Introduction to Computer Vision

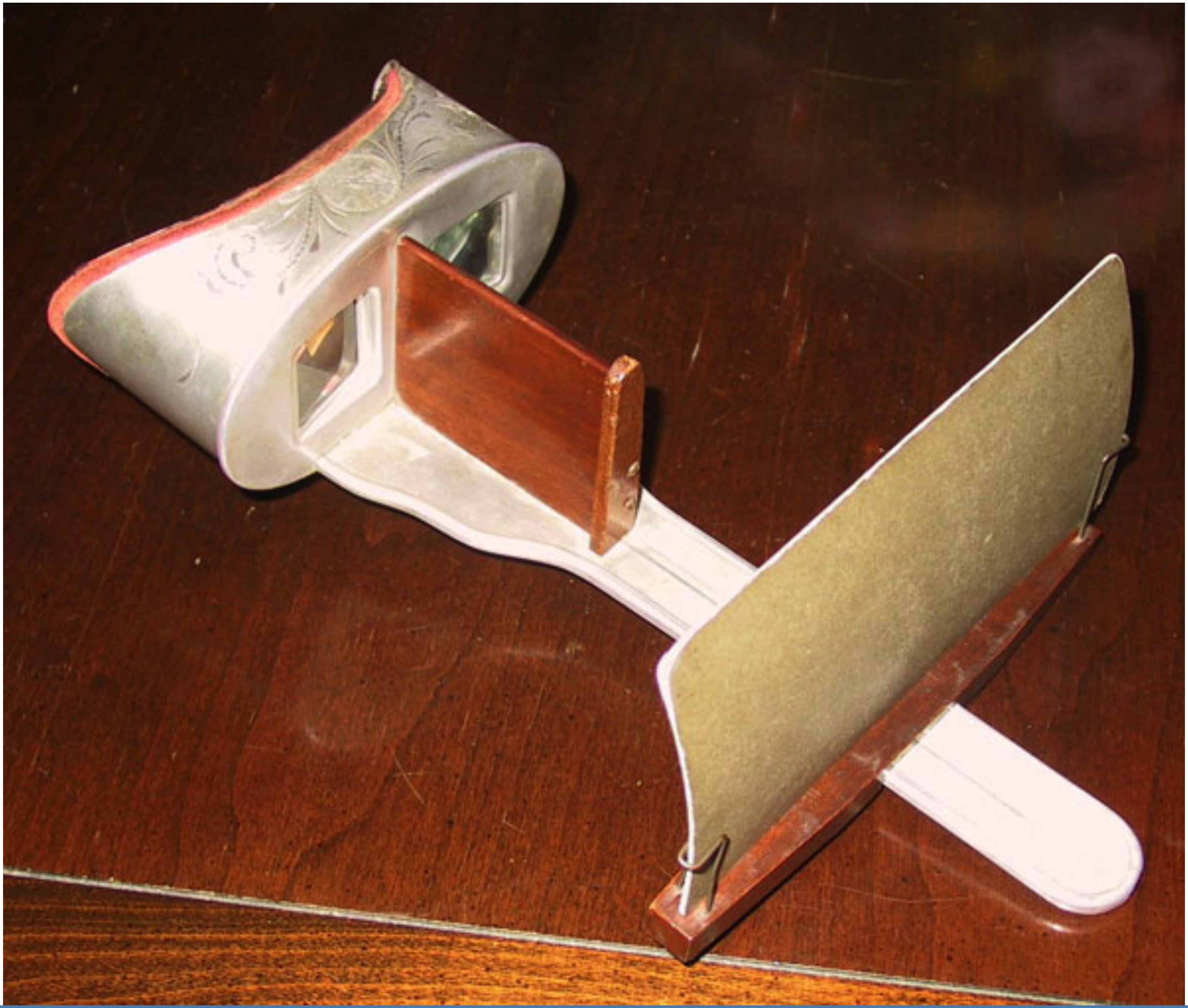
Michael J. Black

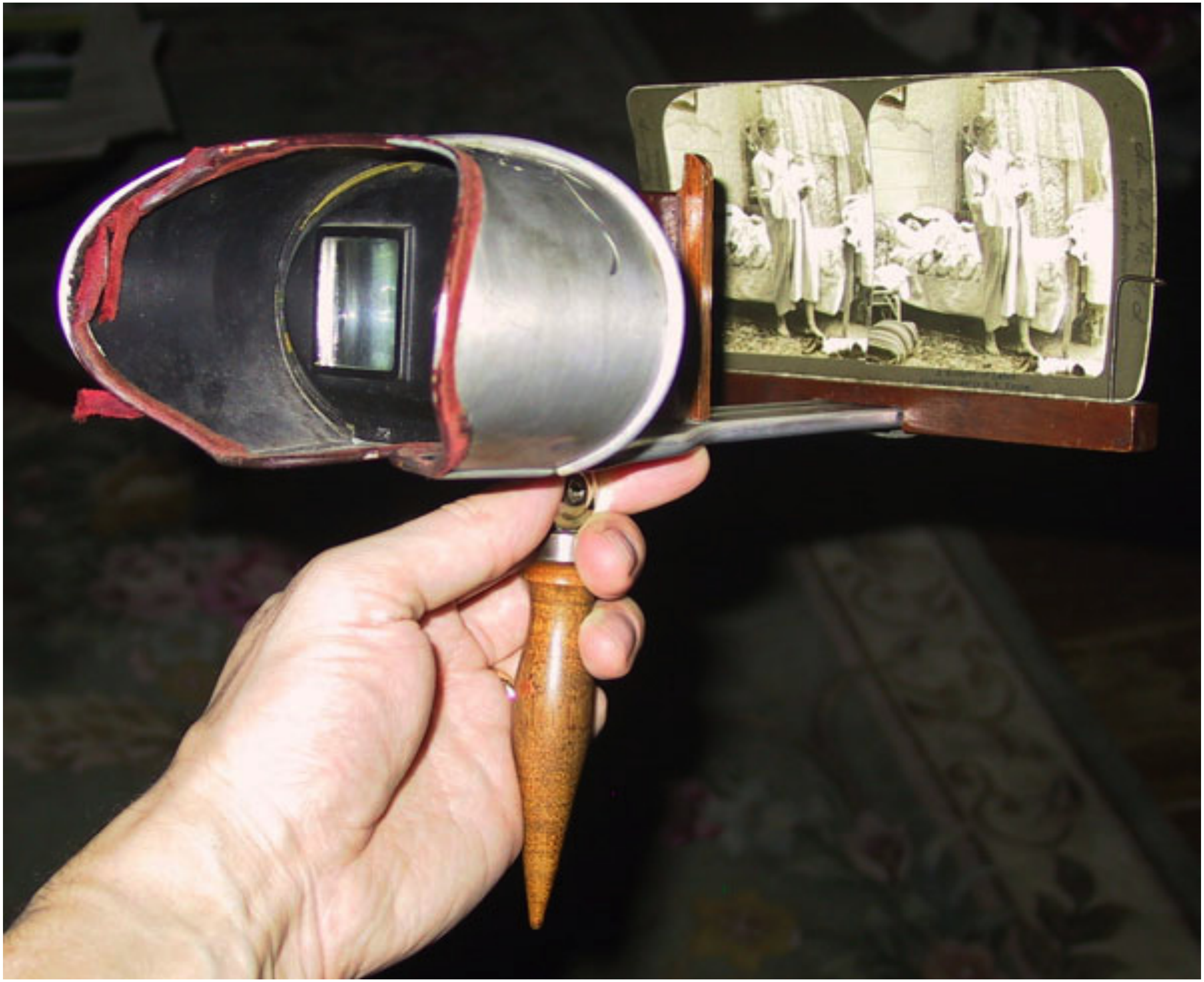
Nov 2009

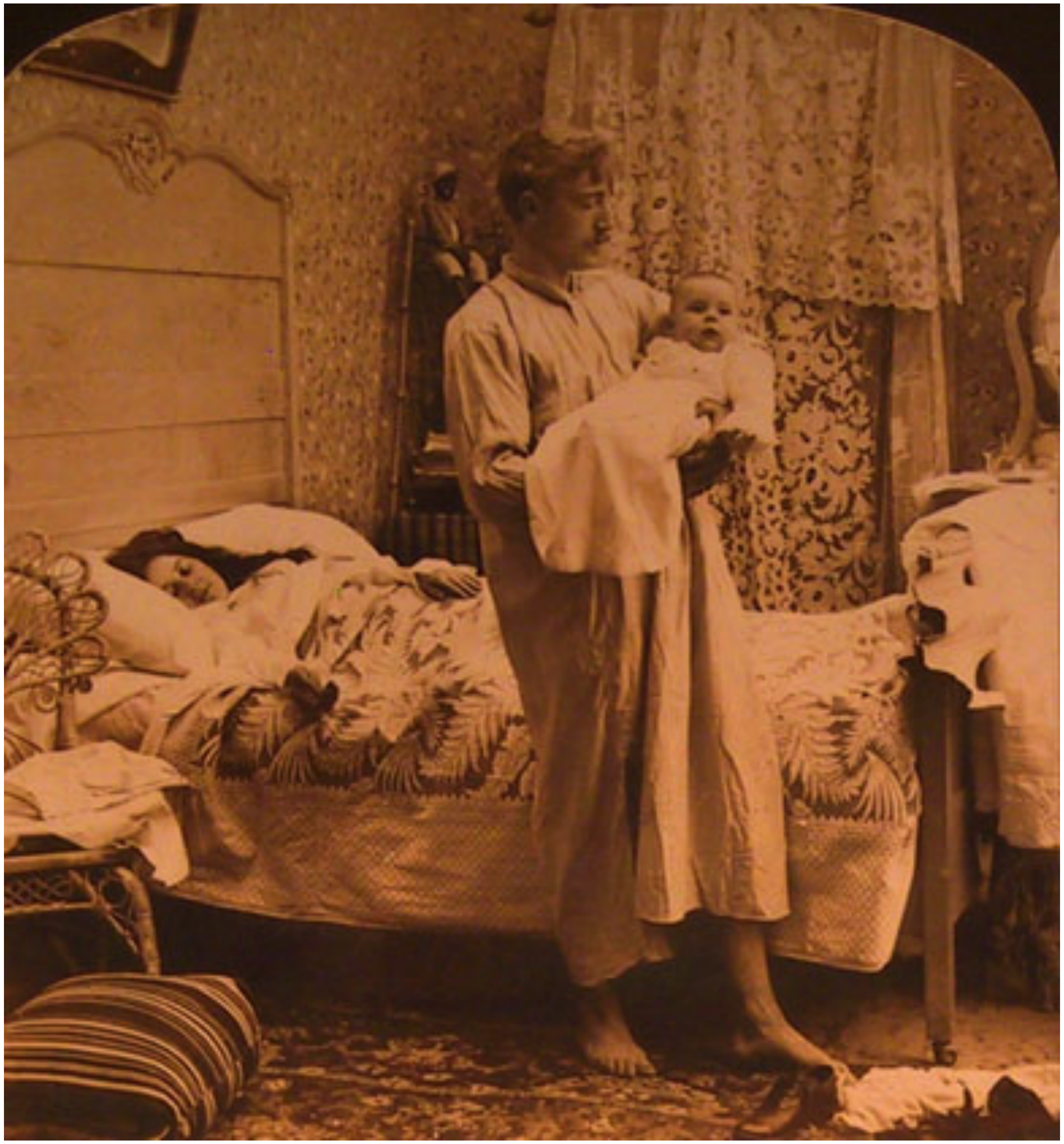
Stereo

Goals

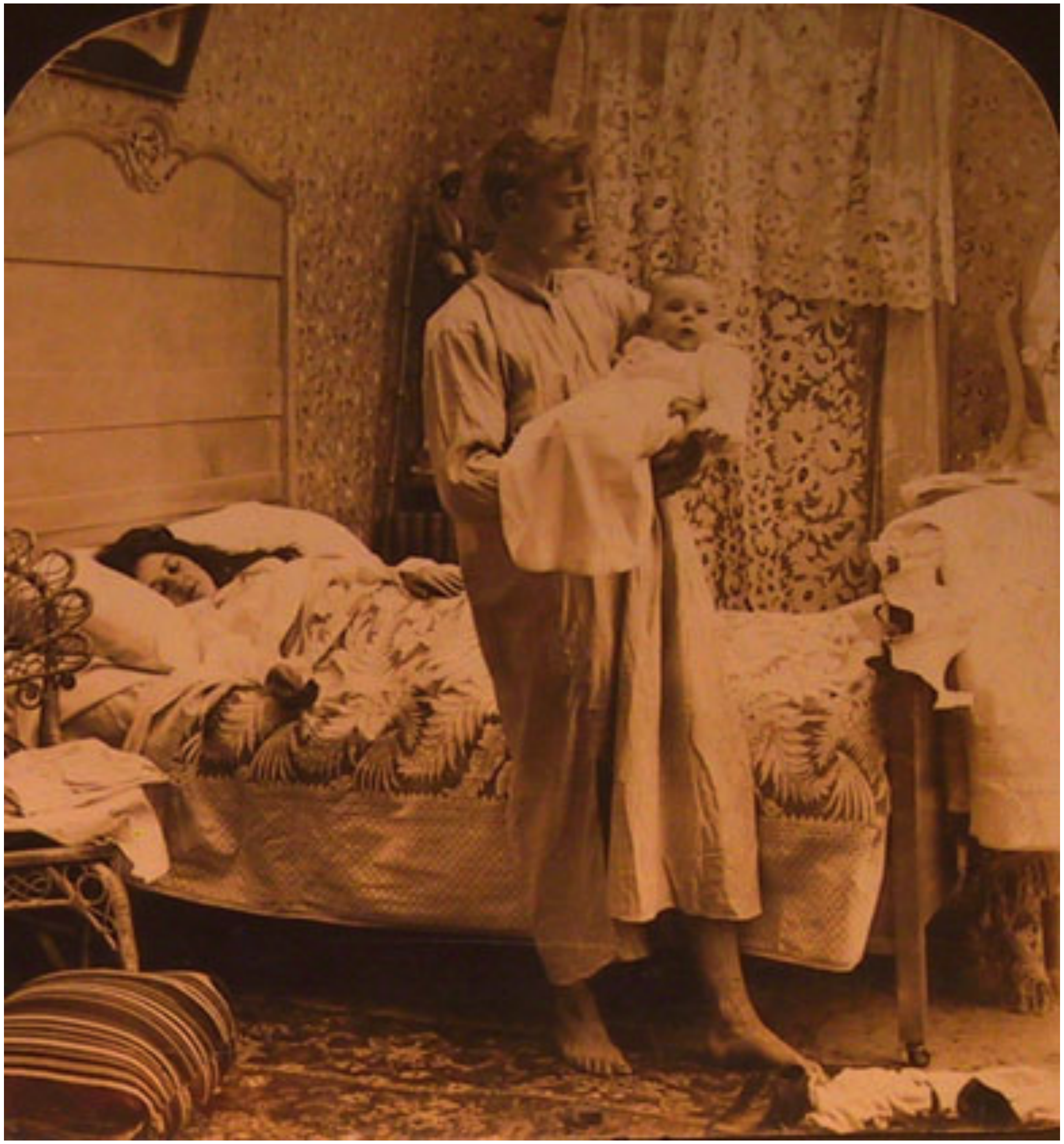
- Today
  - Binocular stereo
- Friday
  - Either object recognition or human shape and pose.





