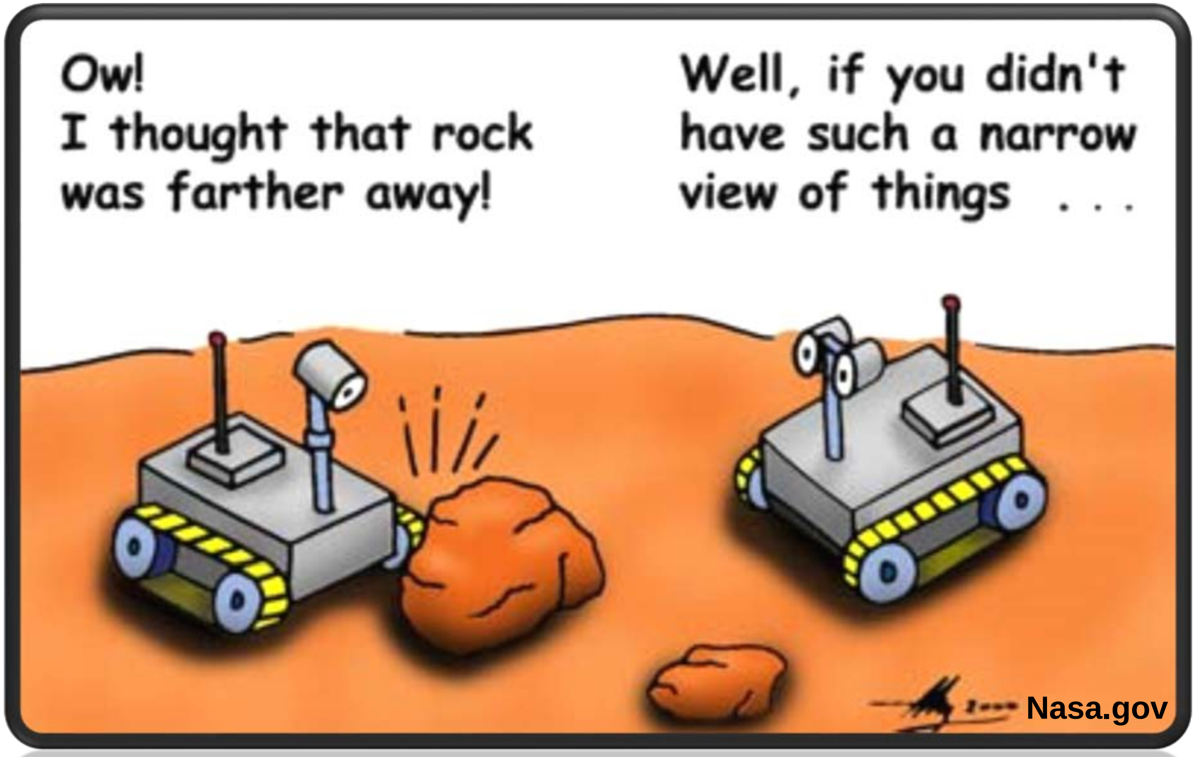
ECSE-626

Statistical Computer Vision



Stereo

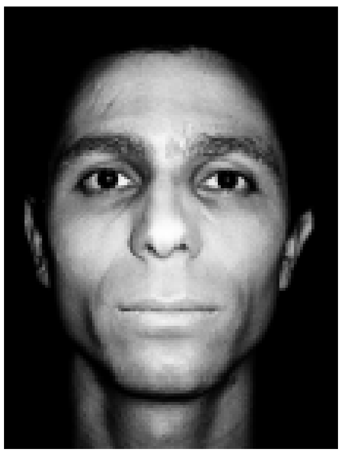
Why do we perceive depth?

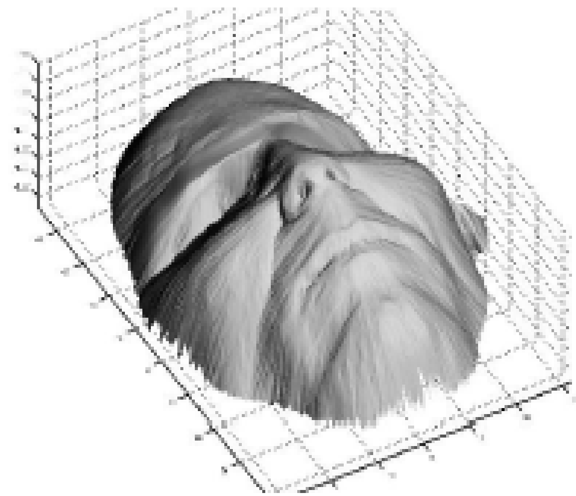
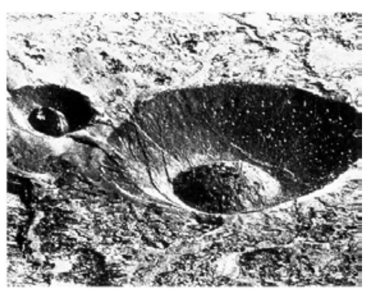
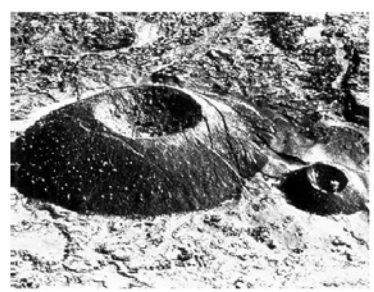
Image from Farhadi, CSE 455, Washington University

What cues help us to perceive 3D shape and depth from one eye?

What cues help us to perceive 3D shape and depth?

- Shading



[Figure from Prados & Faugeras 2006]

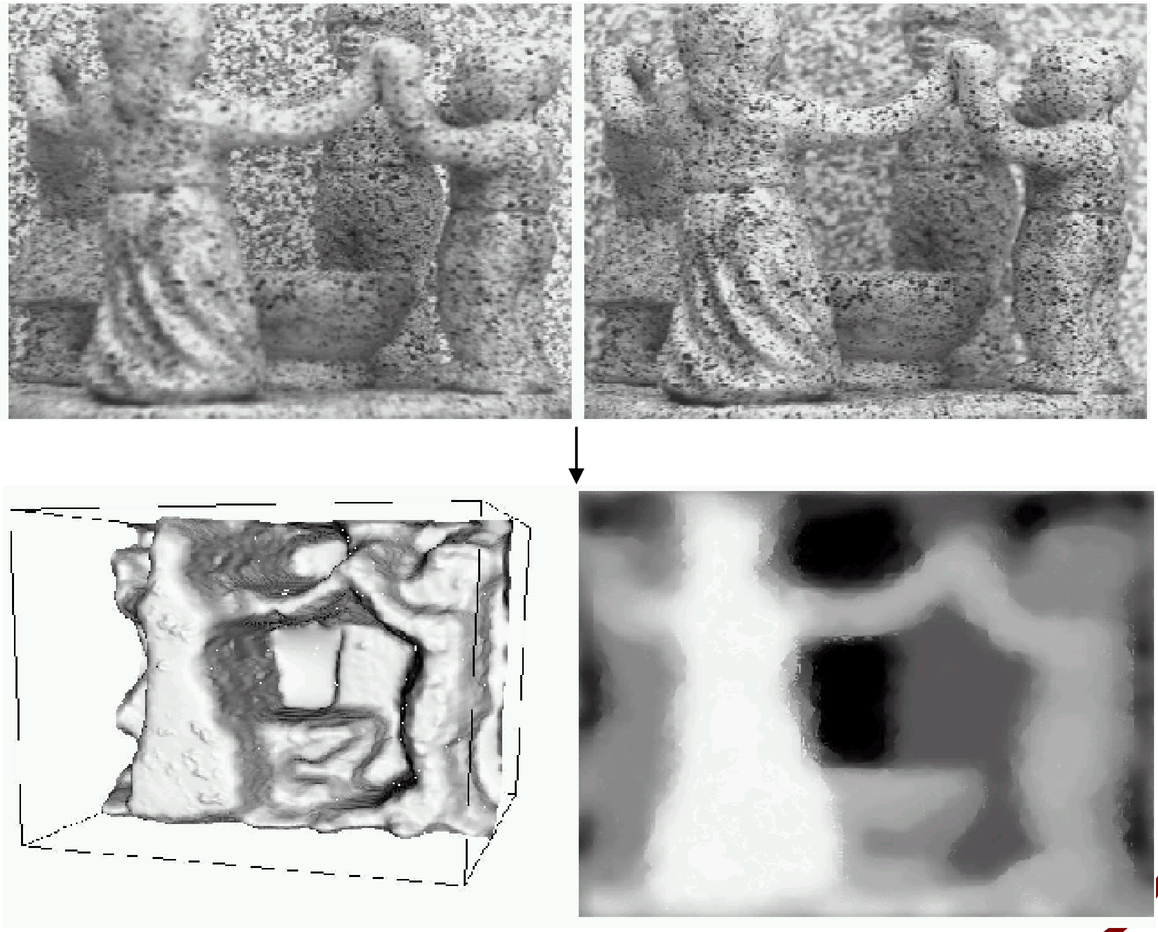
What cues help us to perceive 3D shape and depth?

- Focus/Defocus
  - Amount of blurring depends on depth

http://www.theverge.com/2014/8/22/6055591/iphone-vs-pro-camera

What cues help us to perceive 3D shape and depth?

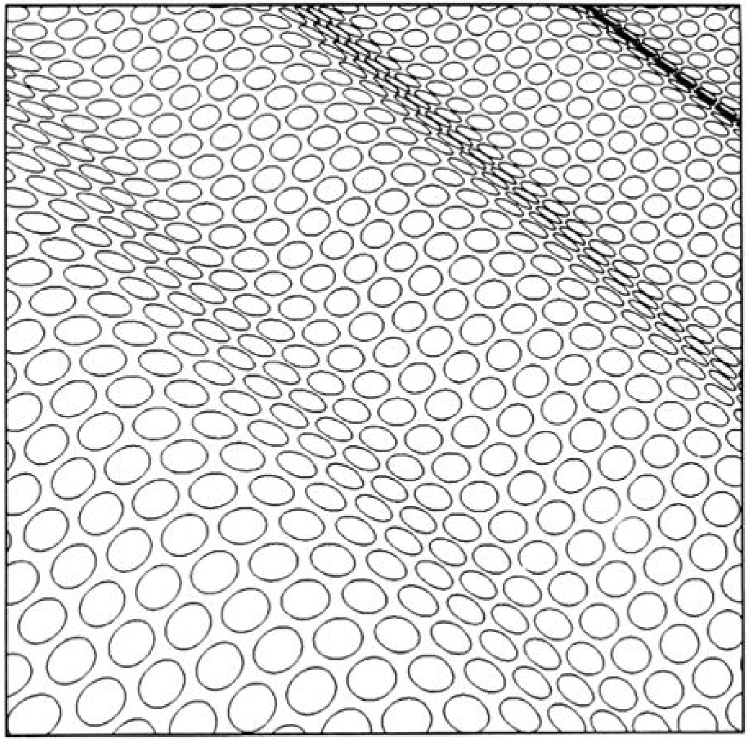
- Focus/Defocus
  - Images from same point of view, different camera parameters
  - 3D shape/depth estimates



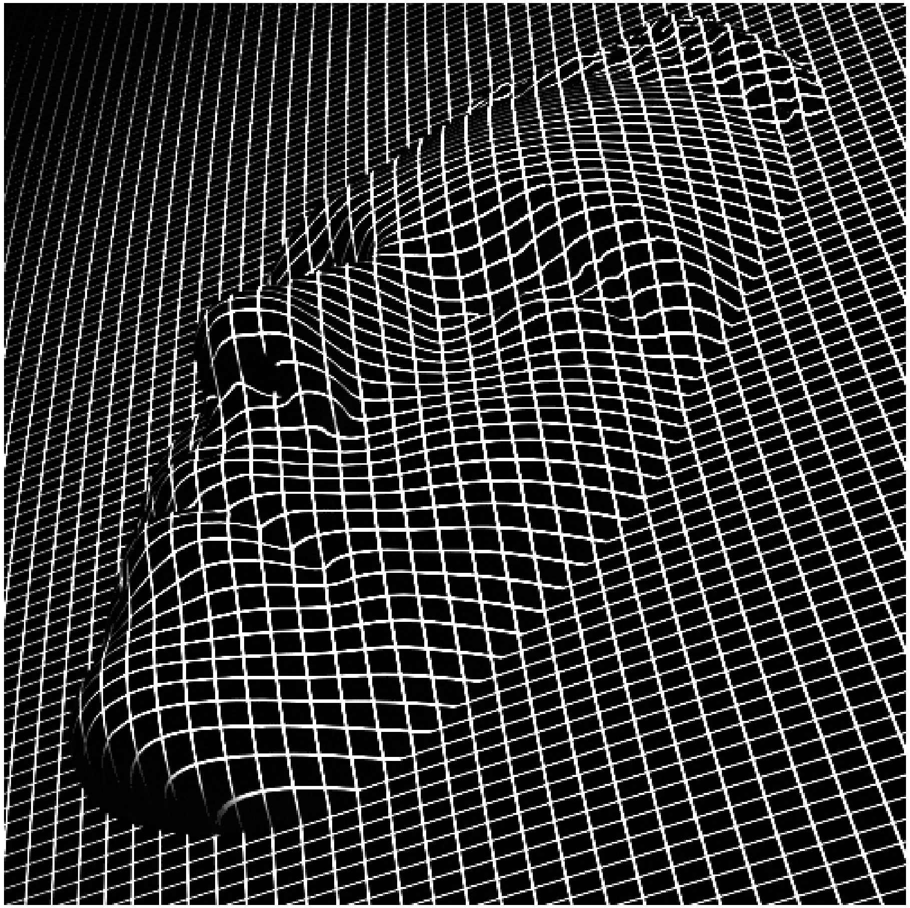
[figs from H. Jin and P. Favaro, 2002]

What cues help us to perceive 3D shape and depth?

- Texture



http://www.cns.nyu.edu/~david/courses/perception/lecturenotes/depth/depth-size.html



http://bigumigu.com/haber/optik-yanilsamali-nu-gif-ler/

What cues help us to perceive 3D shape and depth?

- Perspective effects

Image credit: S. Seitz

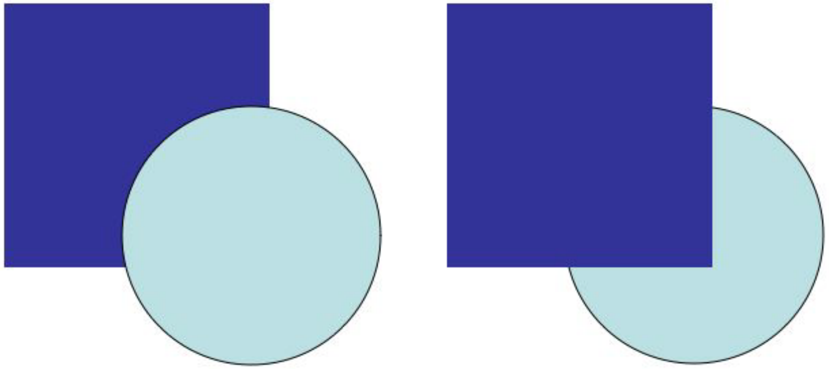
What cues help us to perceive 3D shape and depth?

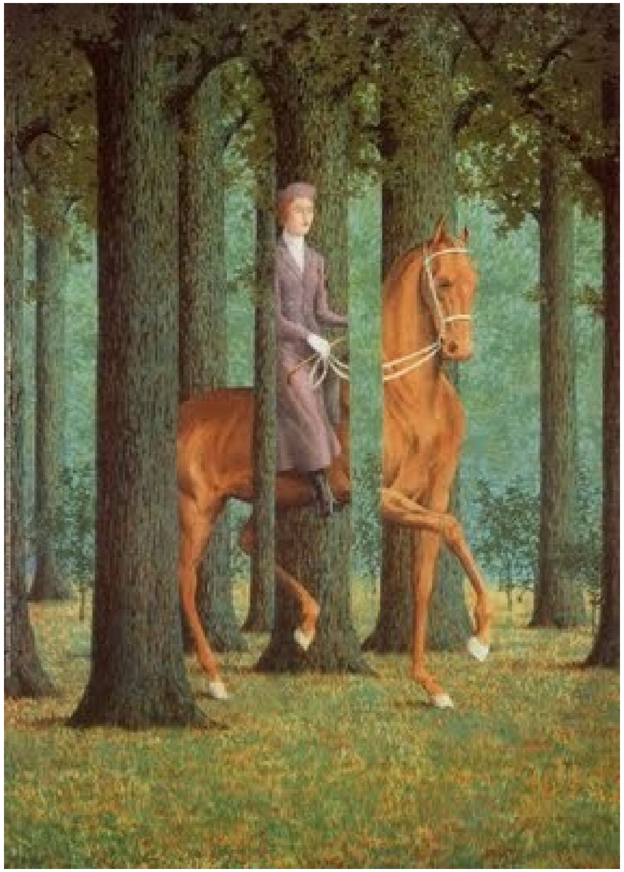
- Motion

Figures from L. Zhang

What cues help us to perceive 3D shape and depth?

