CS 4495 Computer Vision

Stereo: Disparity and Matching

Aaron Bobick
School of Interactive Computing

Administrivia

- PS2 will be out tomrrow. Due Sunday Sept 22nd , 11:55pm

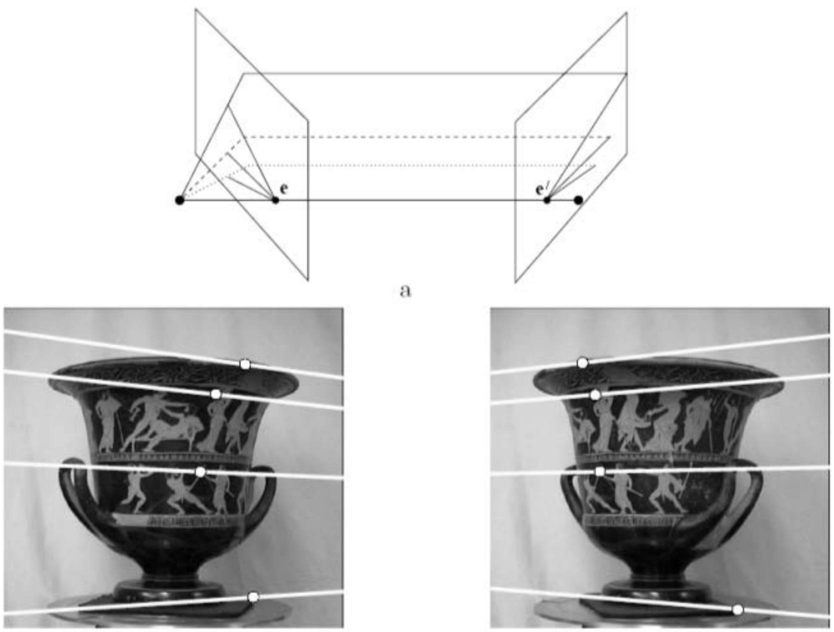
- There is *no* grace period. We can either:
  a) leave submission open and have 50% penalty
  b) or close it, require email and have 50% penalty

You choose…

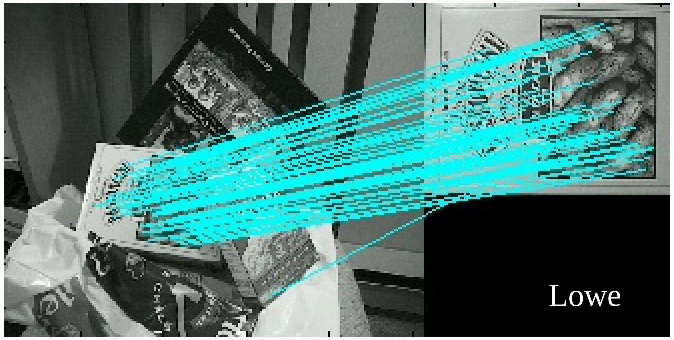
- Read; FP chapter 7

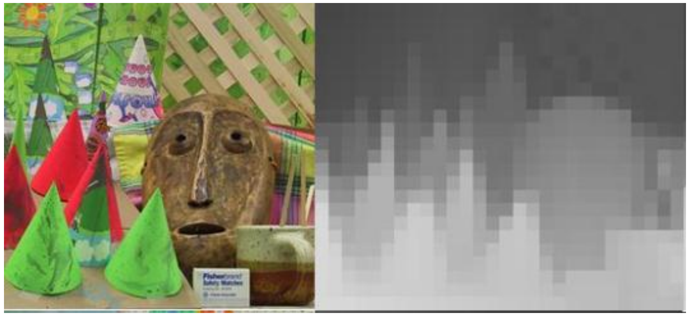
Stereo: A Special case of Multiple views

Multi-view geometry, matching, invariant features, stereo vision



Hartley and Zisserman





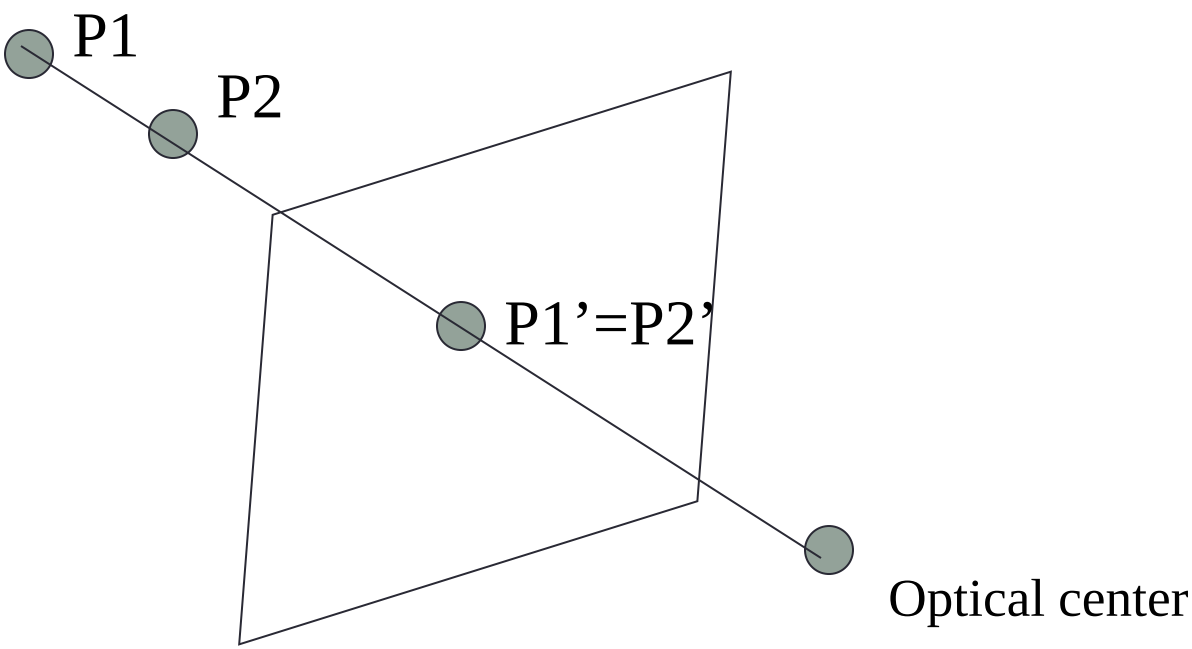
Why multiple views?

- 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.



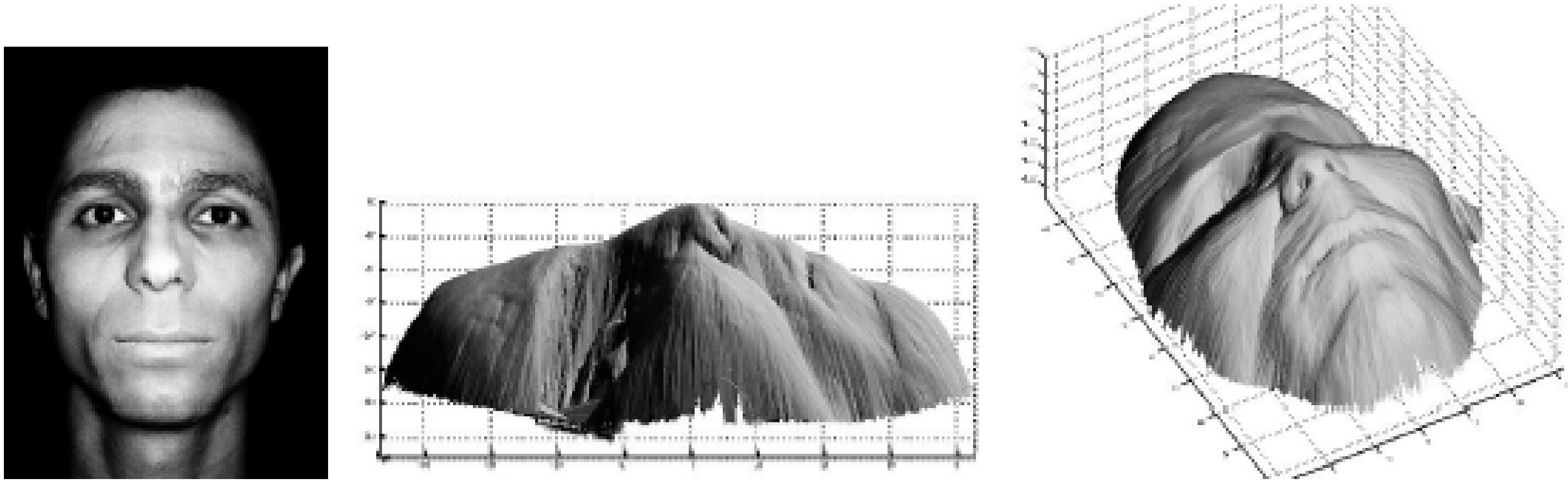
How do we see depth?

- What cues help us to perceive 3d shape and depth?
- What about one eye first?

Perspective effects

S. Seitz

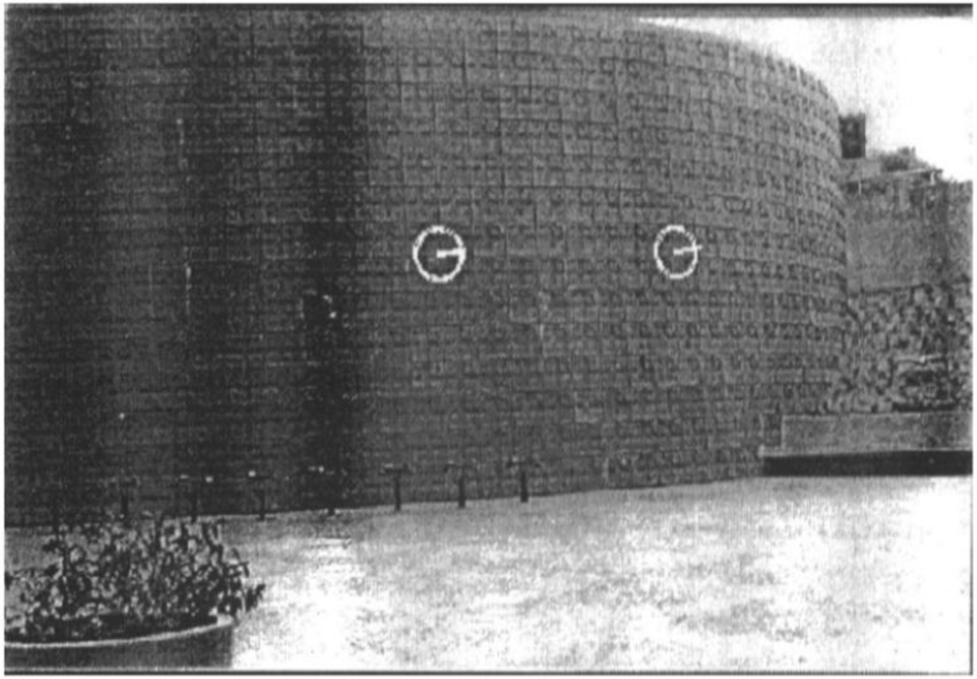
Shading

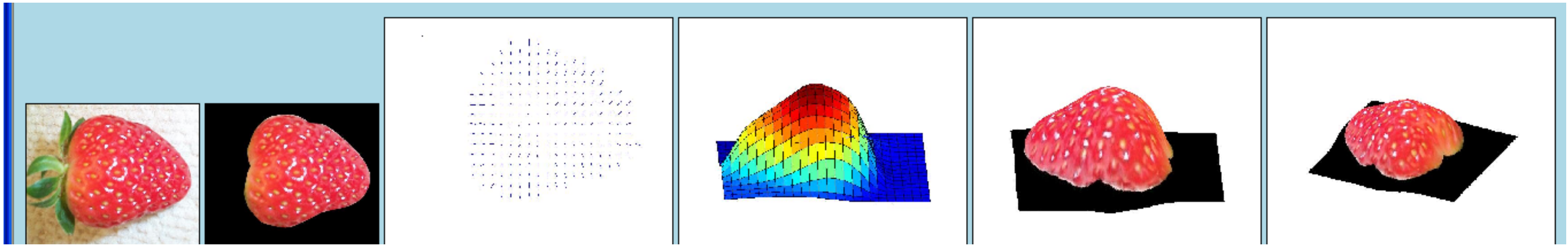


a) b) c)

K. Grauman

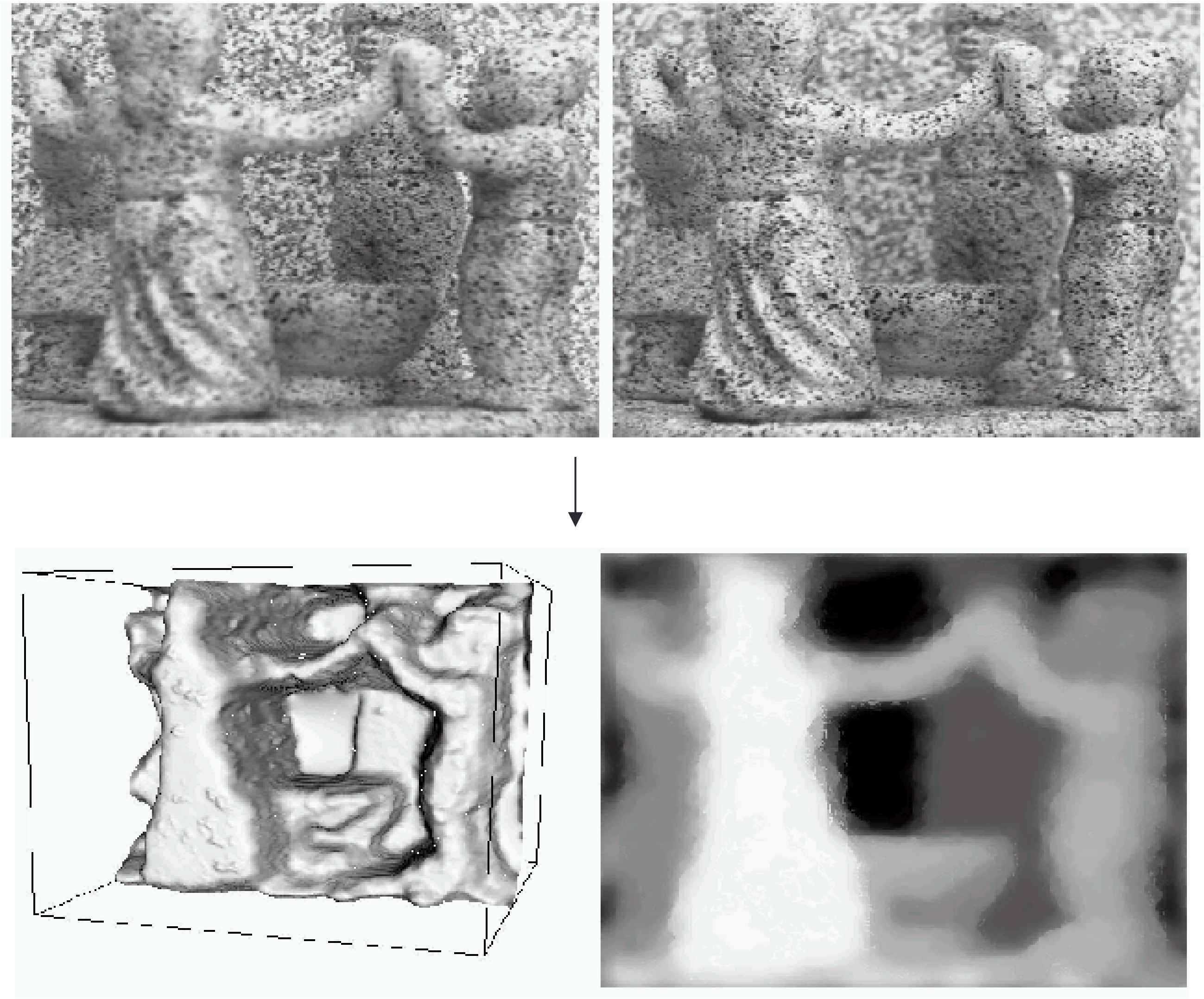
Texture





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

Focus/defocus

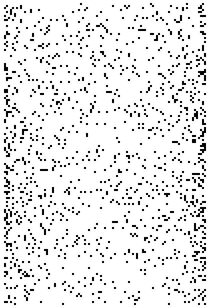


Images from same point of view, different camera parameters

3d shape / depth estimates

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

Motion



Figures from L. Zhang

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

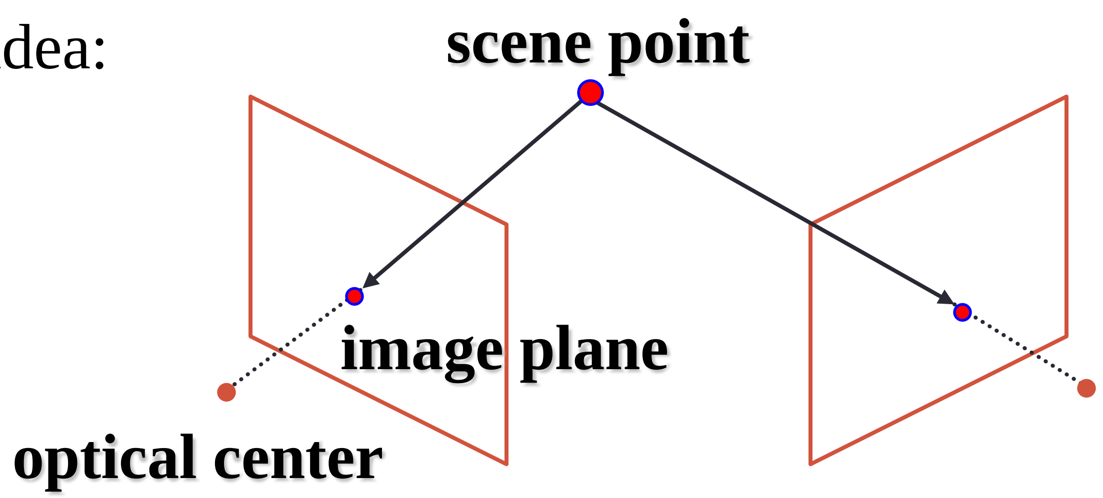
Estimating scene shape from one eye

- “Shape from X”: Shading, Texture, Focus, Motion…
- Very popular circa 1980

But we (and lots of creatures) have two eyes!