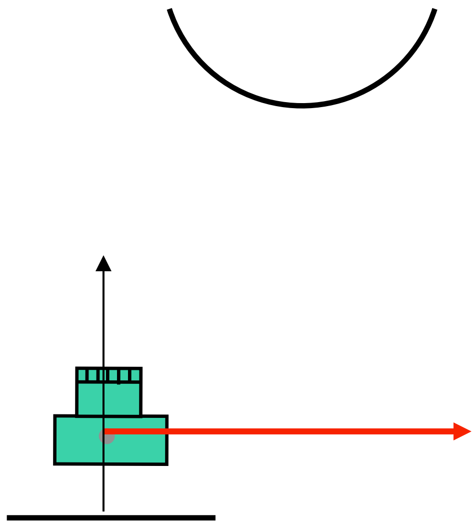


Left image



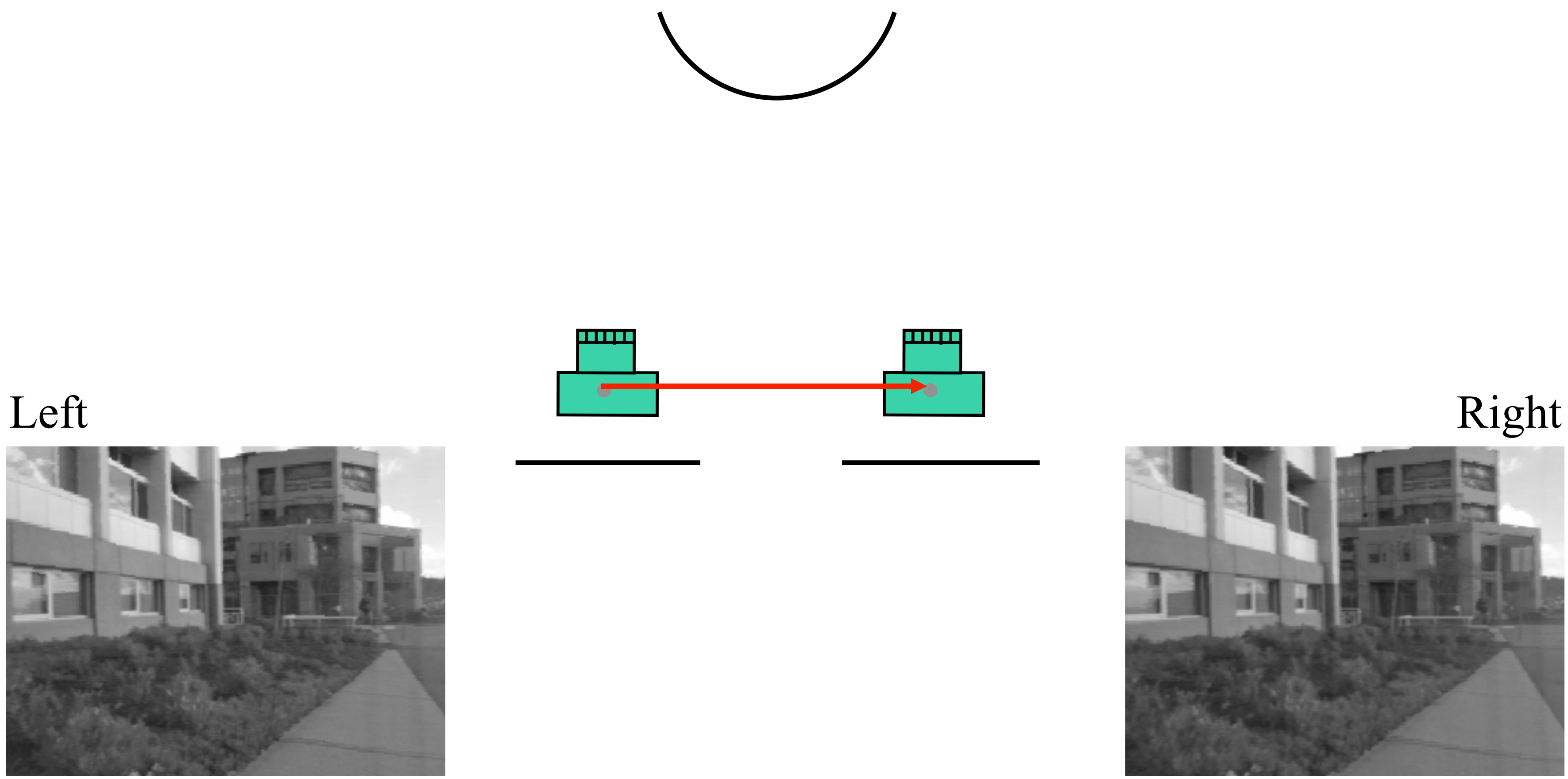
Right image

Binocular Stereo

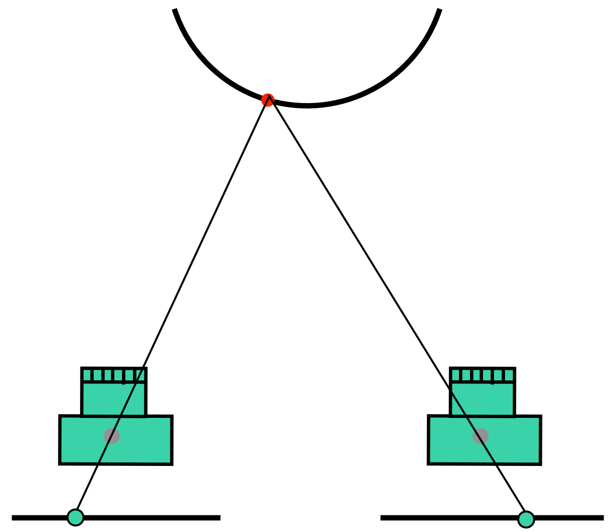


Left

Binocular Stereo



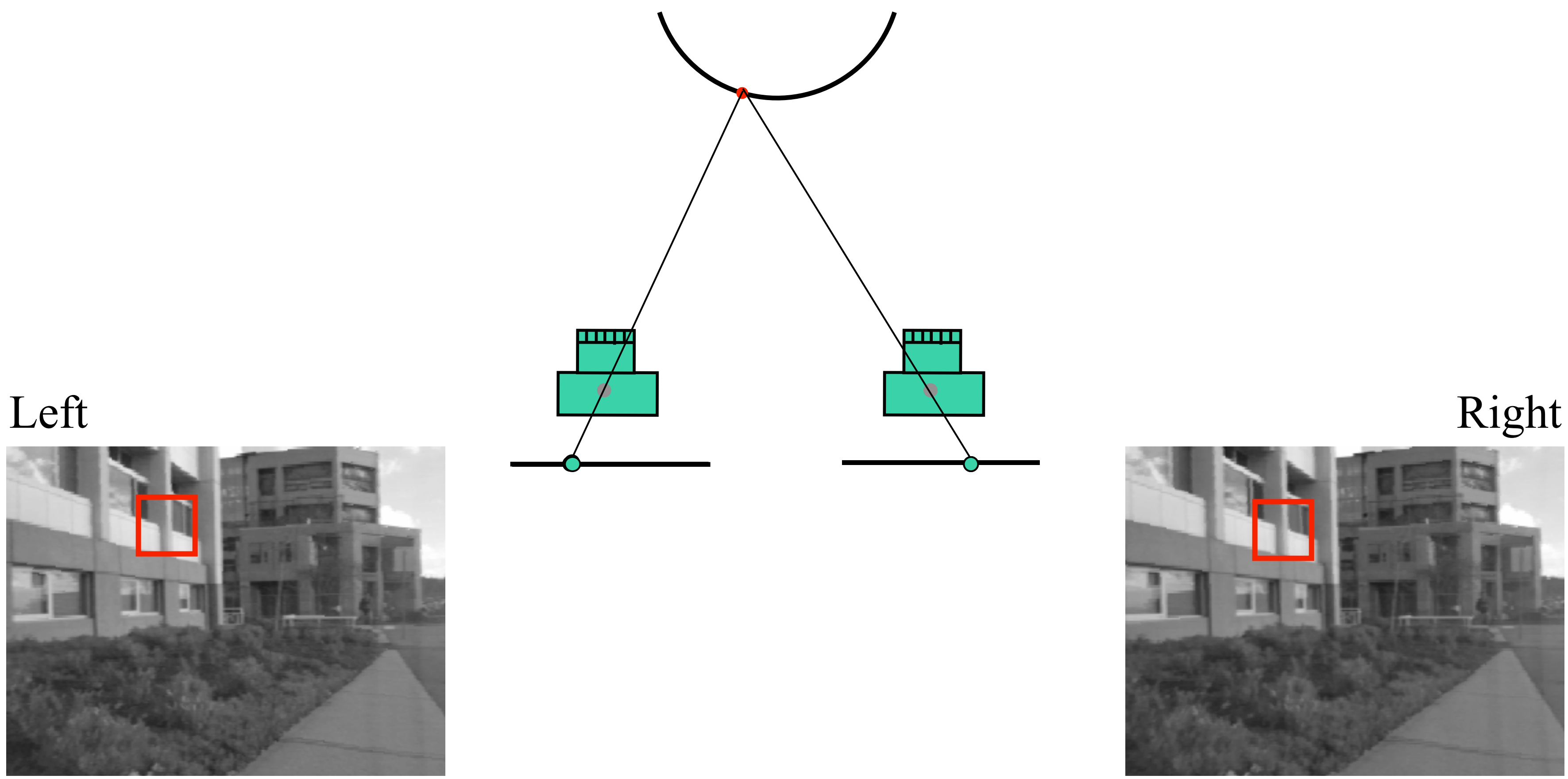
Binocular Stereo



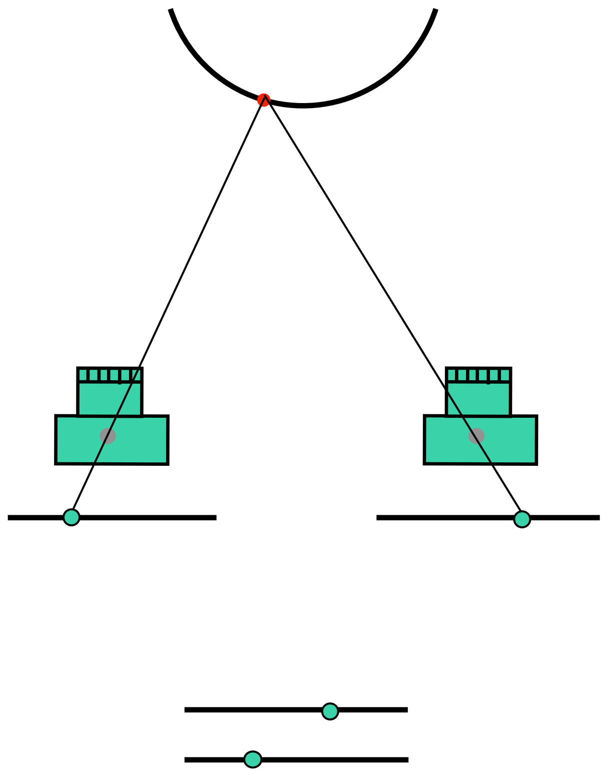
Left

Right

Binocular Stereo



Binocular Stereo

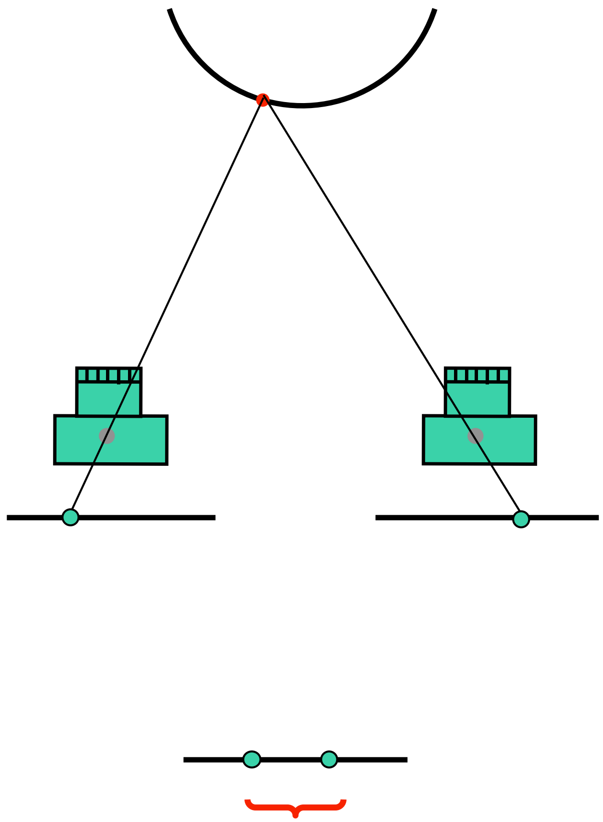


Left

Right

Binocular Stereo

Left



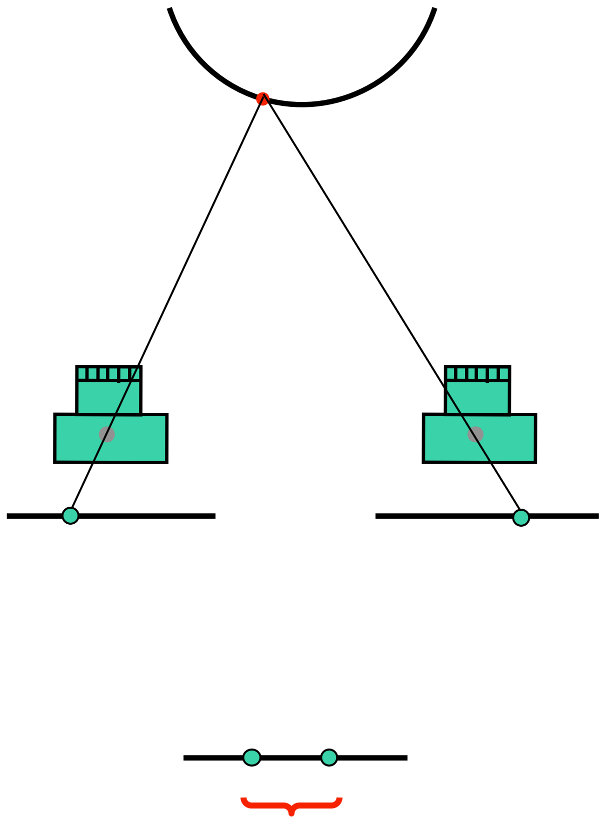
binocular disparity

Right

Binocular Stereo

From known geometry of the cameras and estimated disparity, recover depth in the scene

