Stereo Vision

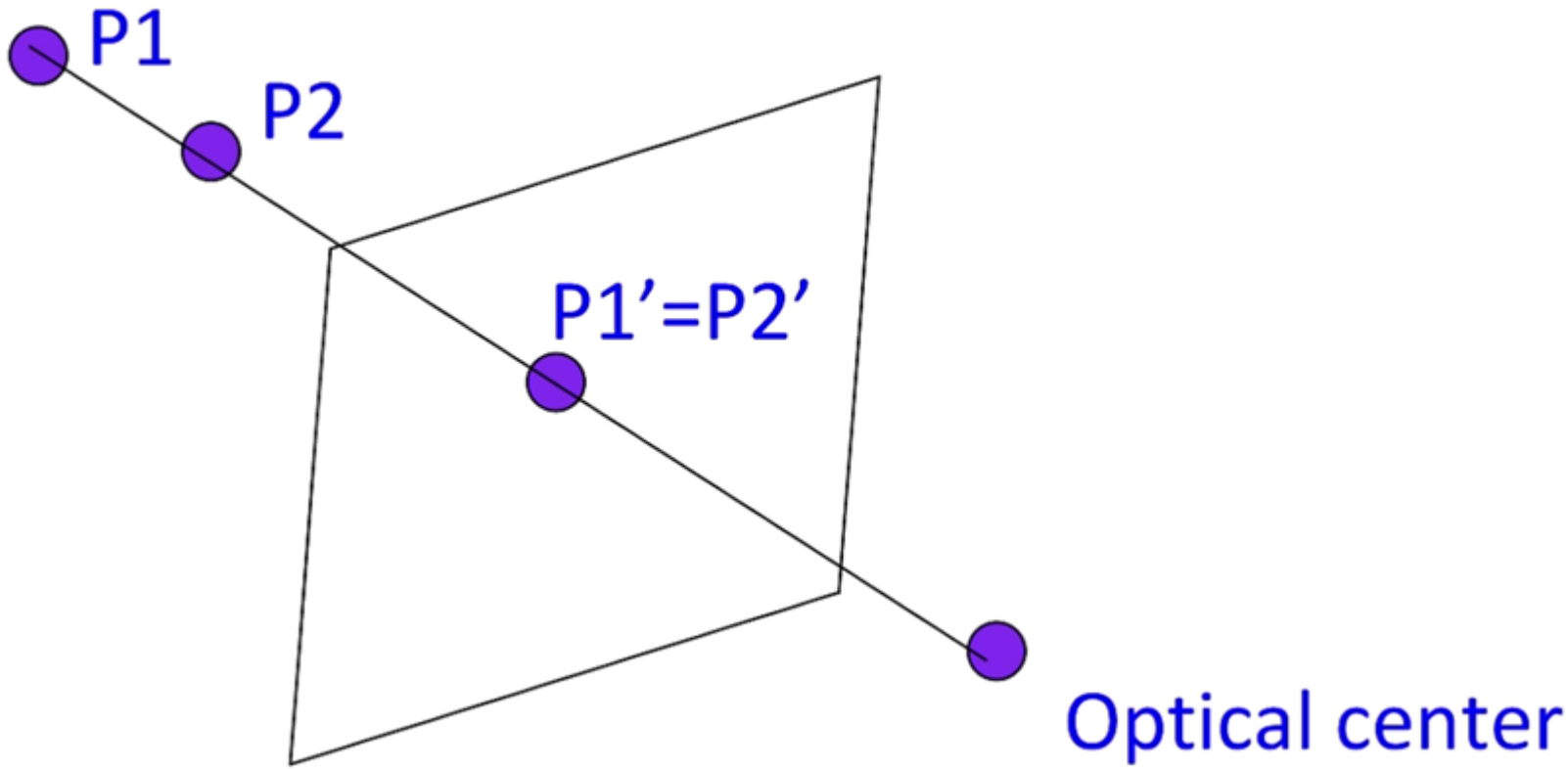
Computer Vision (CS0029)

Why Multiple Views?

- Structure and depth are inherently ambiguous from single views

Why Multiple Views?

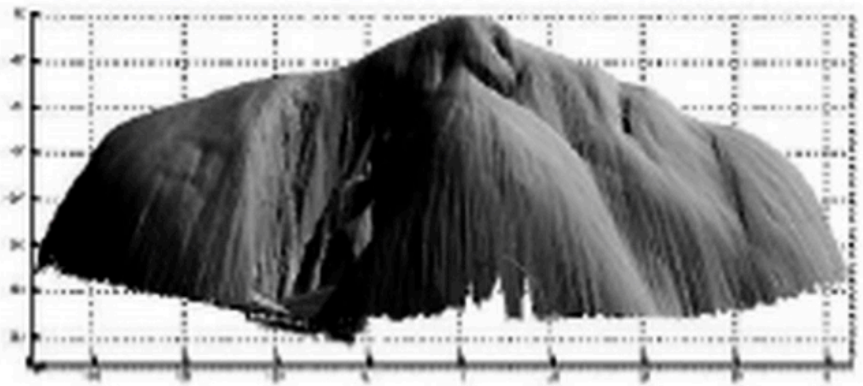
- Structure and depth are inherently ambiguous from single views



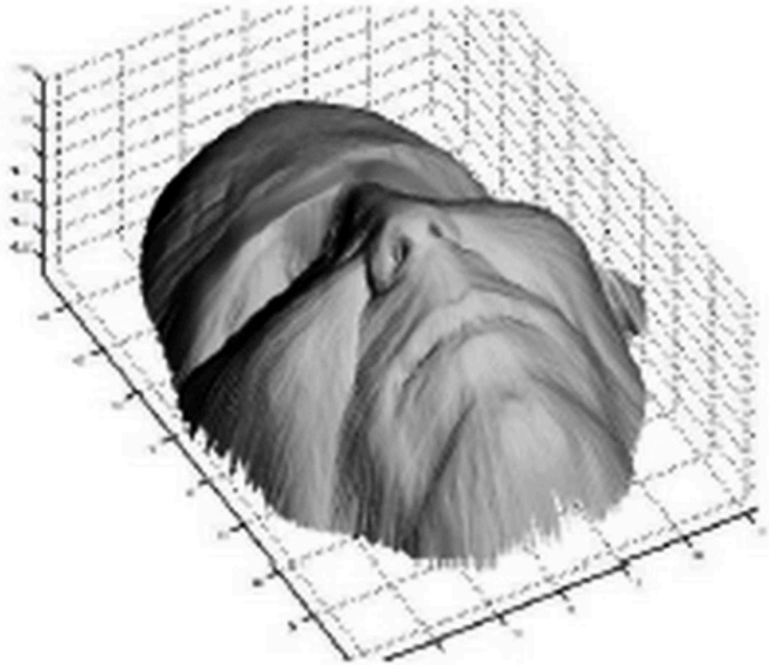
Perspective Effects

Shading

a)

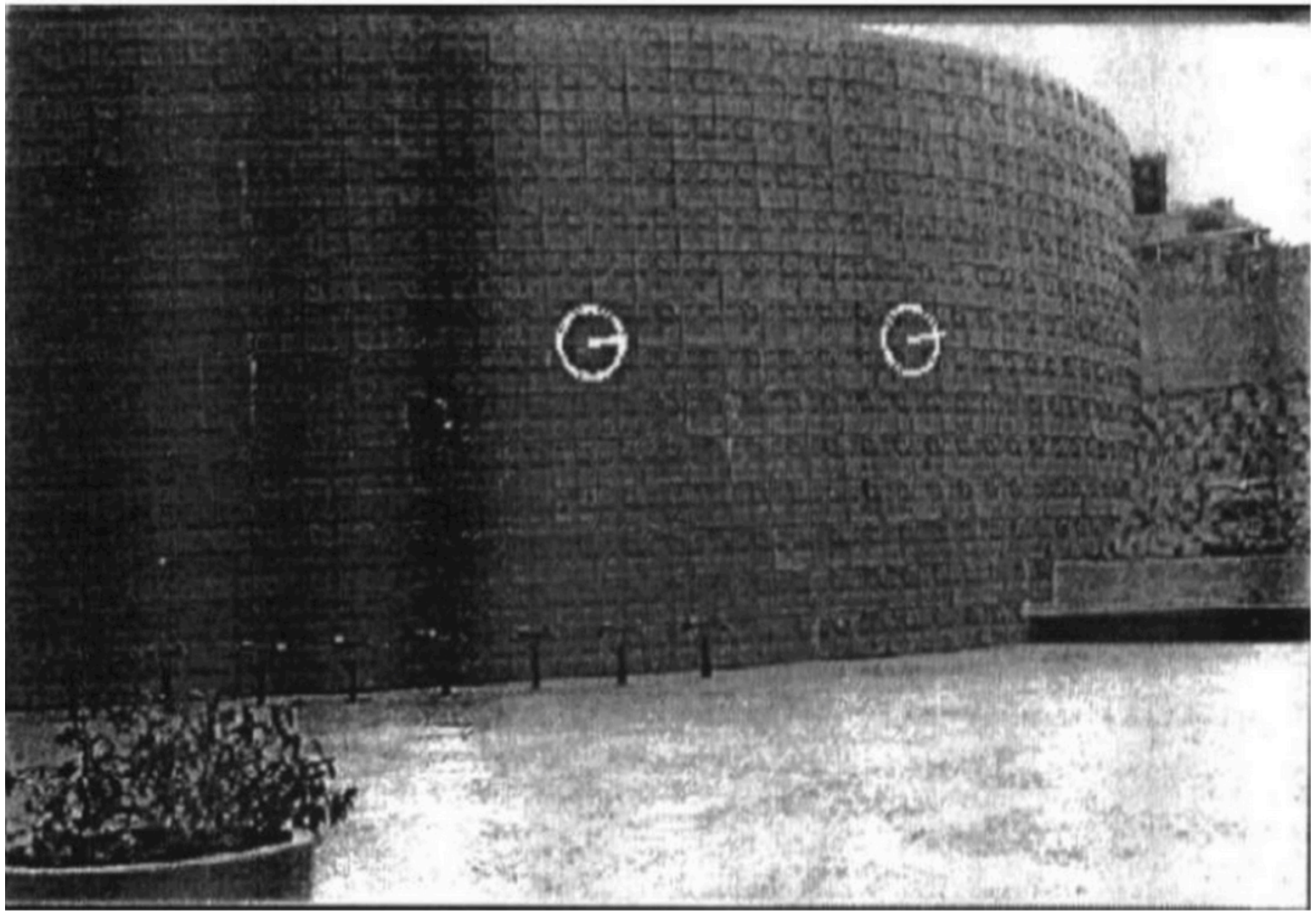


b)

