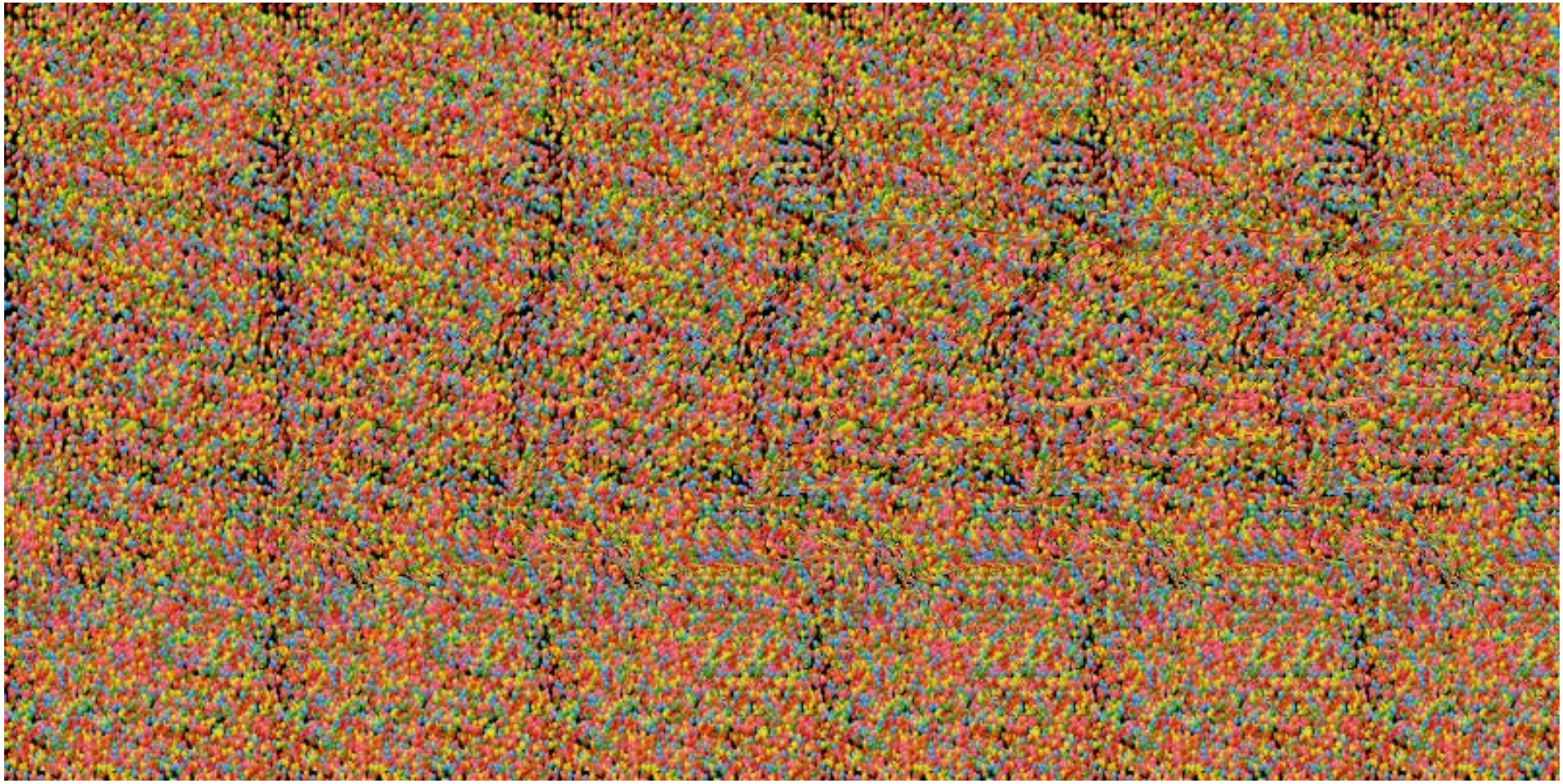
Unlike anaglyphs, you don’t need special glasses to see these, just some practice focusing your eyes behind the page.