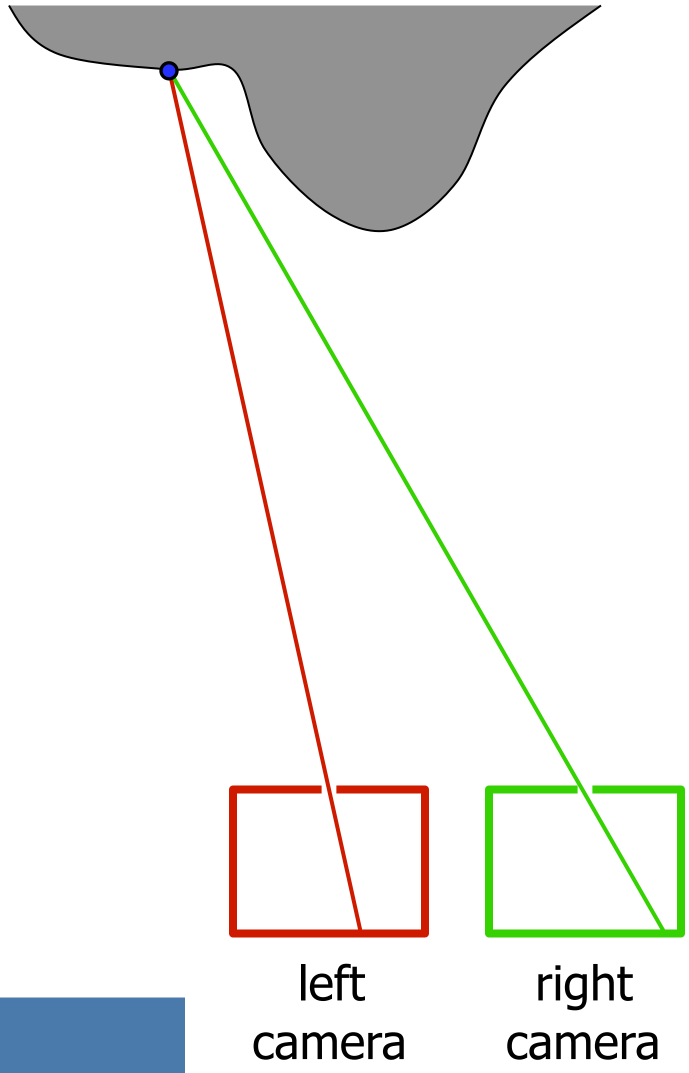
Left



binocular disparity

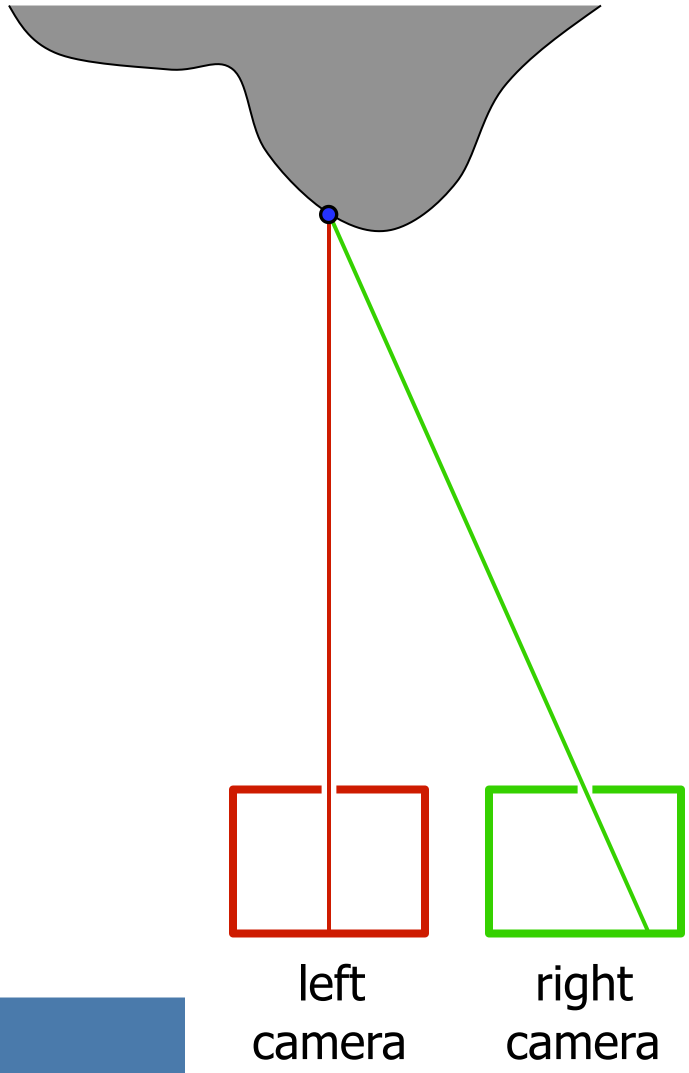
Right

Stereo Geometry

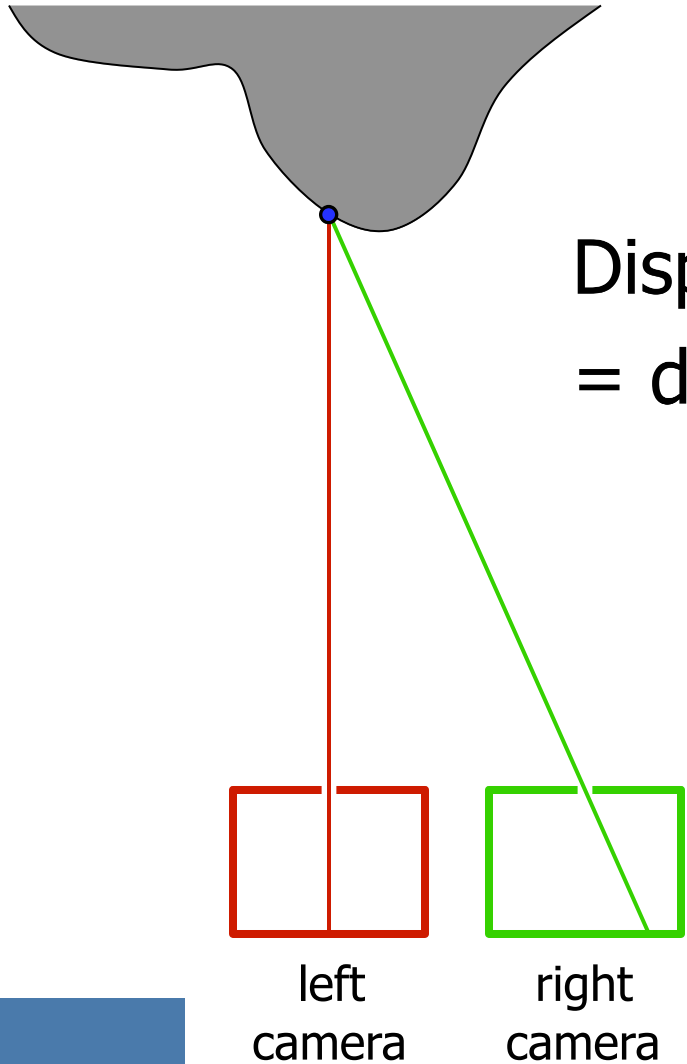


Scharstein

Stereo Geometry

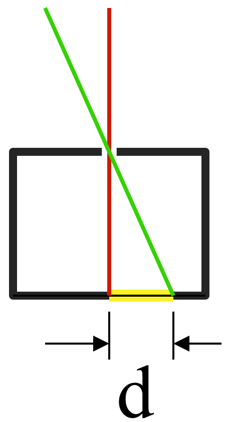


Stereo Geometry

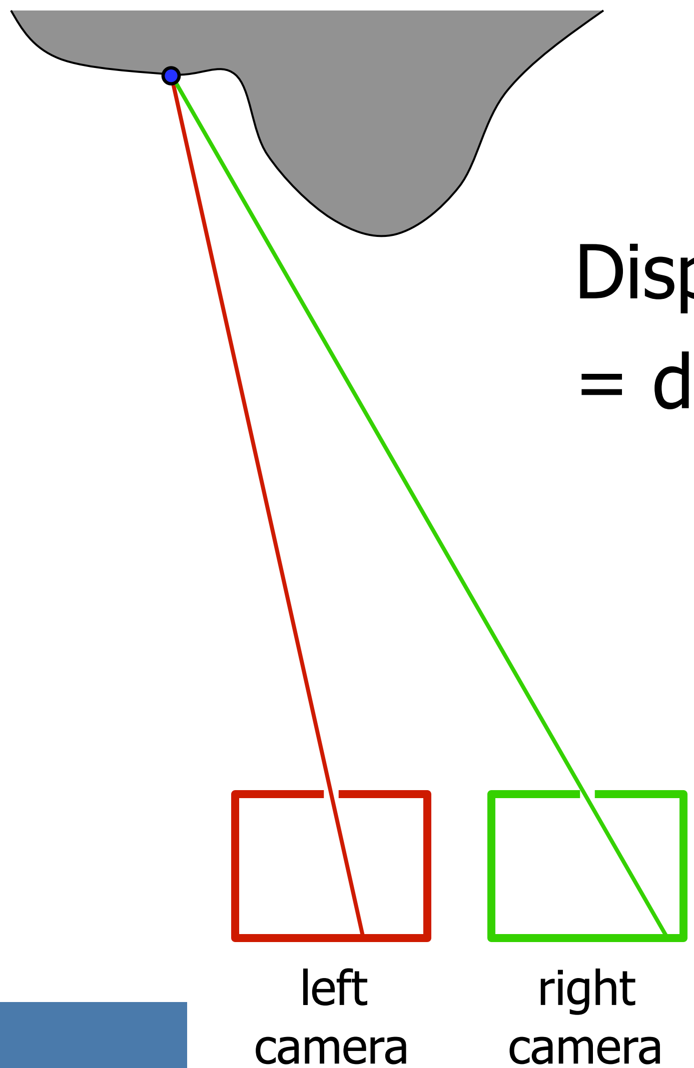


Disparity d

= difference in image position

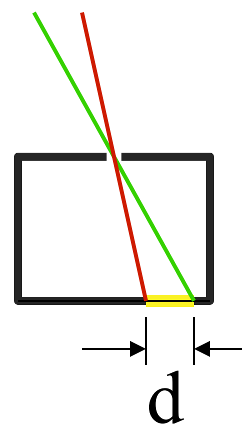


Stereo Geometry