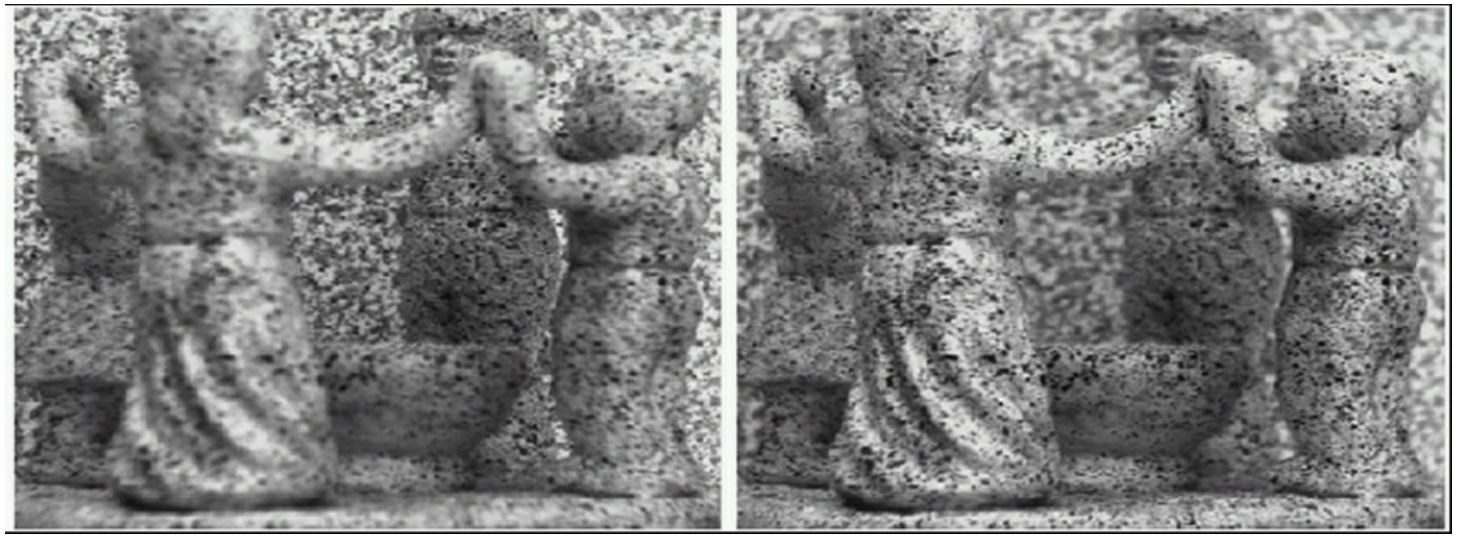


c)

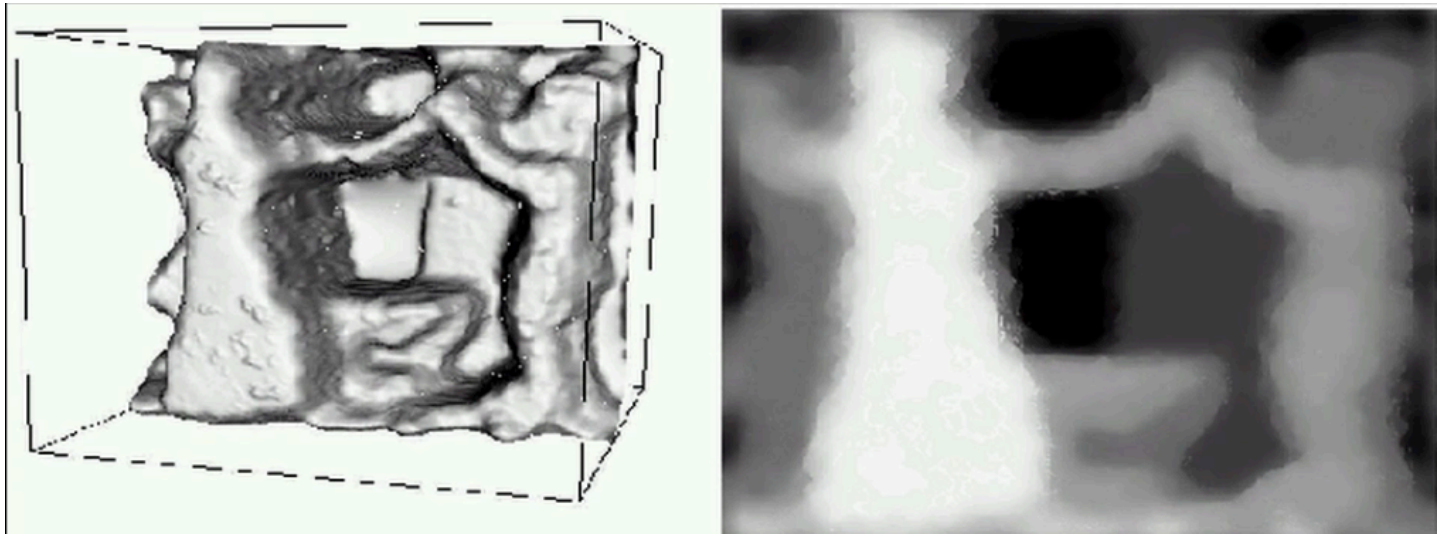
Texture



Focus and Defocus



Images from same point of view, different camera parameters

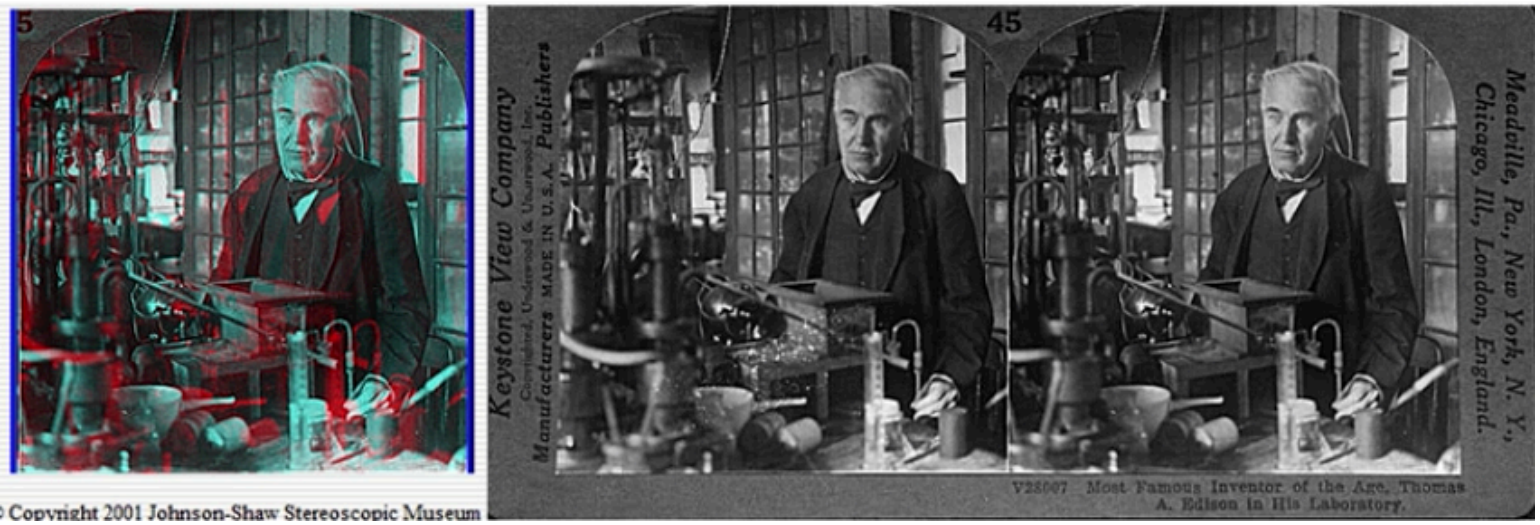


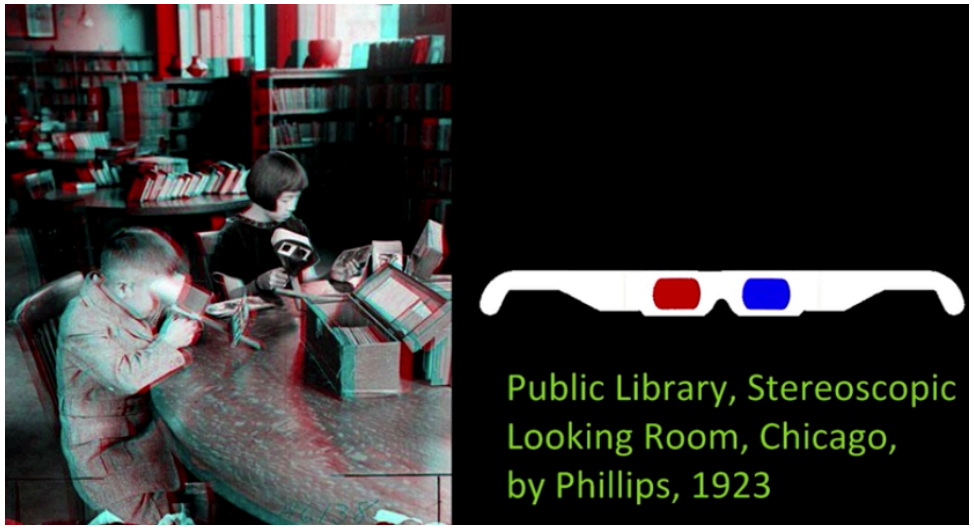
3D shape/depth estimates

Motion

Two Eyes

- Stereo
  - The image from one eye is a little different than the image from the other eye
  - Think of shape from “motion” between two views
  - Infer 3D shape of scene from two (multiple) images from different viewpoints