- **Stereo**:
  - shape from “motion” between two views
  - infer 3d shape of scene from two (multiple) images from different viewpoints

Main idea:

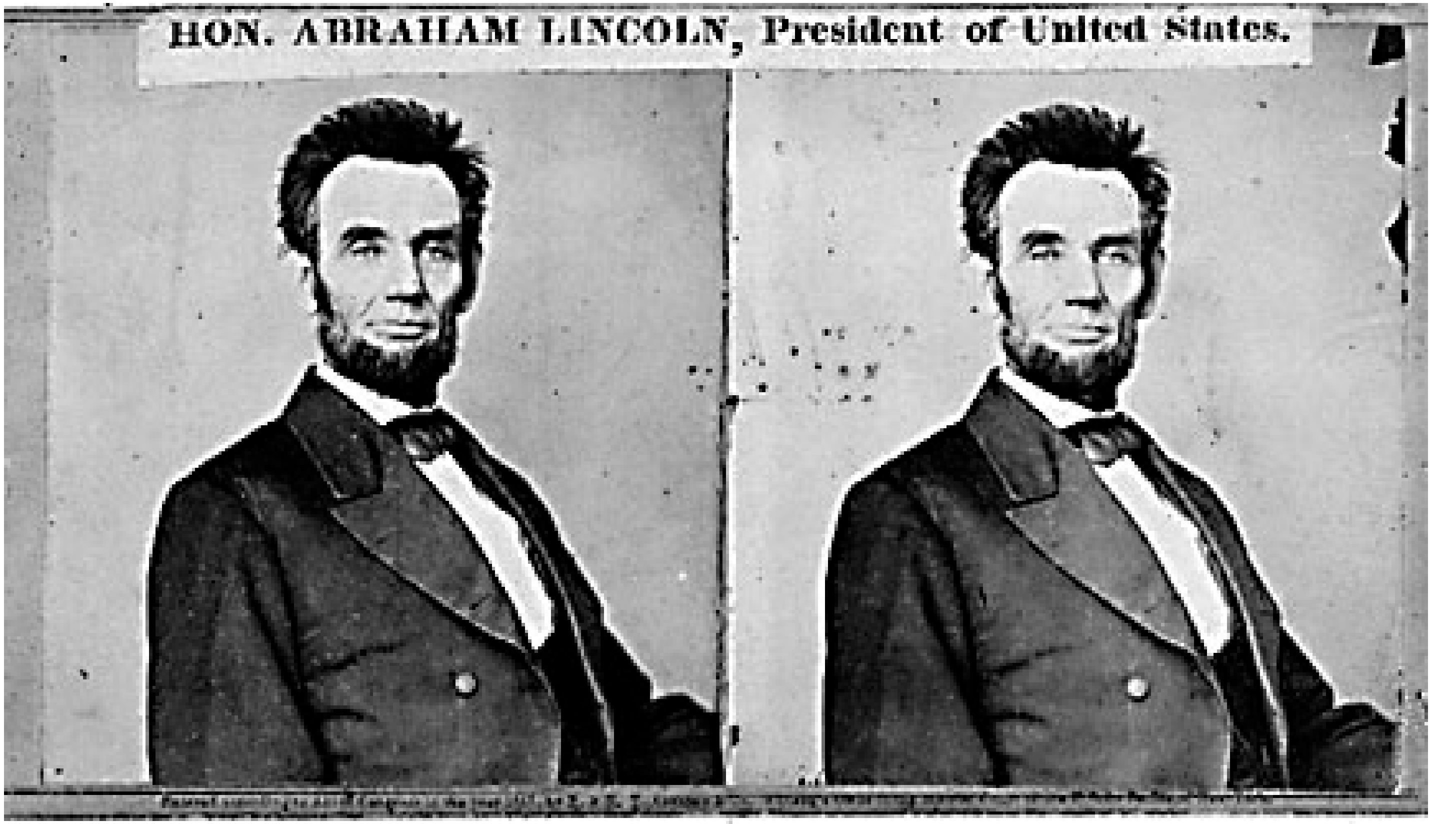


Stereo photography and stereo viewers

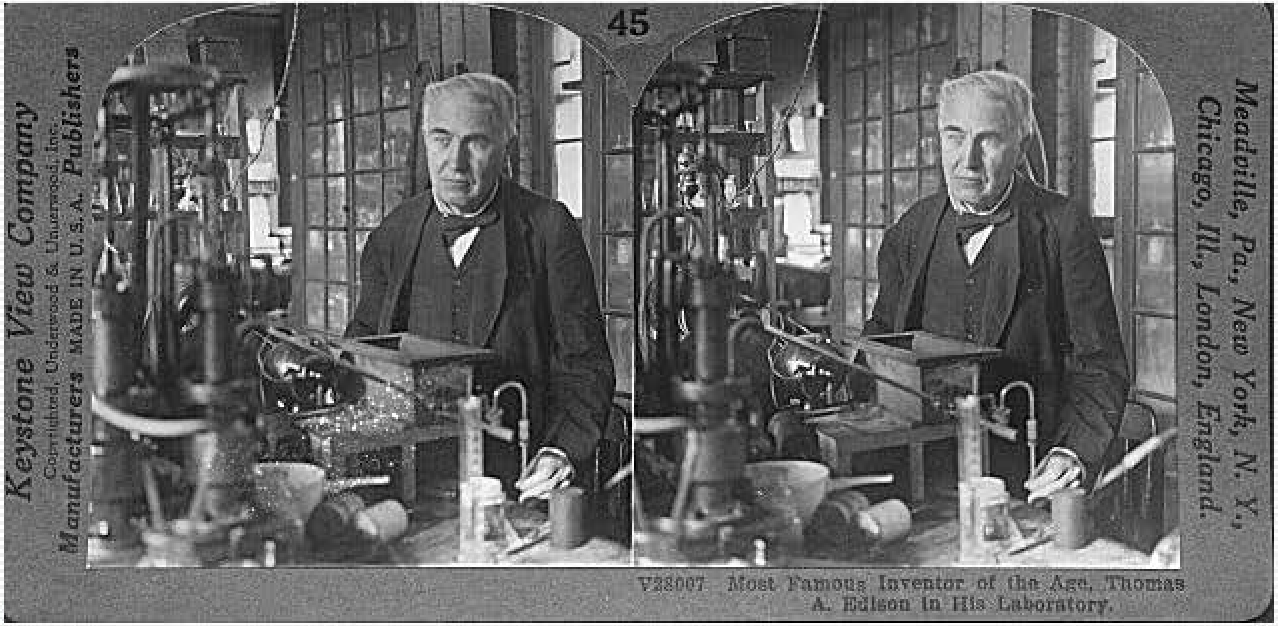
Take two pictures of the same subject from two slightly different viewpoints and display so that each eye sees only one of the images.

Invented by Sir Charles Wheatstone
1838

People fascinated by 3D



© Copyright 2001 Johnson-Shaw Stereoscopic Museum



http://www.johnsonshawmuseum.org

Public Library, Stereoscopic Looking Room, Chicago, by Phillips, 1923

Teesta suspension bridge-Darjeeling, India

Mark Twain at Pool Table", no date, UCR Museum of Photography

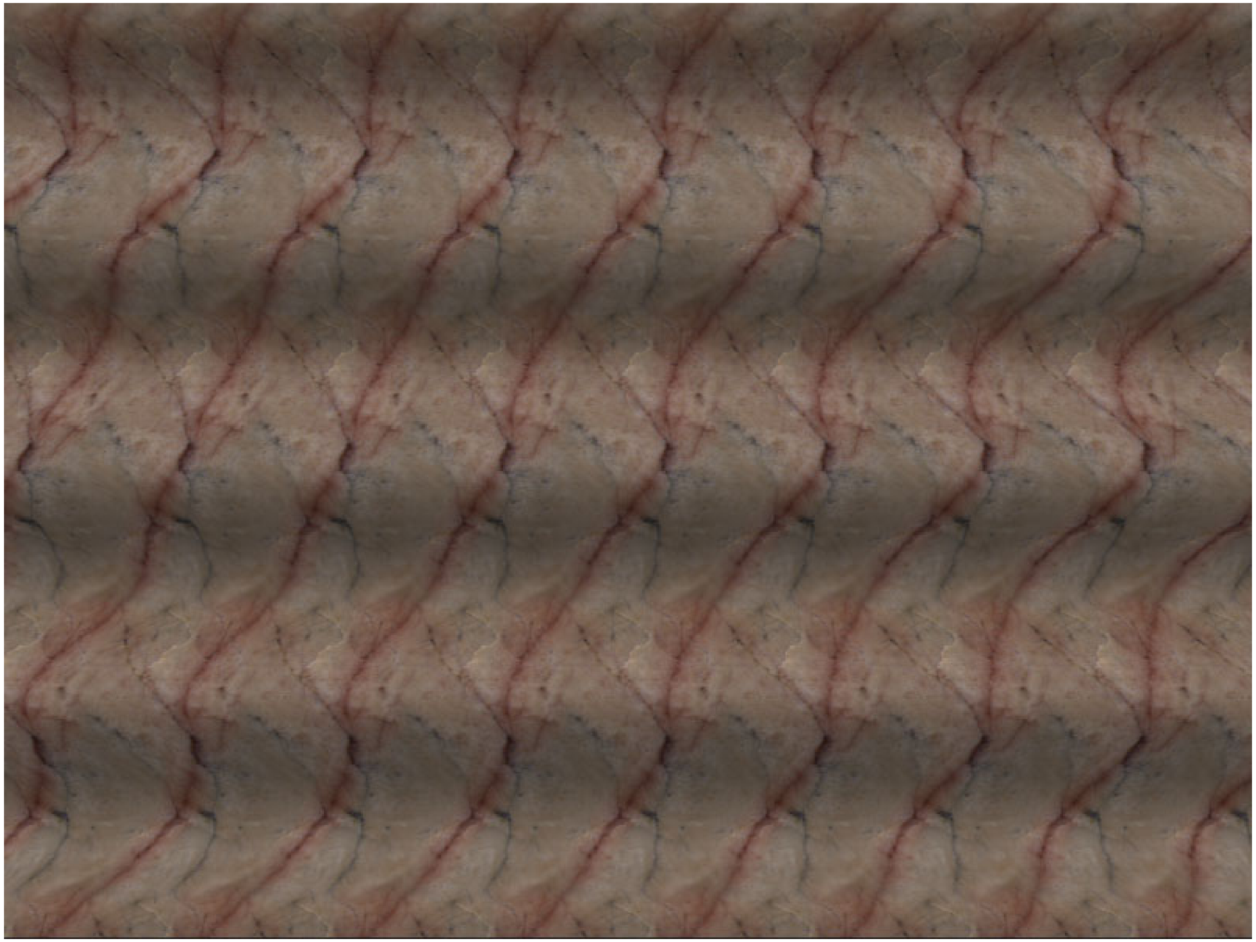
Stereo photography and stereo viewers

When I grew up…

Stereo photography and stereo viewers

If you like to cross (wall-eye) your eyes…

Single Image Stereo: *Autostereogram*



Single image stereogram, by Niklas Een

S. Seitz

The Basic Idea: Two slightly different images

http://www.well.com/~jimg/stereo/stereo_list.html

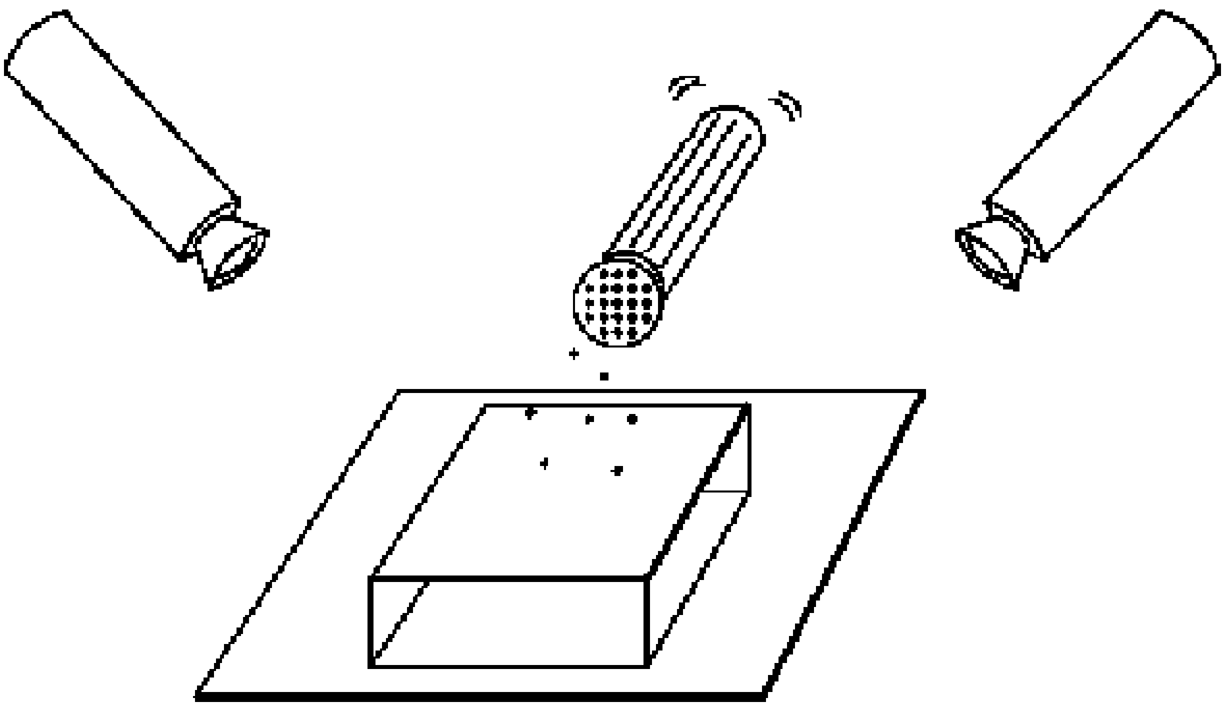
So how do humans do it?

Random dot stereograms

- Julesz 1960: Do we identify local brightness patterns before fusion (monocular process) or after (binocular)?