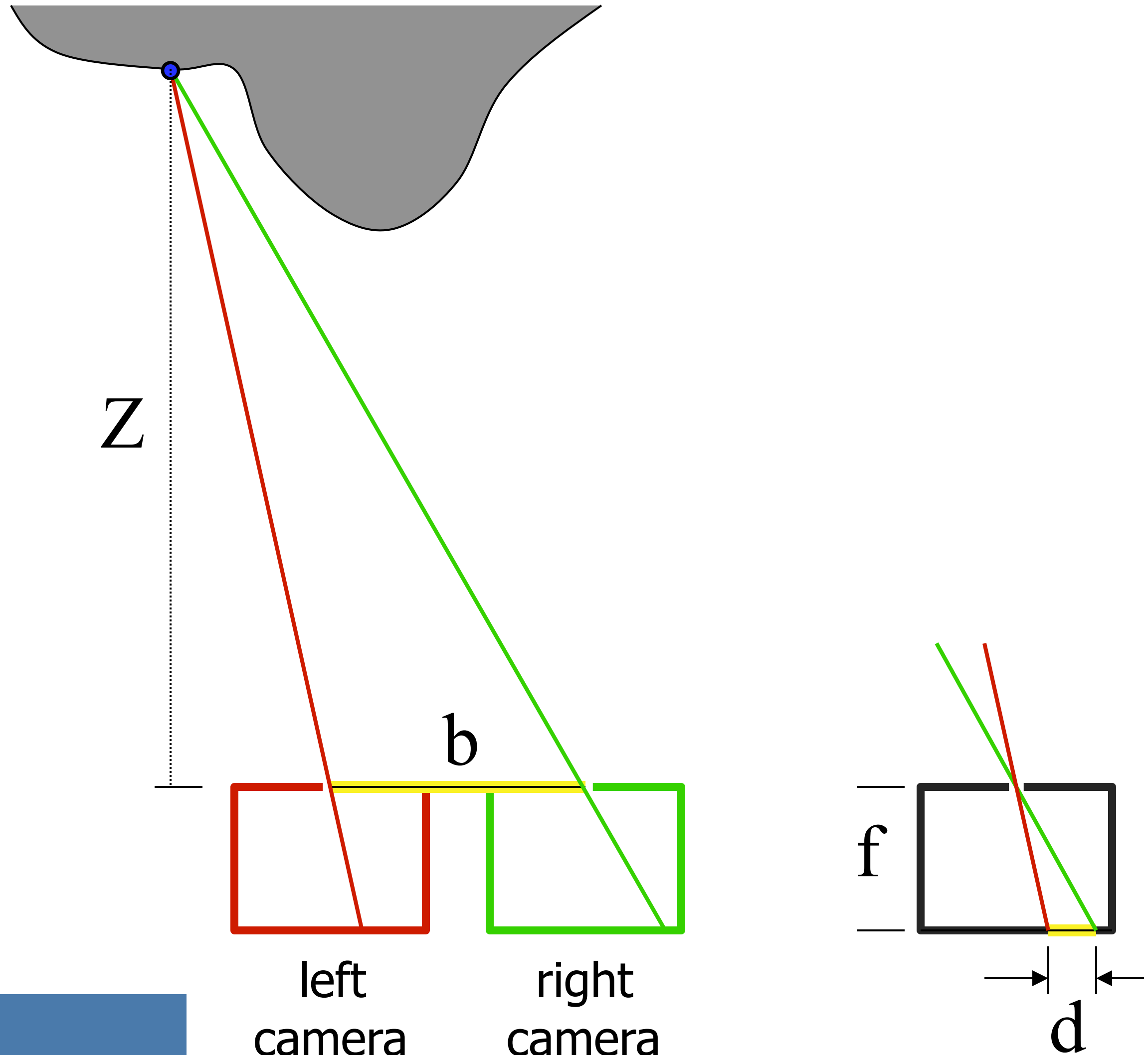


Disparity d

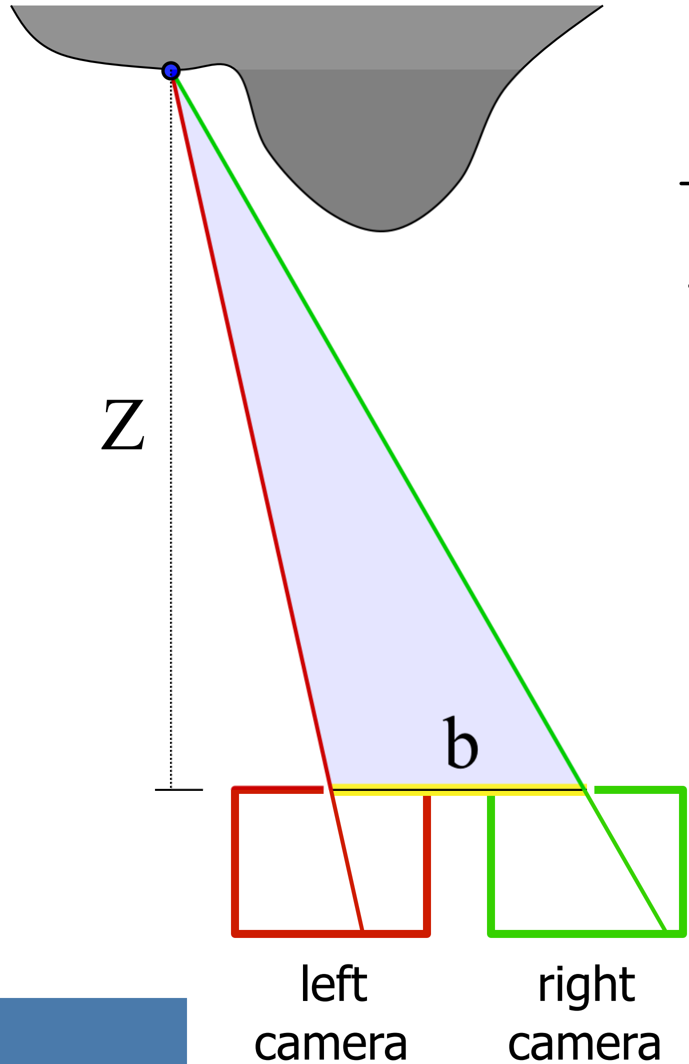
= difference in image position



Stereo Geometry

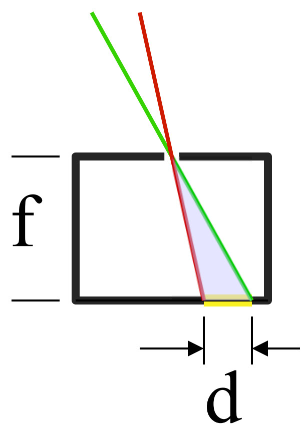


Stereo Geometry



$$\frac{d}{f} = \frac{b}{Z}$$

Disparity $d = bf \frac{1}{Z}$

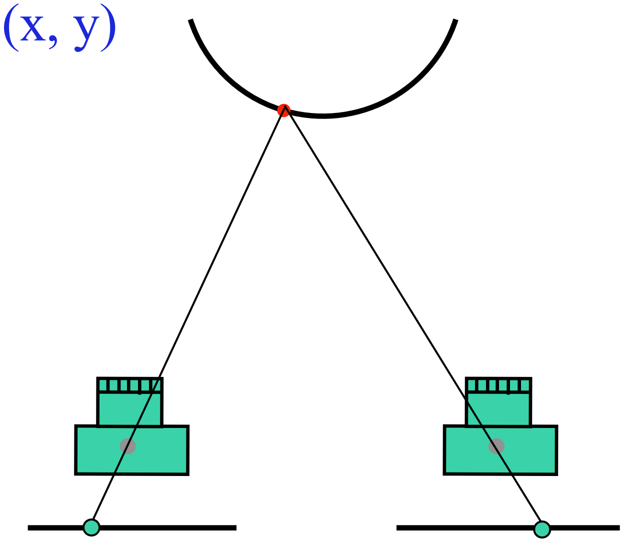


Binocular Disparity

Z(x, y) is depth at pixel (x, y)
d(x, y) is disparity

Estimate:

$$Z(x,y) = \frac{fb}{d(x,y)}$$



Left

Search for best match

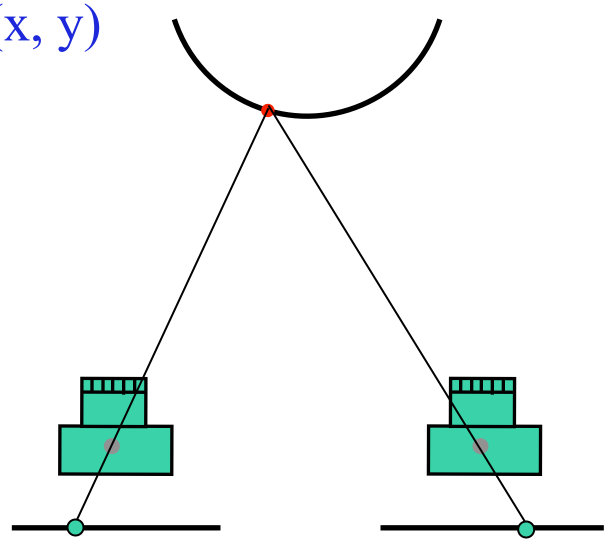
Right

Binocular Disparity

Z(x, y) is depth at pixel (x, y)
d(x, y) is disparity

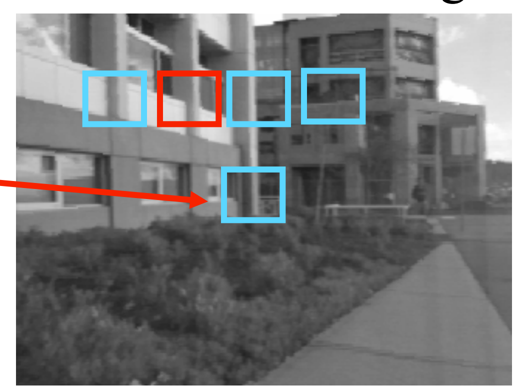
Estimate:

$$Z(x,y) = \frac{fb}{d(x,y)}$$



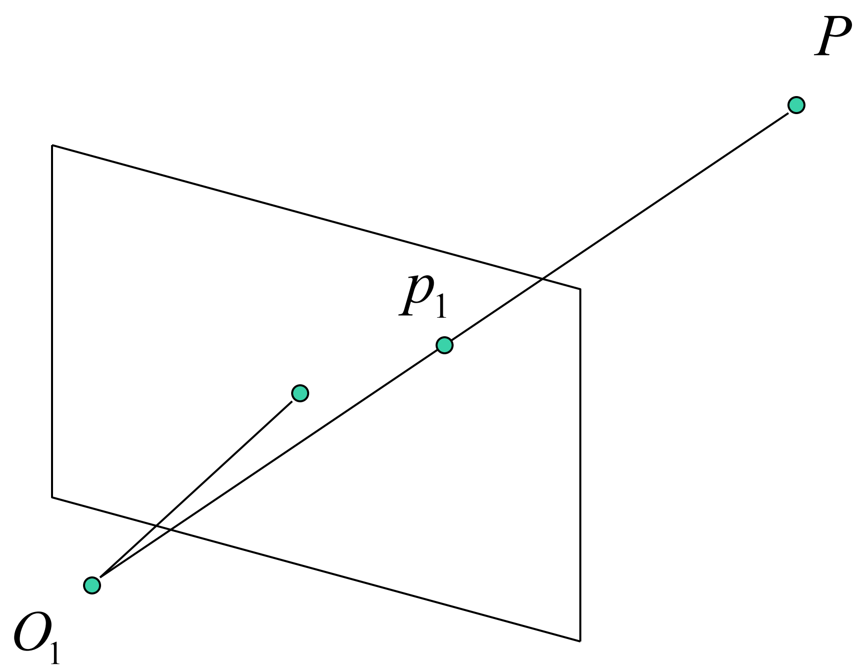
Left

Right

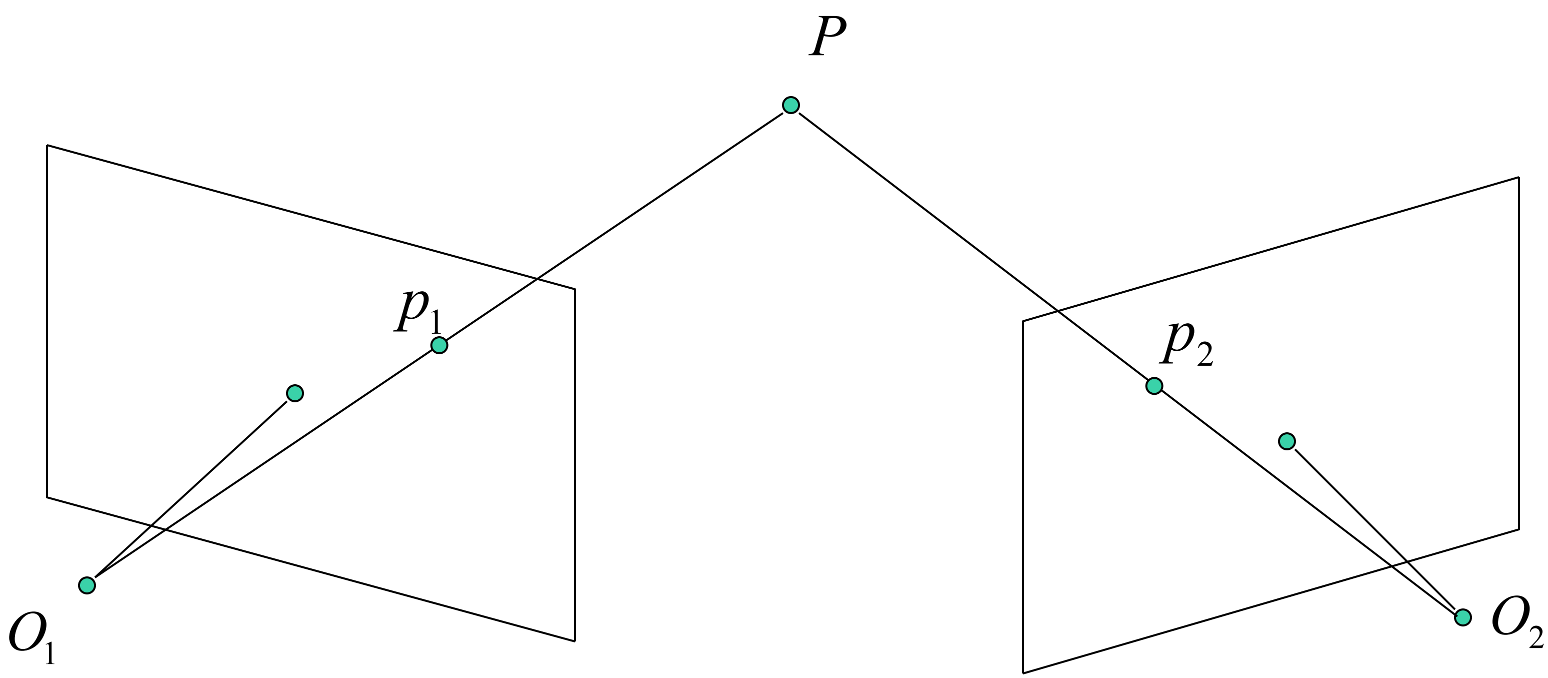


Do I need to consider this region?