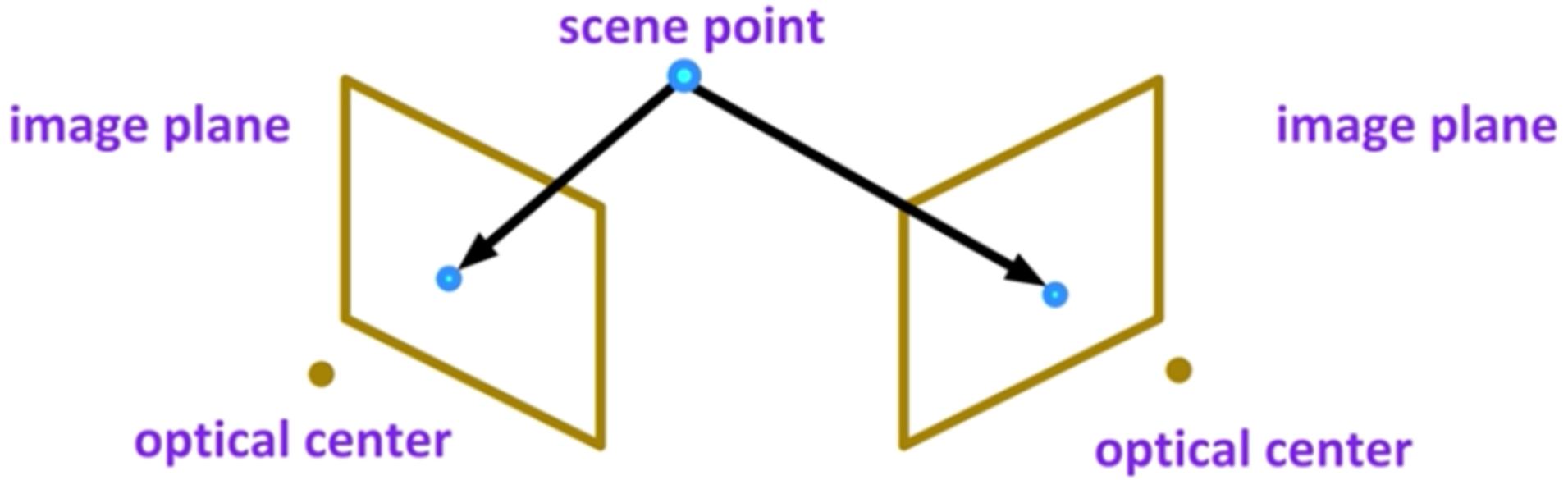
The Basic Idea: Two Slightly Different Image

- https://people.well.com/user/jimg/stereo/stereo_list.html

- Two image from what they move, you get the sense of the depth

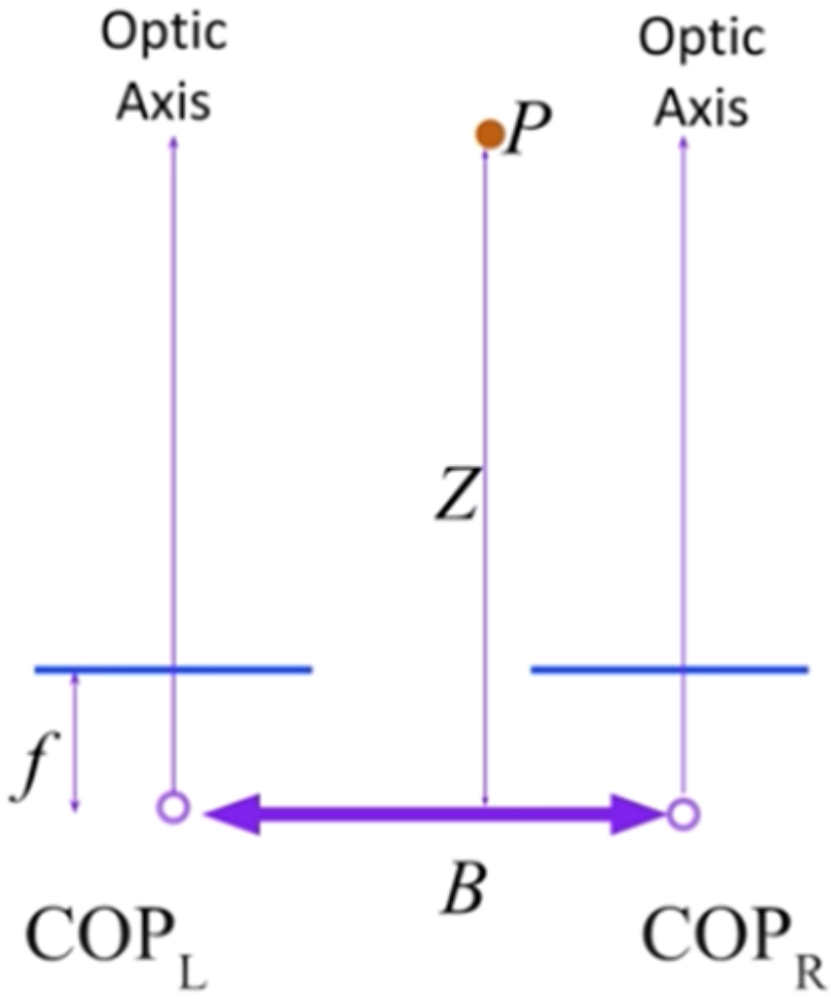
Basic Stereo Geometry

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



Geometry for a Simple Stereo System

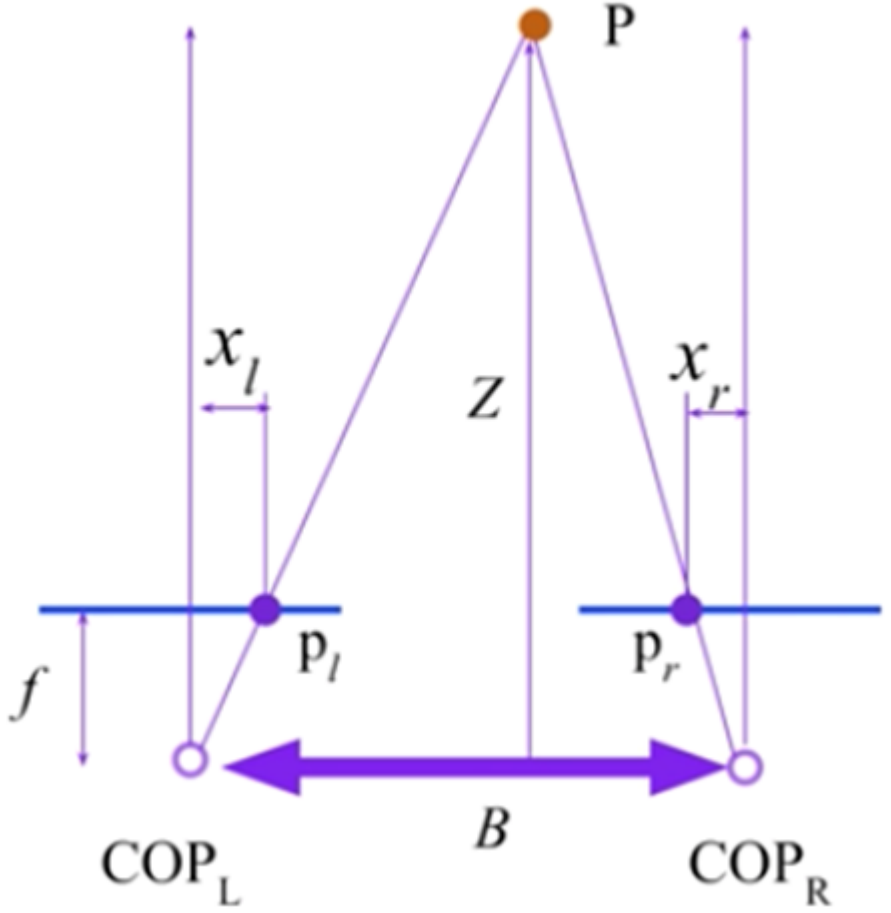
- First, assuming parallel optical axes, known camera parameters (i.e., calibrated cameras)
- Baseline B, focal length f
- Point P is distance Z in camera coordinate systems