- Occlusion

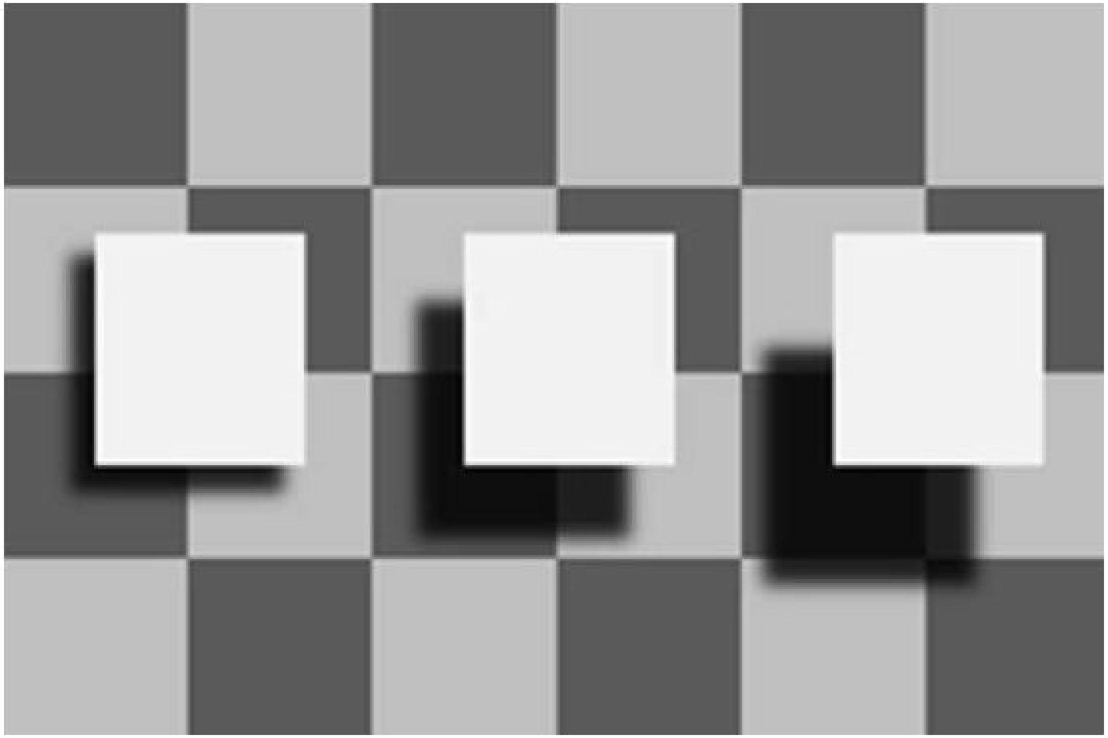




Rene Magritt'e famous painting *Le Blanc-Seing* (literal translation: "The Blank Signature") roughly translates as "free hand" or "free rein".

What cues help us to perceive 3D shape and depth?

- Cast Shadows



http://www.cns.nyu.edu/~david/courses/perception/lecturenotes/depth/depth-size.html

Image from Farhadi, CSE 455, Washington University

Why Stereo Vision?

- Structure and depth are inherently ambiguous from single views

Images from Lana Lazebnik

Stereo Vision

- Seeing the same scene from two eyes at different positions helps resolve ambiguity

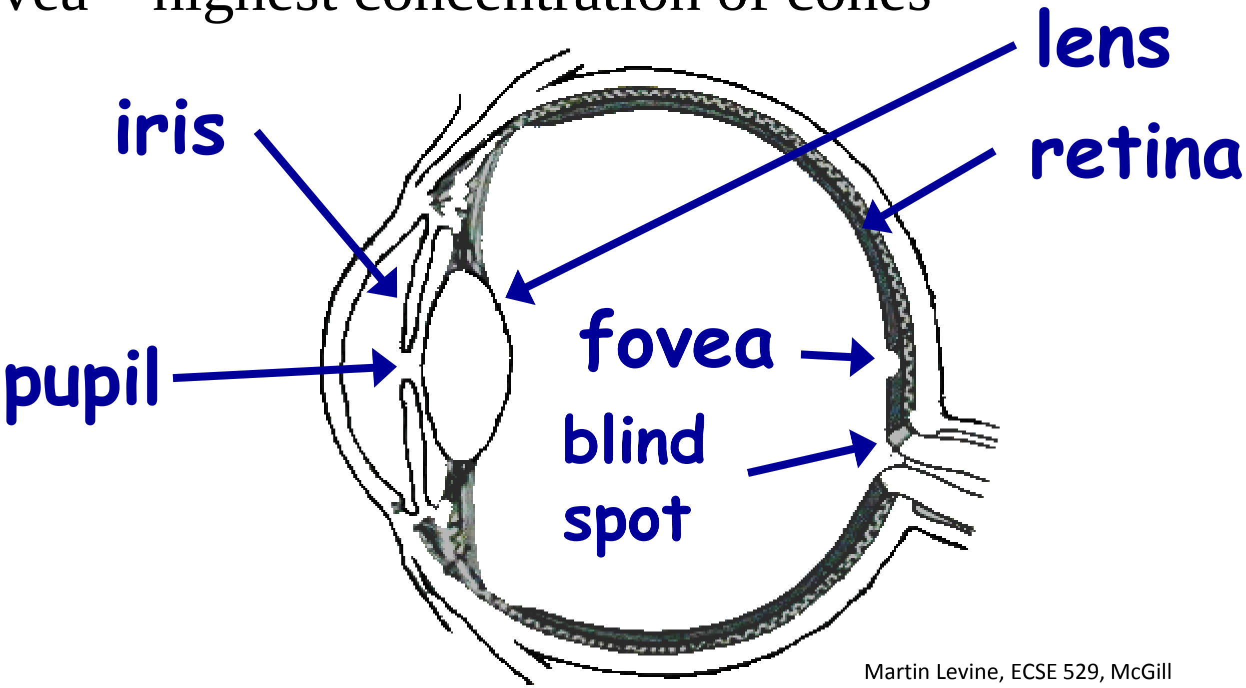
Image from Farhadi, CSE 455, Washington University

Stereopsis

- **Stereopsis**: Process in visual perception leading to perception of depth or distance of objects.
- Depth from stereopsis arises from the slightly different positions each eye occupies in the head (a form of parallax).

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



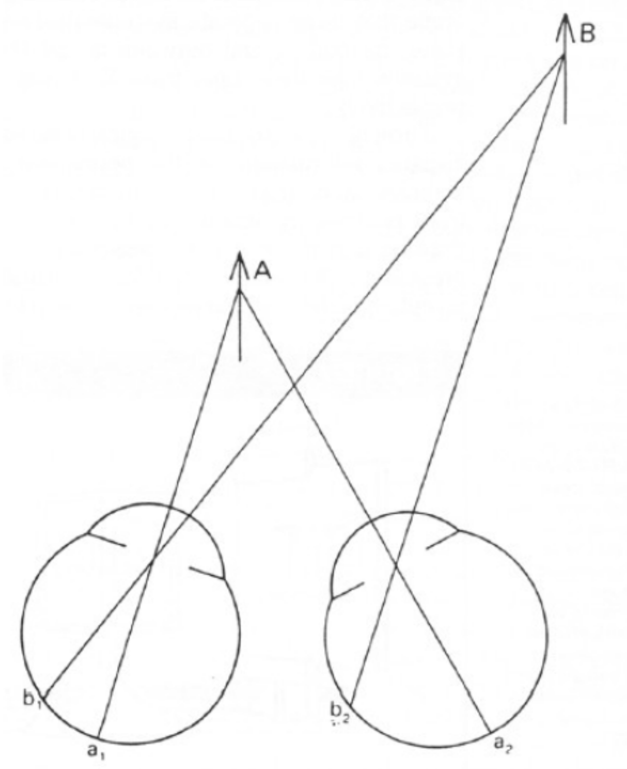
Martin Levine, ECSE 529, McGill

Stereopsis

- Stereopsis appears to be processed in the visual cortex in binocular cells having receptive fields in different horizontal positions in the two eyes.
- Such a cell is active only when its preferred stimulus is in the correct position in the left eye and in the correct position in the right eye, making it a disparity detector.

Stereopsis

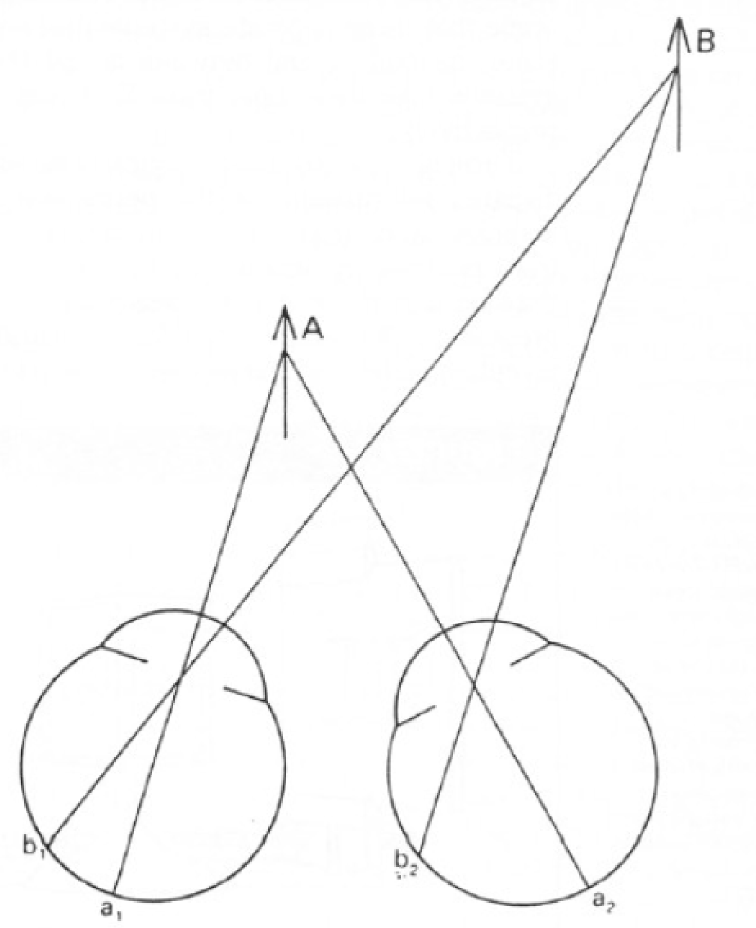
- When a person stares (**fixates**) at an object, the two eyes converge so that the object appears at the center of the retina in both eyes.
    - Eyes rotate so that corresponding images form in centers of fovea



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

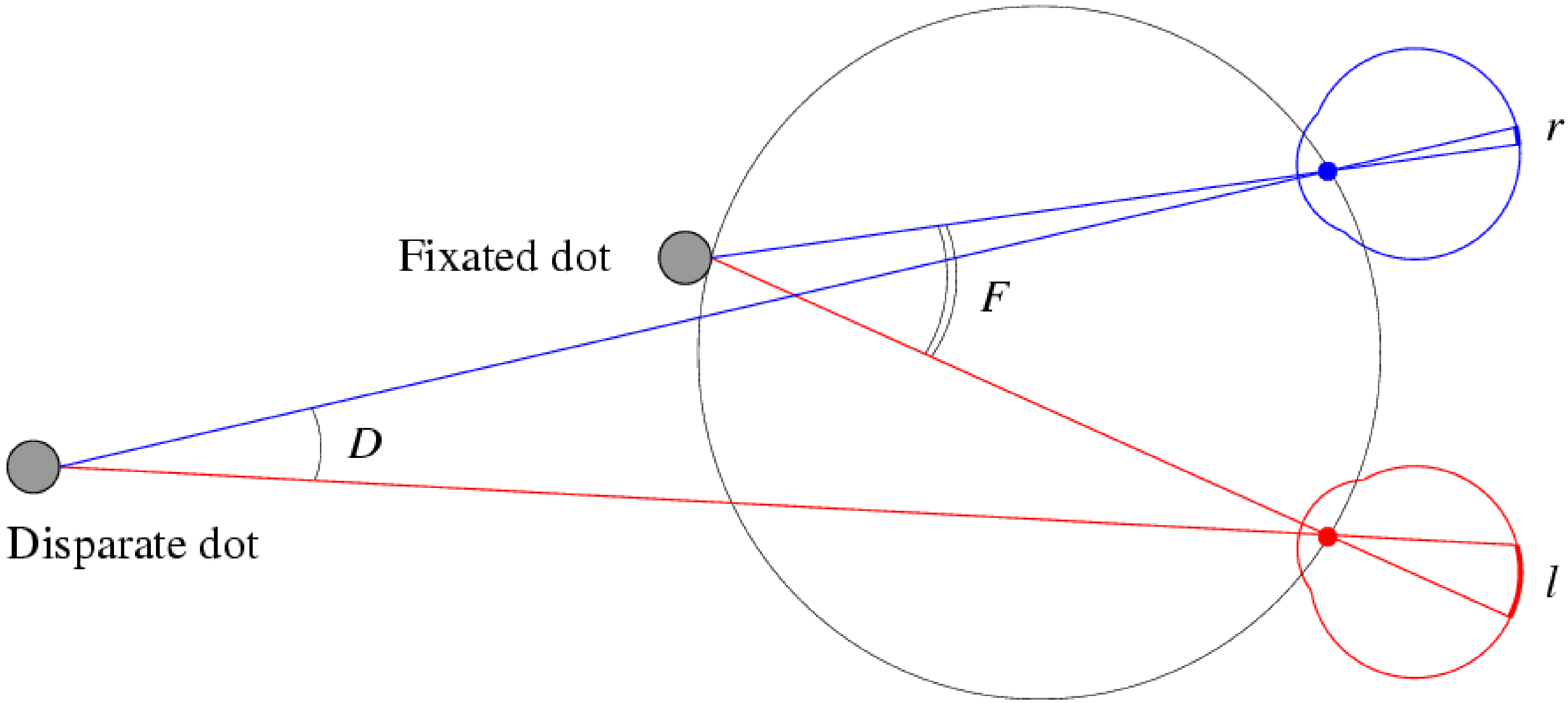
Human Stereopsis: Disparity

- **Disparity** occurs when eyes fixate on one object
  - Other objects around the main object appear shifted in relation to the main object.



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

Human Stereopsis: Disparity



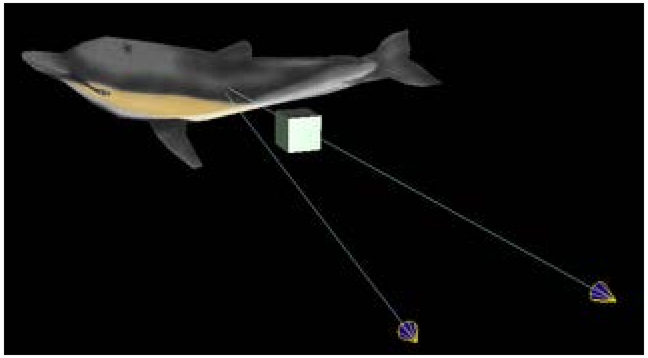
- Disparity $= r - l = D - F$

Forsyth & Ponce

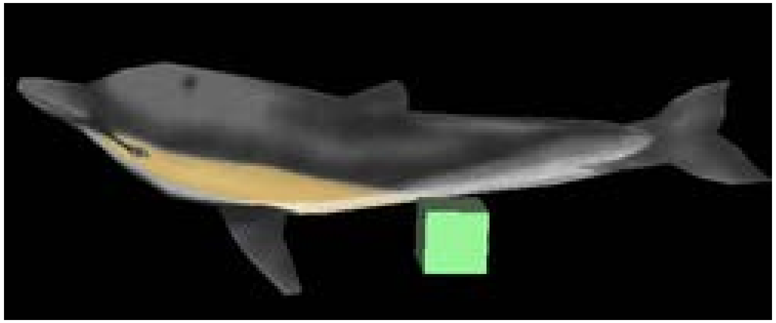
Stereopsis

- In the following example, whereas the main object (dolphin) remains in the center of the two images in the two eyes, the cube is shifted to the right in the left eye's image and is shifted to the left when in the right eye's image.