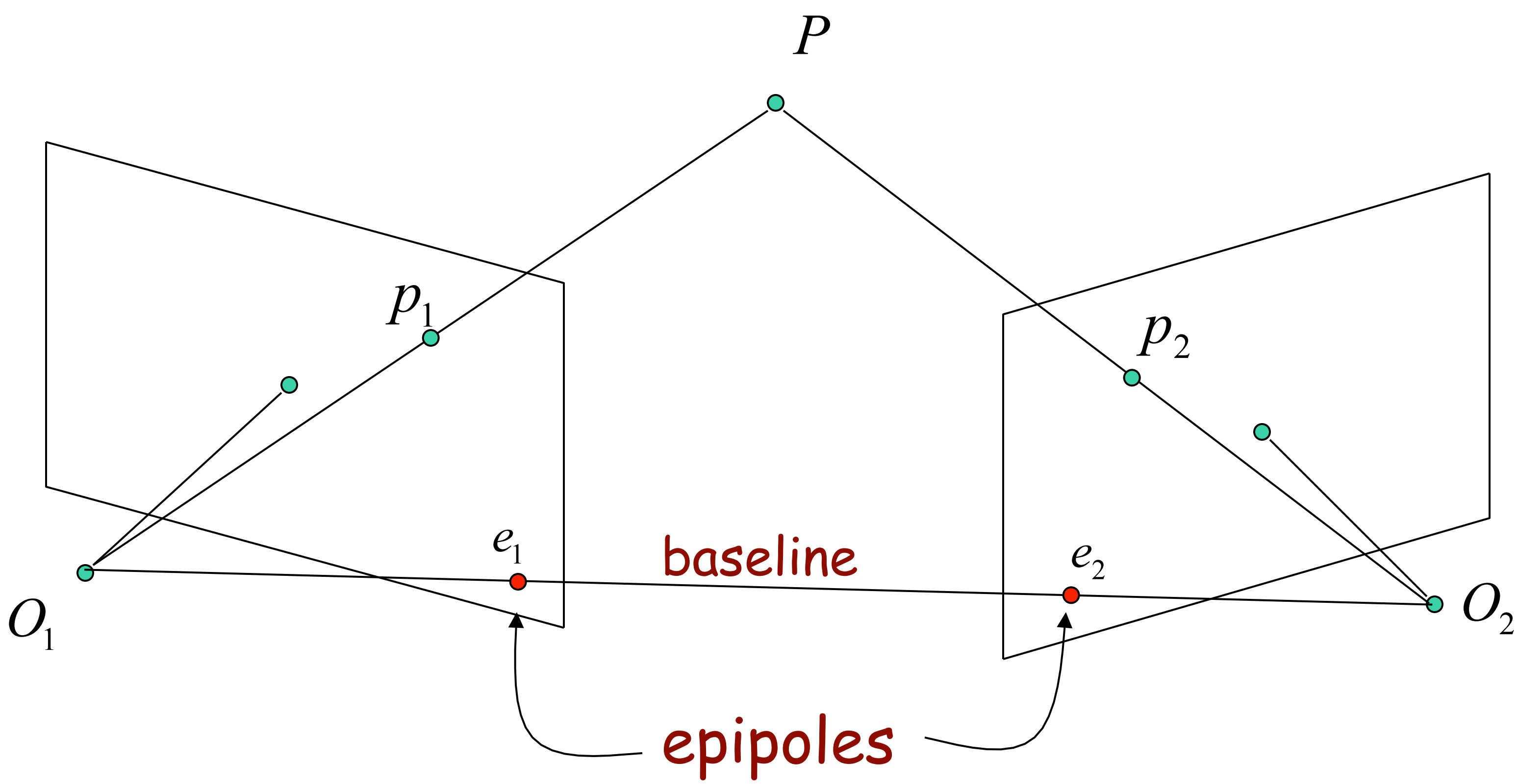
Epipolar Geometry



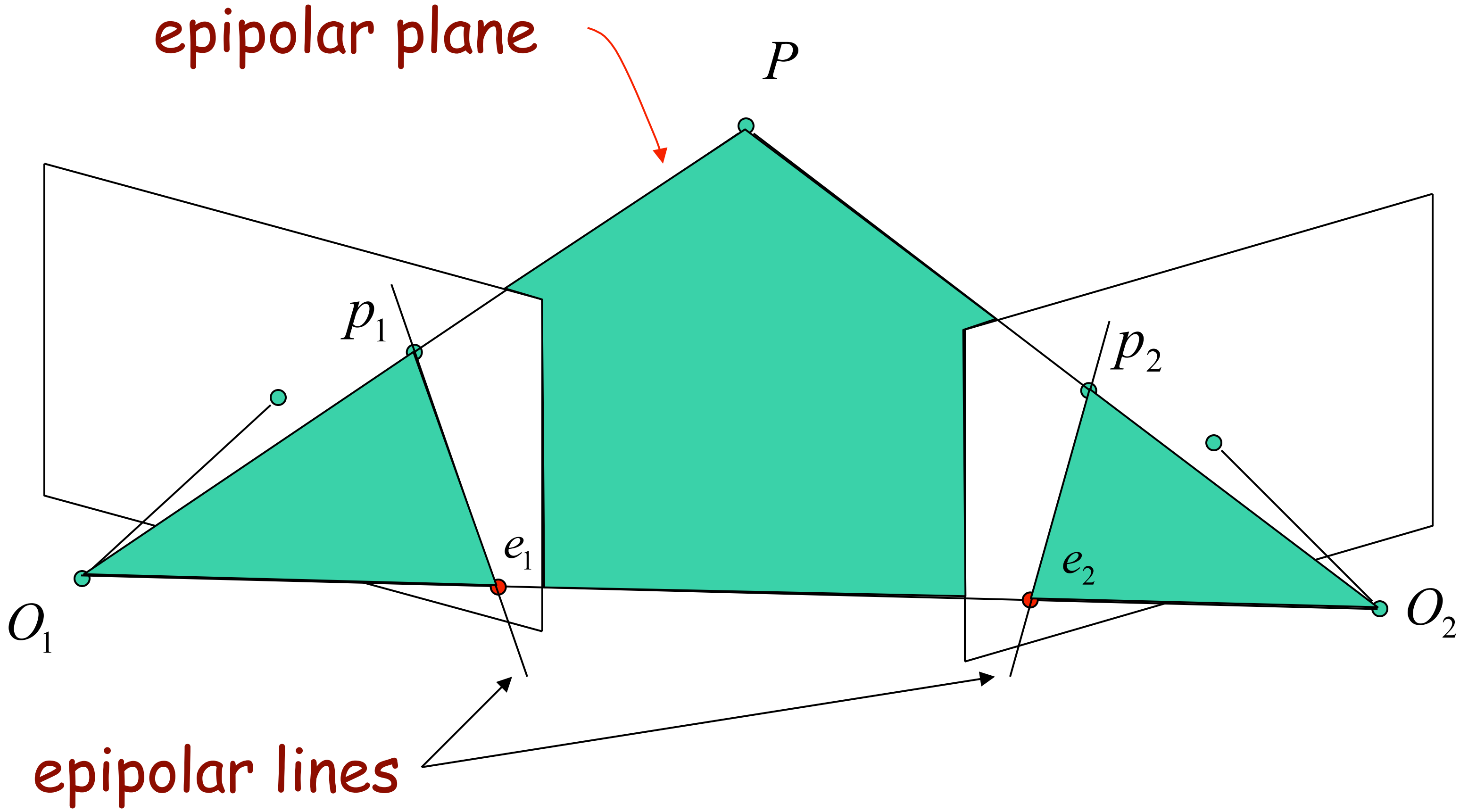
Epipolar Geometry



Epipolar Geometry

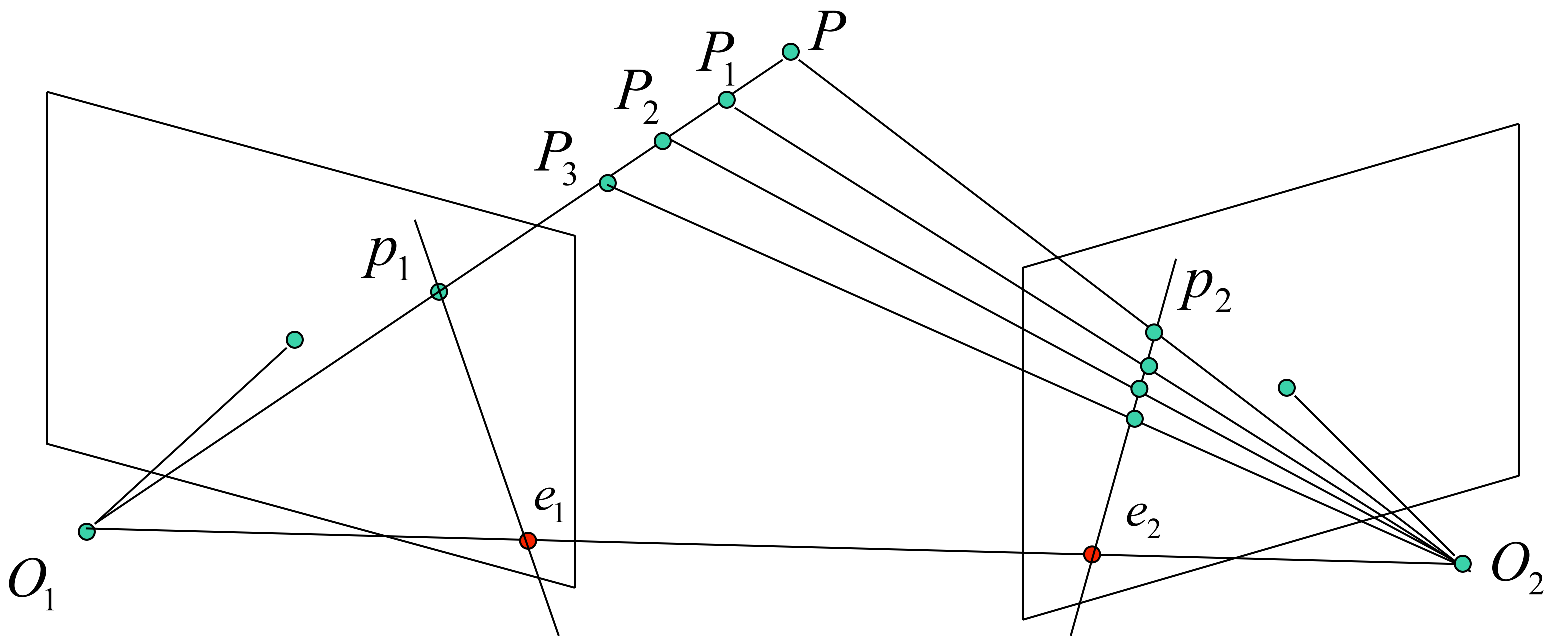


Epipolar Geometry



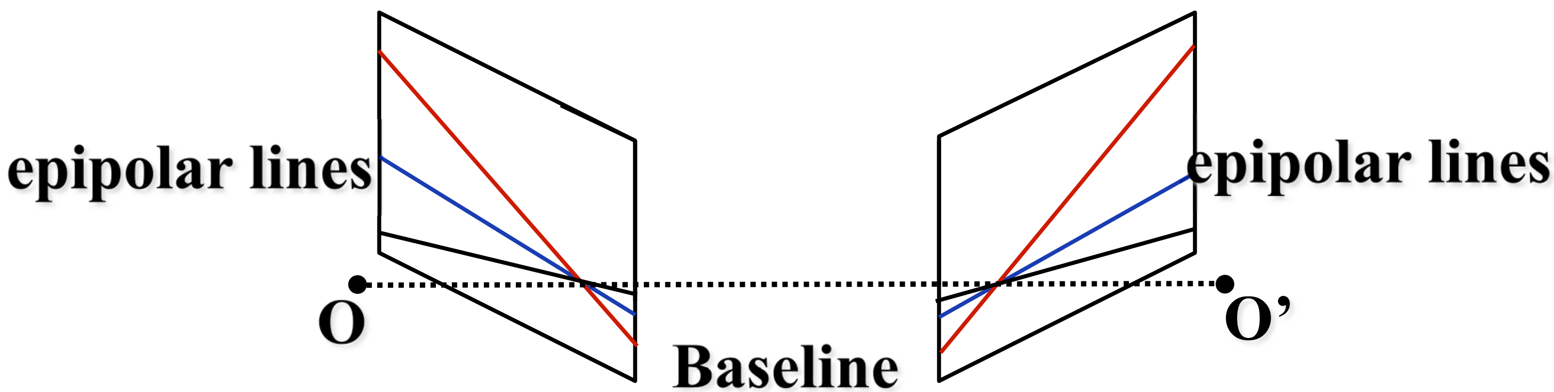
Epipolar Geometry

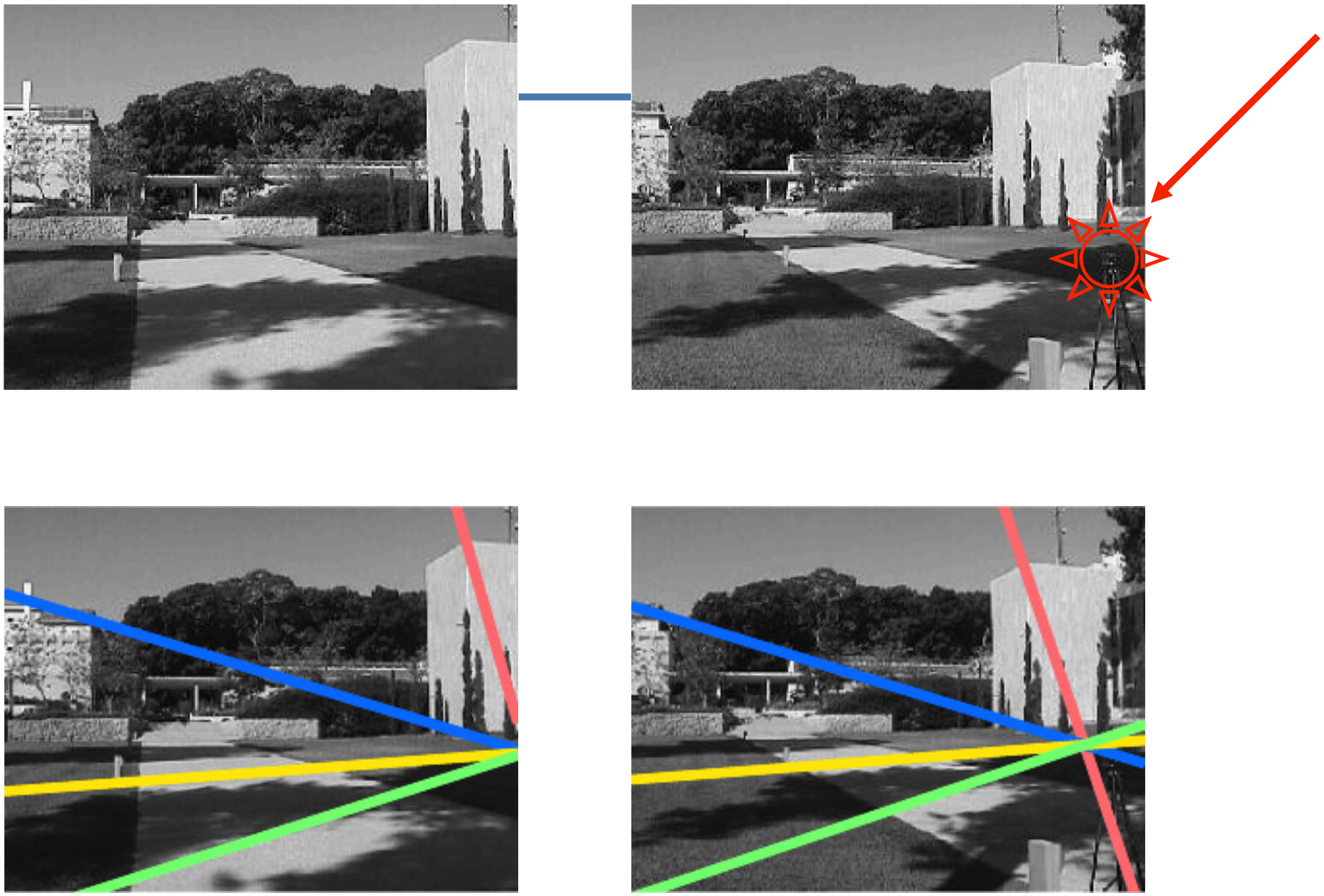
Possible matches for p_1 are constrained to lie along the epipolar line in the other image