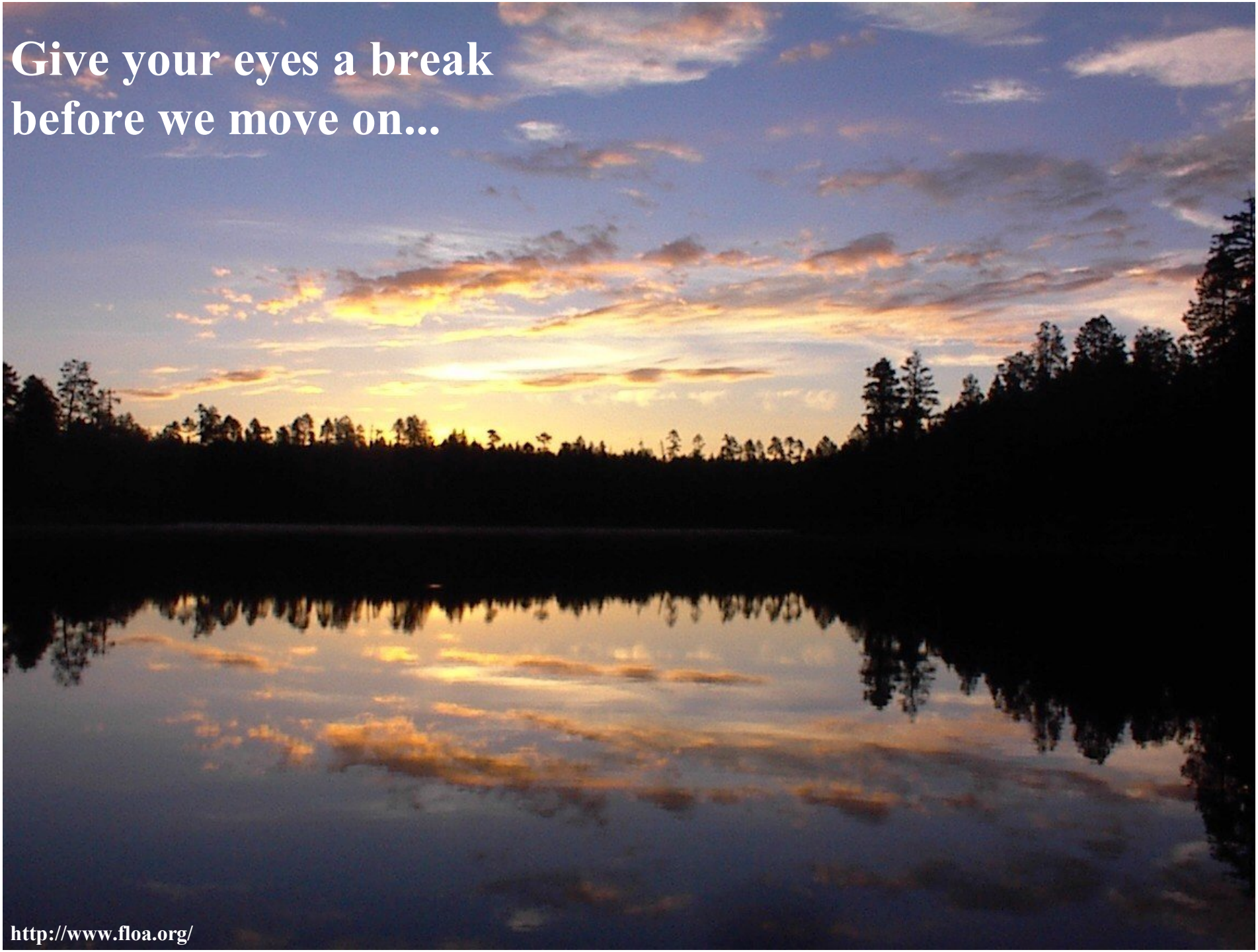






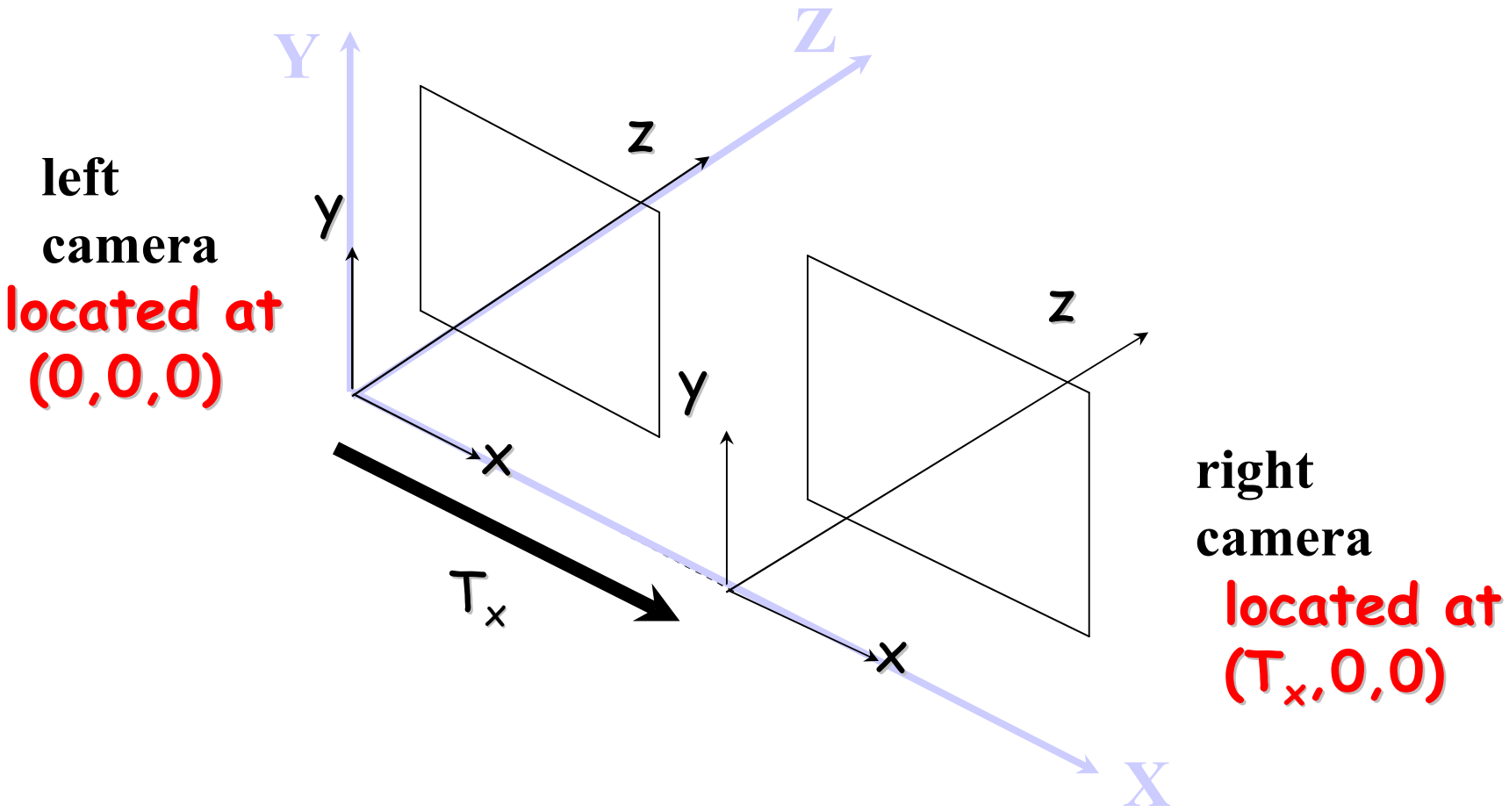





Give your eyes a break
before we move on...