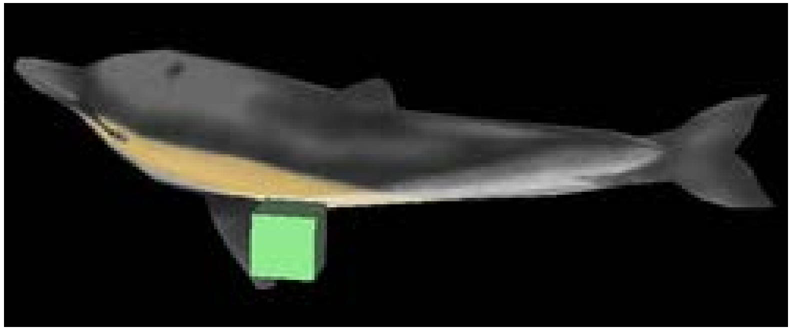
Stereopsis



http://en.wikipedia.org/wiki/Stereopsis

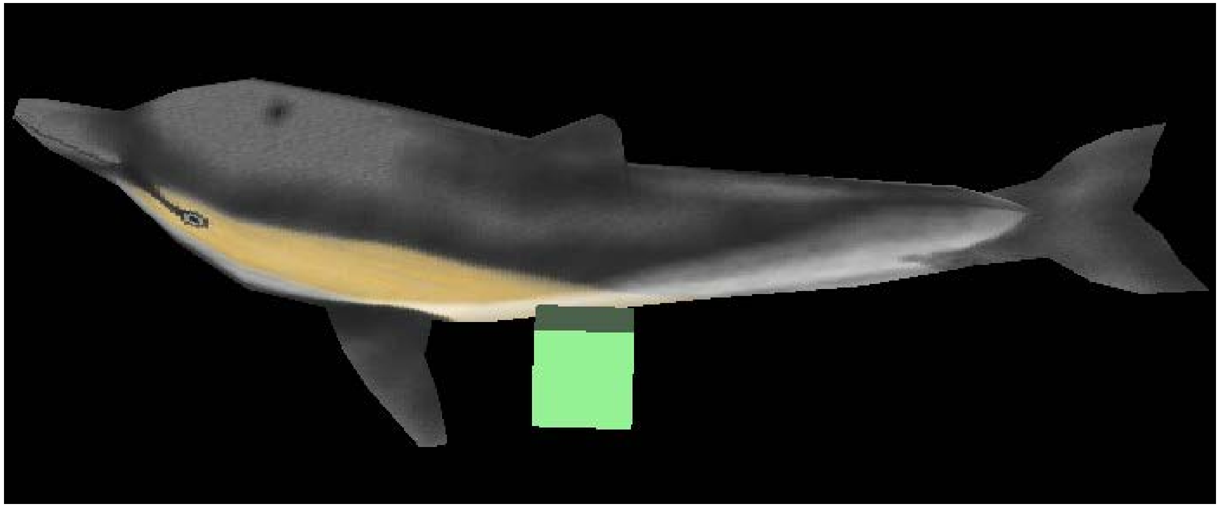


The cube is shifted to the right in left eye's image.



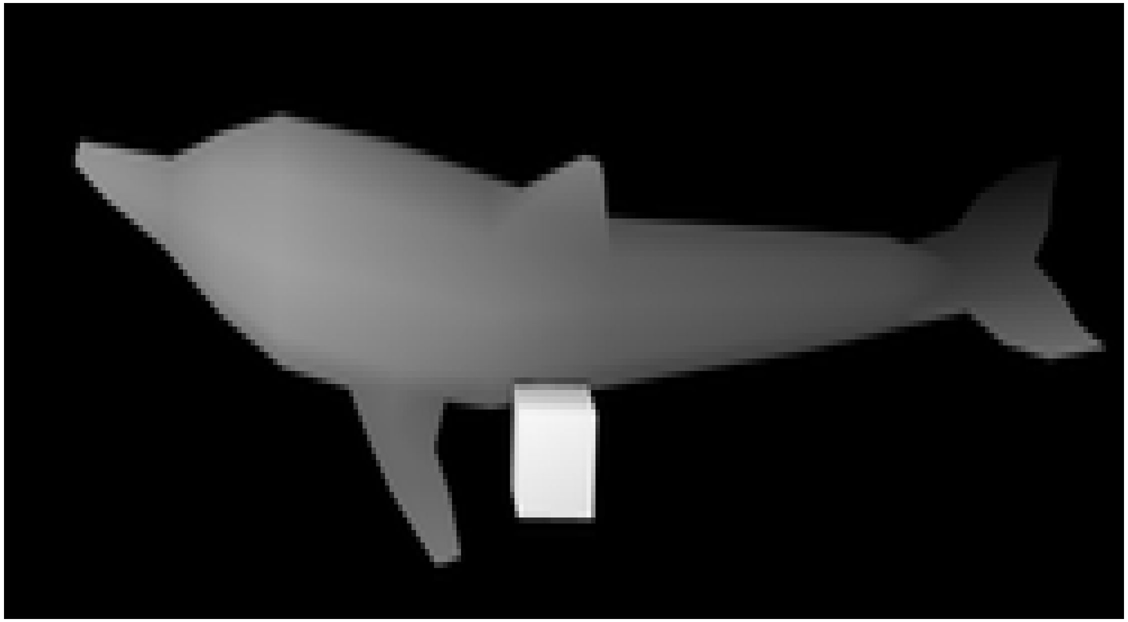
The cube is shifted to the left in the right eye's image.

Stereopsis



We see a single, Cyclopean, image from the two eyes' images.

Stereopsis



The brain gives each point in the Cyclopean image a depth value, represented here by a grayscale depth map.

Steroscope

- The stereoscope is essentially an instrument in which two photographs of the same object, taken from slightly different angles, are simultaneously presented, one to each eye.

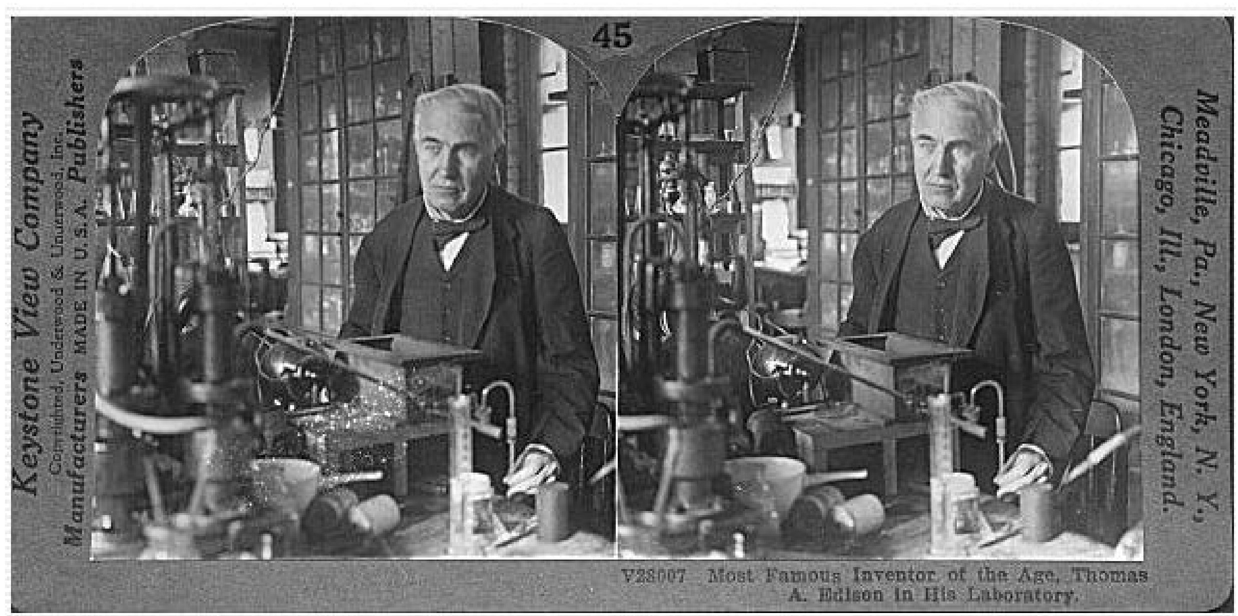
- http://lisahistory.net/hist106/pw/lectures/11LateVic.htm

Steroscope

- Each picture is focused by a separate lens, and the two lenses are inclined so as to shift the images toward each other and thus ensure the visual blending of the two images into one three-dimensional image.

Stereo photography and stereo viewers

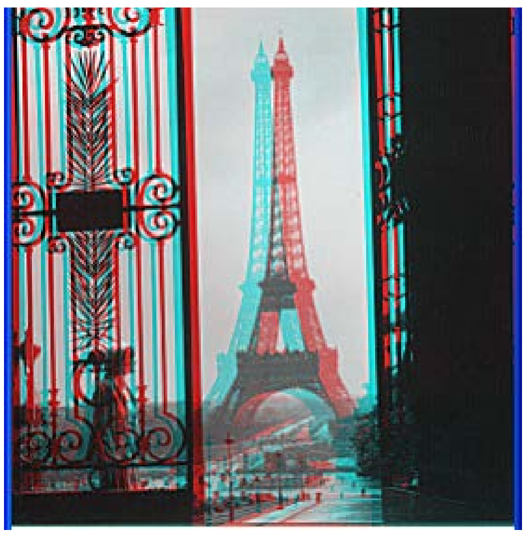
- Take two pictures of the same subject from two slightly different viewpoints
- display so that each eye sees only one of the images



http://www.johnsonshawmuseum.org

Stereo photography and stereo viewers

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



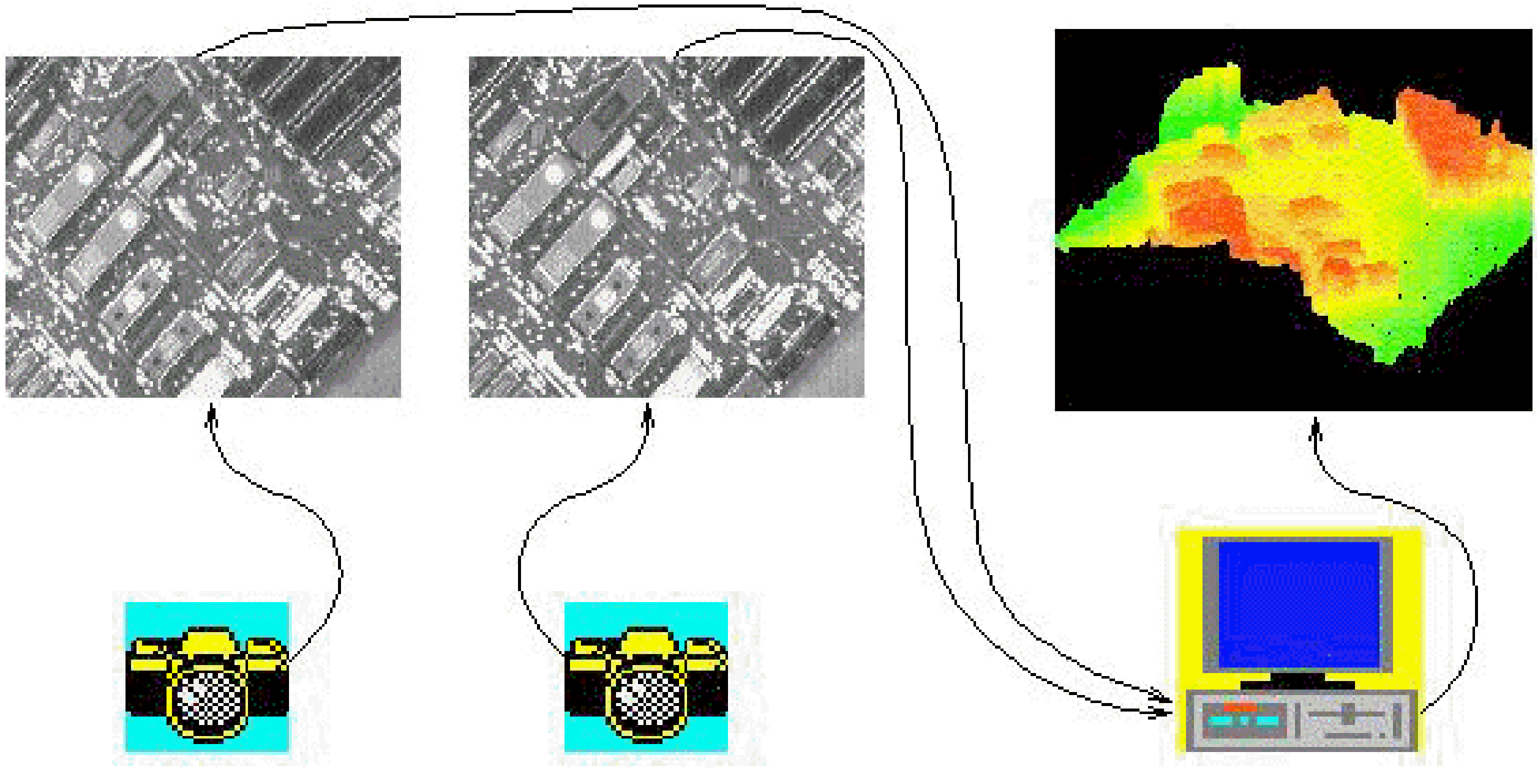
http://www.johnsonshawmuseum.org

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

Computer Vision: Estimating depth from stereo

In computer vision:
Binocular stereo vision is the process of computing depth at points in the scene from 2 images of the points acquired from cameras in different locations.

Computer Vision:
Estimating depth from stereo



http://www.cmis.csiro.au/iap/RecentProjects/stereoEE02.htm

Stereo Vision

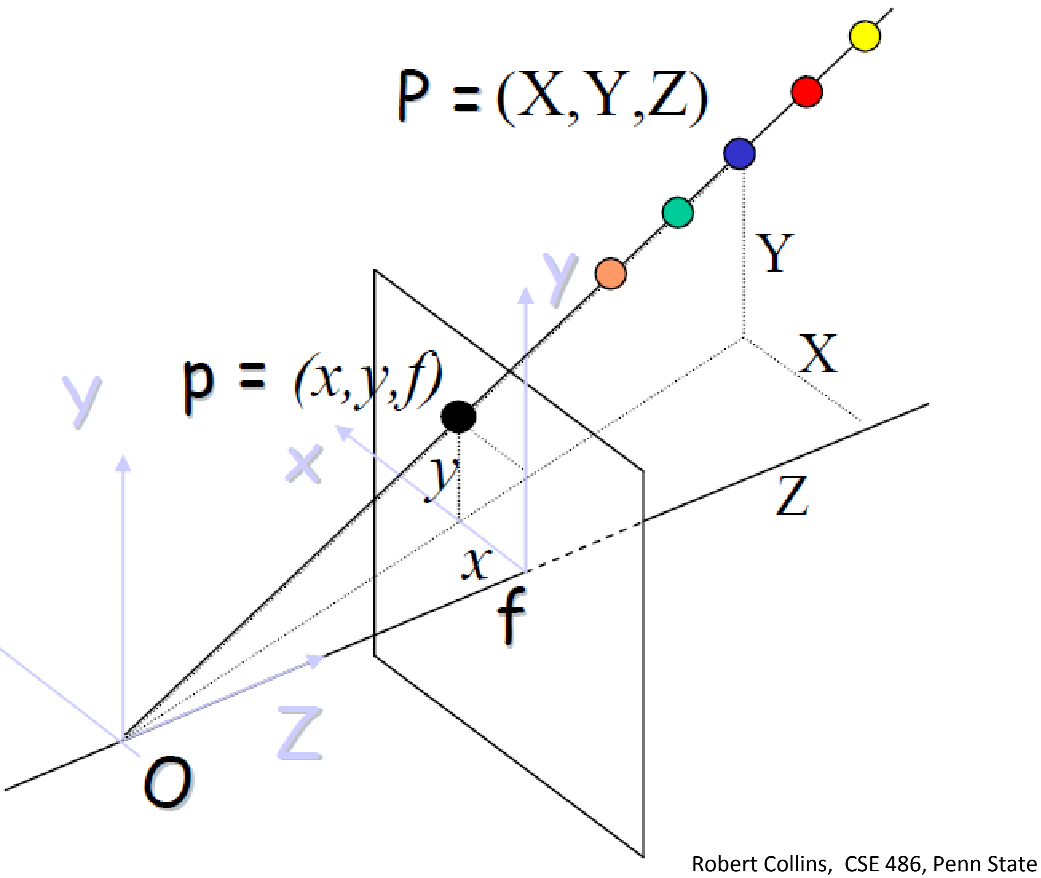
Two cameras, simultaneous views

Point Grey Cameras Inc.