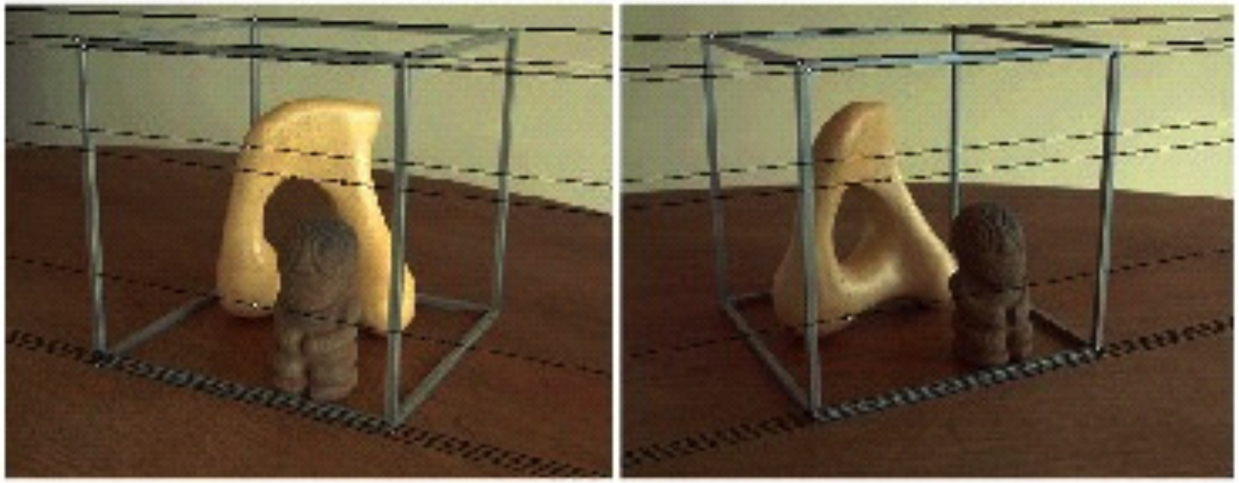
Epipole

- Every plane through the baseline is an epipolar plane, and determines a pair of epipolar lines in the two images
- Two systems of epipolar lines are obtained, each system intersects in a point, the *epipole*
- The epipole is the projection of the center of the other camera





Rectification

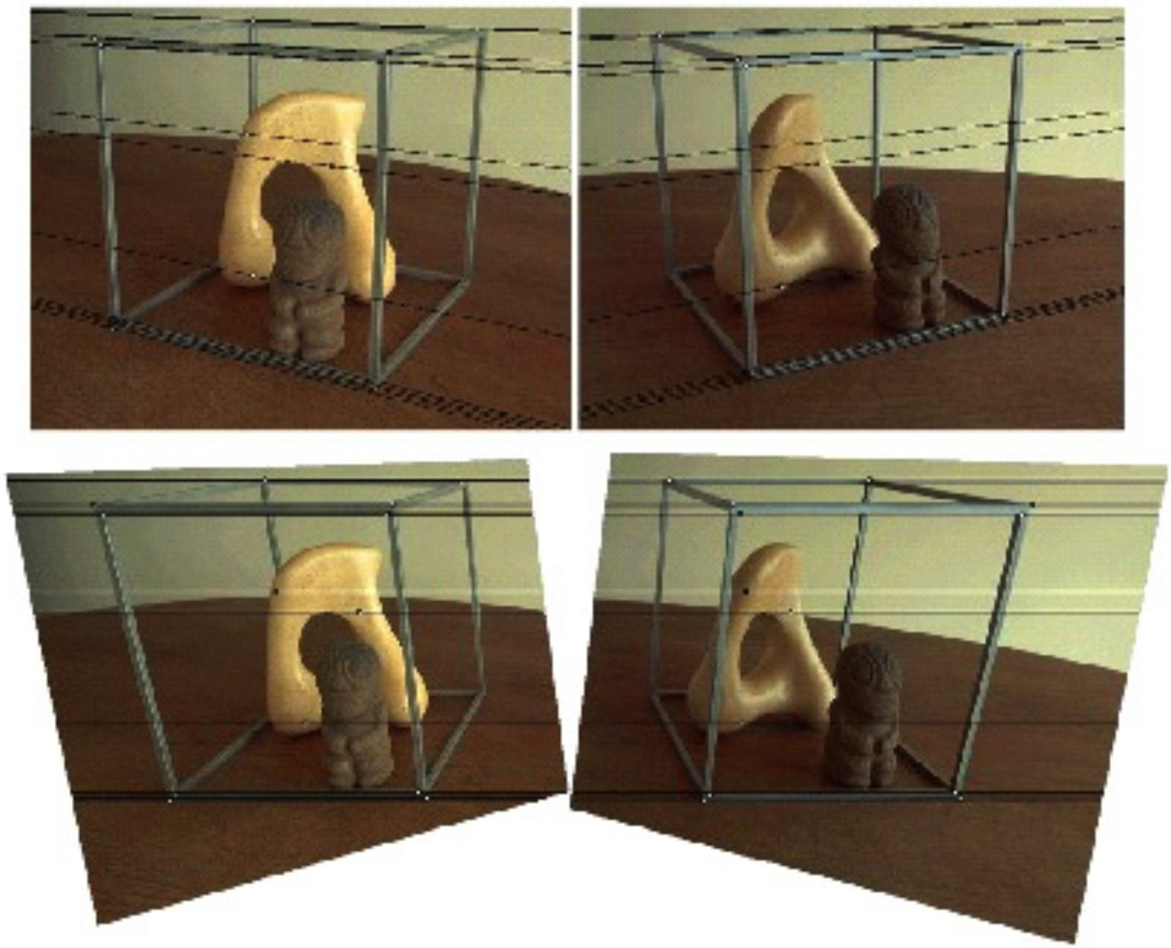


Rectification aligns epipolar lines with scanlines.

- warp images

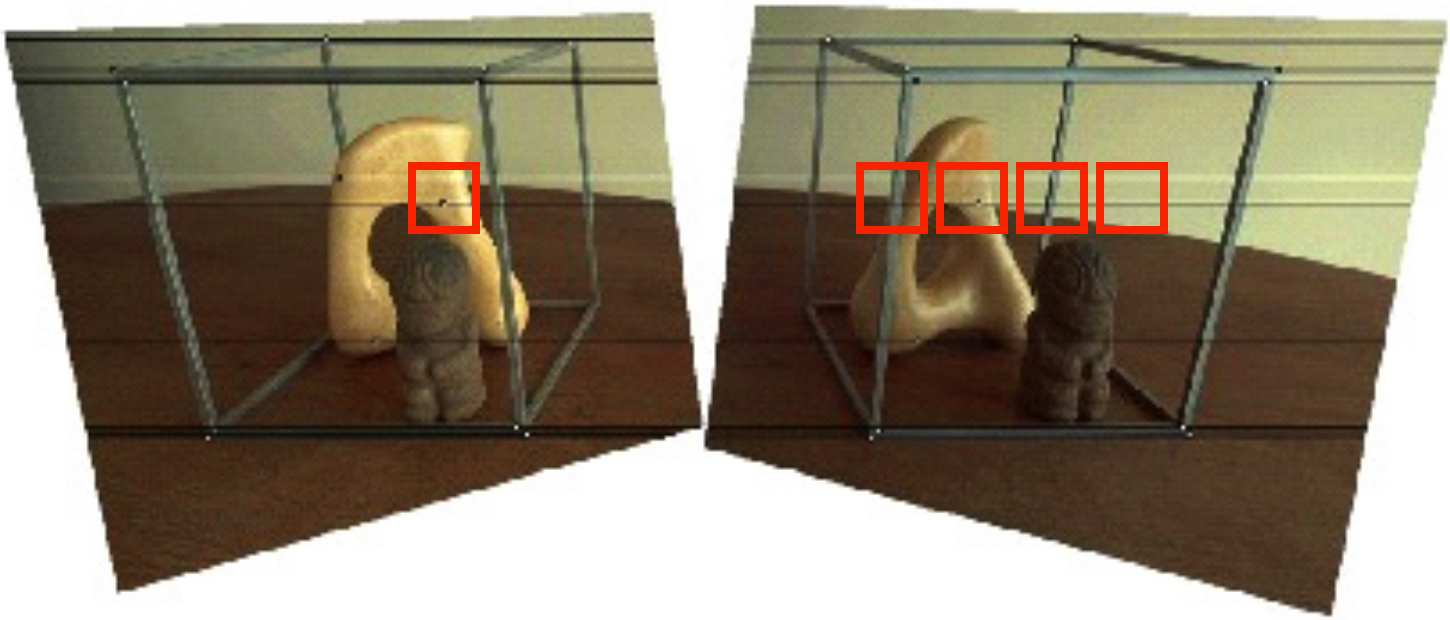
Szeliski and Fleet

Rectification



Szeliski and Fleet

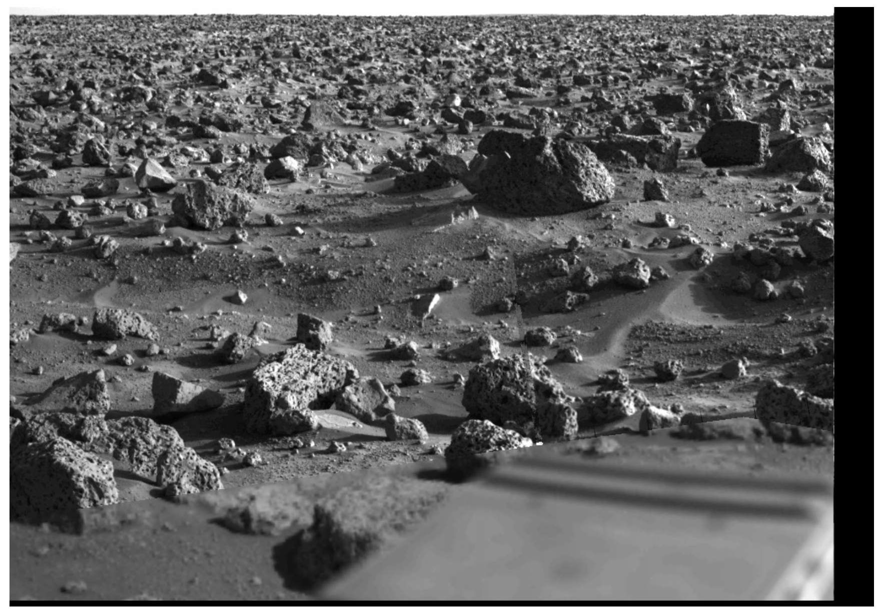
Matching



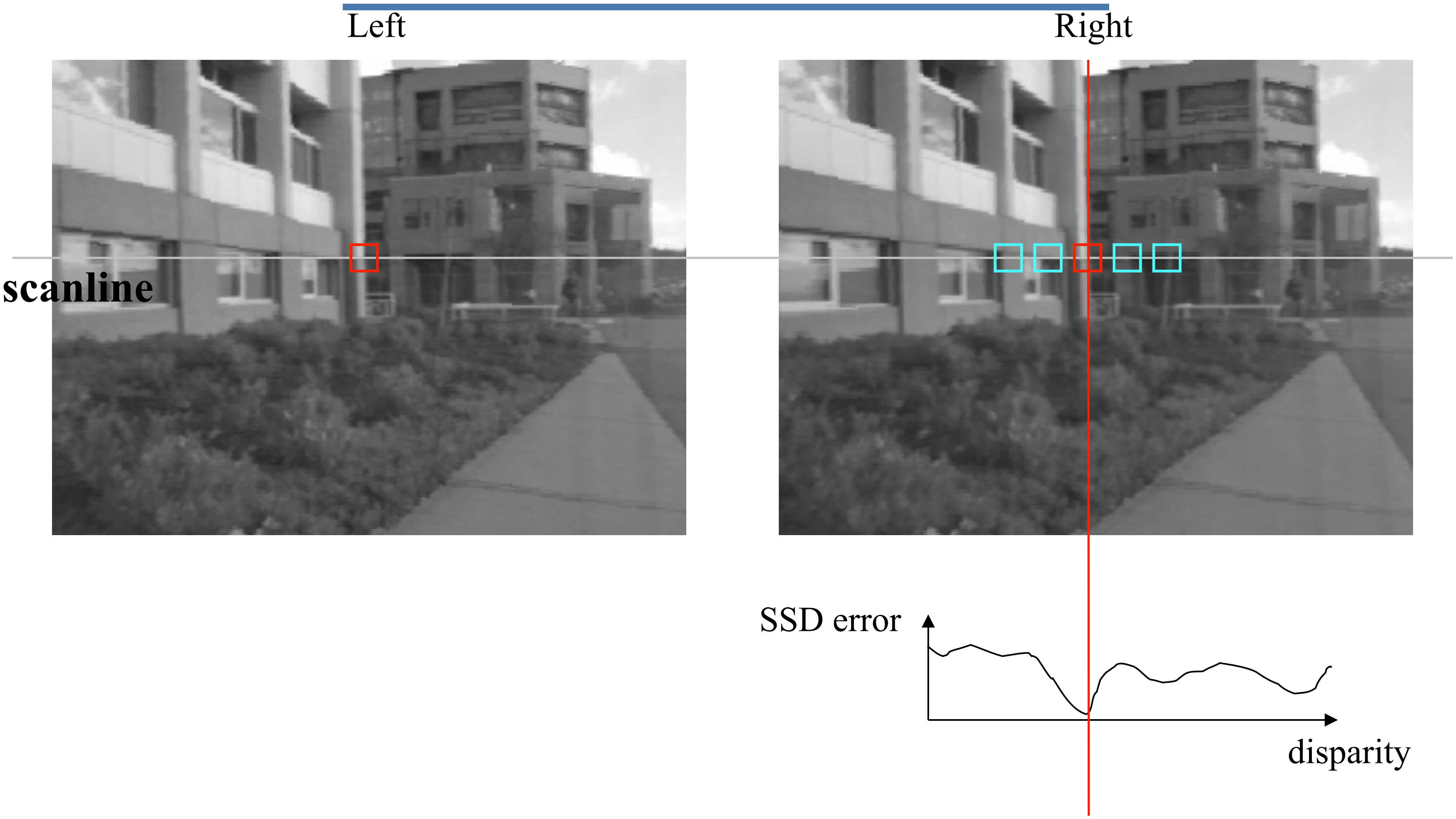
* Matching only has to occur along epipolar lines.
* Now in the simpler binocular case where the cameras are pointing forward.
* Compare with optical flow.