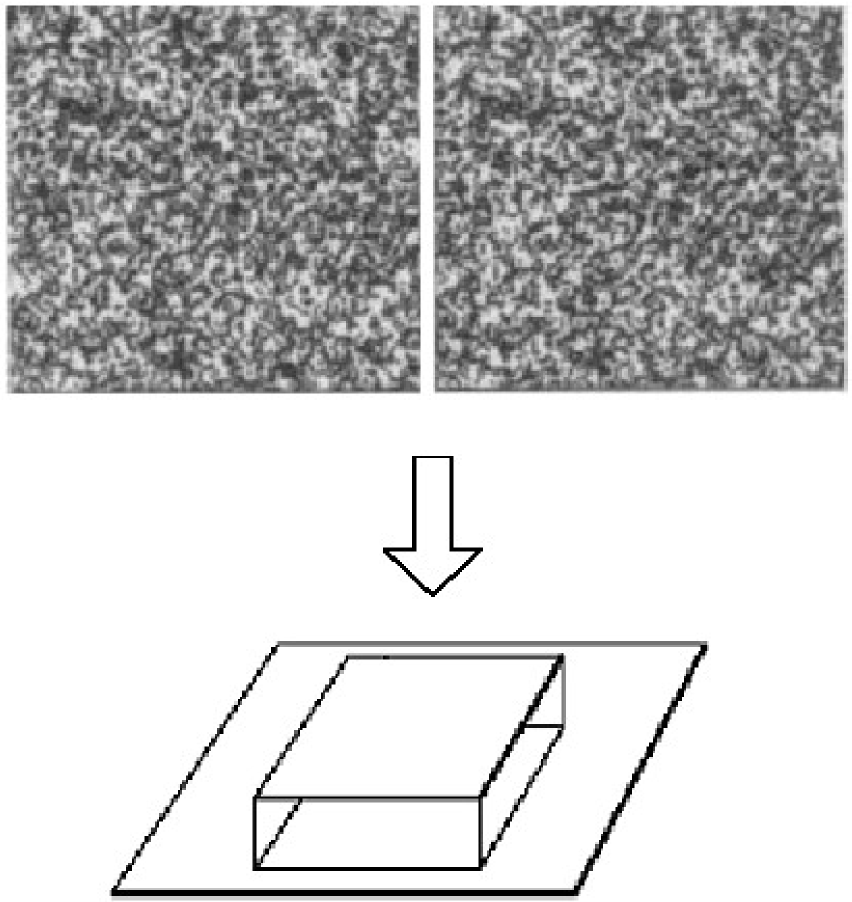
- To test: pair of synthetic images obtained by randomly spraying black dots on white objects

Random dot stereograms



Forsyth & Ponce

Random dot stereograms

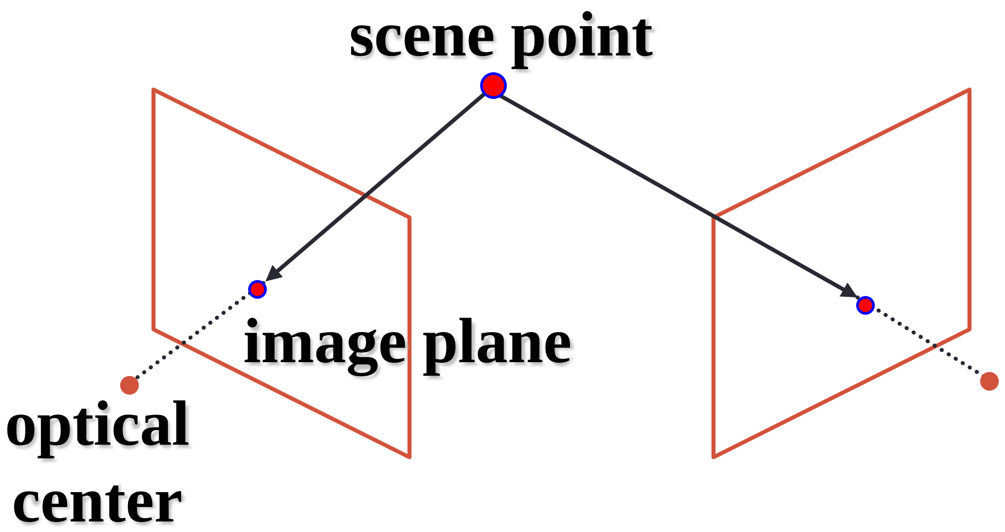


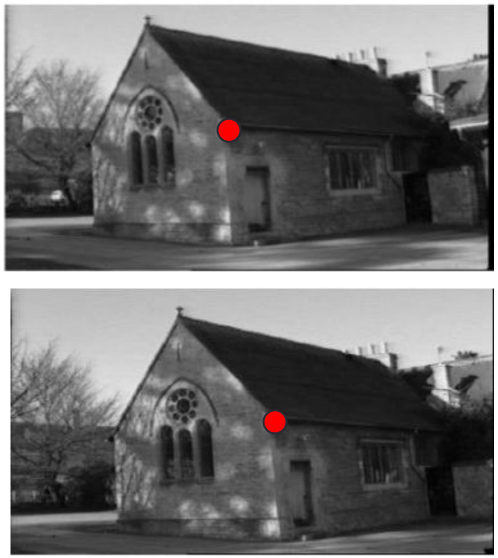
Random dot stereograms

- When viewed monocularly, they appear random; when viewed stereoscopically, see 3d structure.
- Conclusion: human binocular fusion not based upon matching large scale **structures** or any processing of the individual images
- Imaginary "*cyclopean retina*" that combines the left and right image stimuli as a single unit. Later discovered the cells in the brain's visual cortex that create this "percept"

Estimating depth with stereo

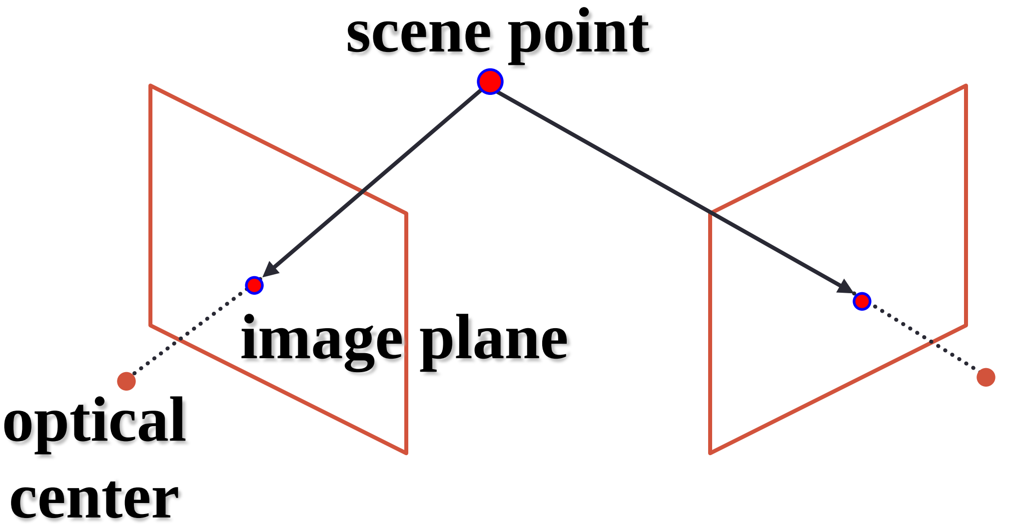
- **Stereo**: shape from "motion" between two views
- We'll need to consider:
  - Info on camera pose ("calibration")
  - Image point correspondences

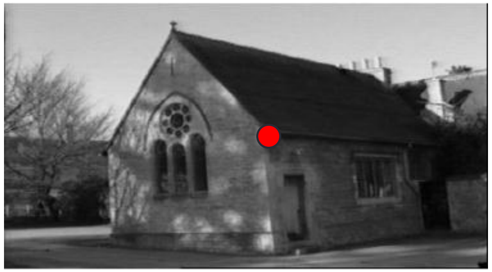




Estimating depth with stereo

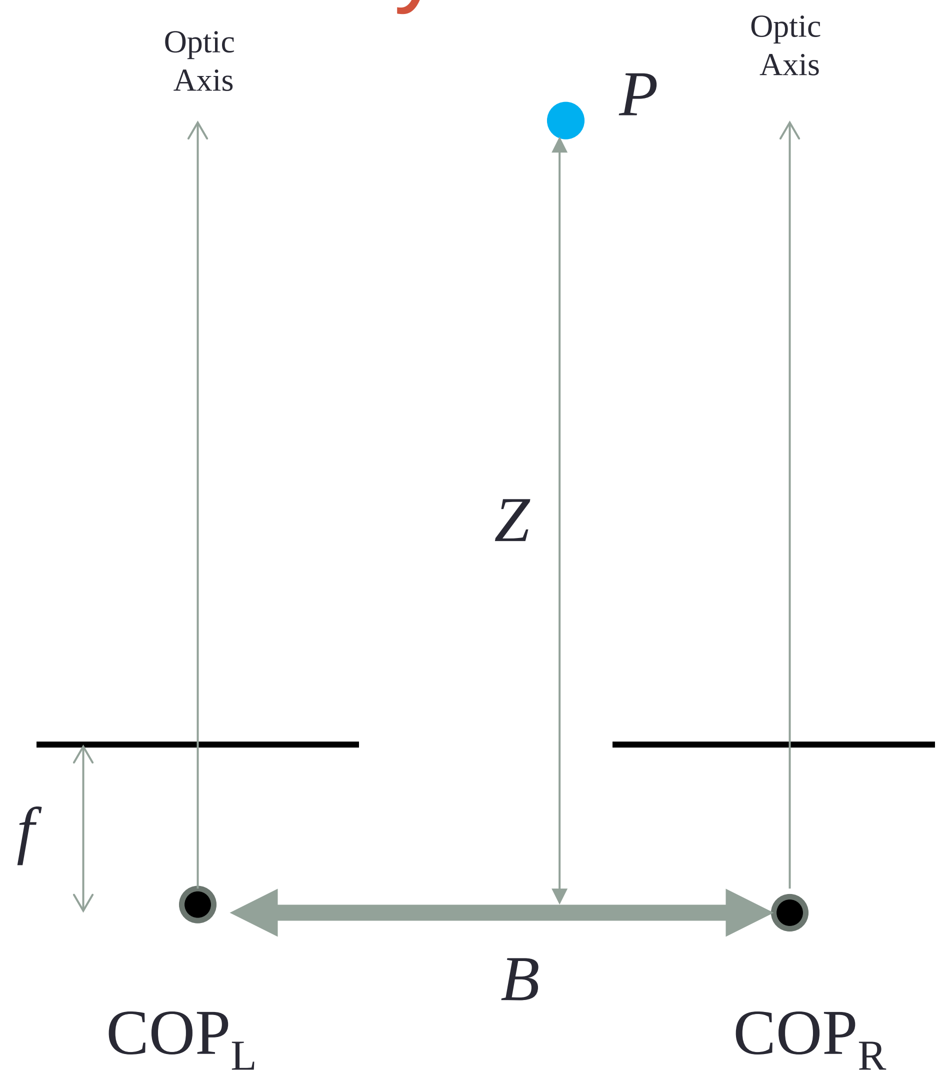
- **Stereo**: shape from “motion” between two views
- We’ll need to consider:
  - Info on camera pose (“calibration”)
  - Image point correspondences





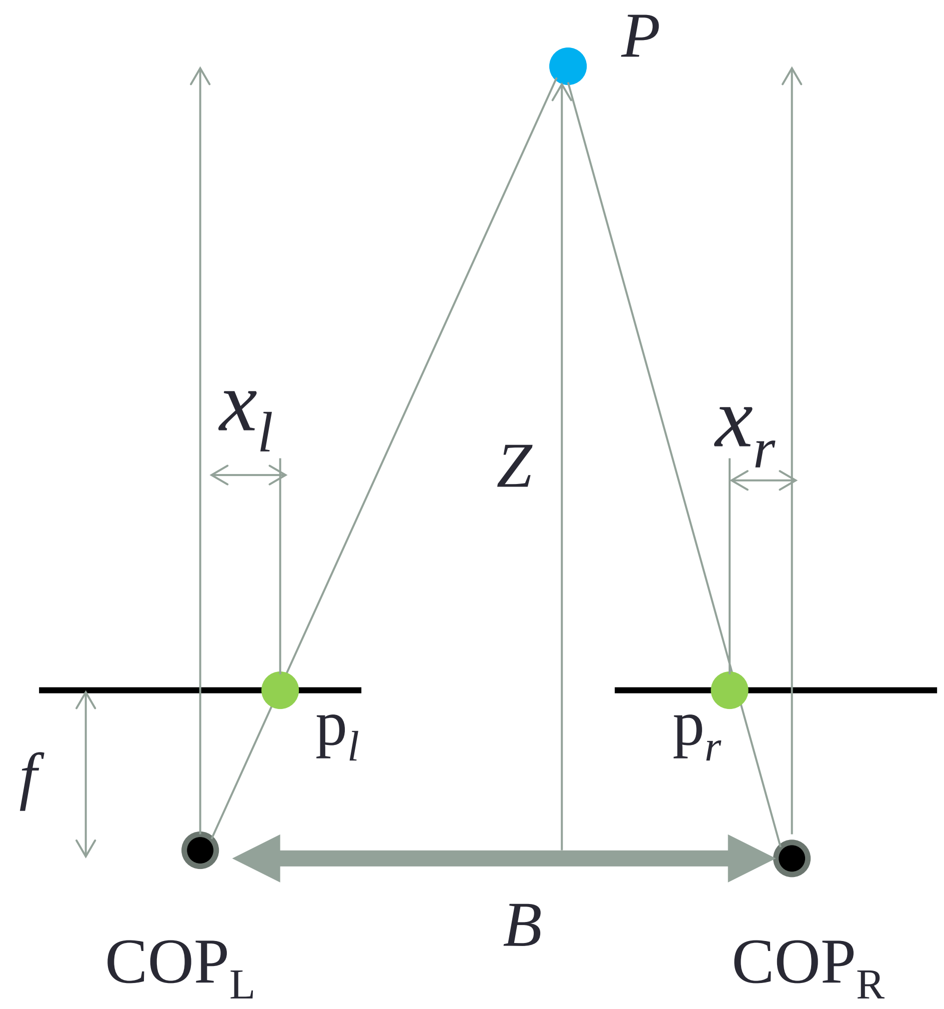
Geometry for a simple stereo system

- First, assuming parallel optical axes, known camera parameters (i.e., calibrated cameras)
- Figure is looking down on the cameras and image planes
- Baseline *B*, focal length *f*
- Point *P* is distance *Z* in camera coordinate systems



Geometry for a simple stereo system

- Point *P* projects into left and right images.
- Distance is positive in left image, and negative in right

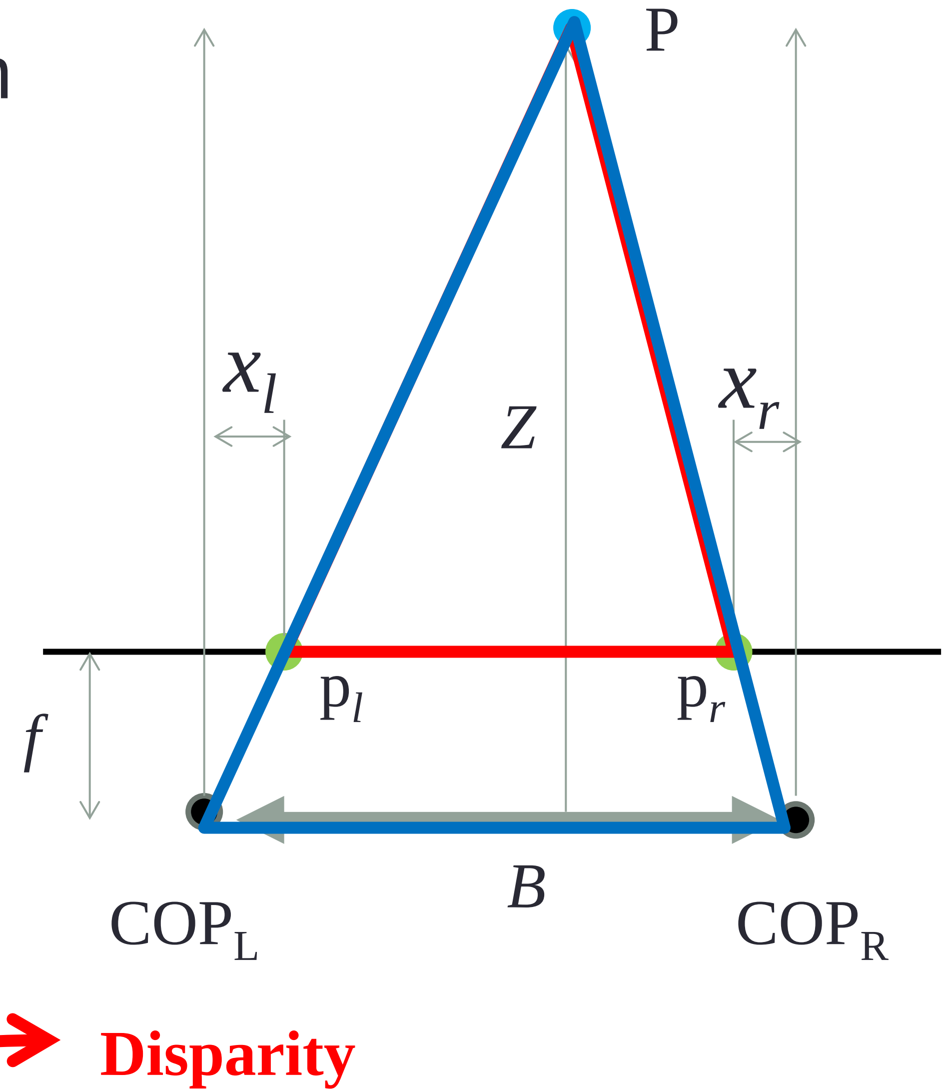


Geometry for a simple stereo system

- What is the expression for Z?
- Similar triangles (p_l, P, p_r) and (C_L,P, C_r):

$$\frac{B - x_l + x_r}{Z - f} = \frac{B}{Z}$$

$$Z = f \frac{B}{x_l - x_r}$$



Depth from disparity

image I(x,y)