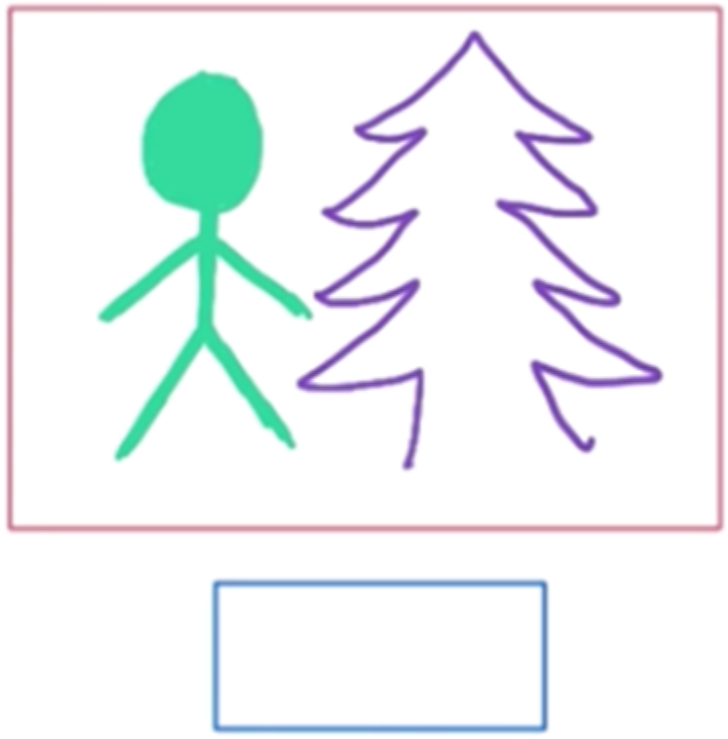
Geometry for a Simple Stereo System

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

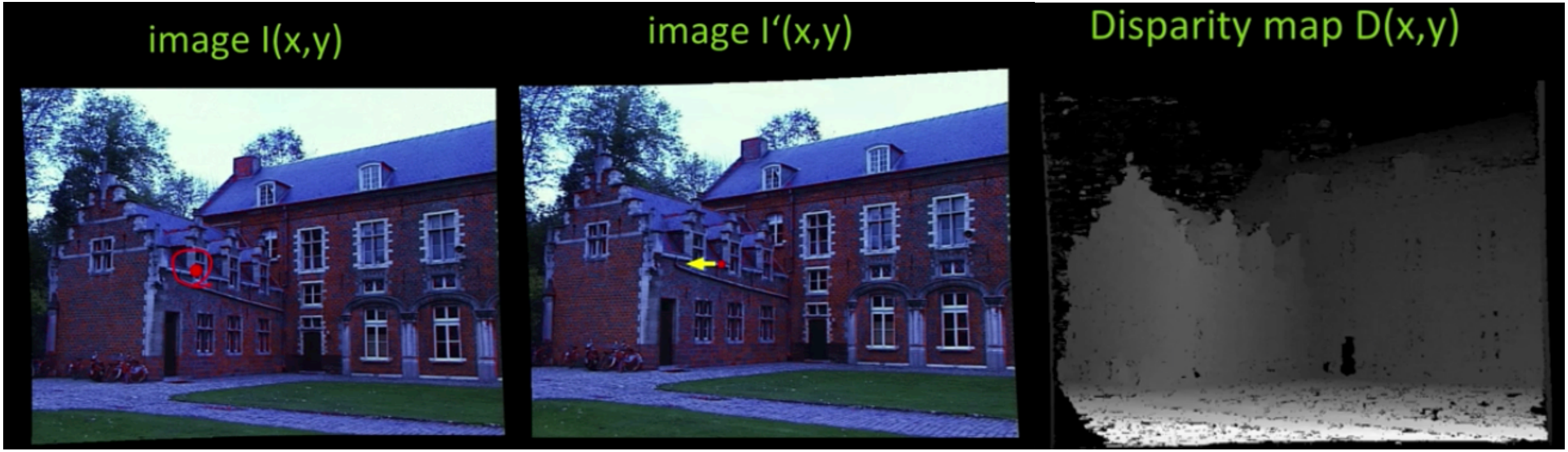
- What is the expression for Z?
- Similar triangles $(p_l, P, p_r)$and (C_l, P, C_r)
  - $\frac{B-x_l+x_r}{Z-f} = \frac{B}{Z}$
  - $Z = f * \frac{B}{x_l - x_r}$ ($x_l - x_r$: disparity)



Practice

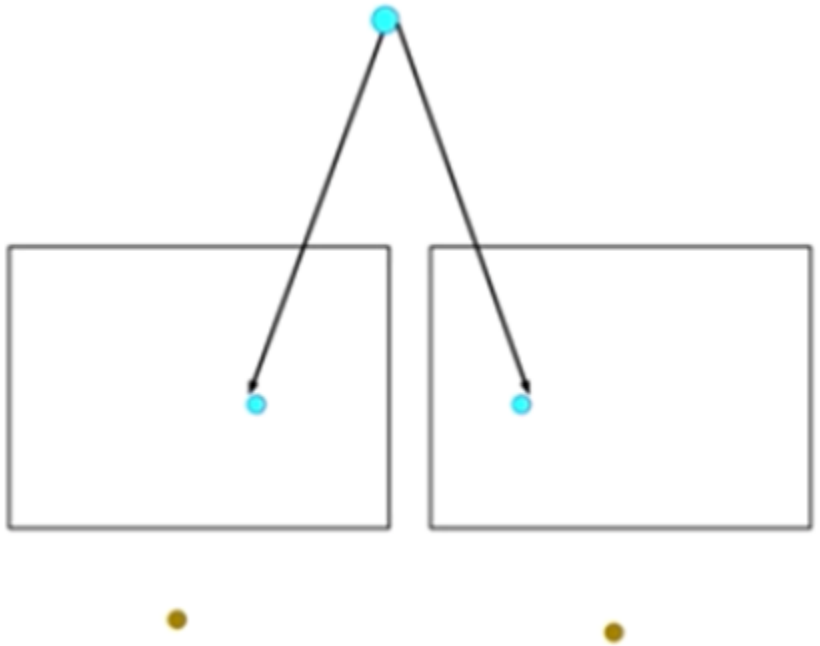


Depth from Disparity

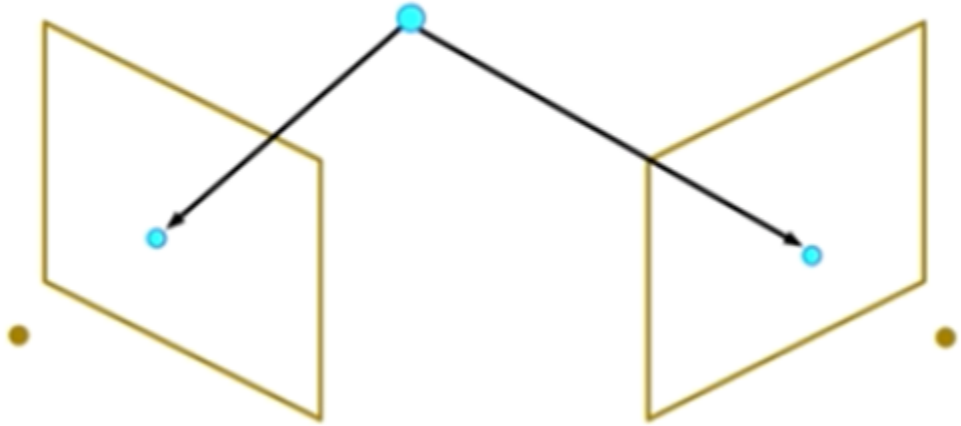


General case, with calibrated cameras

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

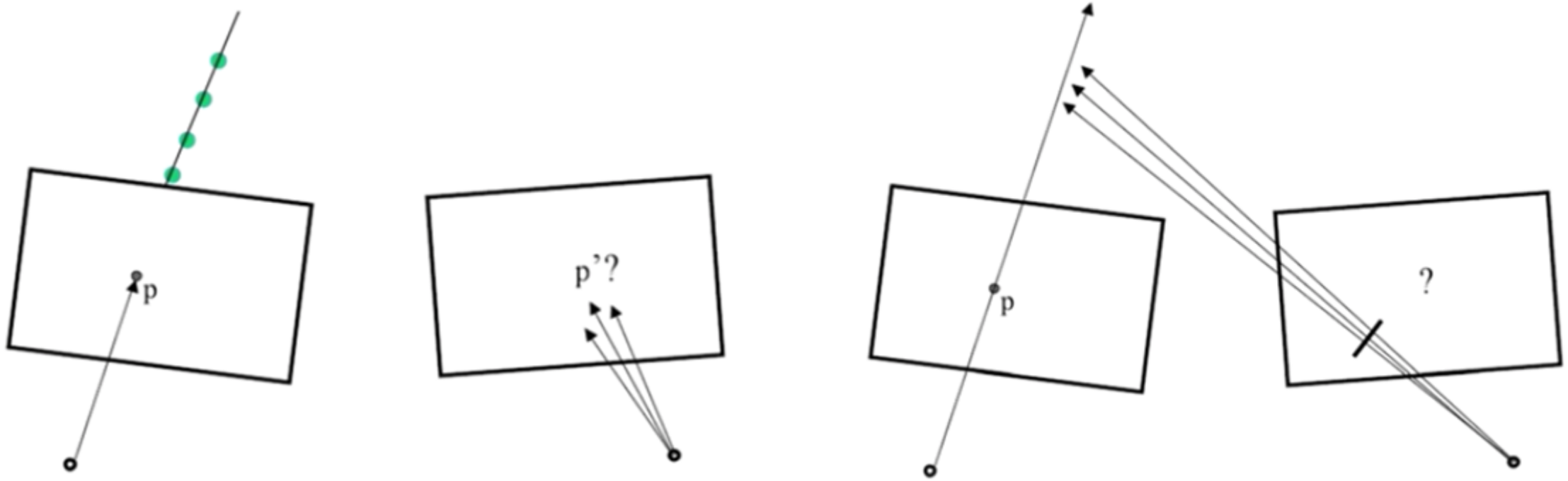


vs.