Michael Black, CS 143, Brown University

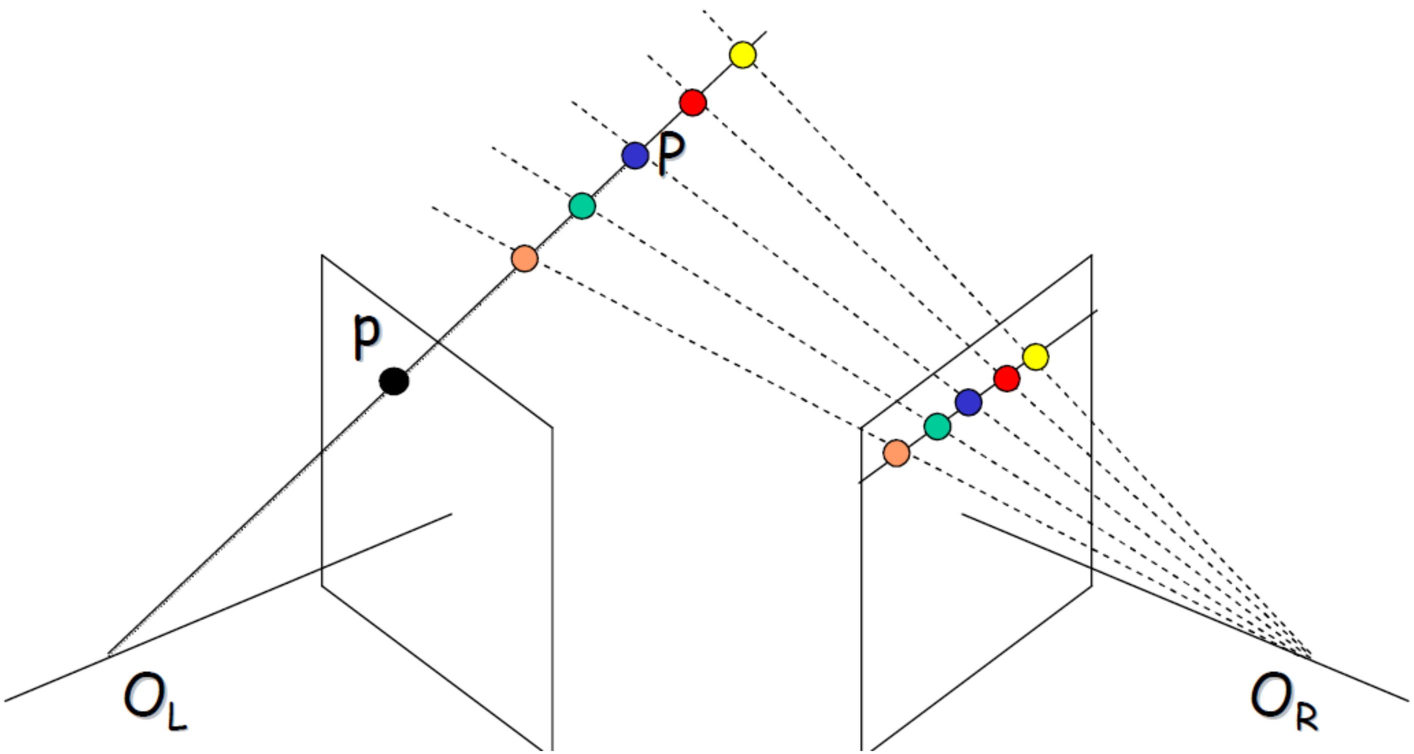
Stereo Vision

- Any point on the ray OP has image p



Stereo Vision

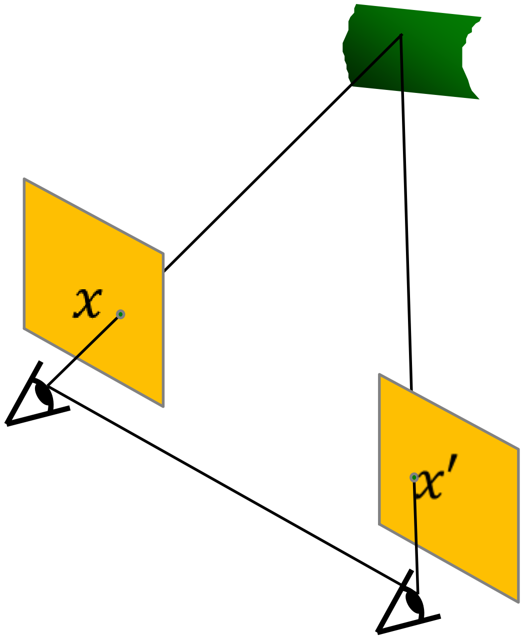
- A second camera can resolve the ambiguity, enabling depth measurement via triangulating

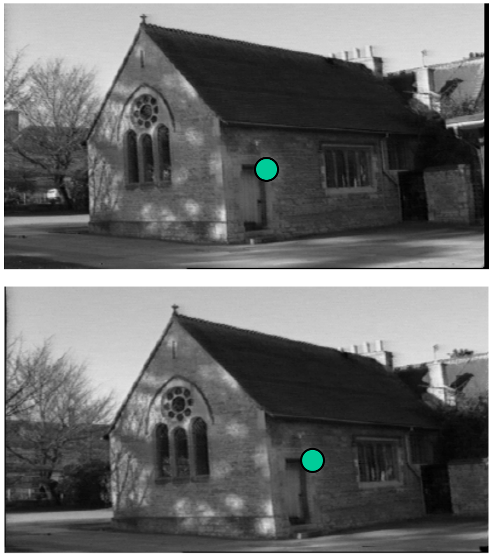


Robert Collins, CSE 486, Penn State

Estimating depth with stereo

- Goal: recover depth by finding image coordinate *x'* that corresponds to *x*.

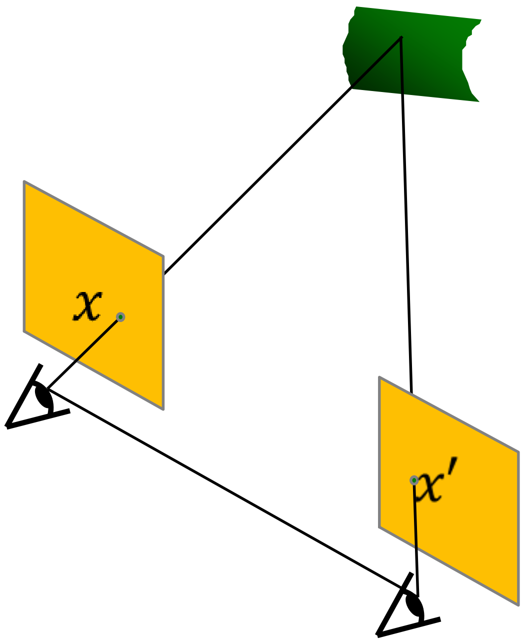




Farhadi, CSE 455, Washington University

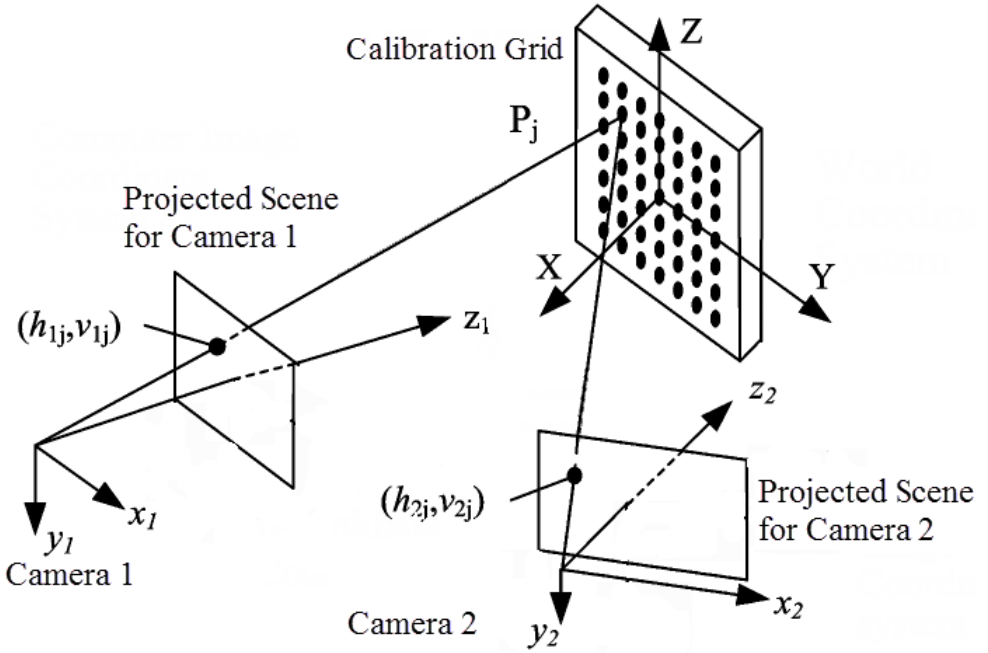
Estimating depth with stereo

- Goal: recover depth by finding image coordinate *x'* that corresponds to *x*.
- Sub-problems:
  - Calibration: How do we recover relationship between cameras (if unknown)?
  - Correspondence: How do search for *x'*?



Farhadi, CSE 455, Washington University

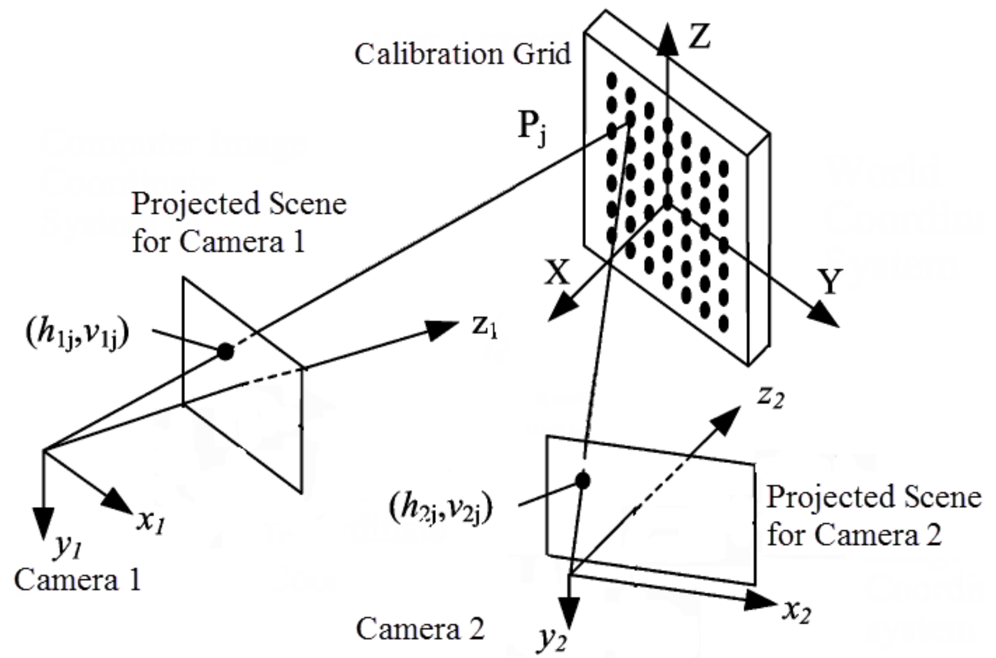
Camera Parameters



- Extrinsic parameters
  - rotation matrix and translation vector
  - Camera frame 1 ↔ Camera frame 2

Image from chriswalkertechblog.blogspot.com

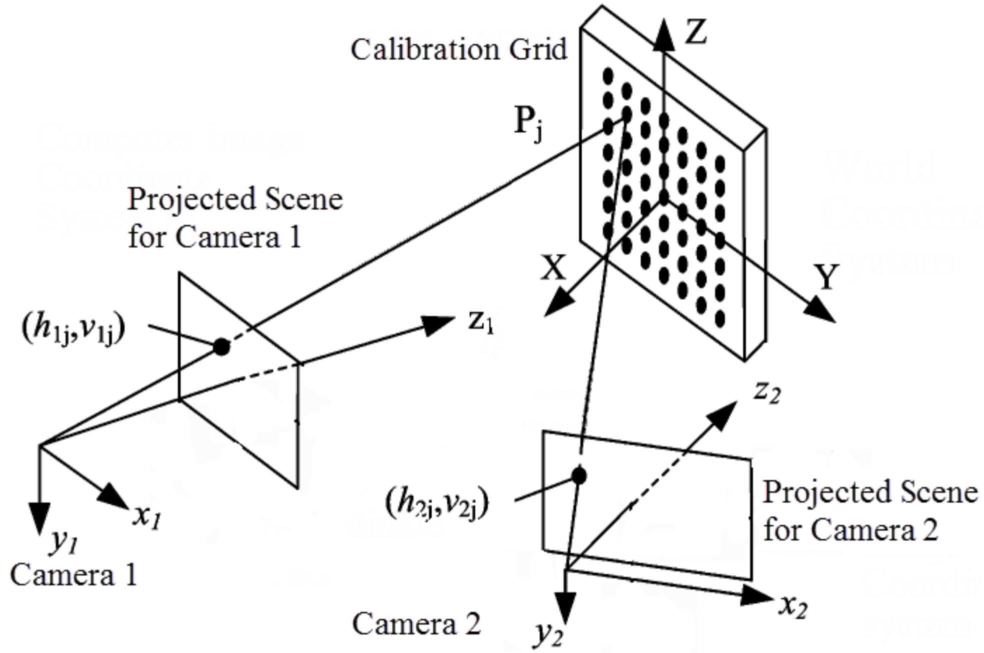
Camera Parameters



- Intrinsic parameters
  - focal length, pixel sizes (mm), image center point, lens distortion parameters
  - Image coordinates relative to camera ↔ pixel coordinates

Image from chriswalkertechblog.blogspot.com

Camera Parameters



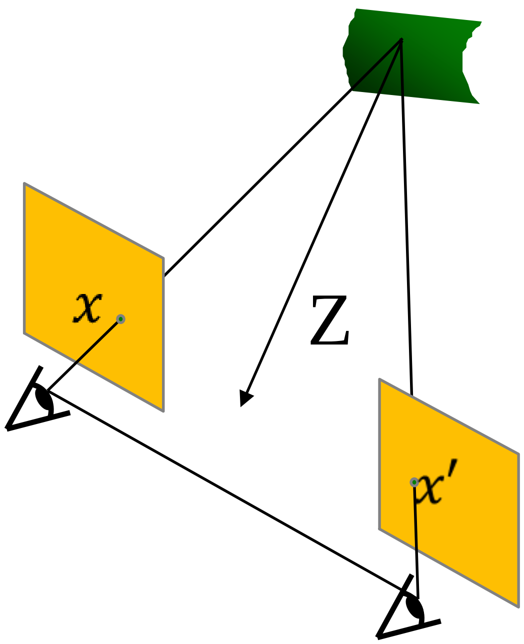
- *We'll assume for now that these parameters are given and fixed*

Image from chriswalkertechblog.blogspot.com

Geometry for a simple stereo system

Assume parallel optical axes, known camera parameters (i.e., calibrated cameras). **We are looking for Z**