Disparity map D(x,y)

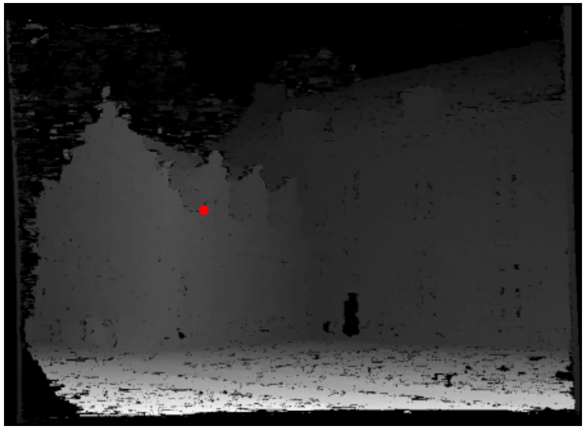


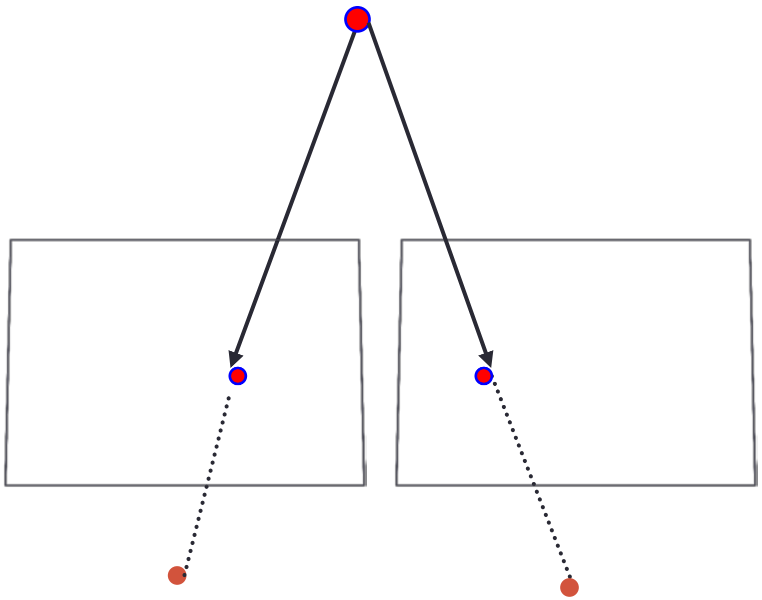
image I´(x´,y´)

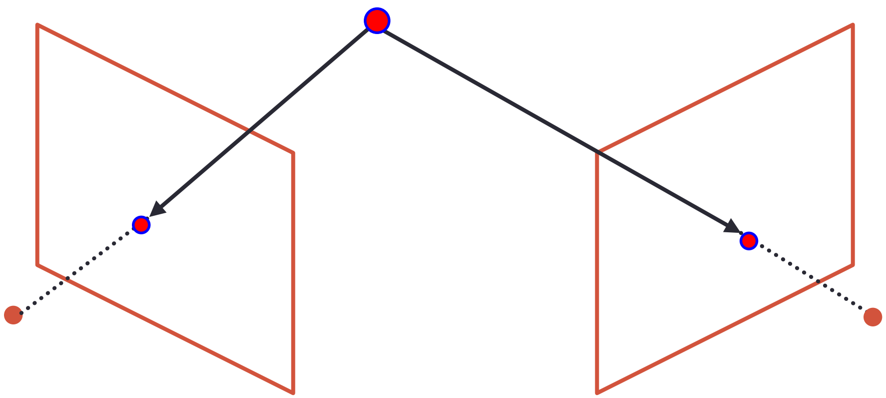
$$(x´,y´)=(x+D(x,y), y)$$

So if we could find the **corresponding points** in two images, we could **estimate relative depth**…

General case, with calibrated cameras

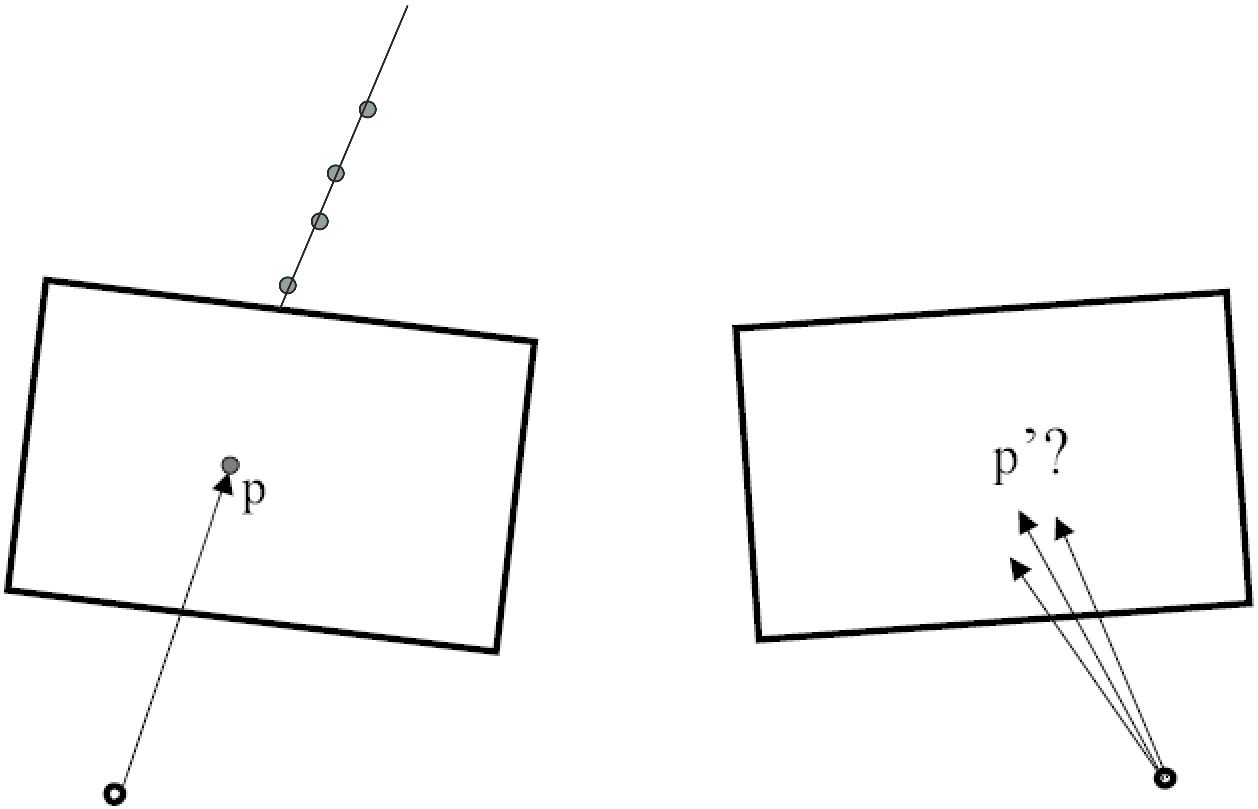
- The two cameras need not have parallel optical axes and image planes.





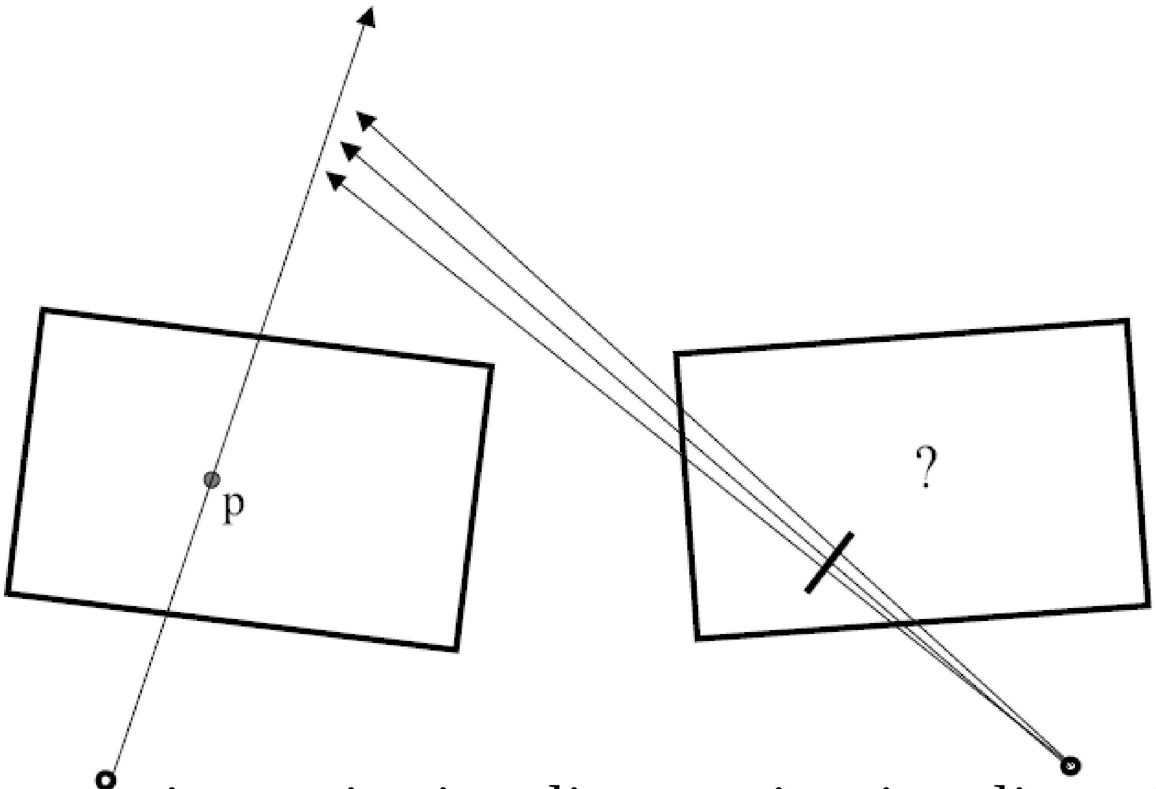
Vs.

Stereo correspondence constraints



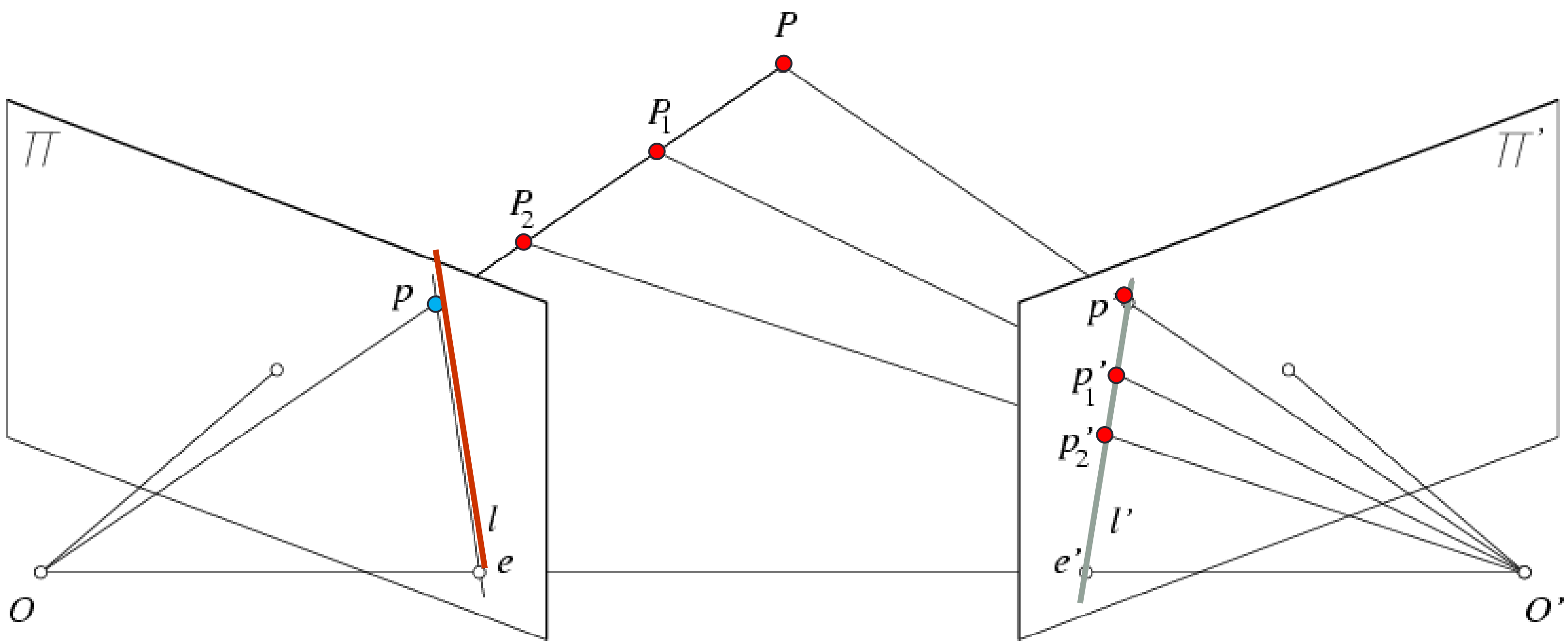
- Given p in left image, where can corresponding point p' be?

Stereo correspondence constraints



- In perspective projection, lines project into lines. So the line containing the center of projection and the point *p* in the left image must project to a ***line*** in the right image.

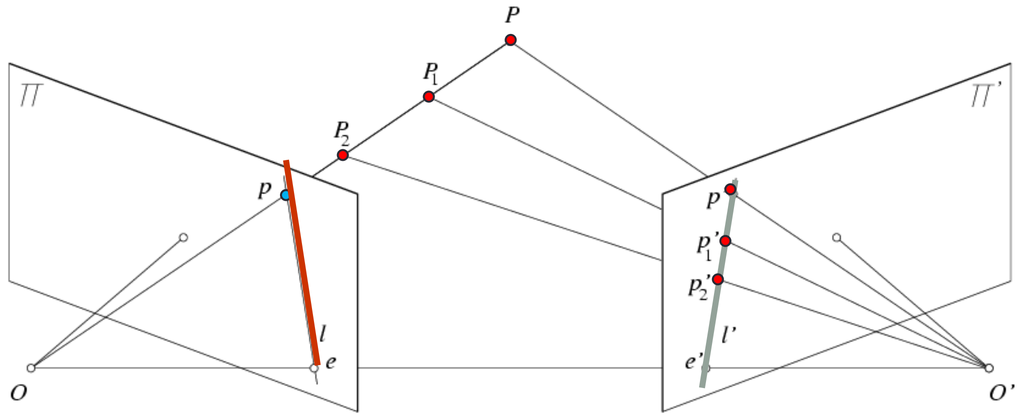
Epipolar constraint



Geometry of two views constrains where the corresponding pixel for some image point in the first view must occur in the second view.