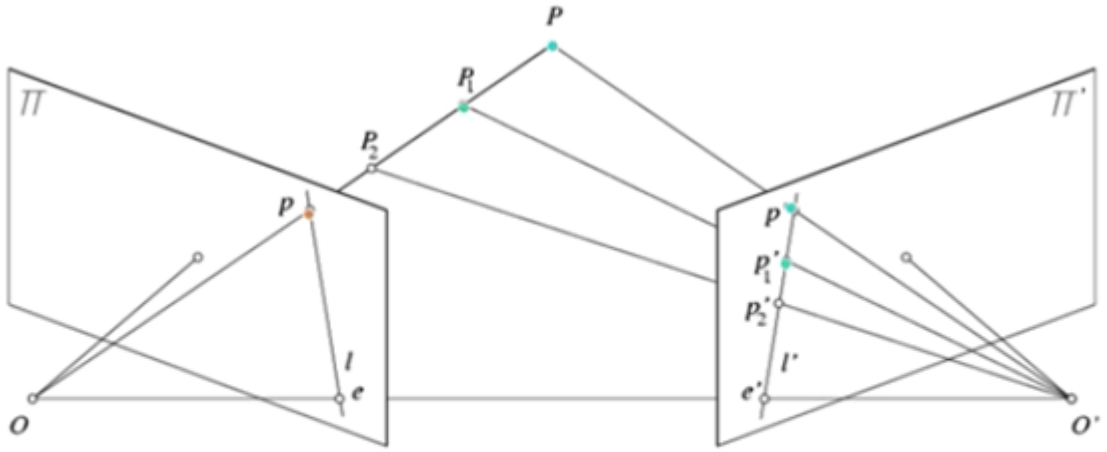
Stereo Correspondence Constraints

- 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

- Baseliine: line joining the camera centers
- Epipolar plane: plane containing baseline and world point
- Epipolar line: intersection of epipolar plane with the image plane –come in pair
- Epipole: point of intersection of baseline with image plane



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

Epipolar Constraint

- The epipolar constraint reduces the correspondence problem to a 1D search along an epipolar line

Example: Converging Cameras

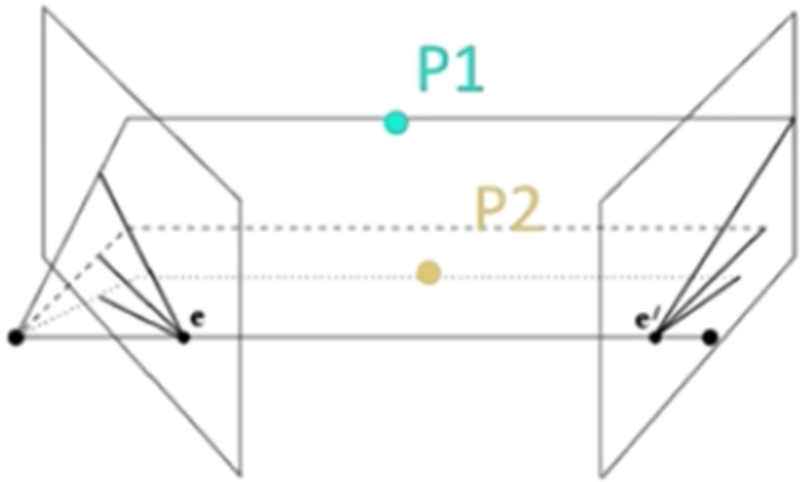
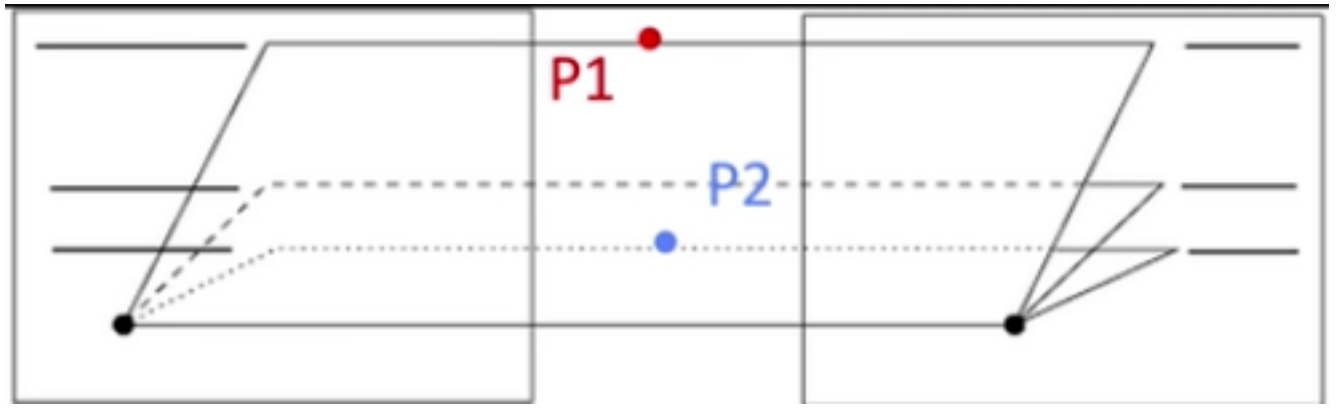