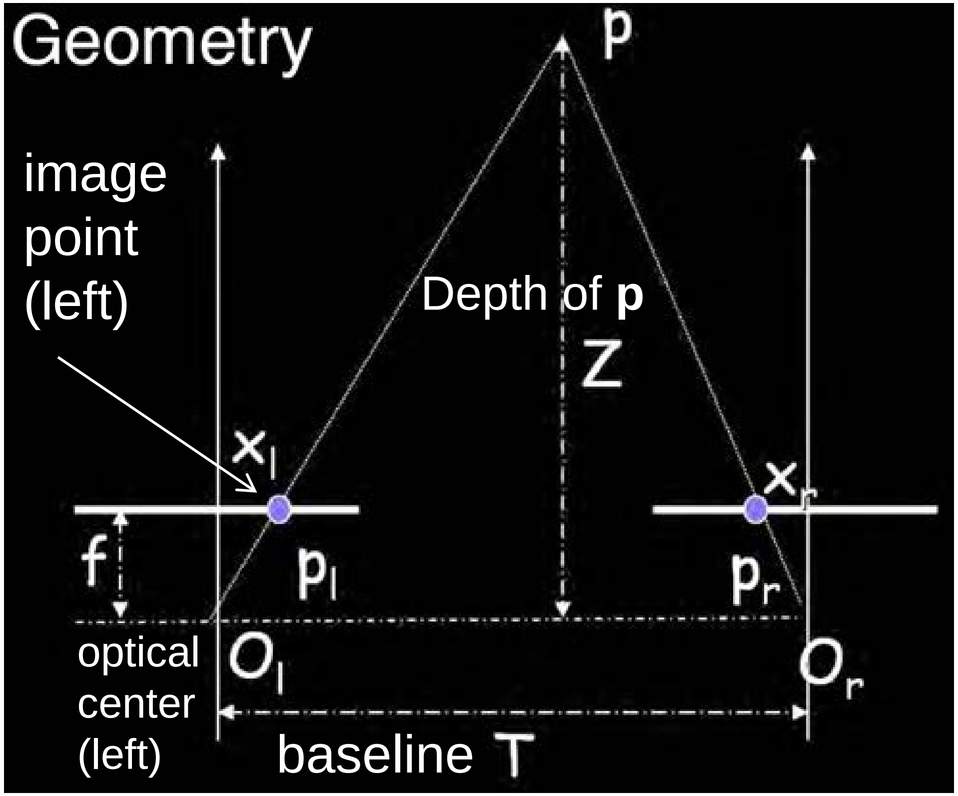
Z is a representation of the distance from a reference point attached to the plane connecting the optical centers of cameras to points in the scene with respect to some world coordinate system.



http://www.cse.psu.edu/~zyin/Demo/Stereo%20geometry.jpg

Geometry for a simple stereo system

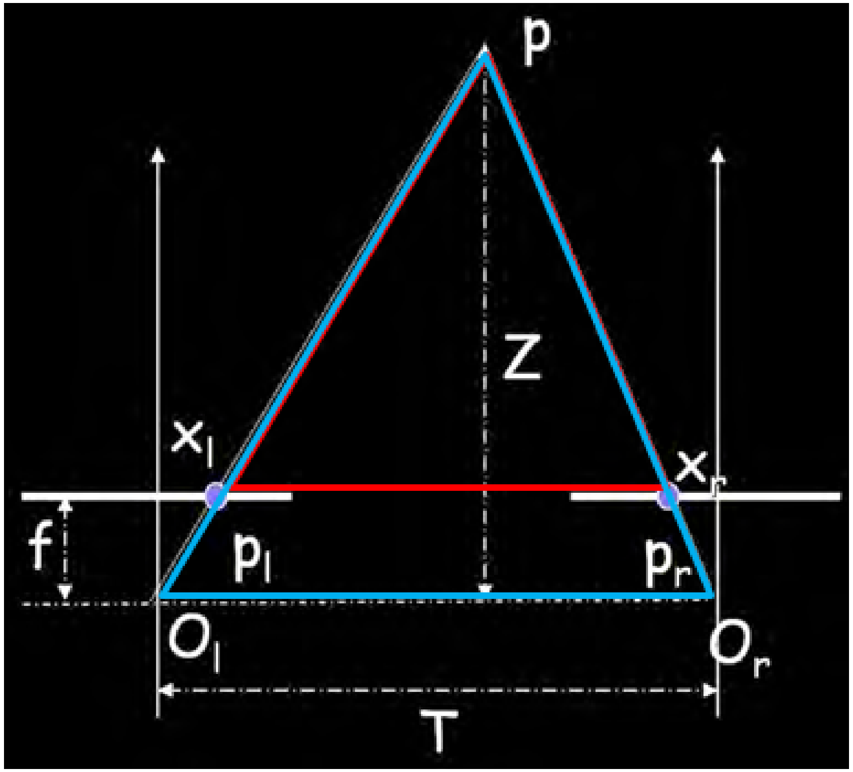
Assume parallel optical axes, known camera parameters (i.e., calibrated cameras). **We are looking for Z**



http://www.cse.psu.edu/~zyin/Demo/Stereo%20geometry.jpg

Geometry for a simple stereo system

- Assume parallel optical axes, known camera parameters (i.e., calibrated cameras). **We are looking for Z**



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

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

disparity

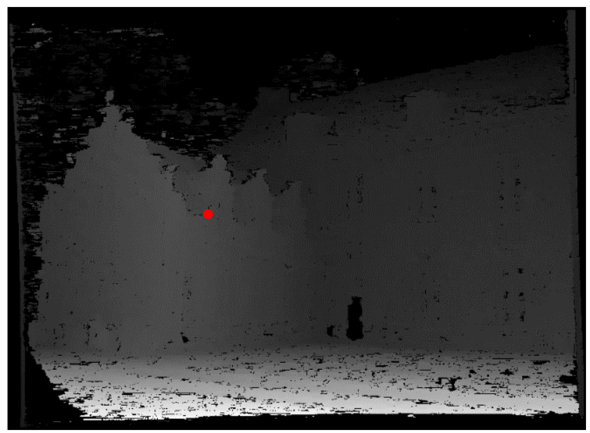
Michael Black, CS 143, Brown University

Depth from disparity

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

$$(x', y') = (x + D(x, y), y)$$

image $I(x, y)$

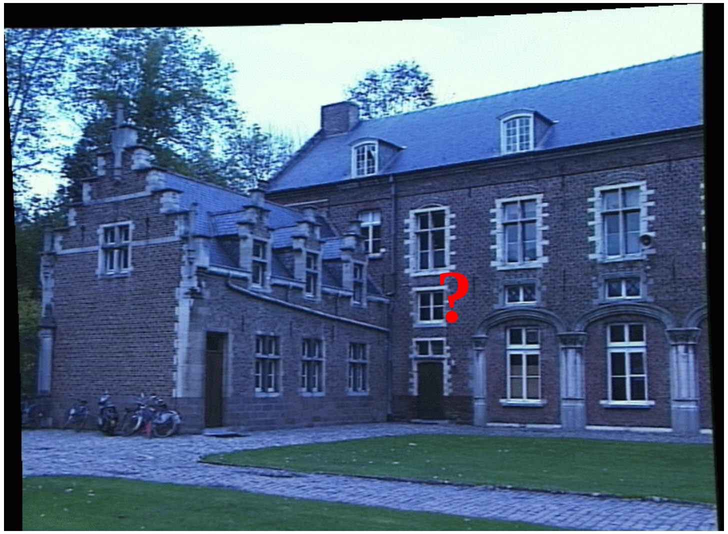


Disparity map $D(x, y)$

image $I'(x', y')$

Michael Black, CS 143, Brown University

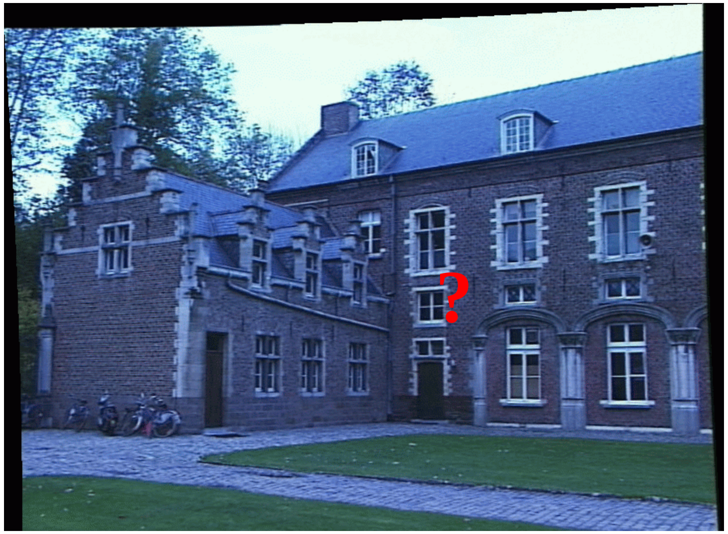
Correspondence Problem



- We need to be able to match points in the set of images which correspond to the same point in the scene.
- This problem, known as the *correspondence problem*, is a very difficult one, in general.

Michael Black, CS 143, Brown University

Correspondence Problem



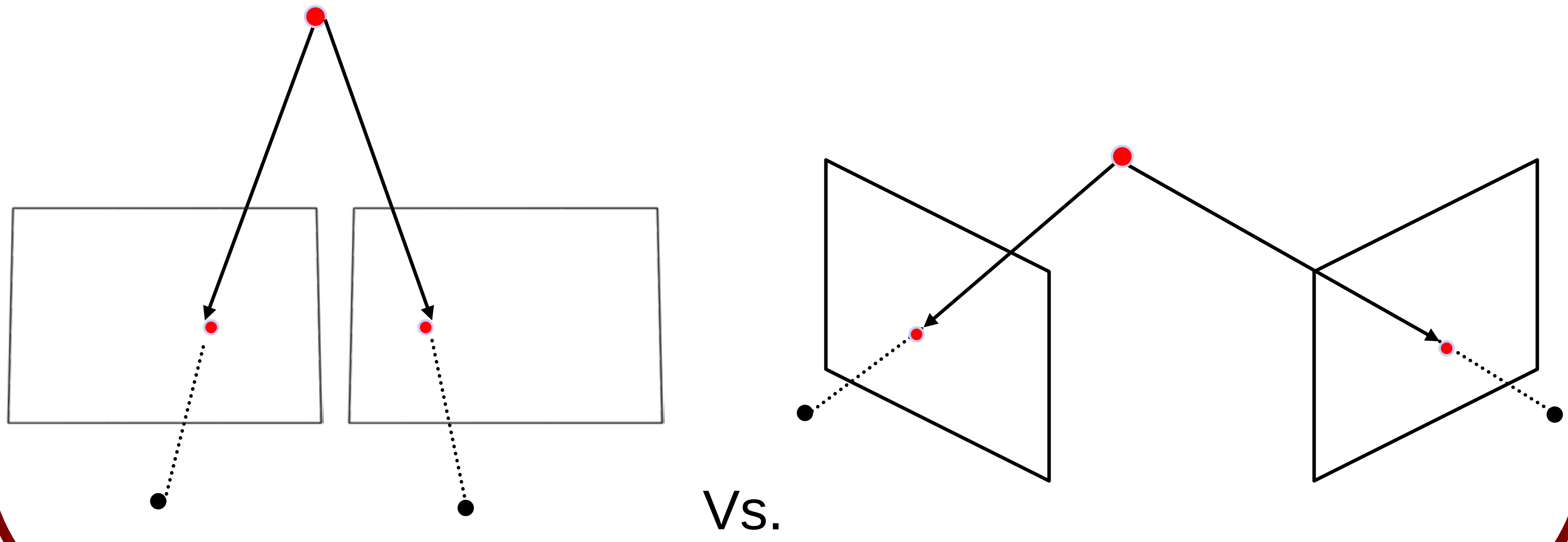
- We have two images taken from cameras with different intrinsic and extrinsic parameters.

- How can we constrain our search?

Michael Black, CS 143, Brown University

General case, with calibrated cameras

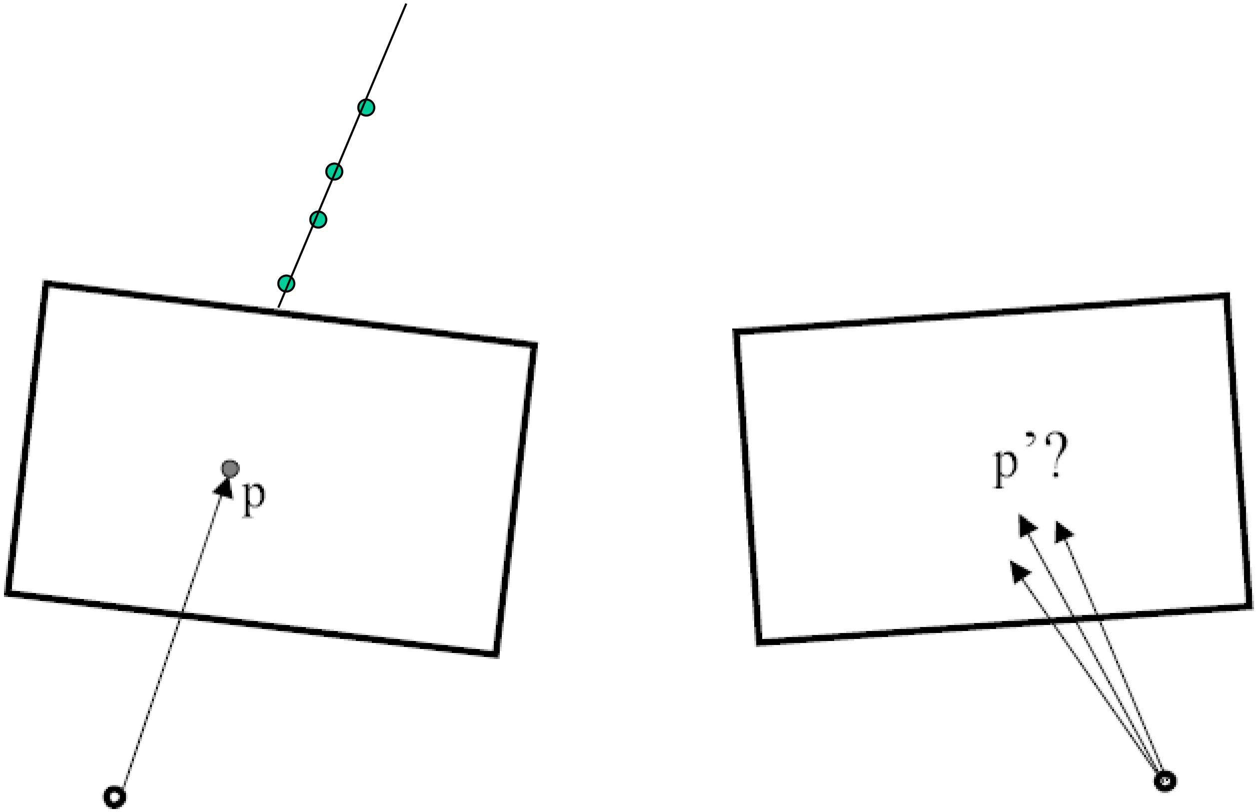
- The two cameras need not have parallel optical axes



James Hays, CS 143, Brown University

General case, with calibrated cameras

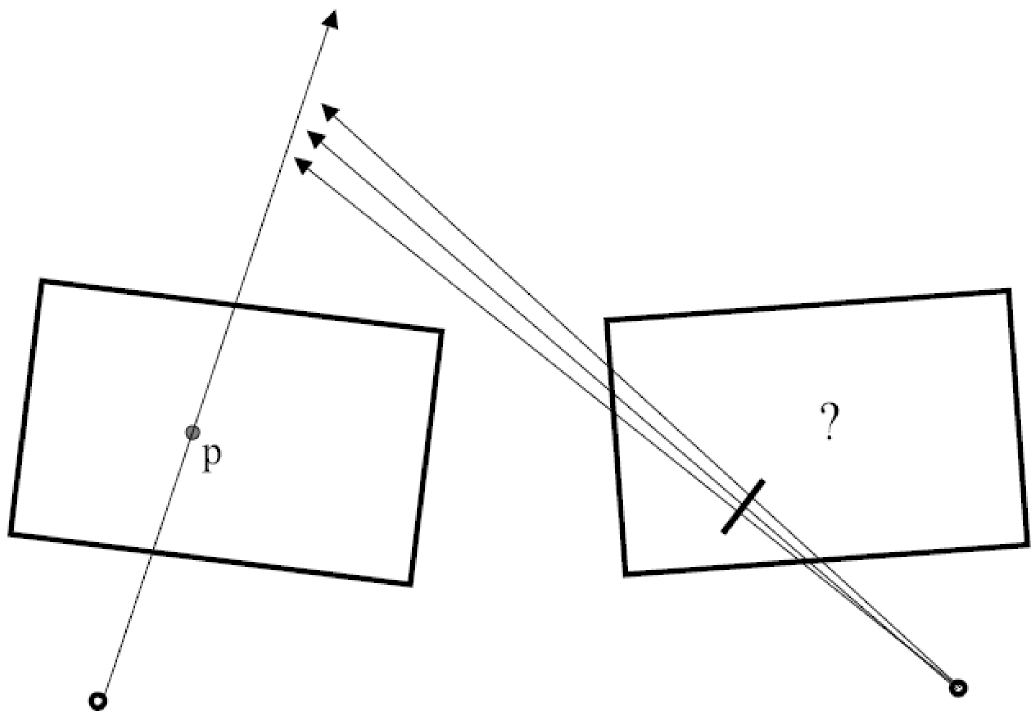
- Given p in left image, where can corresponding point p' be?