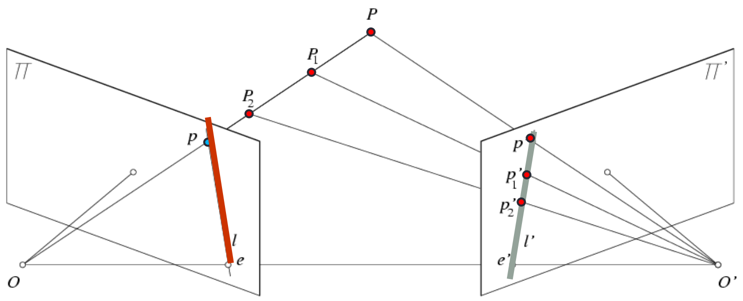
- It must be on the line carved out by a plane connecting the world point and optical centers.

Epipolar constraint



Epipolar geometry: terms

- **Baseline**: line joining the camera centers
- **Epipole**: point of intersection of baseline with image plane
- **Epipolar plane**: plane containing baseline and world point
- **Epipolar line**: intersection of epipolar plane with the image plane



- All epipolar lines intersect at the epipole
- An epipolar plane intersects the left and right image planes in epipolar lines

Why is the epipolar constraint useful?

Epipolar constraint

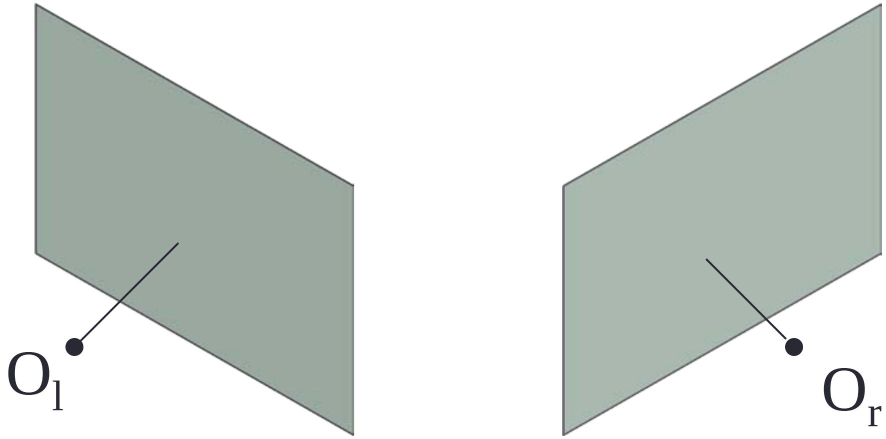
This is useful because it reduces the correspondence problem to a 1D search along an epipolar line.

Image from Andrew Zisserman

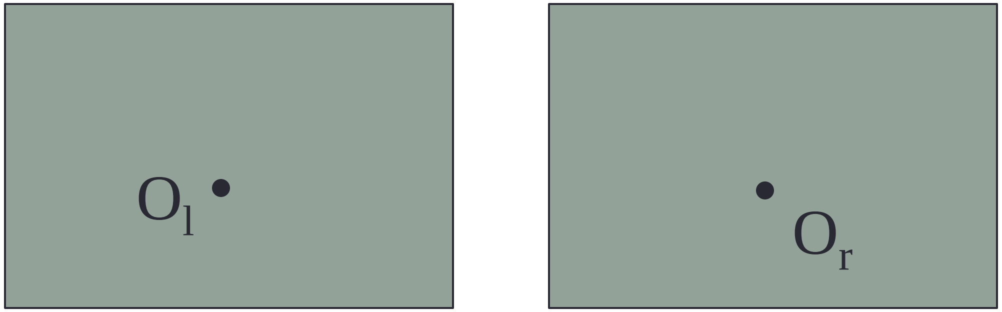
Example

What do the epipolar lines look like?

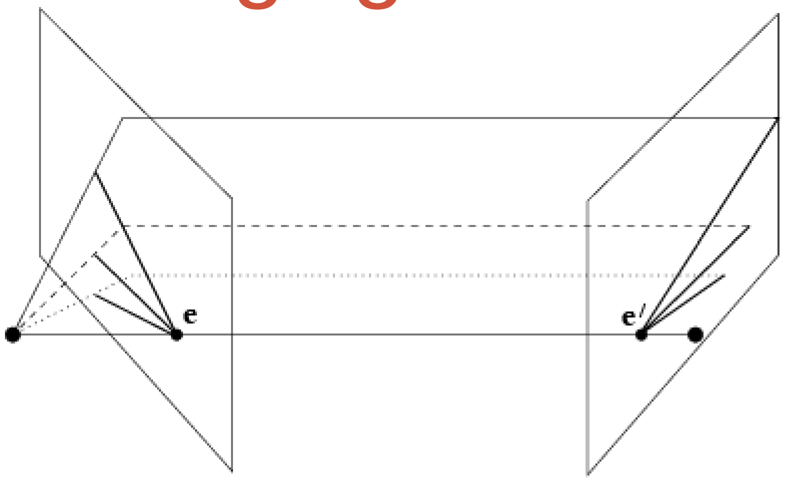
1.



2.



Example: converging cameras



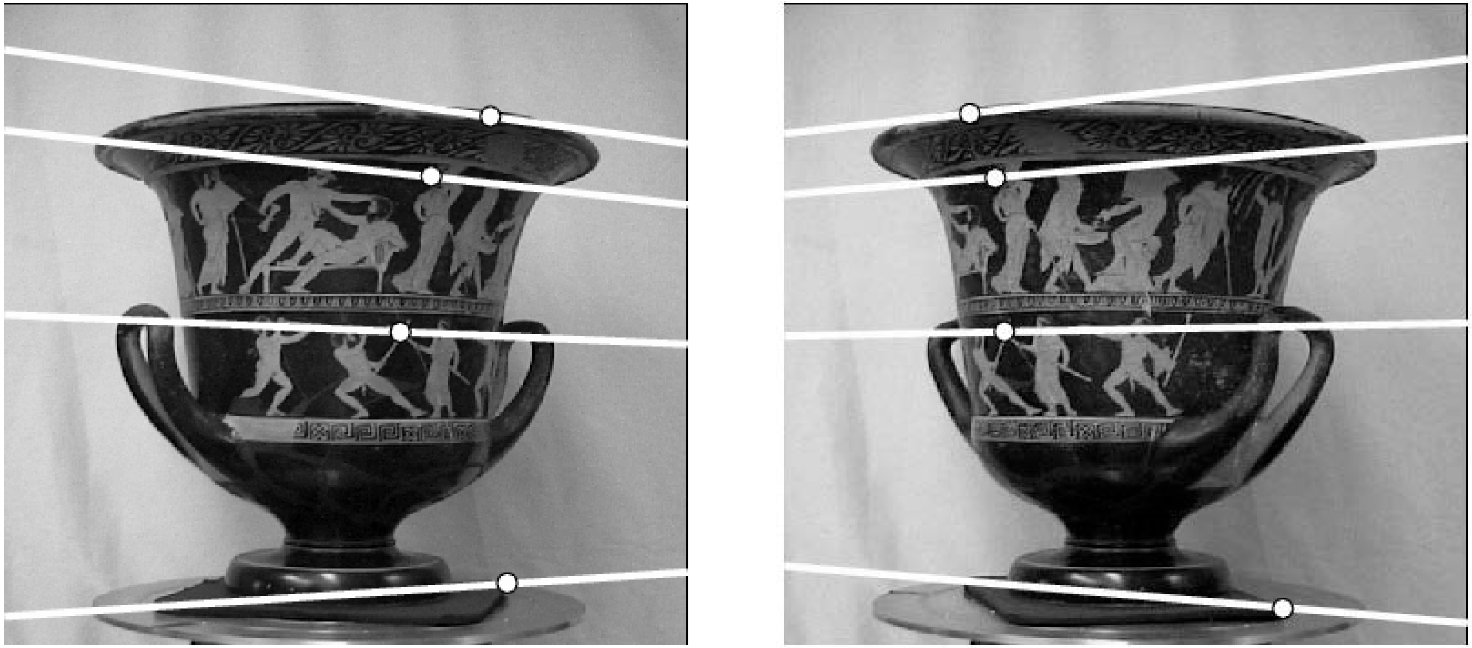
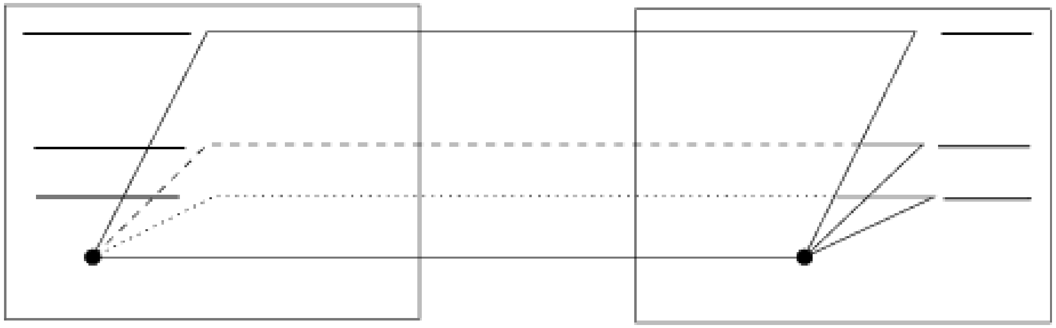


Figure from Hartley & Zisserman

Example: parallel cameras



Where are the epipoles?

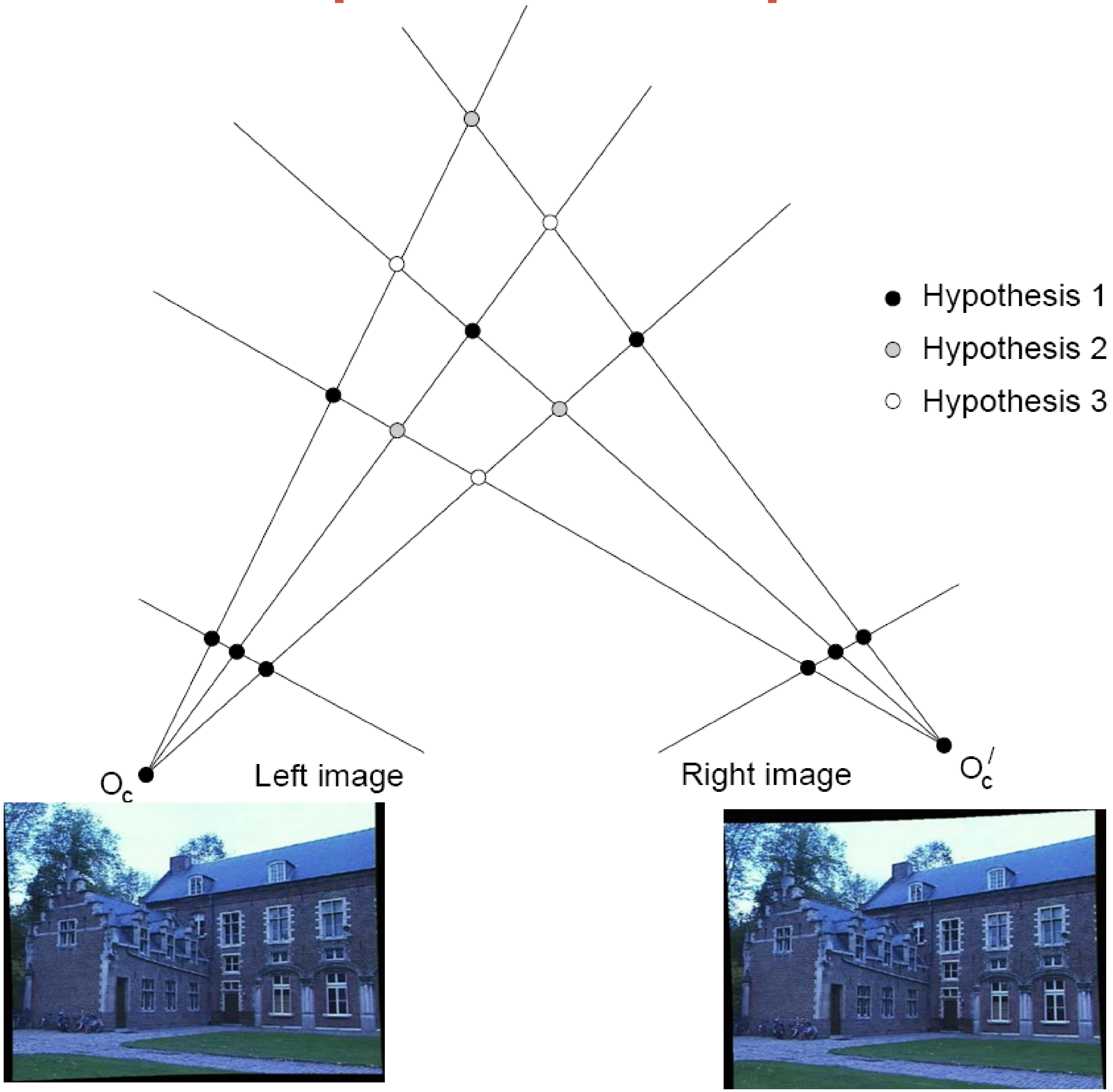
Figure from Hartley & Zisserman

For now assume parallel image planes…

- Assume parallel image planes…
- Assume same focal lengths…
- Assume epipolar lines are horizontal…
- Assume epipolar lines are at the same *y* location in the image…

- That's a lot of assuming, but it allows us to move to the correspondence problem – which you will be solving!

Correspondence problem



Multiple match hypotheses satisfy epipolar constraint, but which is correct?

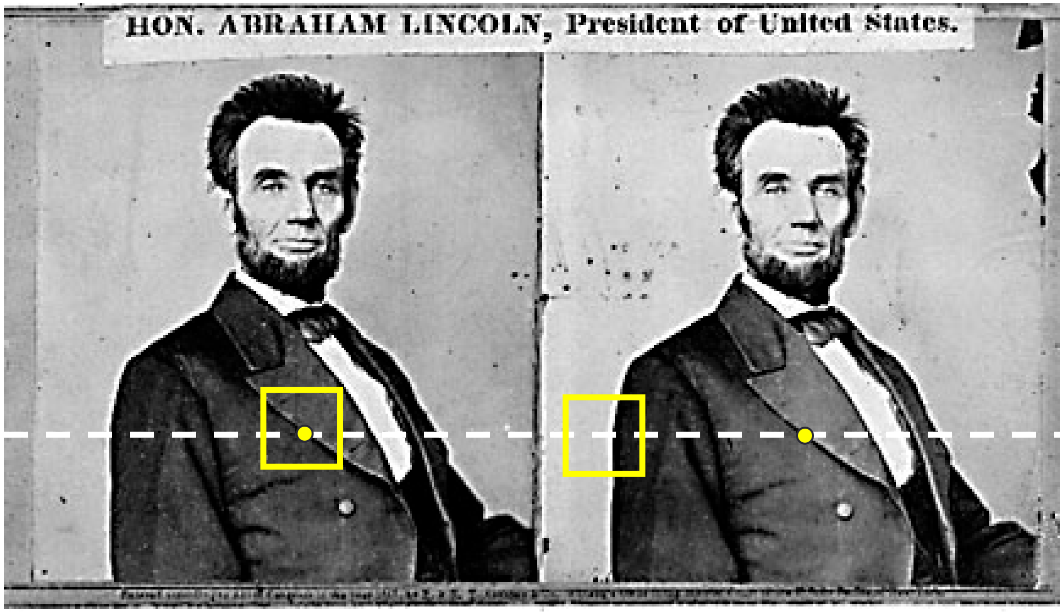
Figure from Gee & Cipolla 1999

Correspondence problem

- Beyond the hard constraint of epipolar geometry, there are “soft” constraints to help identify corresponding points
  - Similarity
  - Uniqueness
  - Ordering
  - Disparity gradient

- To find matches in the image pair, we will assume
  - Most scene points visible from both views
  - Image regions for the matches are similar in appearance