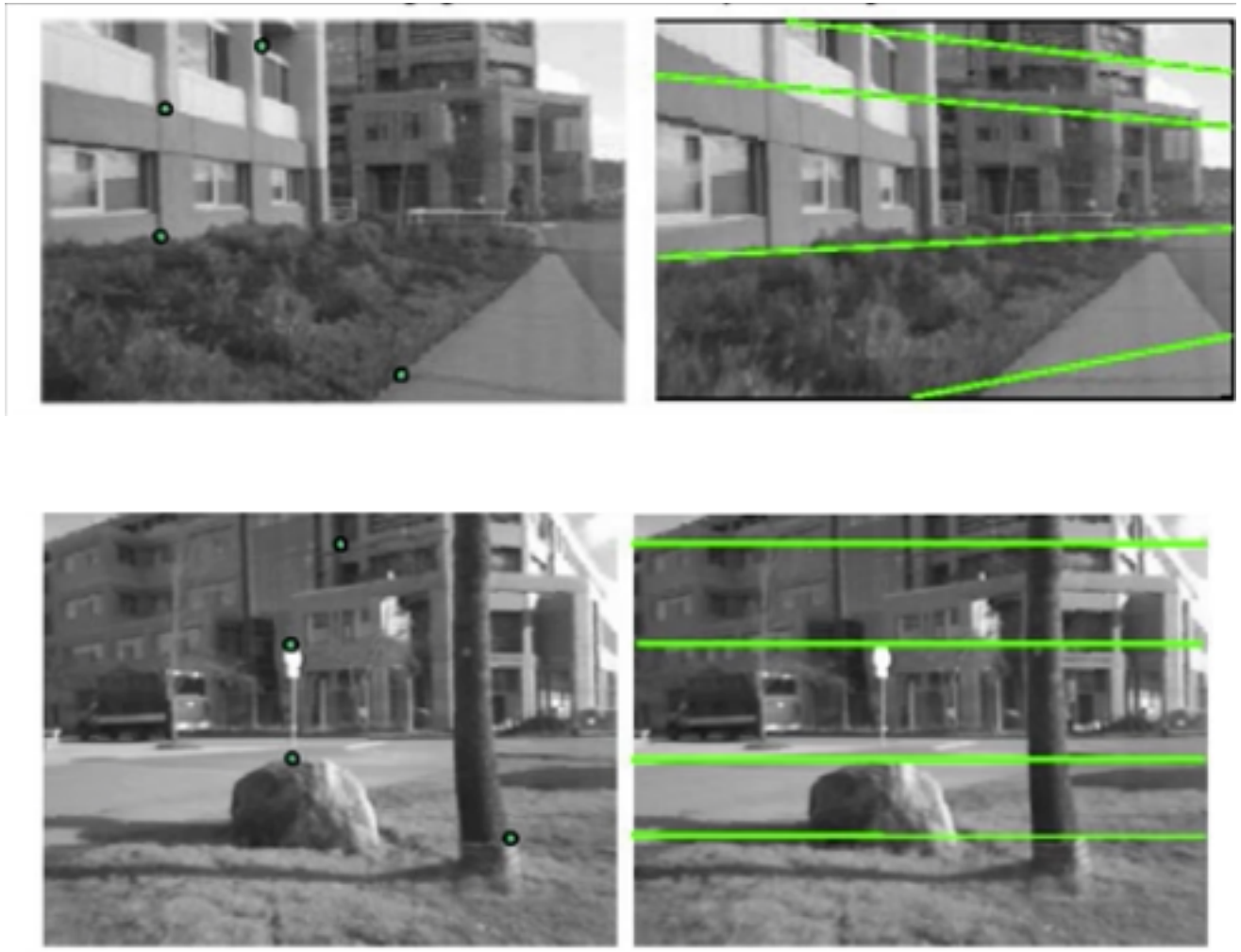
Figure from Hartley & Zisserman

Example: Parallel Image Plane



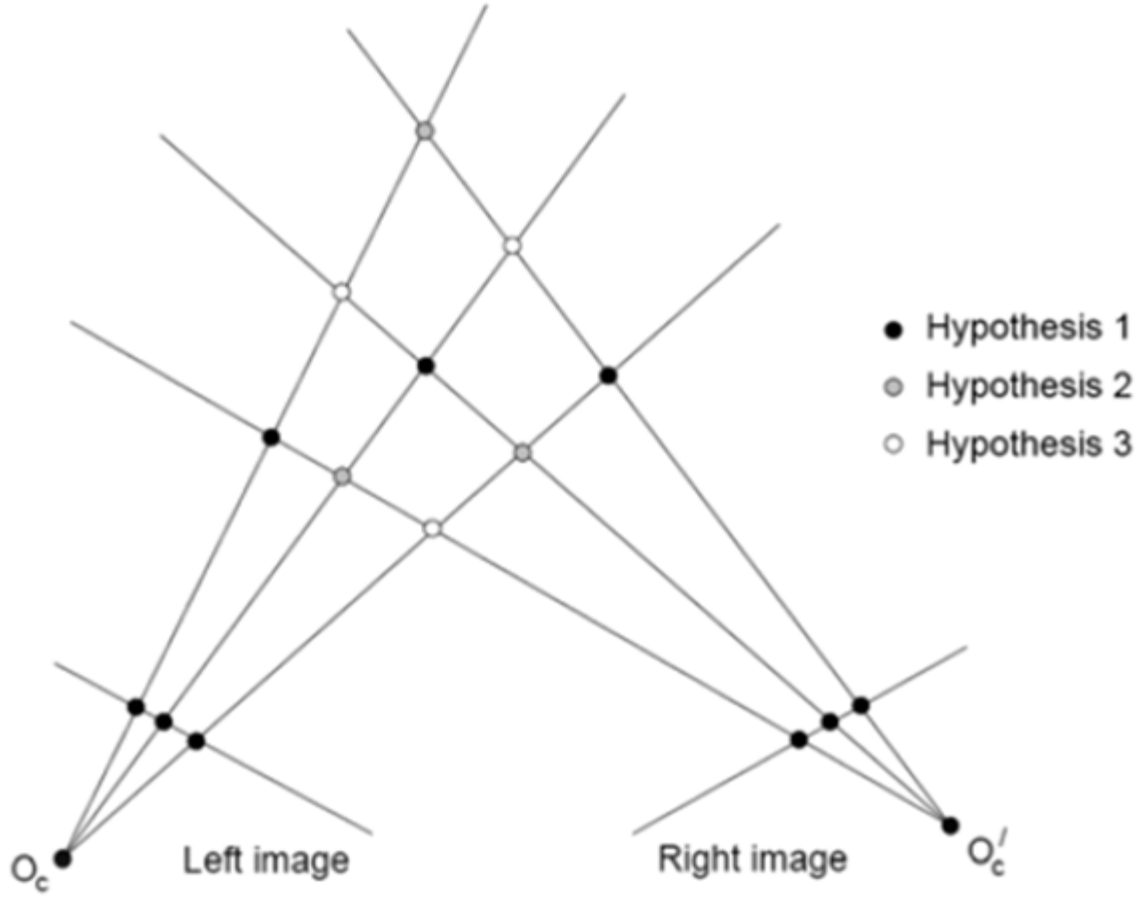
Practice

- How do you know that the bottom one has parallel image planes
- (a) the epipolar lines are horizontal
- (b) the epipolar lines are parallel



Correspondence Problem

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



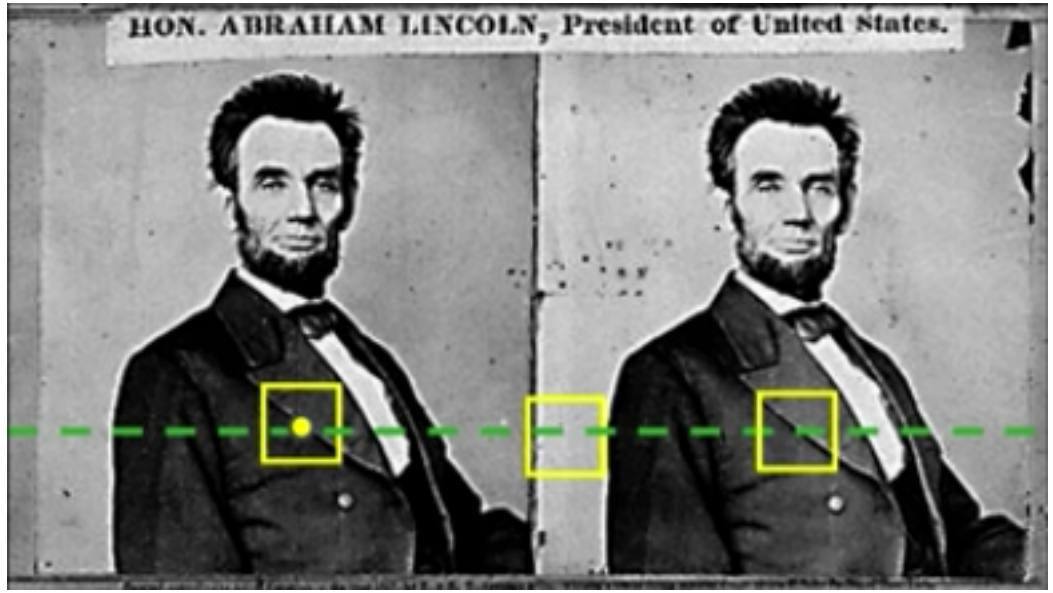
Correspondence Problem

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

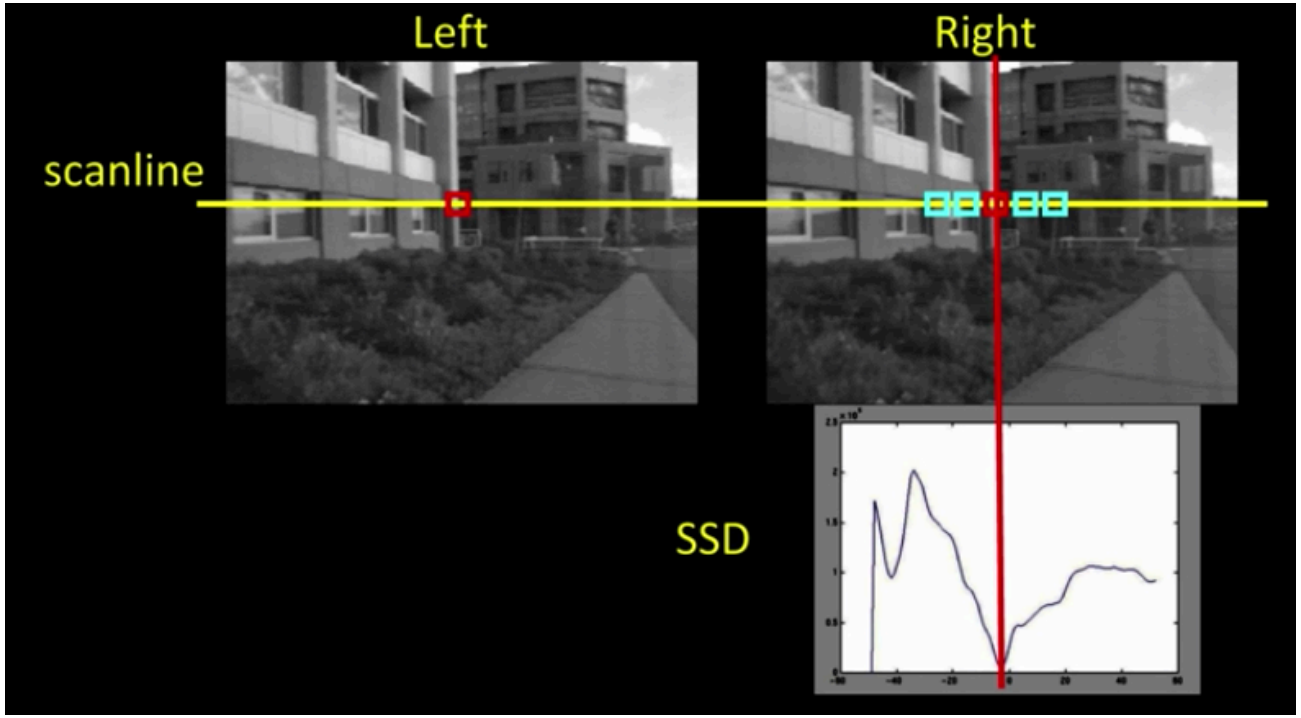
- To find matches in the image pair, we will assume
  - Most scene points visible from both views
  - Image regions for the matches are similar in appearance

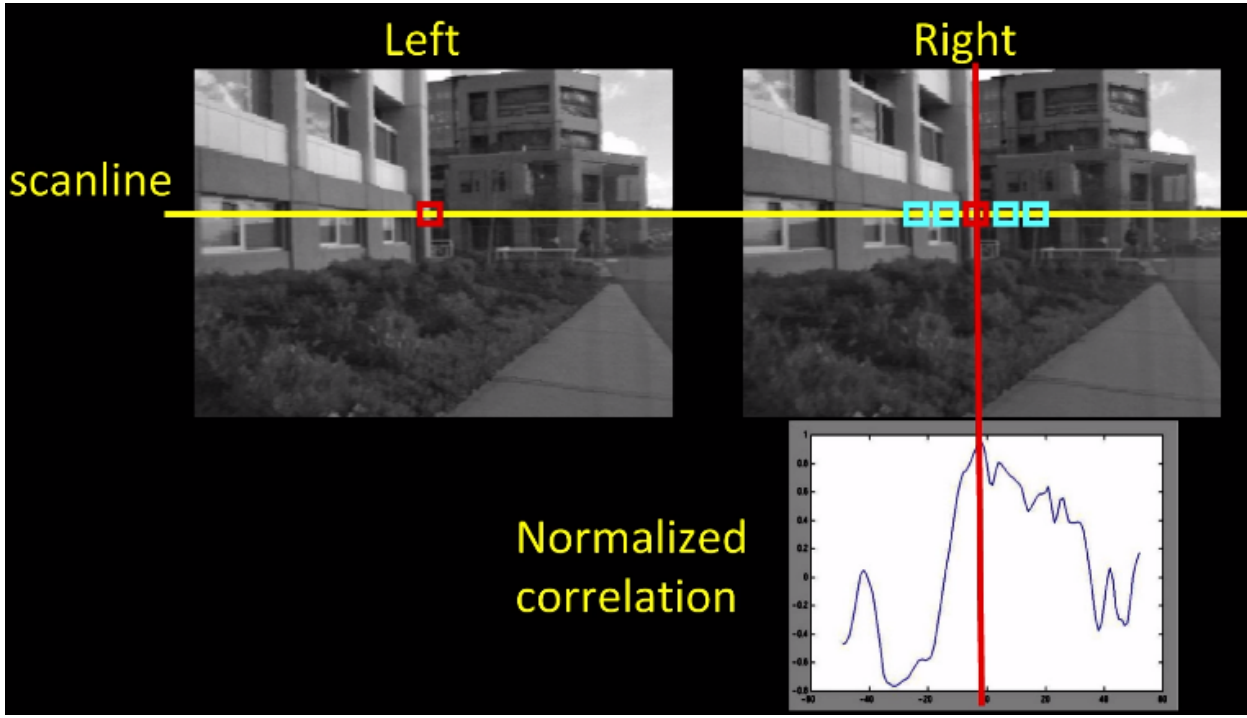
Dense Correspondence Search

- For each 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)