Stereo Correspondence

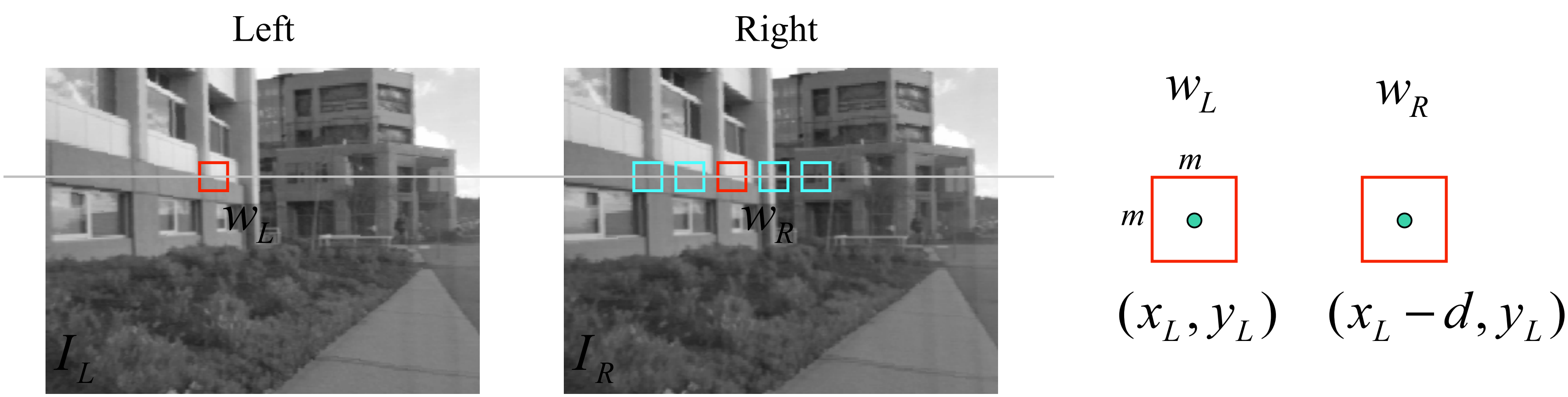
- Search over disparity to find correspondences
- Range of disparities to search over can change dramatically within a single image pair.



Correspondence Using SSD



Sum of Squared (Pixel) Differences



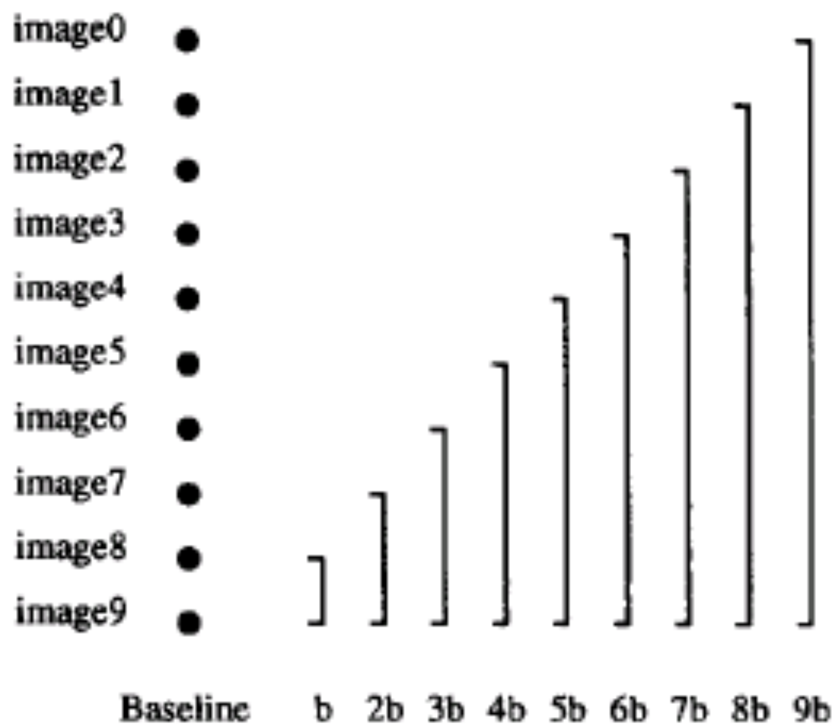
w_L and w_R are corresponding m by m windows of pixels.

The SSD cost measures the intensity difference as a function of disparity :

$$SSD_r(x, y, d) = \sum_{(x',y') \in W_m(x,y)} (I_L(x', y') - I_R(x'-d, y'))^2$$

Dealing with ambiguity

Many repeated structures



* Collect multiple views with different baselines.

M. Okutomi, T. Kanade, Multiple-Baseline Stereo

$$\frac{d}{f\,b} = \frac{1}{Z}$$

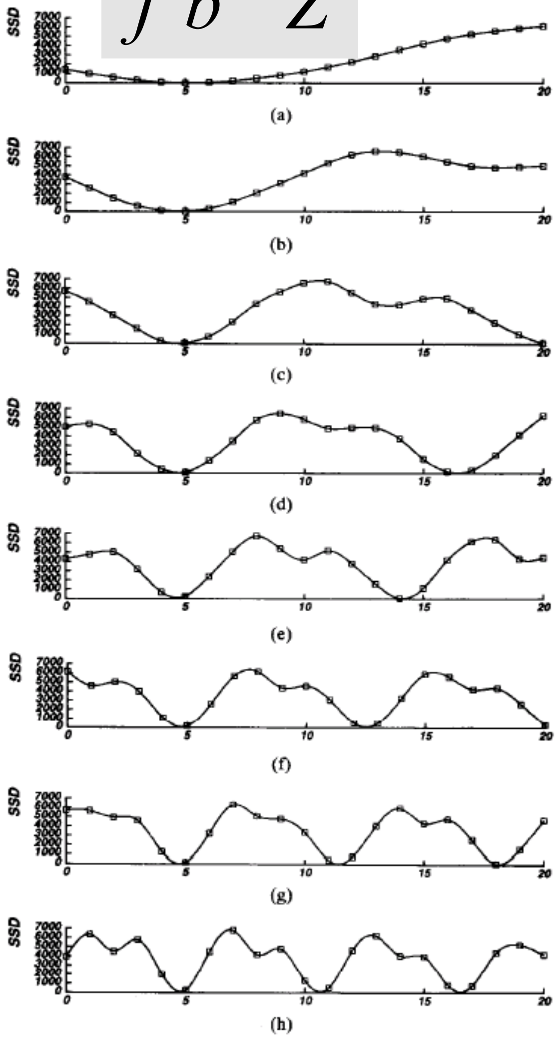


Fig. 5. SSD values versus inverse distance: (a) $B = b$; (b) $B = 2b$; (c) $B = 3b$; (d) $B = 4b$; (e) $B = 5b$; (f) $B = 6b$; (g) $B = 7b$; (h) $B = 8b$. The horizontal axis is normalized such that $8bF = 1$.

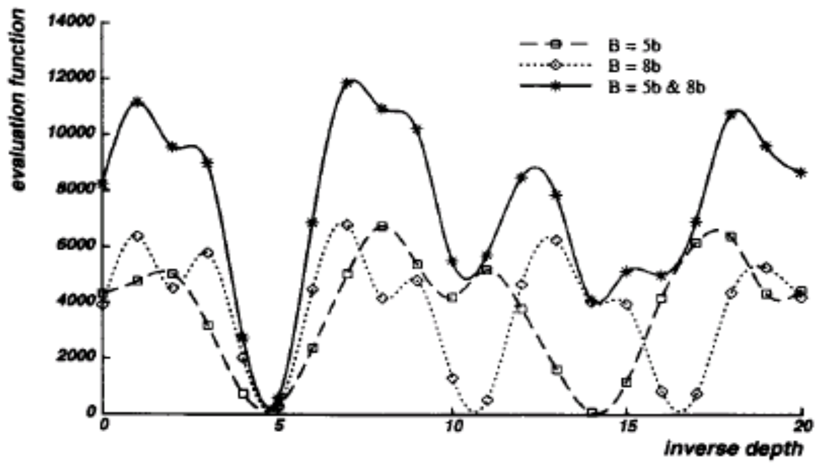


Fig. 6. Combining two stereo pairs with different baselines.

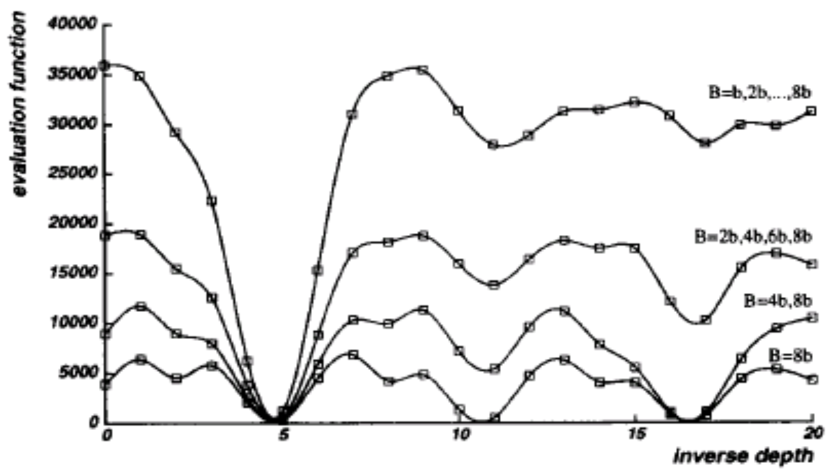


Fig. 7. Combining multiple baseline stereo pairs.

M. Okutomi, T. Kanade, Multiple-Baseline Stereo

Matching

- Even when the cameras are identical models, there can be differences in gain and sensitivity.
- The cameras do not see exactly the same surfaces, so their overall light levels can differ.
  - occlusion

$$E_r(x, y, d) = \sum_{(x', y') \in W_m(x, y)} \rho(I_L(x', y') - I_R(x' - d, y'))$$

Robust matching function.

Looks like optical flow. Why don't we linearize this?

Matching

- Even when the cameras are identical models, there can be differences in gain and sensitivity.
- The cameras do not see exactly the same surfaces, so their overall light levels can differ.
  - occlusion

$$E_r(x,y,d,a,b) = \sum_{(x',y')\in W_m(x,y)} \rho(I_L(x',y') - (aI_R(x'-d,y')+b))$$

Can add parameters to model illumination differences between cameras.

Correspondence Using SSD

Left

Disparity Map

Images courtesy of Point Grey Research

Bayesian Interpretation

$$p_M(\mathbf{d} \mid I_L, I_R)$$

How do we proceed?

Bayesian inference

Prior model $p_P(\boldsymbol{d})$

Likelihood model $p_M(I_L, I_R | \boldsymbol{d})$

Posterior model

$$p(\boldsymbol{d} \mid I_L, I_R) = k\, p_M(I_L, I_R | \boldsymbol{d})\, p_P(\boldsymbol{d})$$

Maximum a Posteriori (MAP estimate):

maximize $p(\boldsymbol{d} \mid I_L, I_R)$