Dense correspondence search



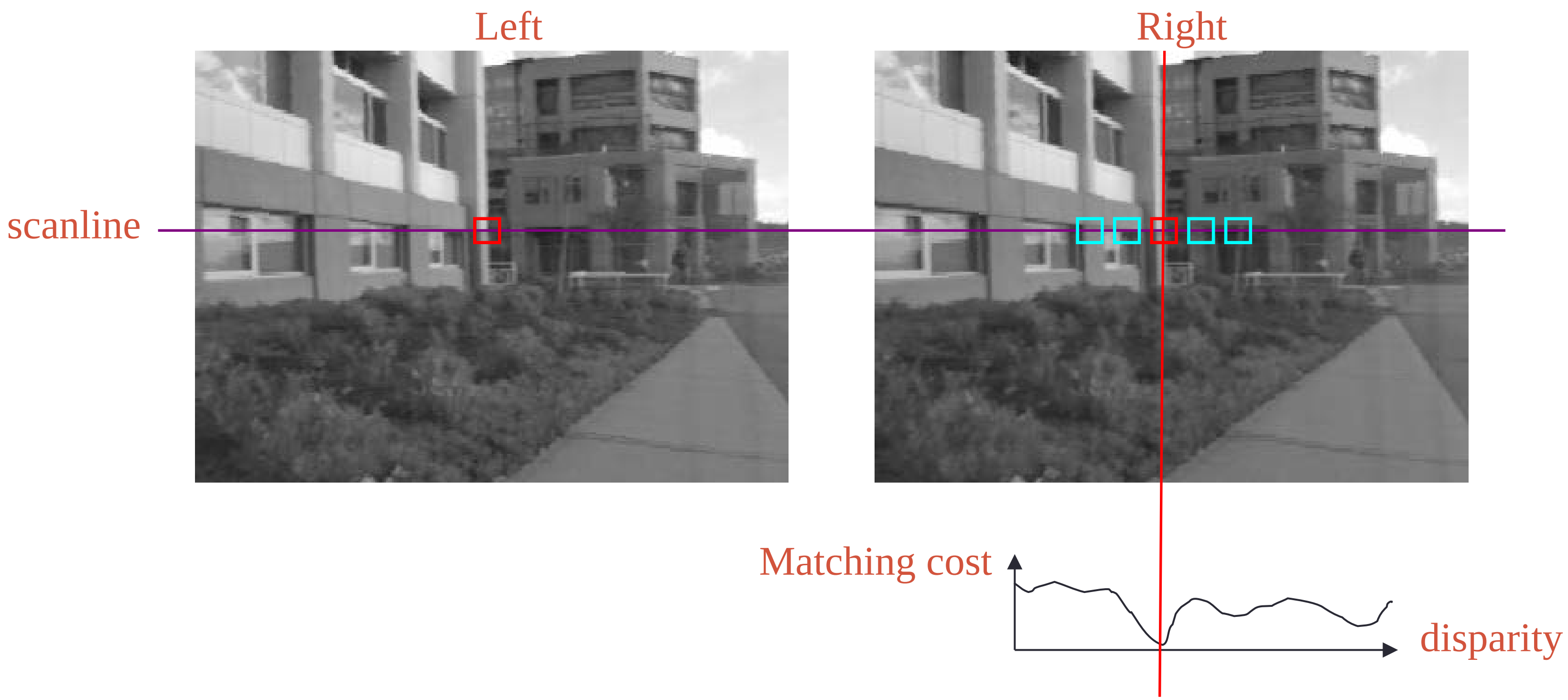
For each epipolar line

For each pixel / window in the left image

- compare with every pixel / window on same epipolar line in right image
- pick position with minimum match cost (e.g., SSD, normalized correlation)

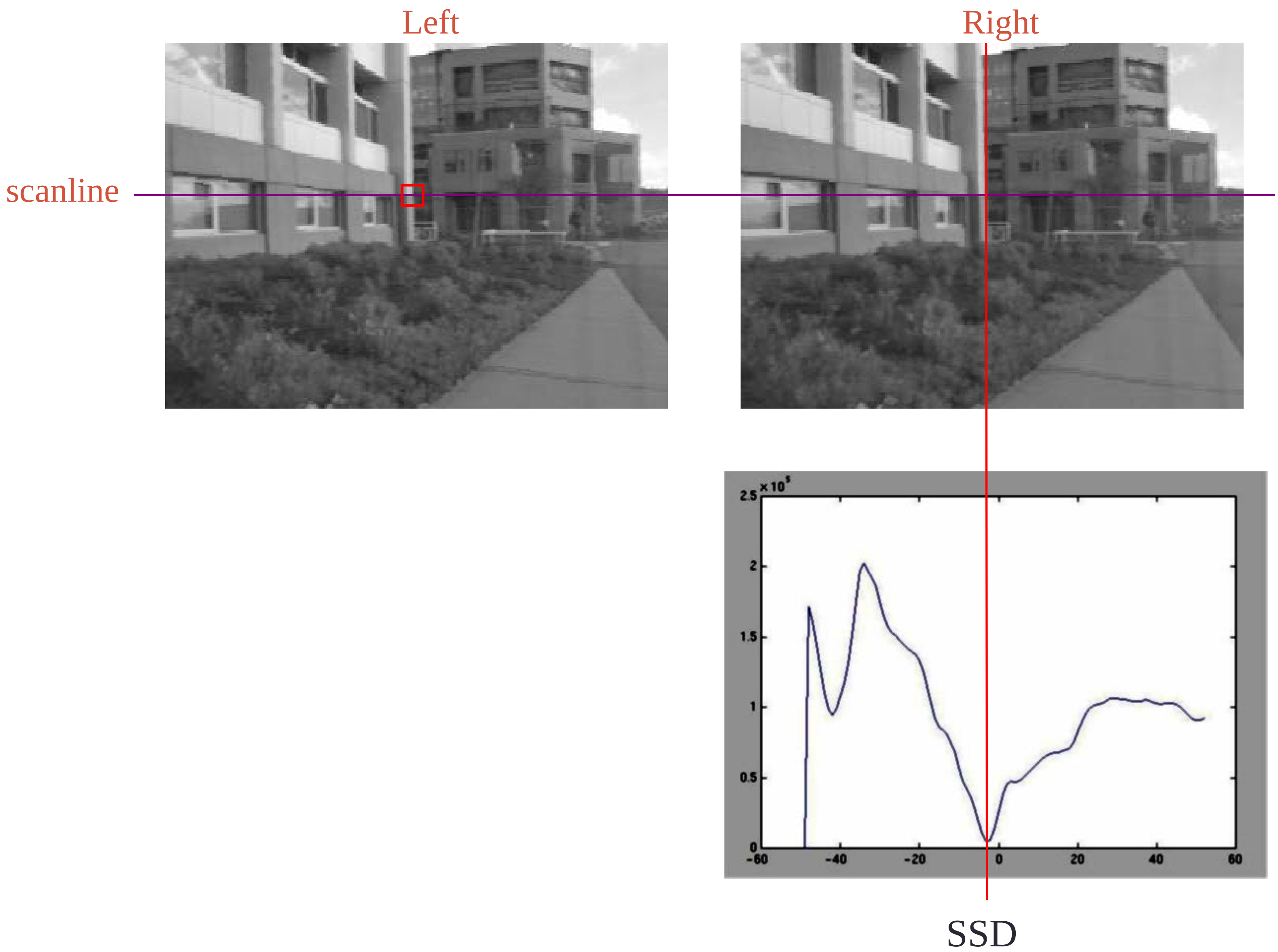
Adapted from Li Zhang

Correspondence search with similarity constraint



- Slide a window along the right scanline and compare contents of that window with the reference window in the left image
- Matching cost: SSD or normalized correlation

Correspondence search with similarity constraint

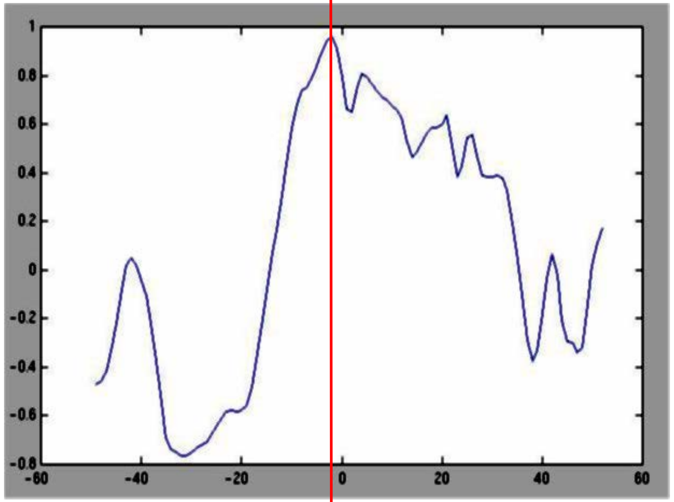


Correspondence search with similarity constraint

Left

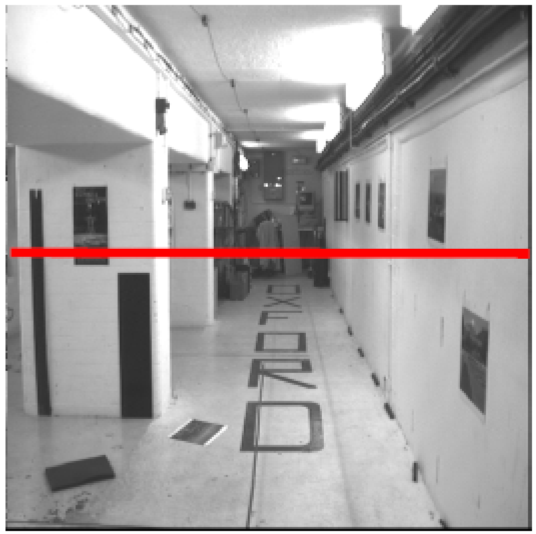
Right

scanline

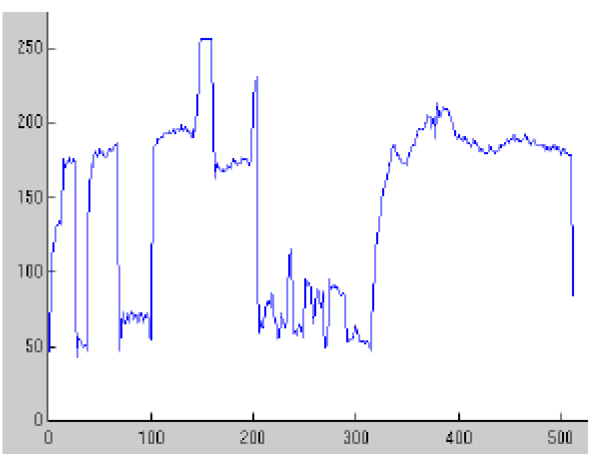


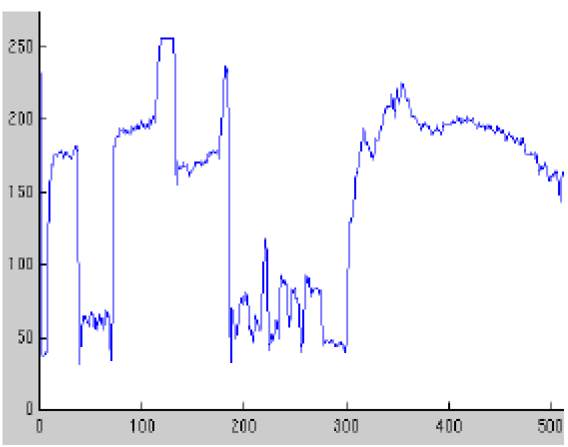
Norm. corr

Correspondence problem



Intensity profiles





- Clear correspondence between intensities, but also noise and ambiguity

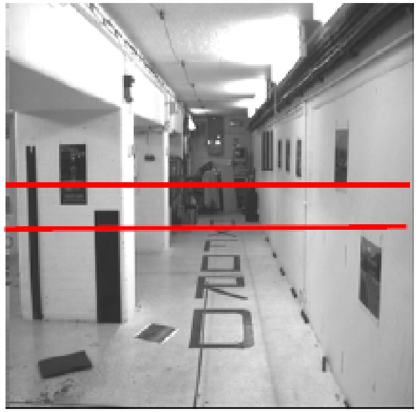
Source: Andrew Zisserman

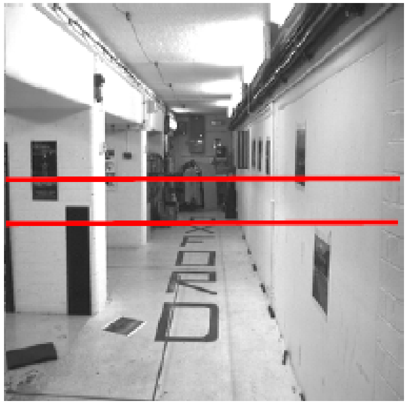
Correspondence problem

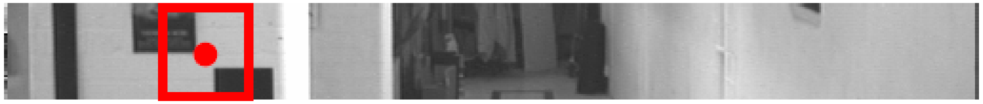
Neighborhoods of corresponding points are similar in intensity patterns.

Source: Andrew Zisserman

Correlation-based window matching



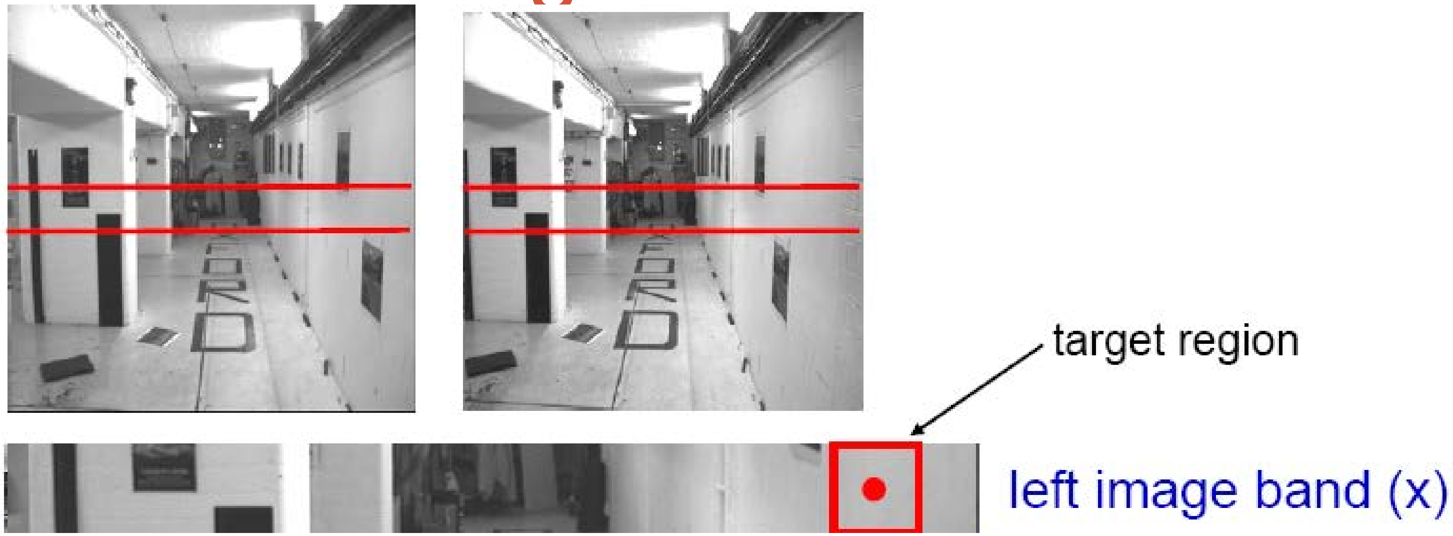




left image band (x)

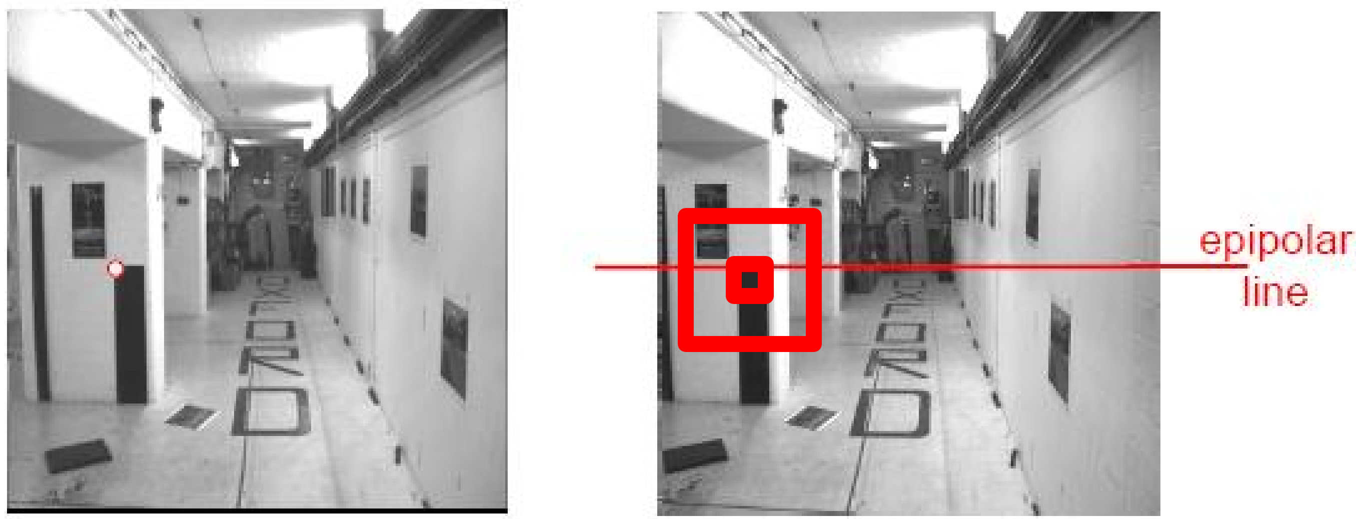
Source: Andrew Zisserman

Textureless regions



Source: Andrew Zisserman

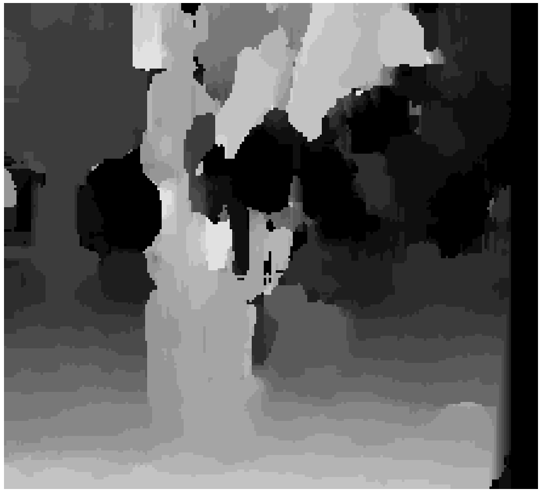
Effect of window size



Source: Andrew Zisserman

Effect of window size

W = 3



W = 20

Want window large enough to have sufficient intensity variation, yet small enough to contain only pixels with about the same disparity.