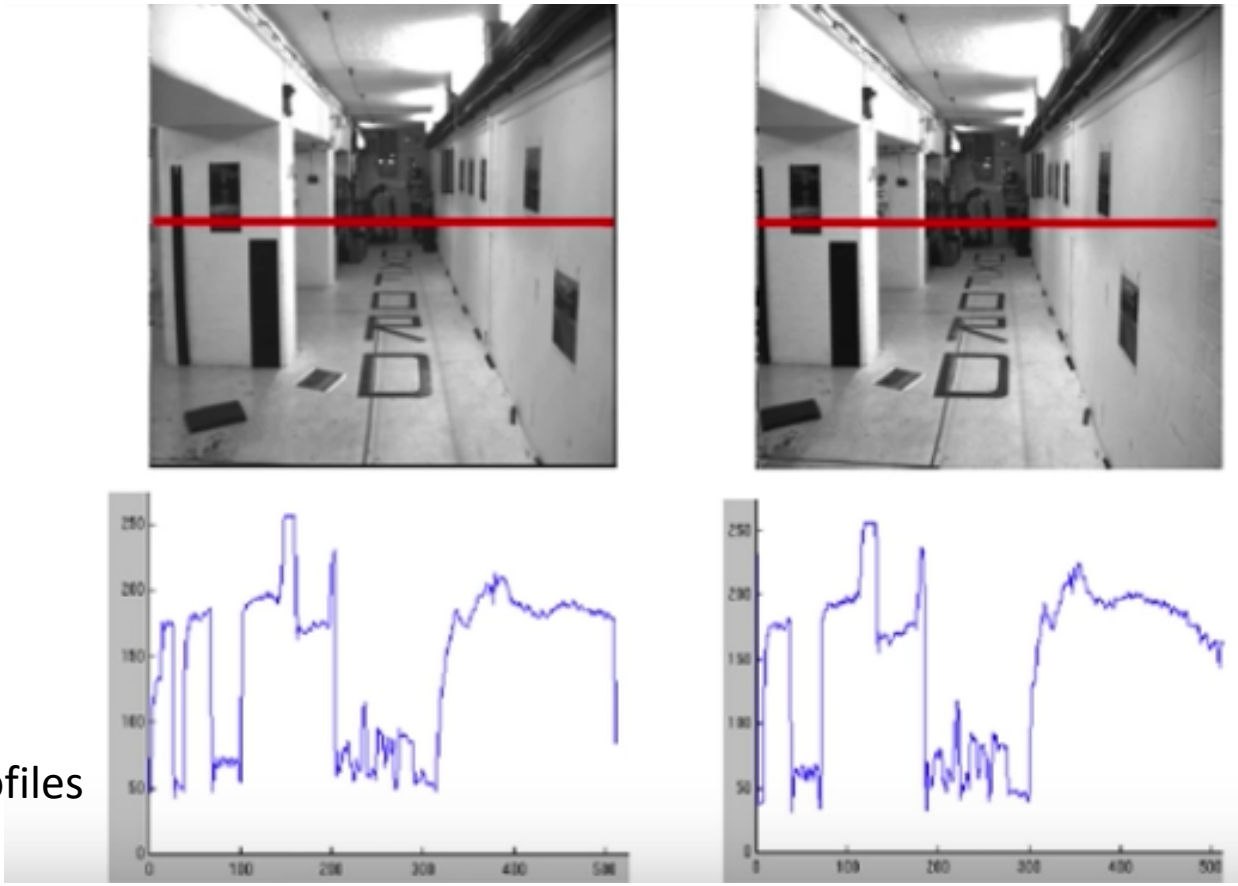


Correspondence Search: Similarity Constraint





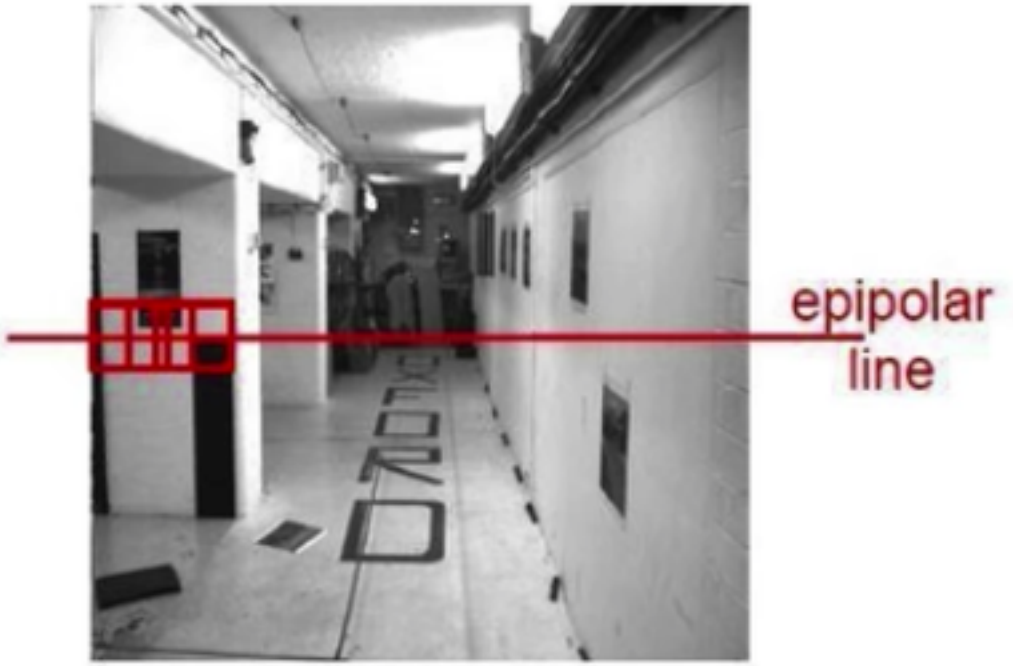
Correspondence Problem



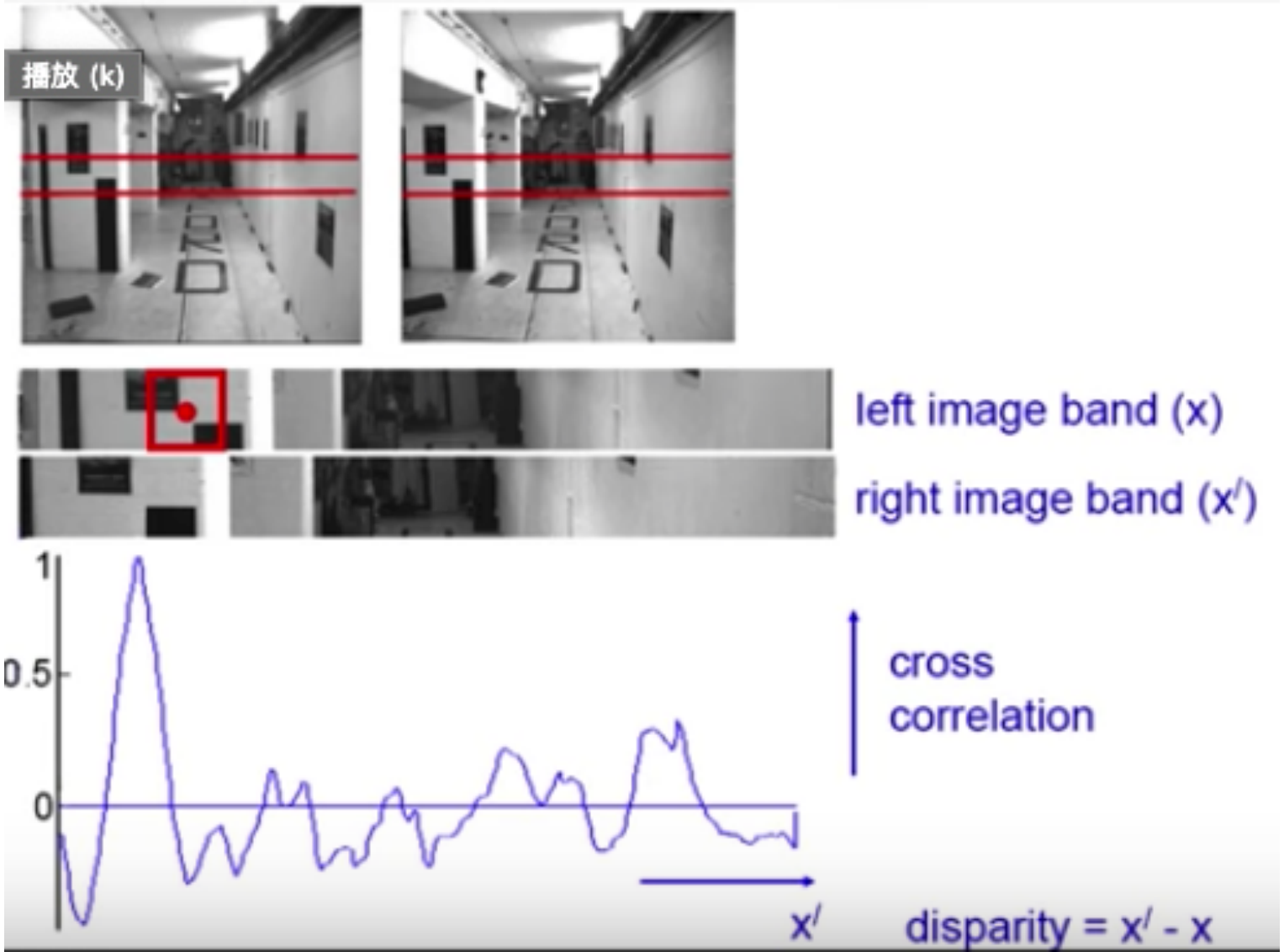
Intensity profiles

Correspondence Problem

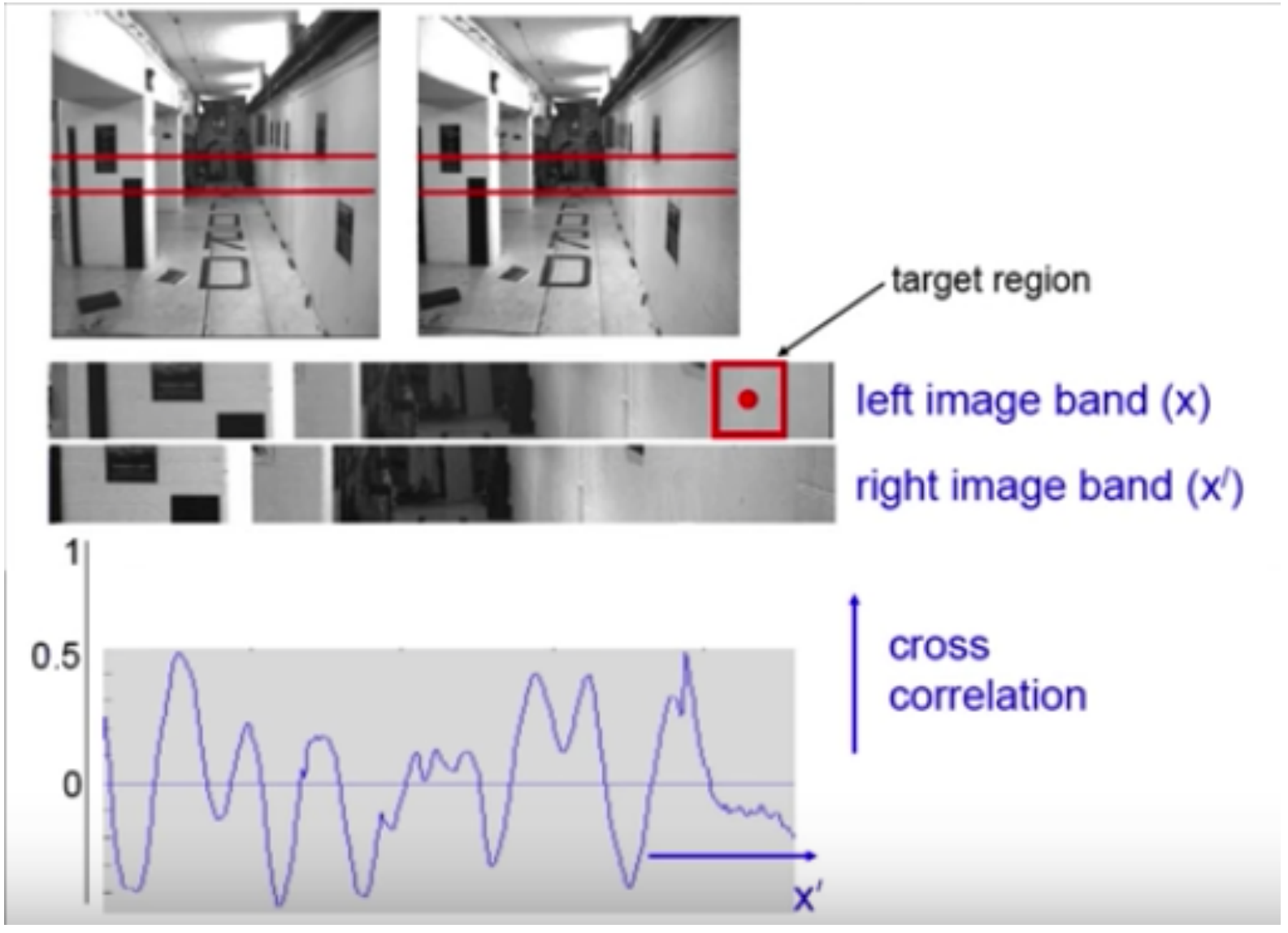