http://www.floa.org/

Robert Collins
CSE486, Penn State

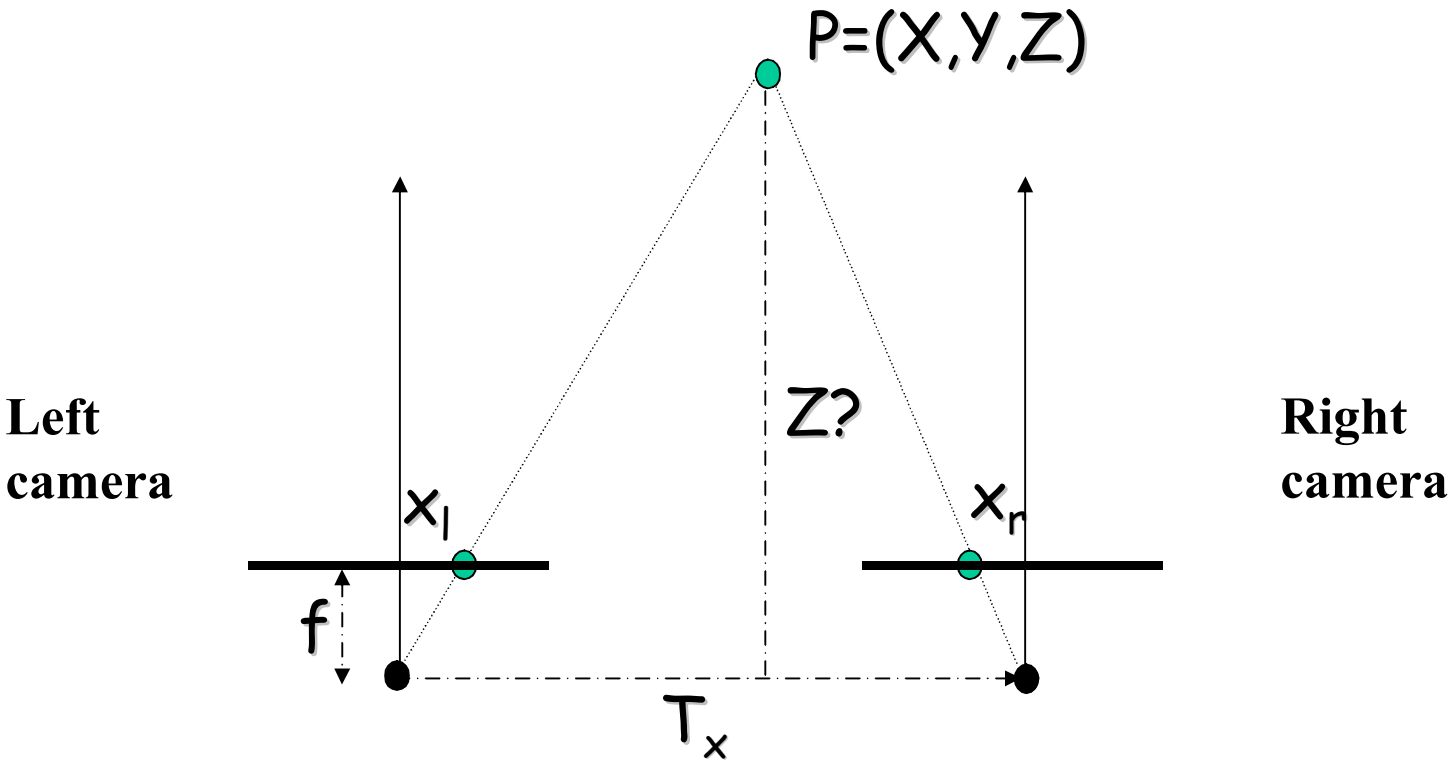
A Simple Stereo System



Right camera is simply shifted by Tx units along the X axis. Otherwise, the cameras are identical (same orientation / focal lengths)