Figures from Li Zhang

Correspondence problem

- Beyond the hard constraint of epipolar geometry, there are "soft" constraints to help identify corresponding points
  - Similarity
  - Disparity gradient – depth doesn't change too quickly.
  - Uniqueness
  - Ordering

Uniqueness constraint

- Up to one match in right image for every point in left image

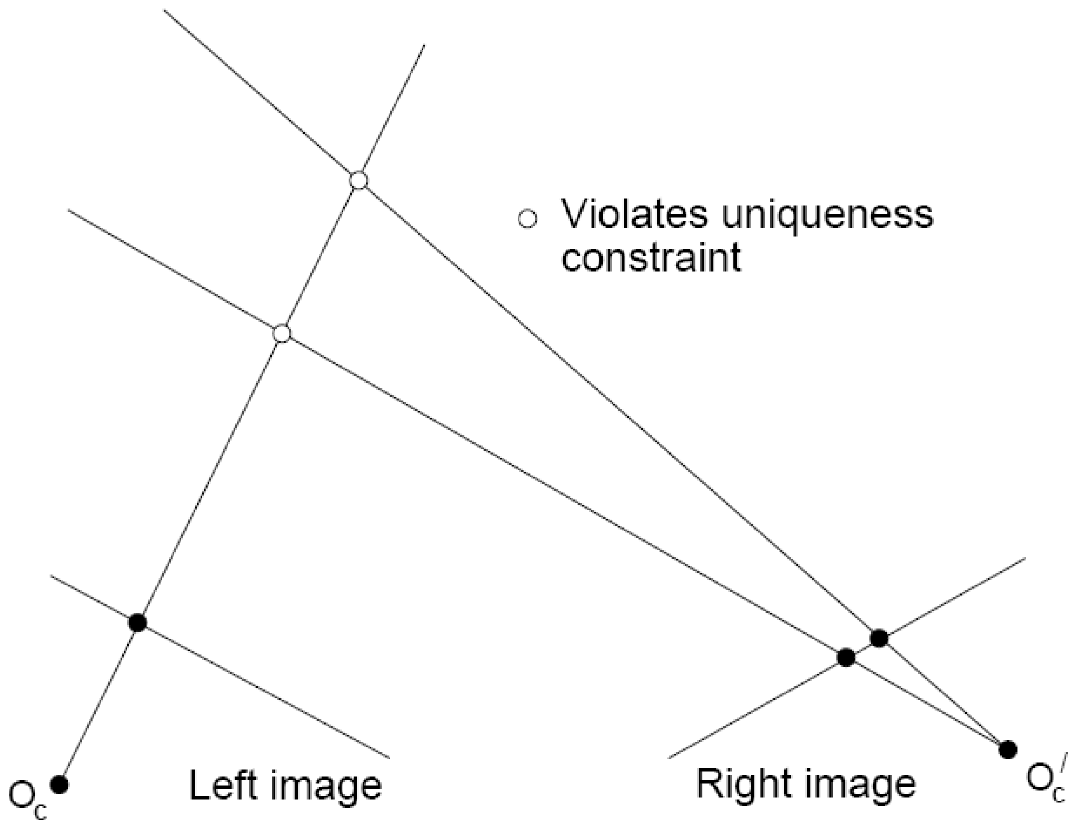
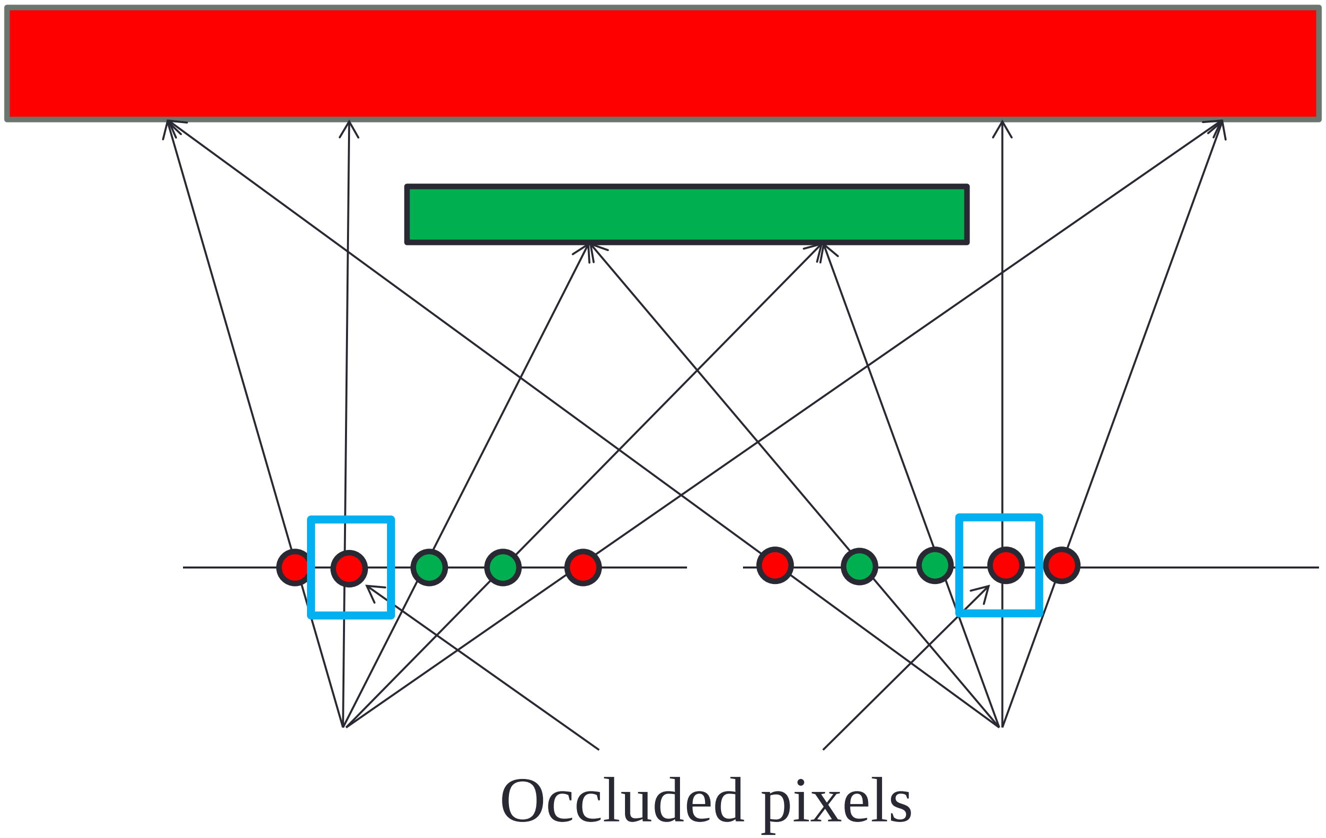


Figure from Gee & Cipolla 1999

Problem: Occlusion

- Uniqueness says "up to match" per pixel
- When is there no match?



Ordering constraint

- Points on **same surface** (opaque object) will be in same order in both views

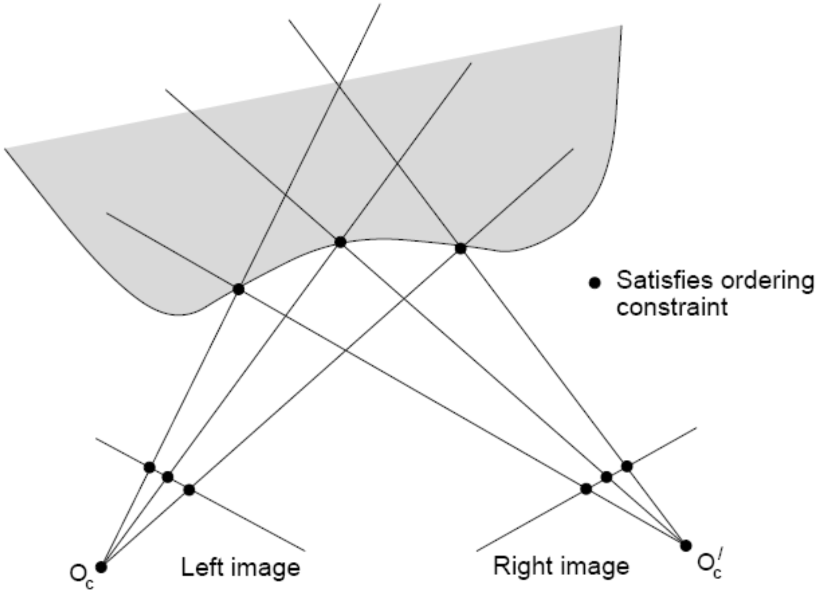
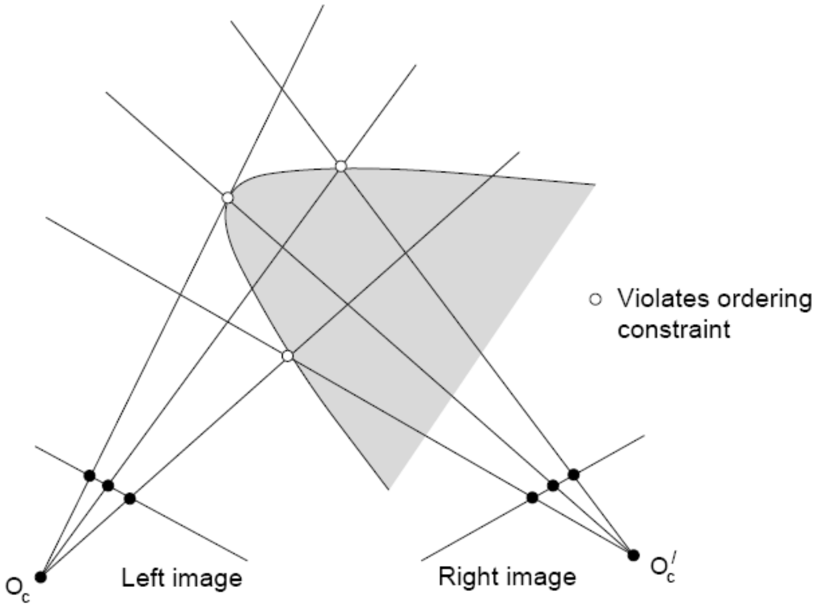
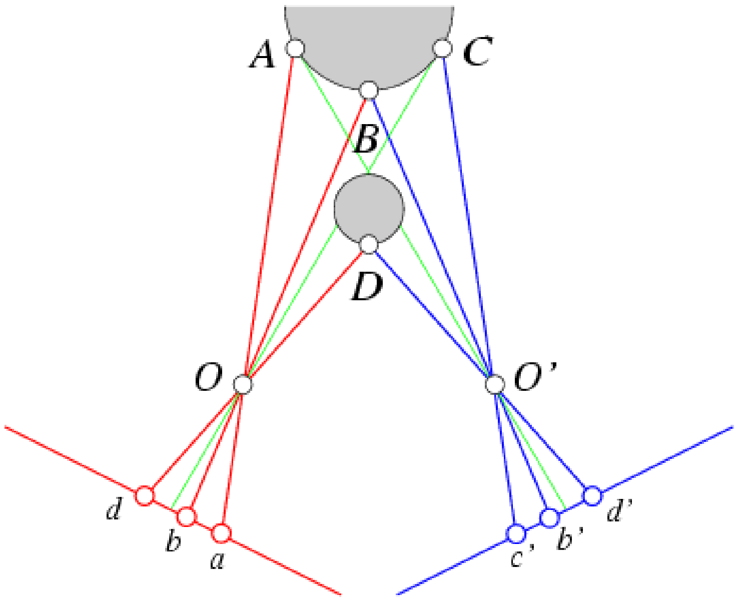


Figure from Gee & Cipolla 1999

Ordering constraint

- Won't always hold, e.g. consider transparent object, or an occluding surface



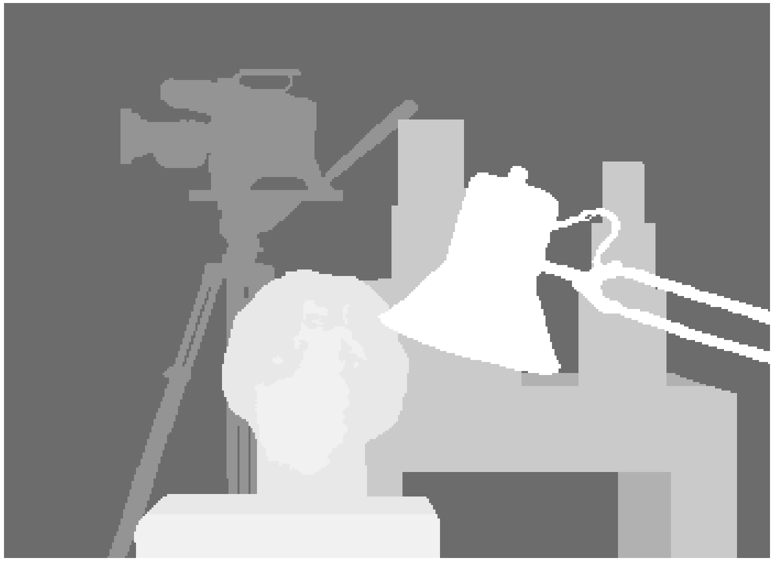


Figures from Forsyth & Ponce

Stereo results

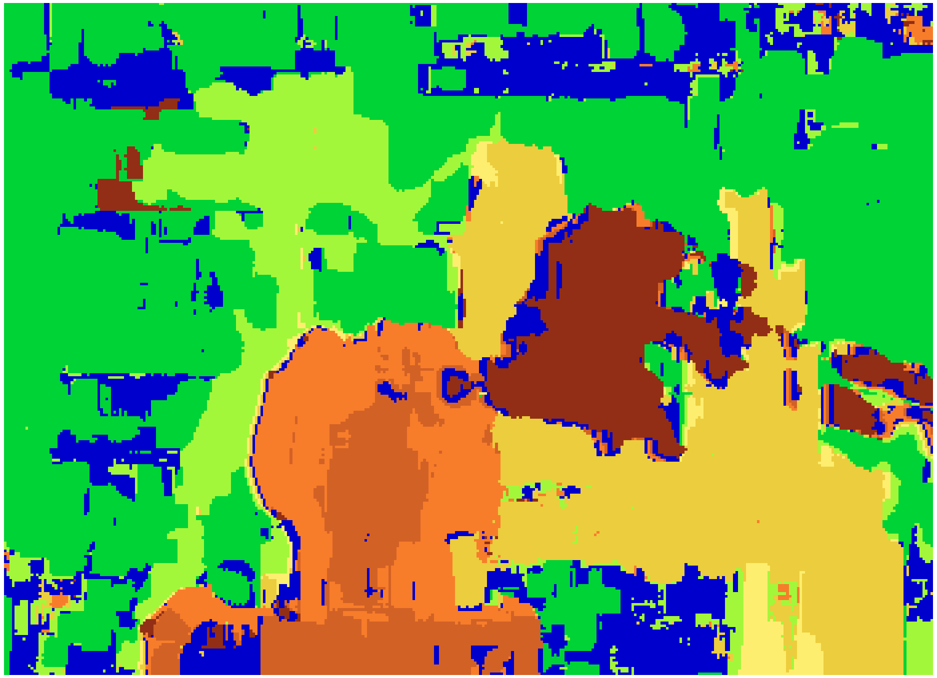
- Data from University of Tsukuba
- Similar results on other images without ground truth