James Hays, CS 143, Brown University

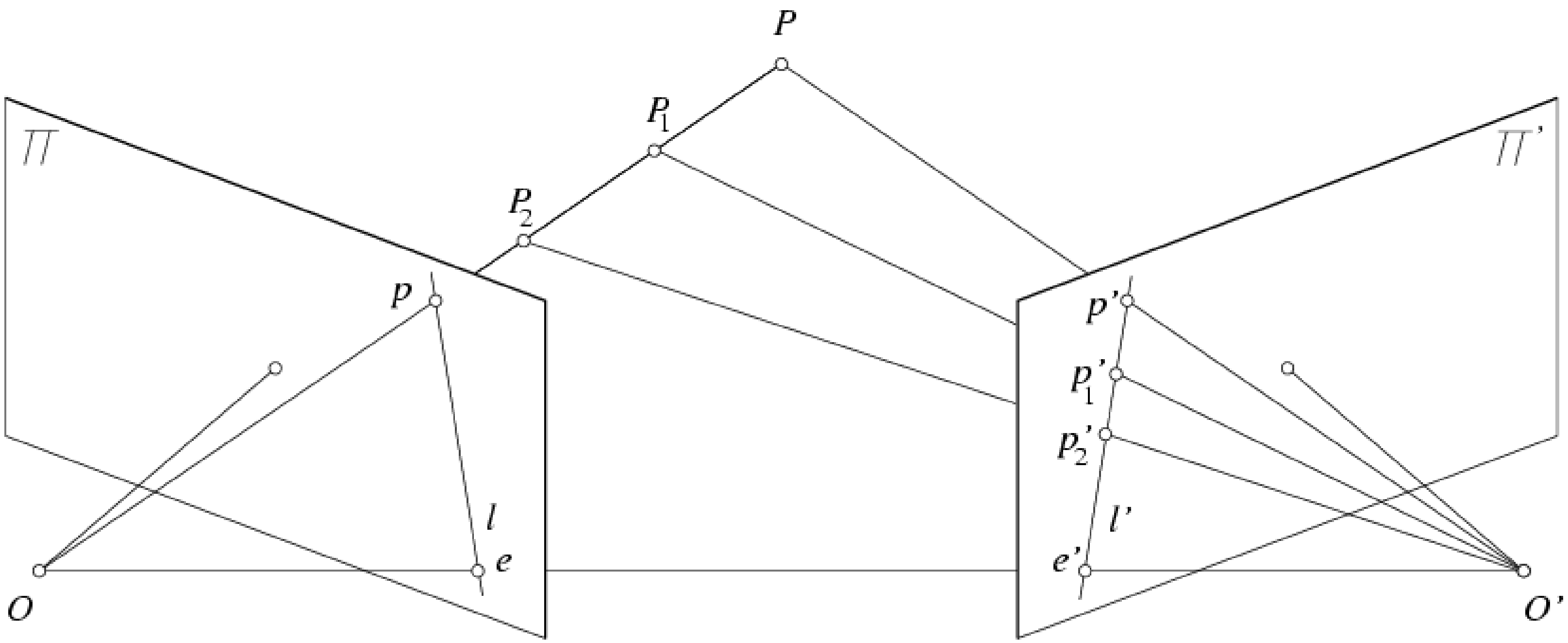
General case, with calibrated cameras

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



James Hays, CS 143, Brown University

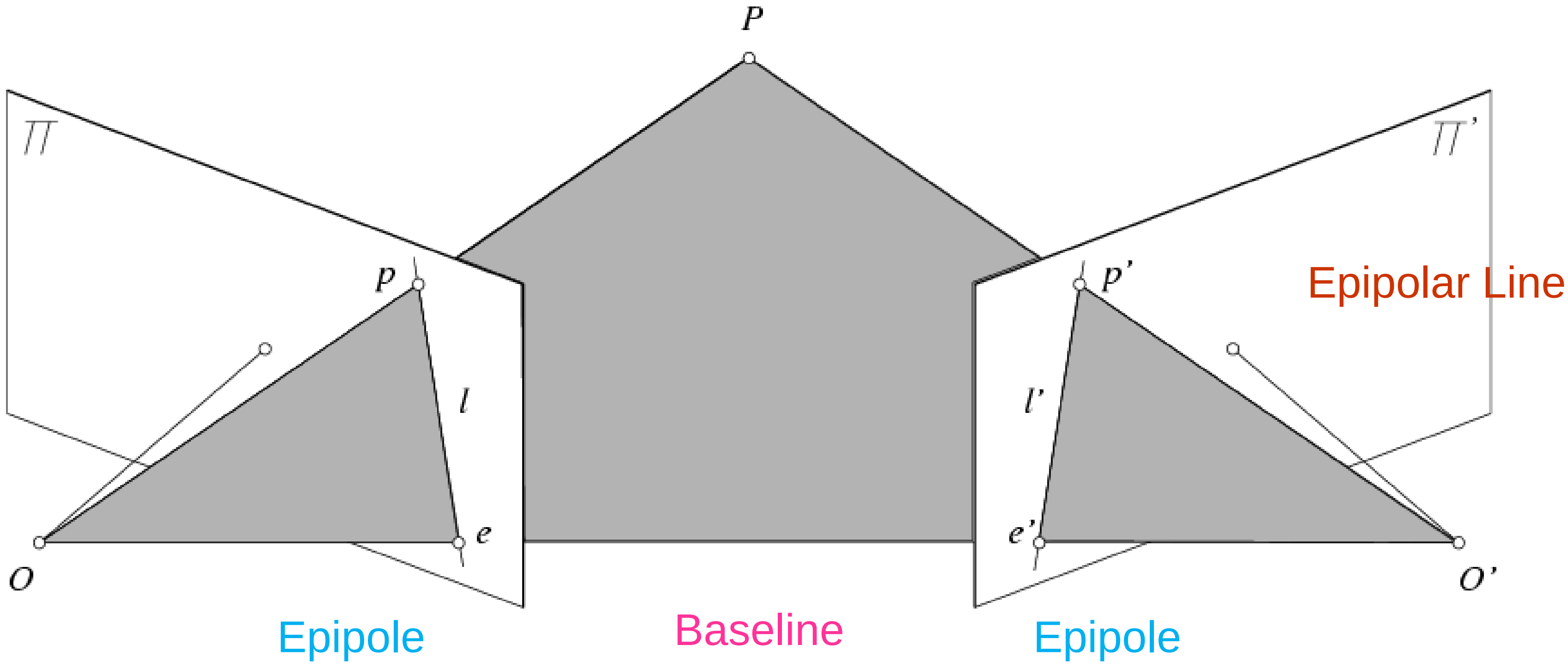
Epipolar constraint



- 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

James Hays, CS 143, Brown University

Epipolar geometry



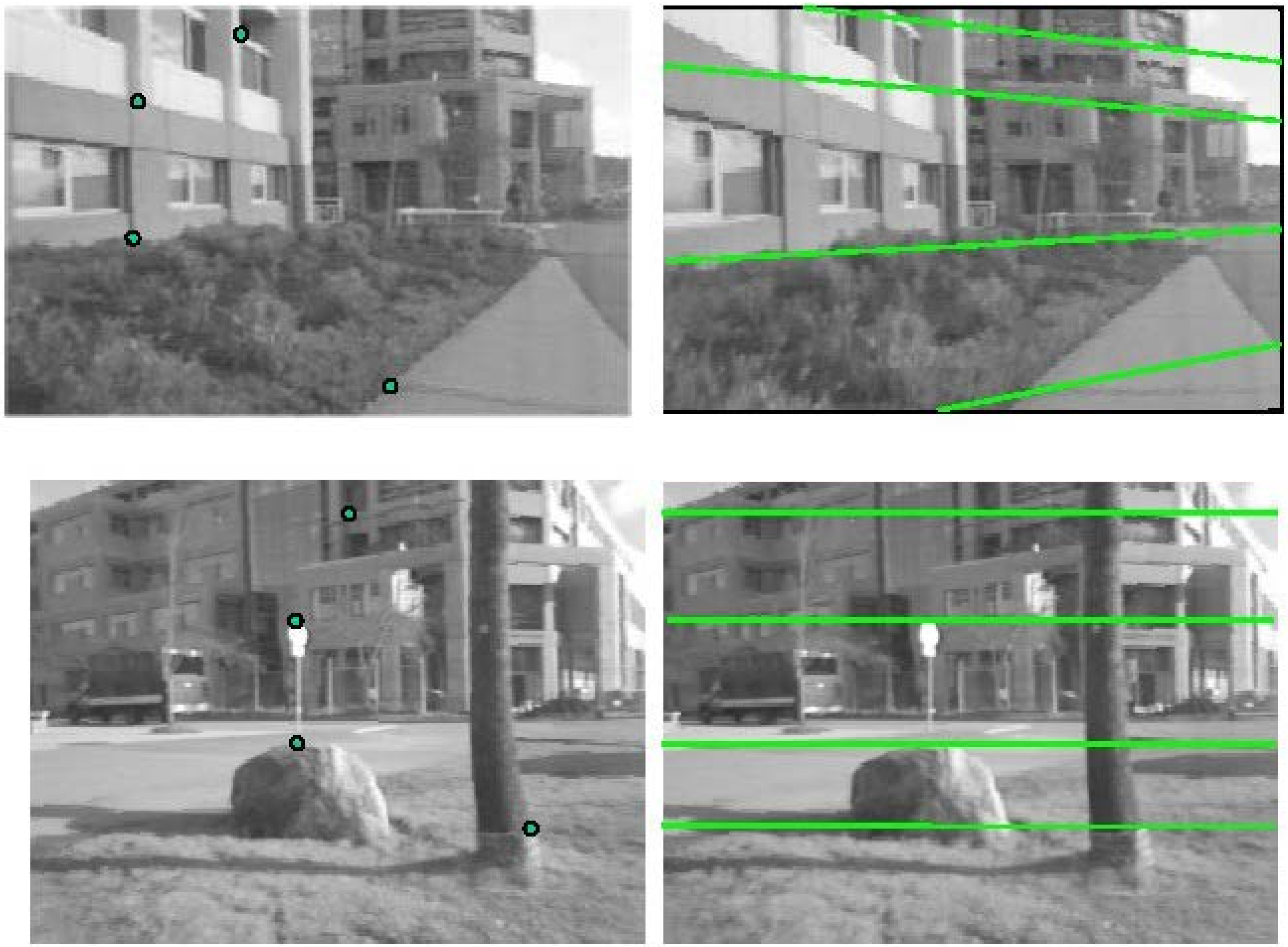
http://www.ai.sri.com/~luong/research/Meta3DViewer/EpipolarGeo.html

Why is the epipolar constraint useful?

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

Image from Andrew Zisserman

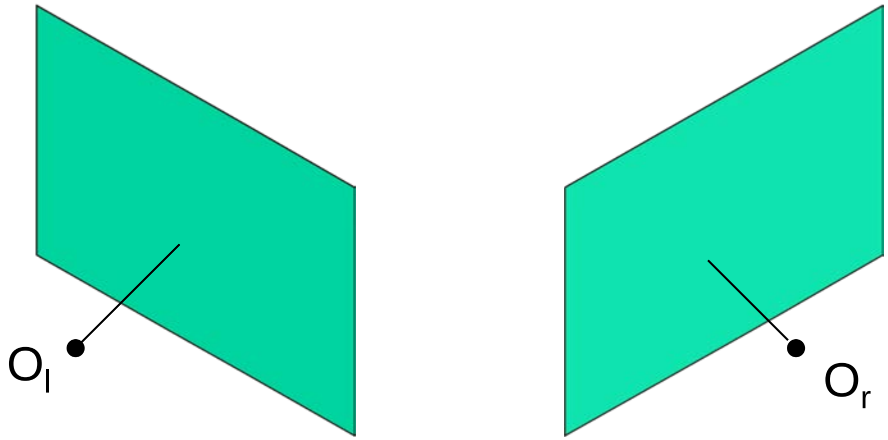
Why is the epipolar constraint useful?



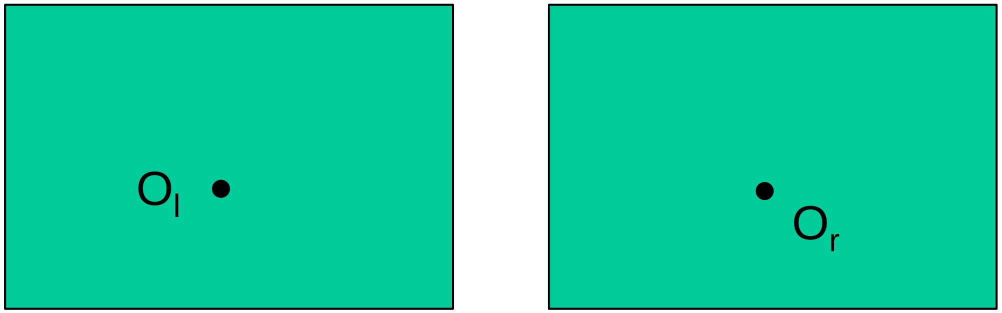
James Hays, CS 143, Brown University

What do the epipolar lines look like?

1.



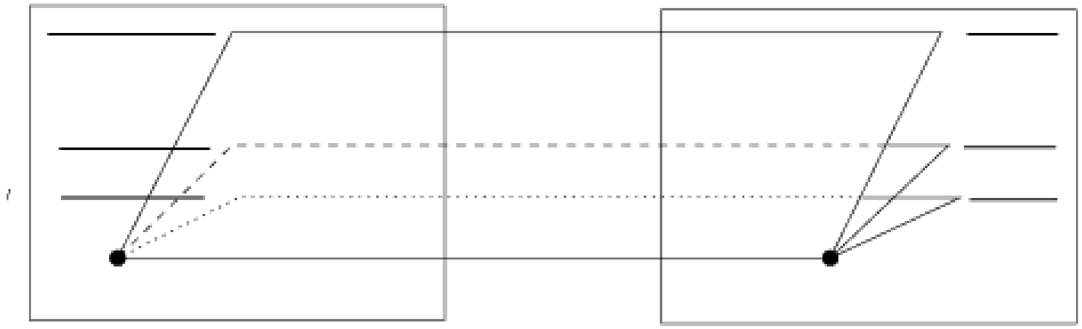
2.



James Hays, CS 143, Brown University

What do the epipolar lines look like?

Parallel cameras



Where are the epipoles?