Robert Collins
CSE486, Penn State

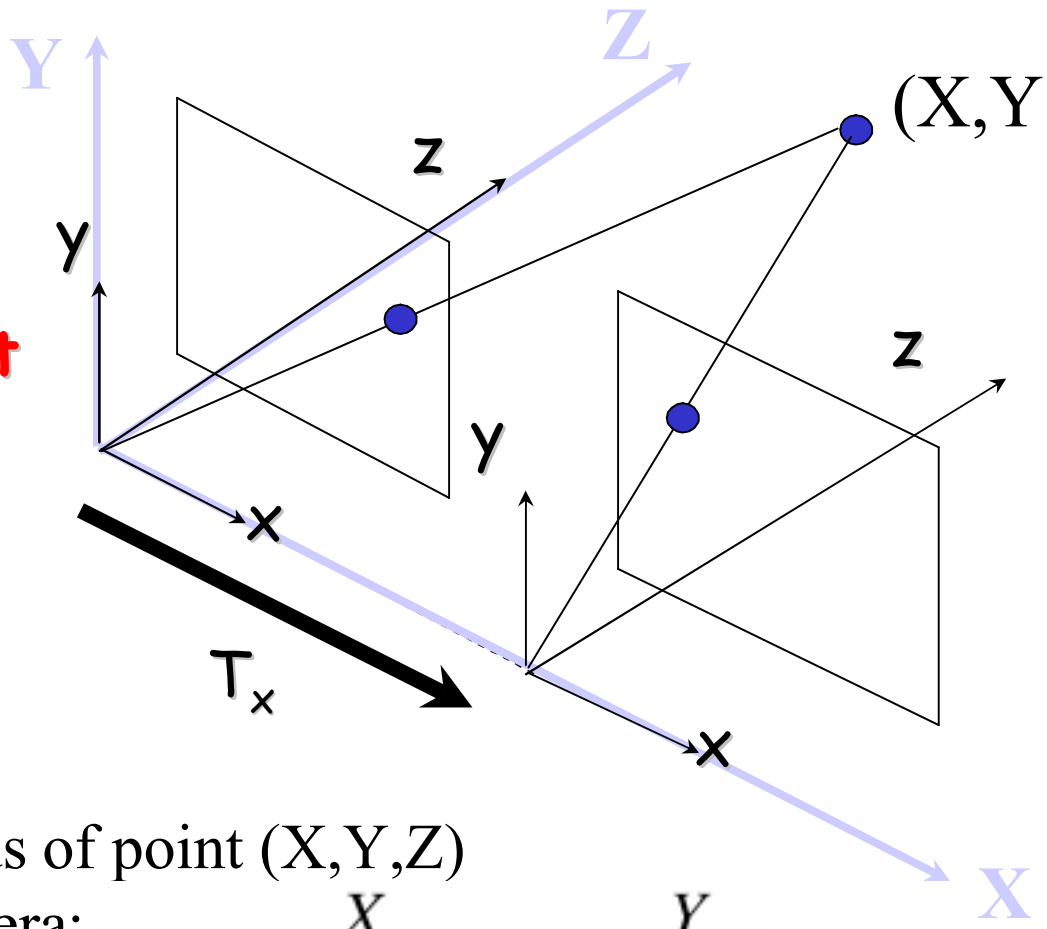
A Simple Stereo System
Top Down View (XZ plane)



Translated by a distance T_x along X axis
(T_x is also called the stereo "baseline")

Robert Collins
CSE486, Penn State

A Simple Stereo System



left camera located at (0,0,0)

right camera located at (T_x,0,0)