Scene



Ground truth

Results with window search



Window-based matching (best window size)

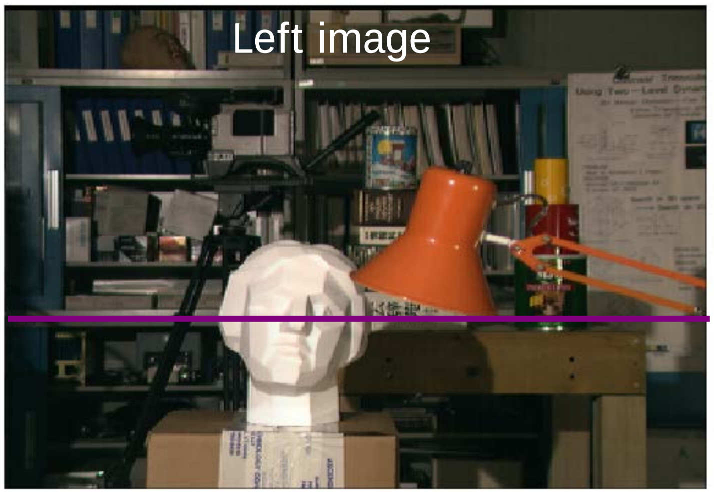
Ground truth

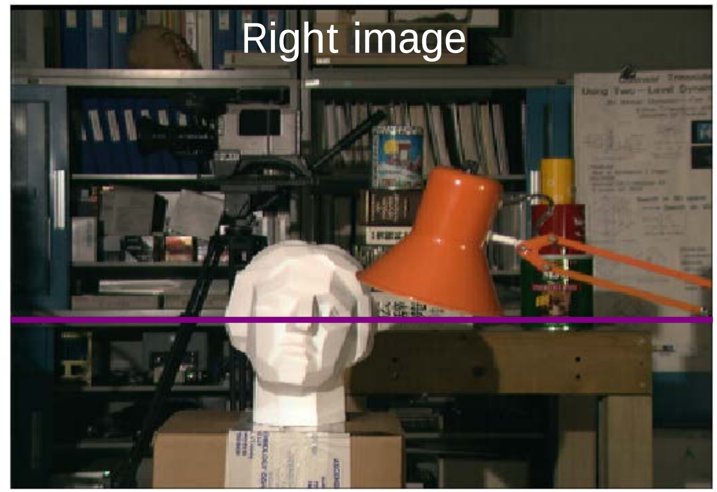
Better solutions

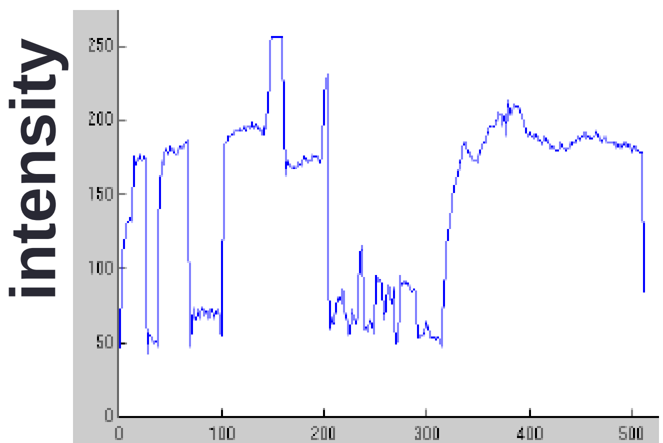
- Beyond individual correspondences to estimate disparities:
- Optimize correspondence assignments jointly
  - Scanline at a time (DP)
  - Full 2D grid (graph cuts)

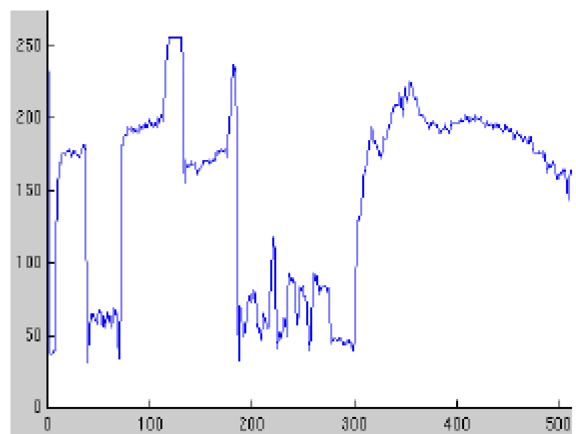
Scanline stereo

- Try to coherently match pixels on the entire scanline
- s

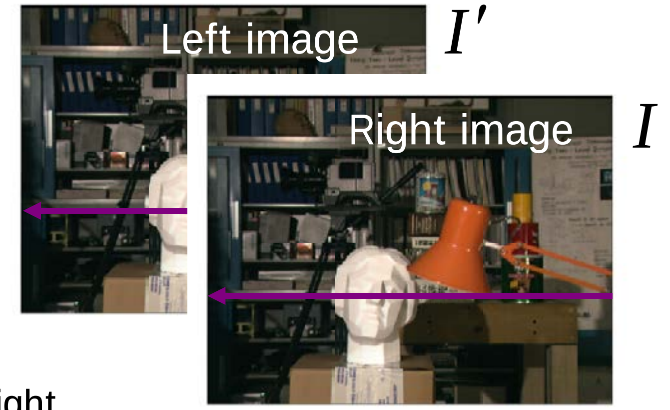


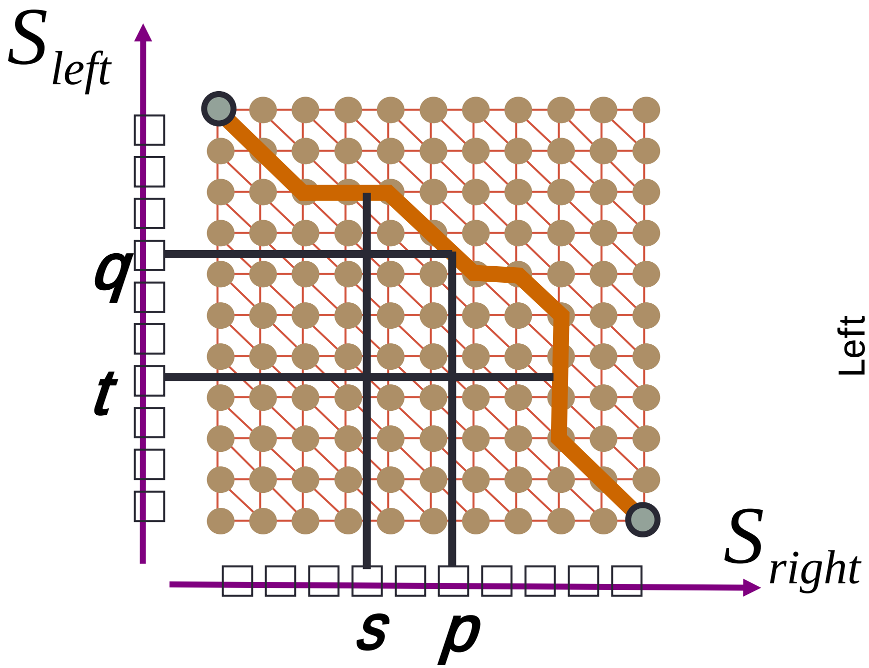


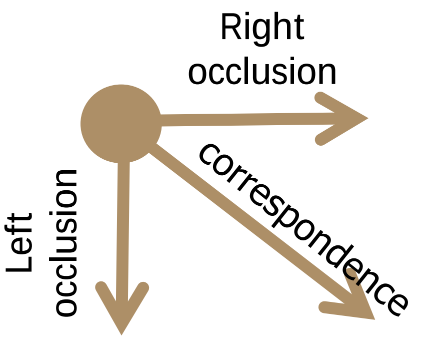




"Shortest paths" for scan-line stereo







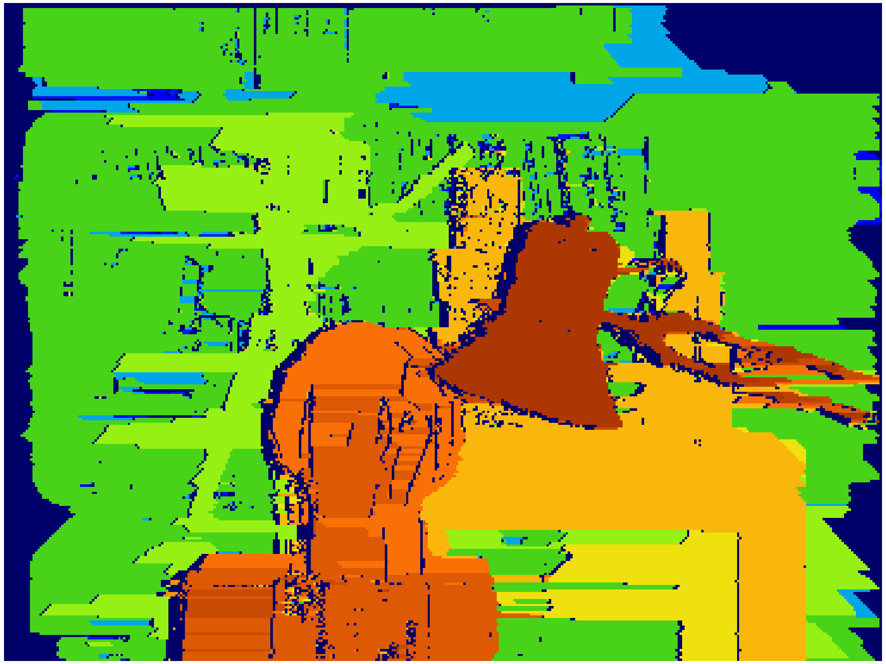
Can be implemented with dynamic programming

Ohta & Kanade '85, Cox et al. '96, Intille & Bobick, '01

Slide credit: Y. Boykov

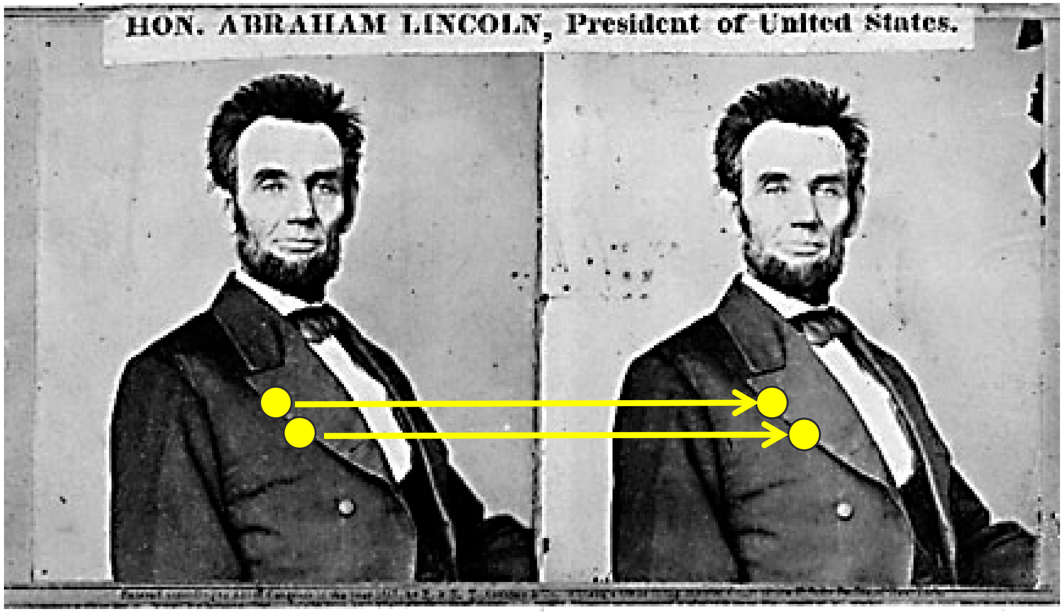
Coherent stereo on 2D grid

- Scanline stereo generates streaking artifacts



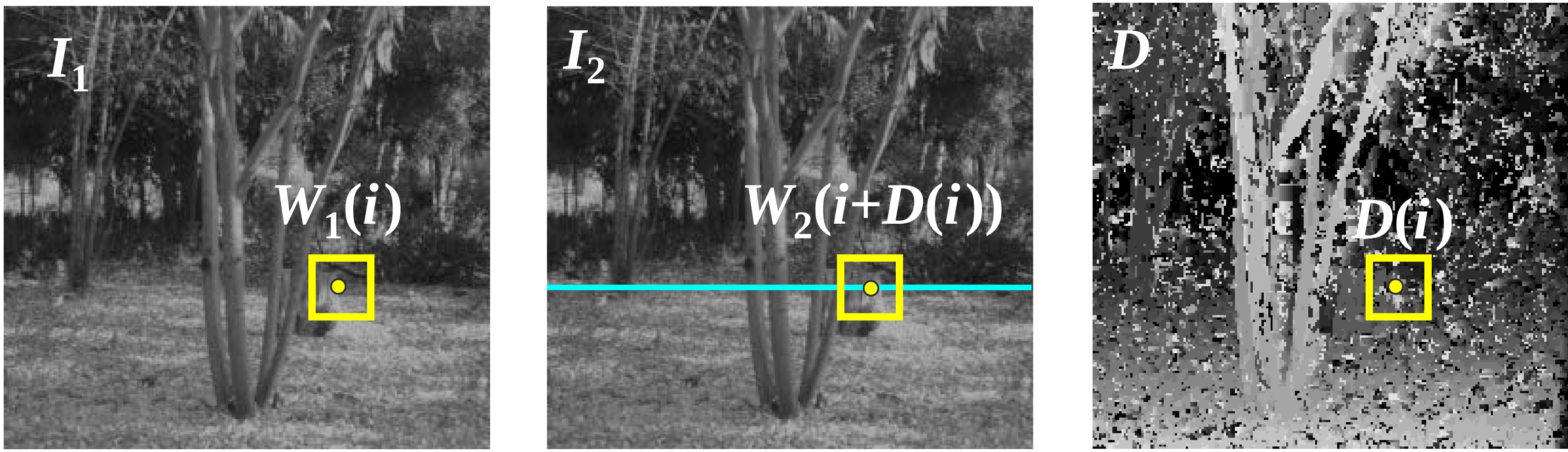
- Can't use dynamic programming to find spatially coherent disparities/ correspondences on a 2D grid

Stereo as energy minimization



- What defines a good stereo correspondence?
  1. Match quality
     - Want each pixel to find a good match in the other image
  2. Smoothness
     - If two pixels are adjacent, they should (usually) move about the same amount

Stereo matching as energy minimization



$$E = \alpha E_{\text{data}}(I_1, I_2, D) + \beta E_{\text{smooth}}(D)$$

$$E_{\text{data}} = \sum_i \left(W_1(i) - W_2(i + D(i))\right)^2$$

$$E_{\text{smooth}} = \sum_{\text{neighbors } i,j} \rho\left(D(i) - D(j)\right)$$

- Energy functions of this form can be minimized using *graph cuts*

Y. Boykov, O. Veksler, and R. Zabih, [Fast Approximate Energy Minimization via Graph Cuts](), PAMI 2001

Source: Steve Seitz

Better results…

State of the art method

Ground truth

Boykov et al., [Fast Approximate Energy Minimization via Graph Cuts](), International Conference on Computer Vision, September 1999.

For the latest and greatest: http://www.middlebury.edu/stereo/

Challenges

- Low-contrast ; textureless image regions
- Occlusions
- Violations of brightness constancy (e.g., specular reflections)
- Really large baselines (foreshortening and appearance change)
- Camera calibration errors