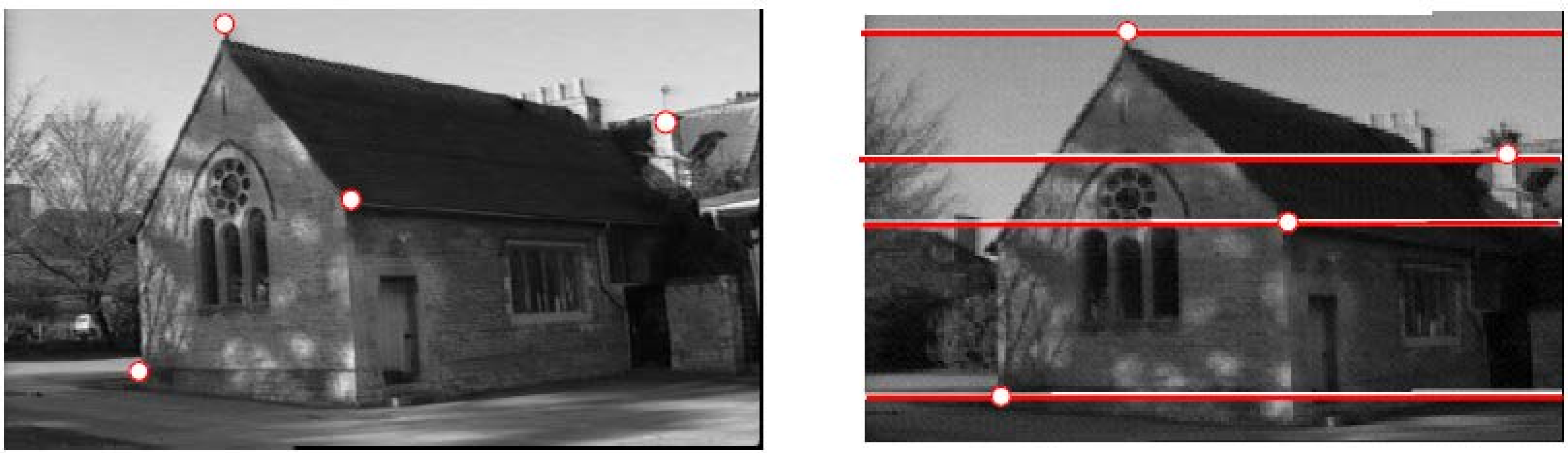
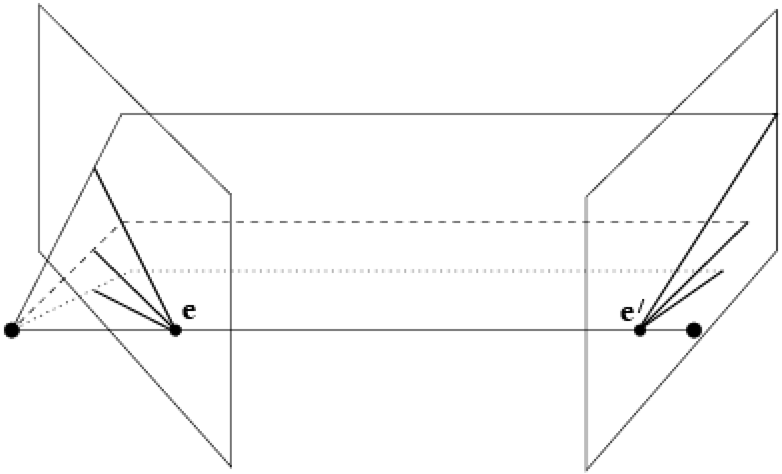
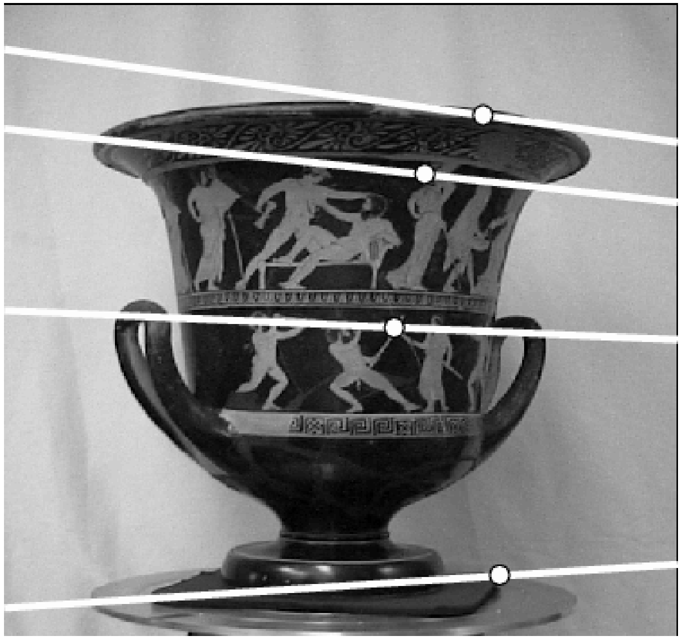


Figure from Hartley & Zisserman

What do the epipolar lines look like?

converging cameras





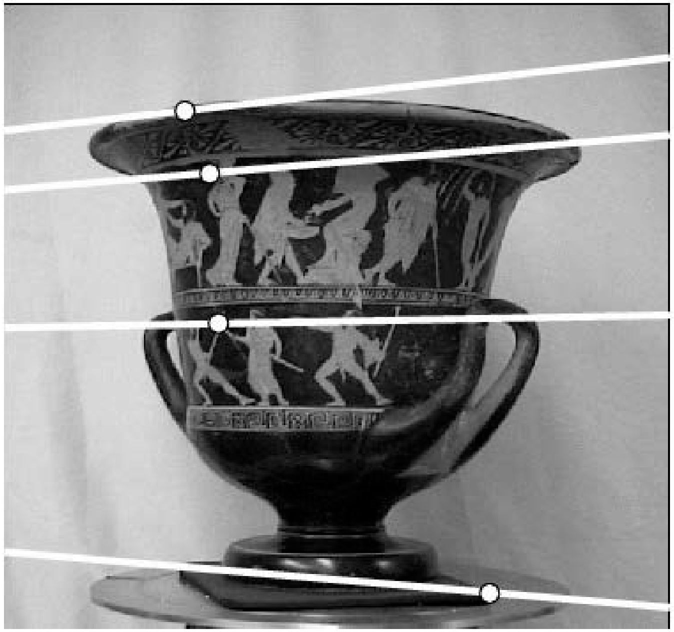
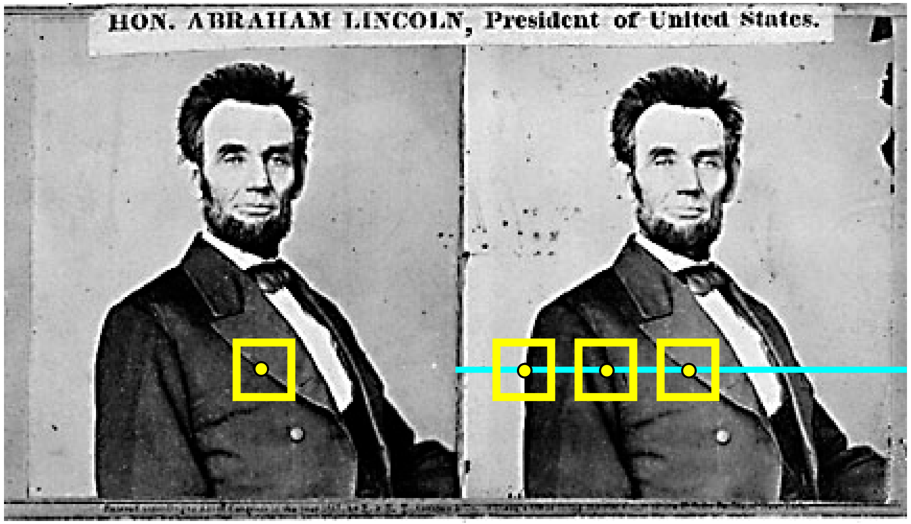


Figure from Hartley & Zisserman

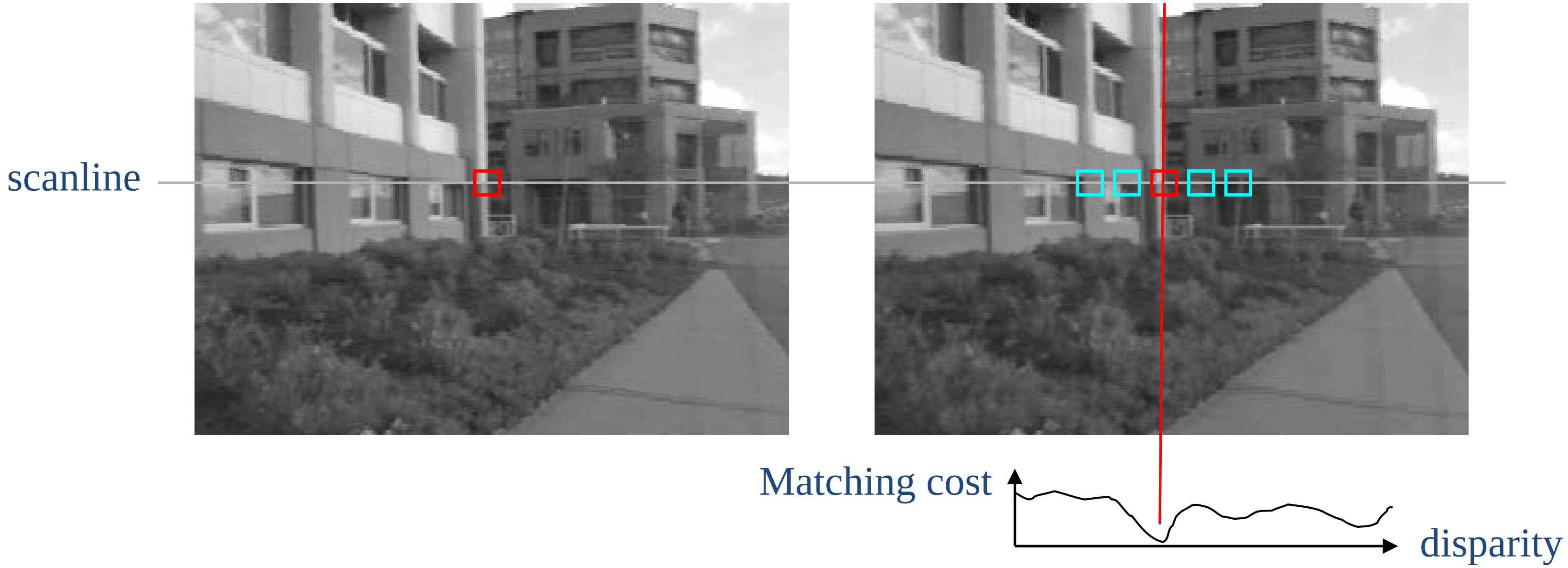
Basic stereo matching algorithm



- For each pixel x in the first image
  - Find corresponding epipolar scanline in the right image
  - Examine all pixels on the scanline and pick the best match x'
  - Compute disparity $x - x'$ and set $Z(x) \leftarrow \frac{fT}{(x-x')}$

Michael Black, CS 143, Brown University

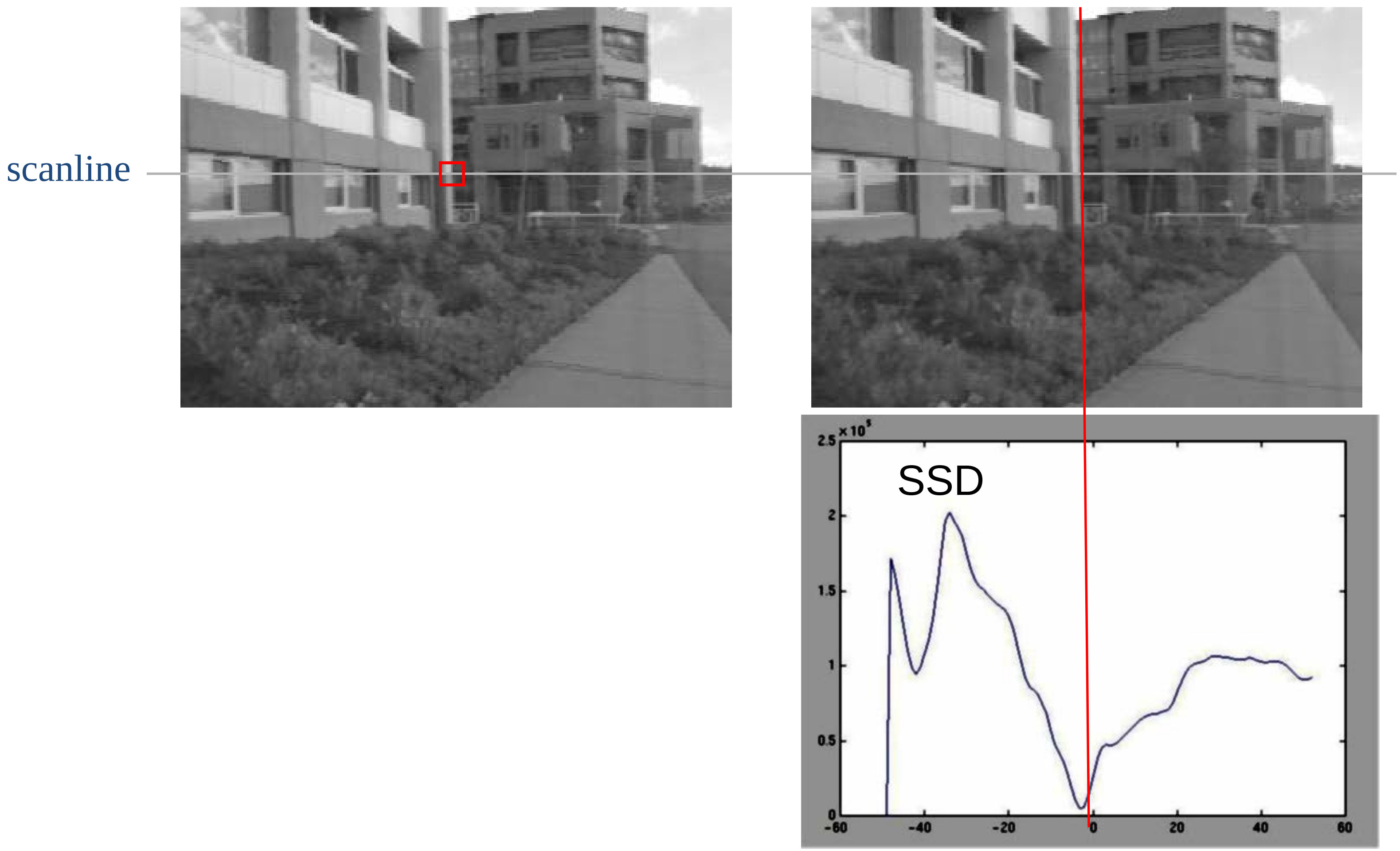
Correspondence search



- 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

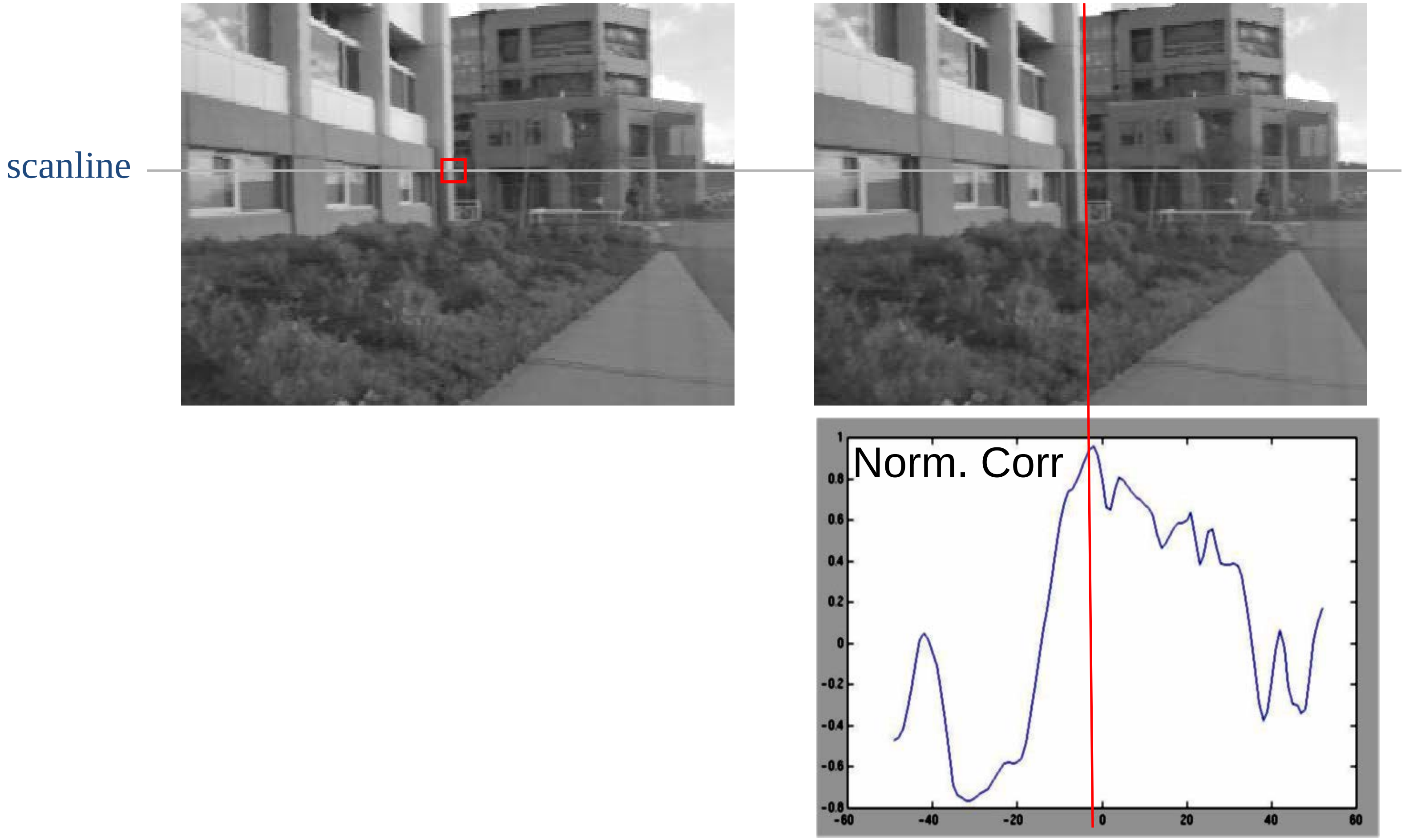
Michael Black, CS 143, Brown University

Correspondence search



Michael Black, CS 143, Brown University

Correspondence search



Michael Black, CS 143, Brown University

Stereo as an Inverse Problem

Is this problem *well-posed*?