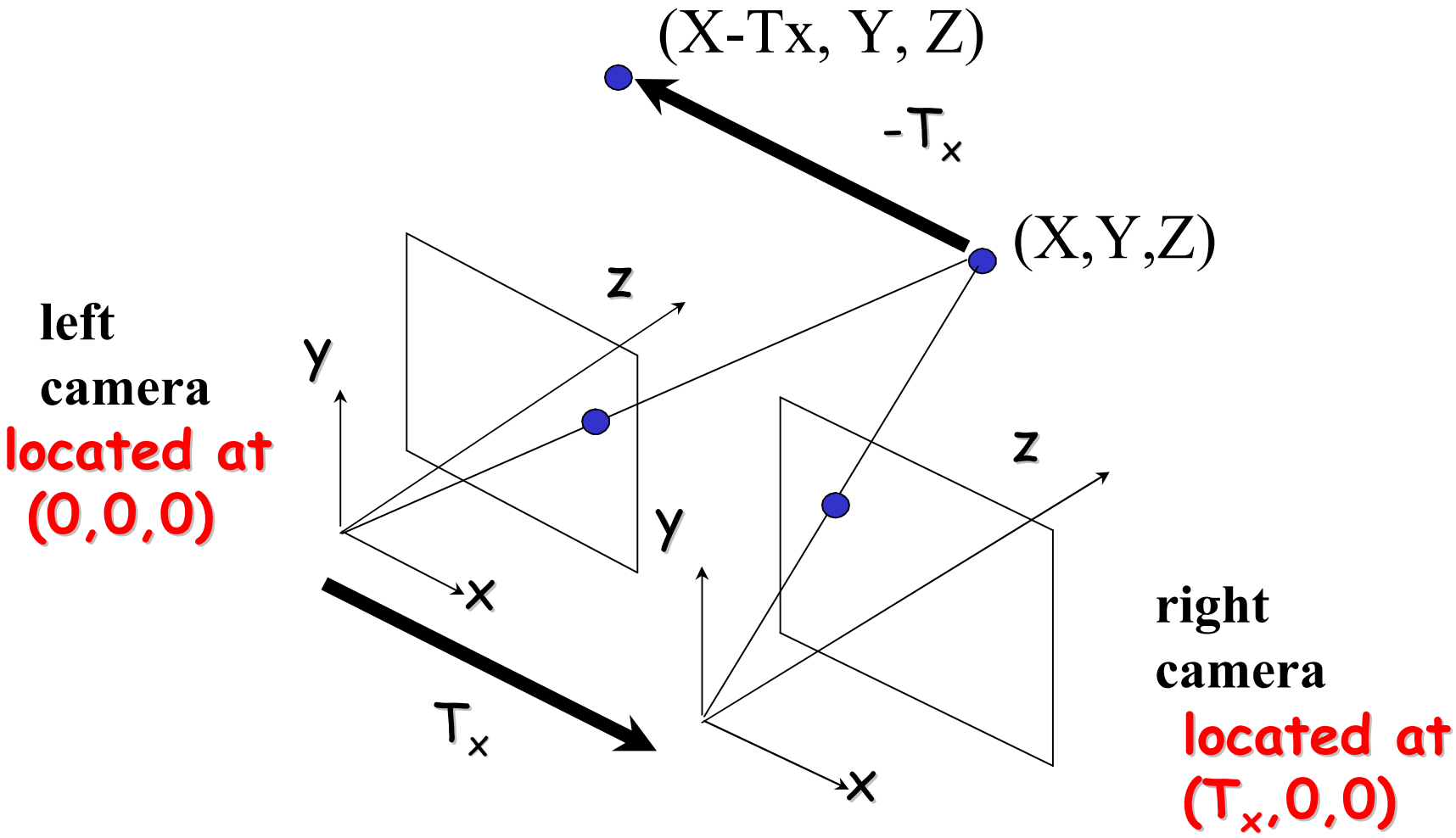
Image coords of point (X,Y,Z) in Left Camera:

$$x_l = f\frac{X}{Z} \qquad y_l = f\frac{Y}{Z}$$

What are image coords of that same point in the Right Camera?

Robert Collins
CSE486, Penn State

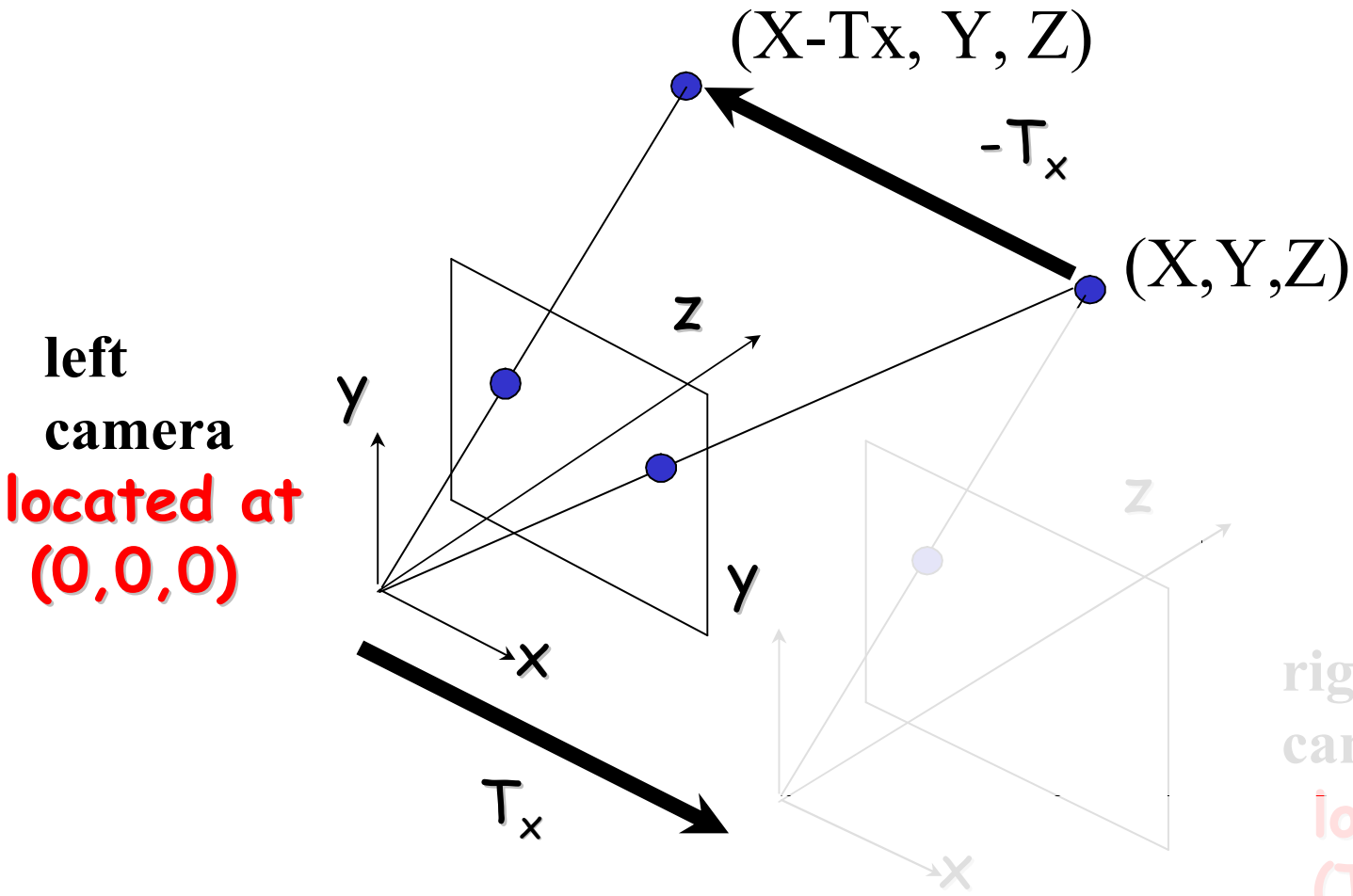
A Simple Stereo System



Insight: translating camera to the right by Tx is equivalent to leaving the camera stationary and translating the world to the left by Tx.