Is there a unique solution?

Inverse Problem

We need to be able to match points in the set of images which correspond to the same point in the scene.

It may appear at first glance that we have enough information in the images to uniquely determine the depth values at each point.

This is not true in general, however, due to ambiguities in the correspondence between the images of scene points.

The correspondence problem

- Epipolar geometry constrains our search, but we still have a difficult correspondence problem.
- In general, the *correspondence problem* does not have a unique solution.
- Therefore the stereo vision problem is ill-posed.

James Hays, CS 143, Brown University

Regularization

We could consider regularization as a means of handling the ill-posedness of the stereo problem.

For example, suppose we try to find the disparity between the images, which can be defined as the vector in the image plane coordinates between the images of corresponding points.

Regularization

That is, suppose a scene point projects to a point (x_1,y_1) in image 1 and to (x_2,y_2) in image 2.

Then the disparity vector will be

$$d = (x_2-x_1,\ y_2-y_1).$$

The disparity can be seen to be related to the distance D.

Regularization

Iso-brightness constraint: We could make an assumption that the brightness of corresponding points in all images is the same.

Thus we could try to find the D(x,y) for which $I_1(x,y) = I_2((x,y) + D(x,y))$.

Regularization

In two-dimensions there are no choices of I_1 and I_2 for which such a D(x,y) is unique.

If we assume that D(x,y) = D(x) (i.e. that the disparity varies only along one direction, which will always be the case when we have only two images) then there will be a unique D(x,y) that satisfies the *iso-brightness* constraint only if :
$I_1(x,y)$ and $I_2(x,y)$ are strictly *monotonic* functions of x.

Regularization

In general, the matching problem admits many possible solutions, so that the problem is ill-posed.

So we could try to regularize the problem, and find the disparity function which minimizes the following functional:

Regularization

In general, the matching problem admits many possible solutions, so that the problem is ill-posed.

So we could try to regularize the problem, and find the disparity function which minimizes the following functional:

$$\iint (I_1(x,y) - I_2(x + D(x), y))^2 + \lambda \left| \frac{dD(x)}{dx} \right|^2 \, dxdy$$

Regularization

- The “smoothness” constraint in effect limits sudden big changes in disparity.

- What are the implications of this constraint?
- In what way does this constrain our physical world and how plausible is the solution?

Problems with regularization

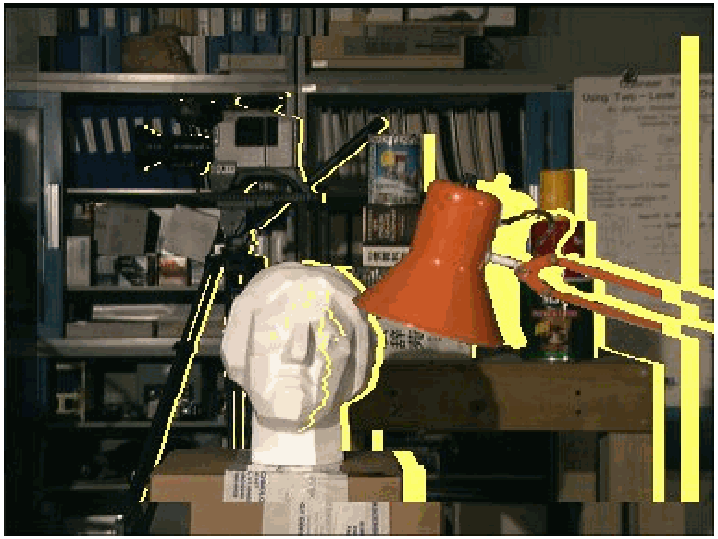
There are a number of difficulties with this approach:

- The true physical disparity is usually discontinuous.

Problems with regularization

There are a number of difficulties with this approach:

• The true physical disparity is undefined where occlusions are present between objects, as a point in one image may have no corresponding point in another image.



http://vision.middlebury.edu/stereo/data/scenes2001/data/imagehtml/tsukuba.html

Failures of correspondence search

- Image Brightness Constancy: Assuming Lambertian surfaces, the brightness of corresponding points in stereo images are the same

Farhadi, CSE 455, Washington University

Why does stereo fail?

- Monotonic Ordering - Points along an epipolar scanline appear in the same order in both stereo images

- Occlusion – All points are visible in each image

Farhadi, CSE 455, Washington University

Why does stereo fail?

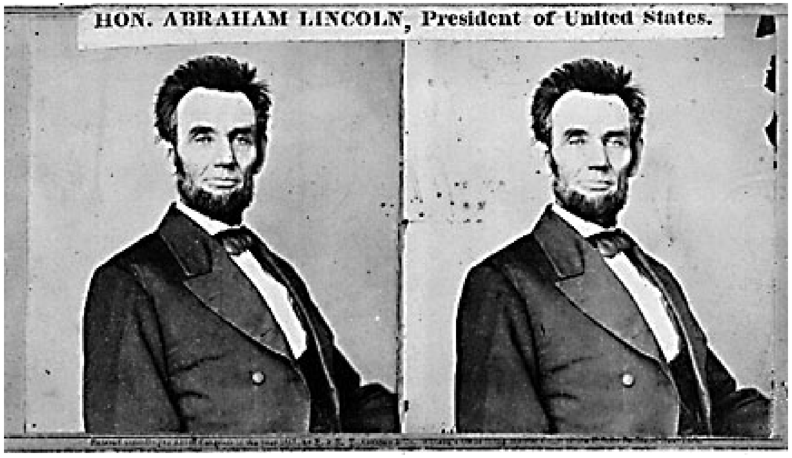
- Fronto-Parallel Surfaces: Depth is constant within the region of local support

Farhadi, CSE 455, Washington University

Problems with regularization

• In general, the solution provided by the regularization method will not correspond to the physical reality.

Failures of correspondence search



Textureless surfaces

Occlusions, repetition

Non-Lambertian surfaces, specularities

Problems with regularization

Because of the extreme ambiguity imposed by the correspondence problem, the functional shown earlier is often very difficult to minimize.

There are usually a large number of local minima.

The discontinuity of D(x) can be handled by the use of *"line processes".* These are binary valued functions which are used to "break" the smoothness constraint. That is, we allow the solution to violate smoothness at a small set of isolated points. To do this, we minimize the following functional:

$$\iint\left((I_1(x,y) - I_2(x + D(x), y))^2 + \lambda(1 - l(x,y)) \left| \frac{dD(x)}{dx} \right|^2 + l(x,y) \right) dxdy$$

$$\iint \left((I_1(x,y) - I_2(x + D(x), y))^2 + \lambda(1 - l(x,y)) \left| \frac{dD(x)}{dx} \right|^2 + l(x,y) \right) dx dy$$

This functional is minimized jointly with respect to the disparity function *D(x)* and the line process function *l(x)*. Note that when the line process is equal to one, the smoothness constraint is multiplied by zero, effectively removing it. We have added an extra term which measures how many breaks we have created. We want this to be small, otherwise we would have breaks everywhere!

Problems with regularization

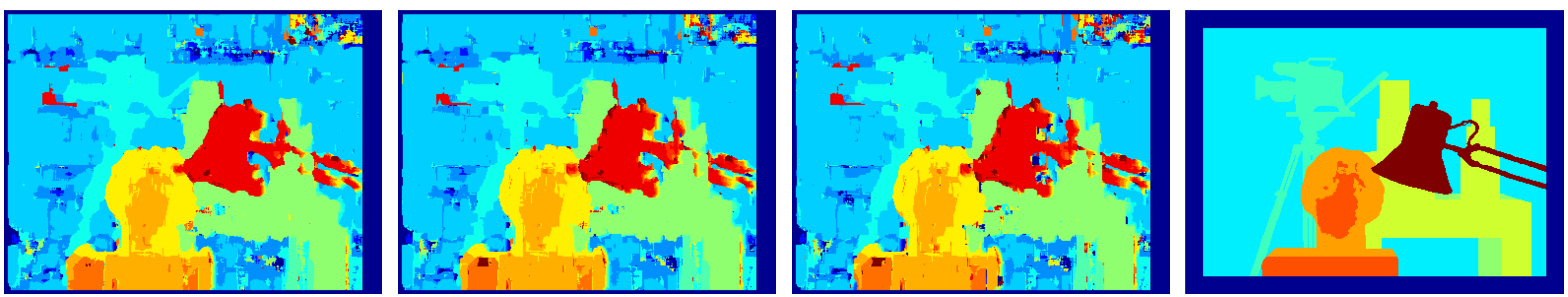
Current research looks into more sophisticated ways to impose more physical 2D constraints into global optimization such as:

- Considering occlusions
- Performing 3D modeling of objects and surfaces

Matching Functions

Similarity Measure	Formula
Sum of Absolute Differences (SAD)	$\sum_{(i,j)\in W} \lvert I_1(i,j) - I_2(x+i, y+j)\rvert$
Sum of Squared Differences (SSD)	$\sum_{(i,j)\in W} \left(I_1(i,j) - I_2(x+i, y+j)\right)^2$
Zero-mean SAD	$\sum_{(i,j)\in W} \lvert I_1(i,j) - \bar{I}_1(i,j) - I_2(x+i,y+j) + \bar{I}_2(x+i,y+j)\rvert$
Locally scaled SAD	$\sum_{(i,j)\in W} \lvert I_1(i,j) - \frac{\bar{I}_1(i,j)}{\bar{I}_2(x+i,y+j)} I_2(x+i,y+j)\rvert$
Normalized Cross Correlation (NCC)	$\frac{\sum_{(i,j)\in W} I_1(i,j).I_2(x+i,y+j)}{\sqrt[2]{\sum_{(i,j)\in W} I_1^2(i,j).\sum_{(i,j)\in W} I_2^2(x+i,y+j)}}$

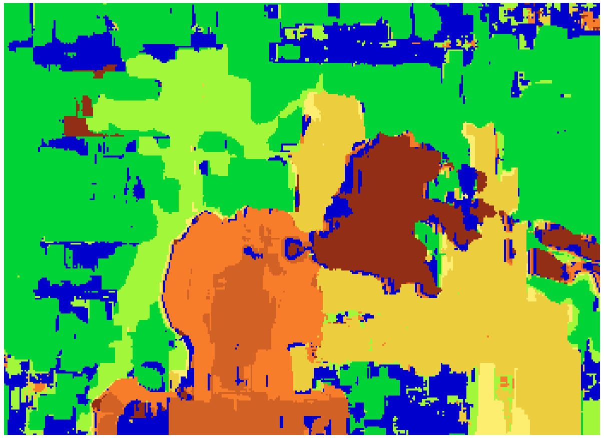
Matching Functions



SAD SSD NCC Ground truth

Kemelmacher, CSE 576, Washington University

Results with window search method

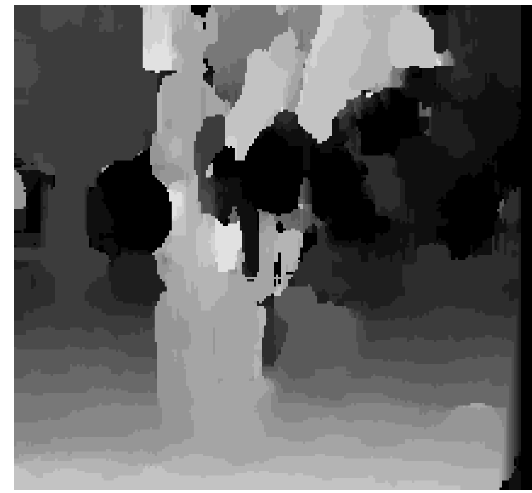


Window-based matching

Ground truth

Effect of window size

W = 3



W = 20

- Smaller window
  - More detail
  - More noise

- Larger window
  - Smoother disparity maps
  - Less detail

Better methods exist…

Graph cuts method

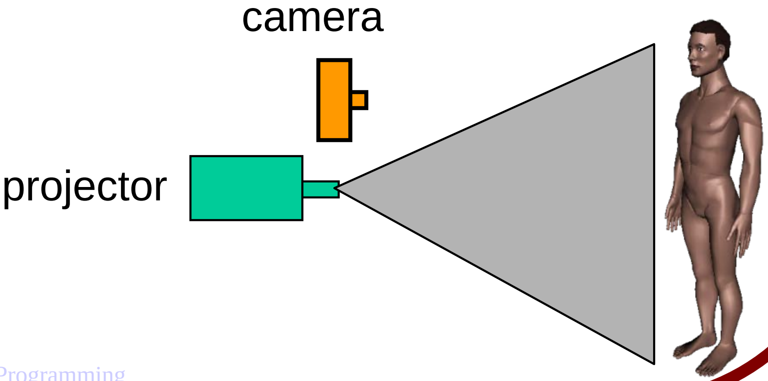
Ground truth

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

What has been done to get accurate depth measurements?

Active stereo with structured light

- Project "structured" light patterns onto the object
  - Simplifies the correspondence problem
  - Allows us to use only one camera



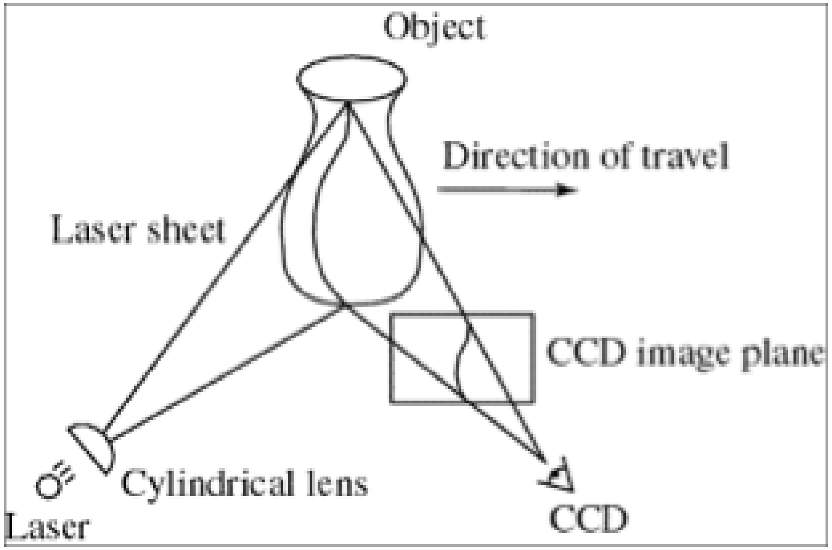
Rapid Shape Acquisition Using Color Structured Light and Multi-pass Dynamic Programming

Kinect: Structured infrared light

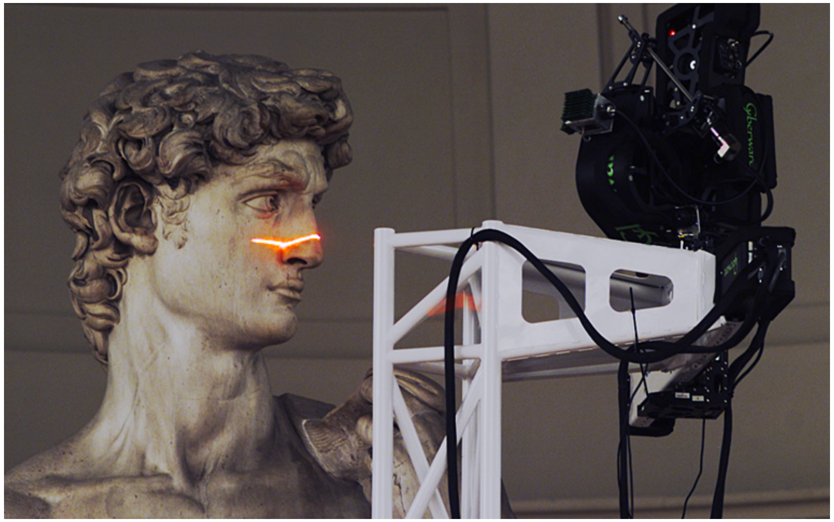
http://bbzippo.wordpress.com/2010/11/28/kinect-in-infrared/

Laser scanned models

- Optical triangulation
  - Project a single stripe of laser light
  - Scan it across the surface of the object
  - This is a very precise version of structured light scanning



Kemelmacher, CSE 576, Washington University



Digital Michelangelo Project
http://graphics.stanford.edu/projects/mich/

Laser scanned models

The Digital Michelangelo Project, Levoy et al.

Kemelmacher, CSE 576, Washington University

Laser scanned models

The Digital Michelangelo Project, Levoy et al.

Kemelmacher, CSE 576, Washington University