Robert Collins
CSE486, Penn State

A Simple Stereo System



$$x_r = f\frac{X - T_x}{Z} \qquad y_r = f\frac{Y}{Z}$$

Robert Collins
CSE486, Penn State

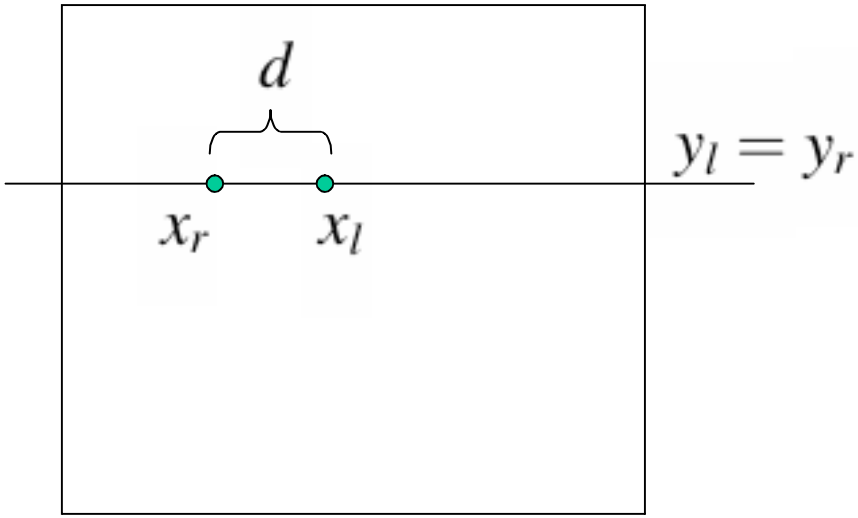
Stereo Disparity

Left camera

$$x_l = f\frac{X}{Z} \qquad y_l = f\frac{Y}{Z}$$

Right camera

$$x_r = f\frac{X - T_x}{Z} \qquad y_r = f\frac{Y}{Z}$$



Stereo Disparity

$$d = x_l - x_r = f\frac{X}{Z} - \left(f\frac{X}{Z} - f\frac{T_x}{Z}\right)$$

$$d = \frac{f\,T_x}{Z}$$

depth **baseline**

$$Z = \frac{f\,T_x}{d}$$

disparity

Important equation!

Robert Collins
CSE486, Penn State

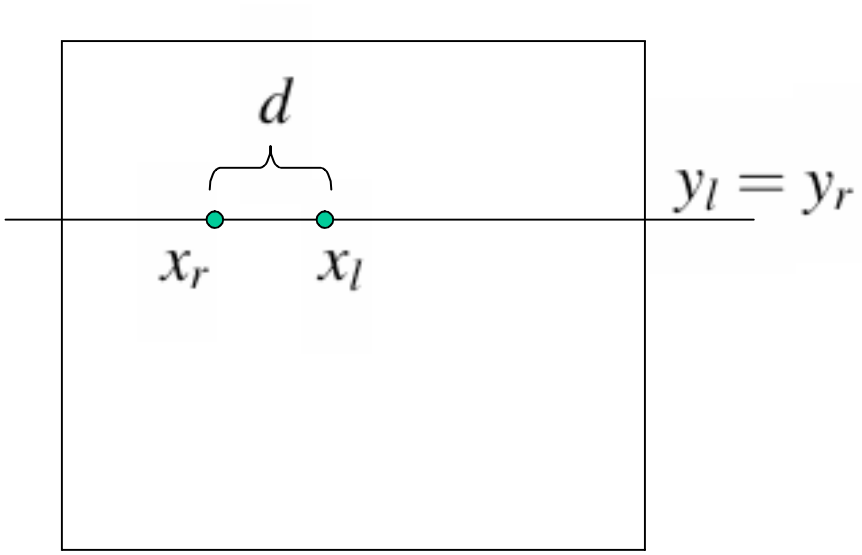
Stereo Disparity

Left camera

$$x_l = f\frac{X}{Z} \qquad y_l = f\frac{Y}{Z}$$

Right camera

$$x_r = f\frac{X - T_x}{Z} \qquad y_r = f\frac{Y}{Z}$$



Note: Depth and stereo disparity are inversely proportional

depth

$$Z = \frac{f\,T_x}{d}$$

disparity

Important equation!

Robert Collins
CSE486, Penn State

Stereo Disparity / Parallax

Tie in with Intro: for our purposes

Disparity = Parallax

$\Rightarrow$Disparity/Parallax inversely proportional to depth

$\Rightarrow$ this is why near objects appear to move more than far away ones when the camera translates sideways