- Neighborhoods of corresponding points are similar in intensity patterns



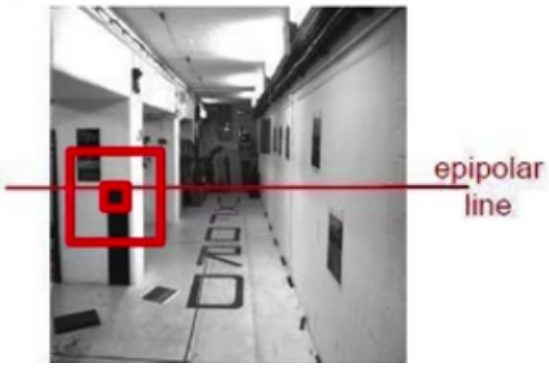
Correlation-based Window Matching



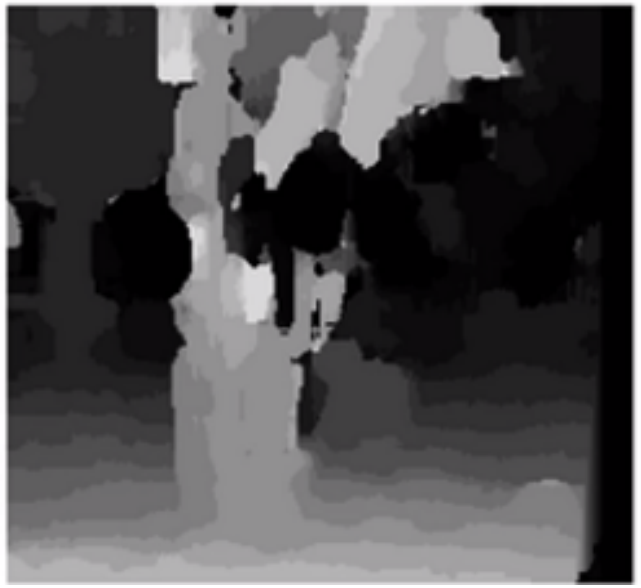
Correlation-based Window Matching



Effect of Window Size



W=3



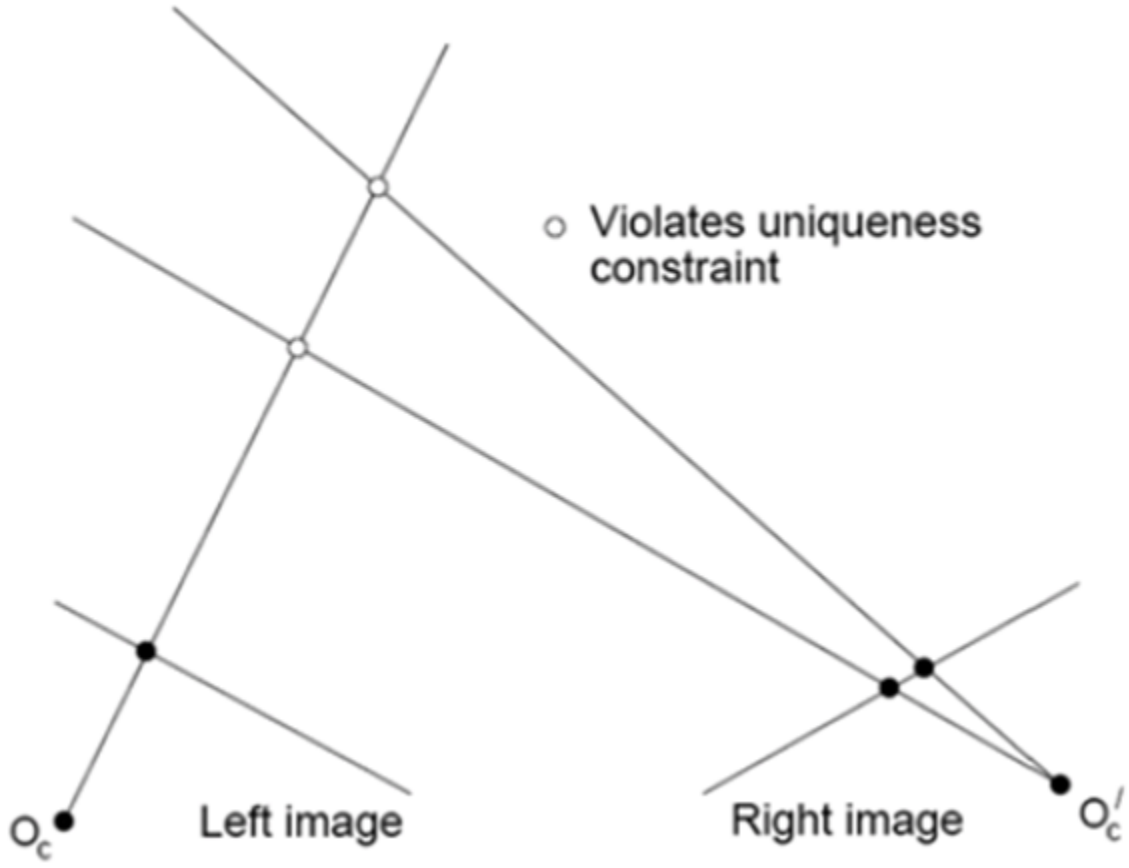
W=20

Correspondence Problem

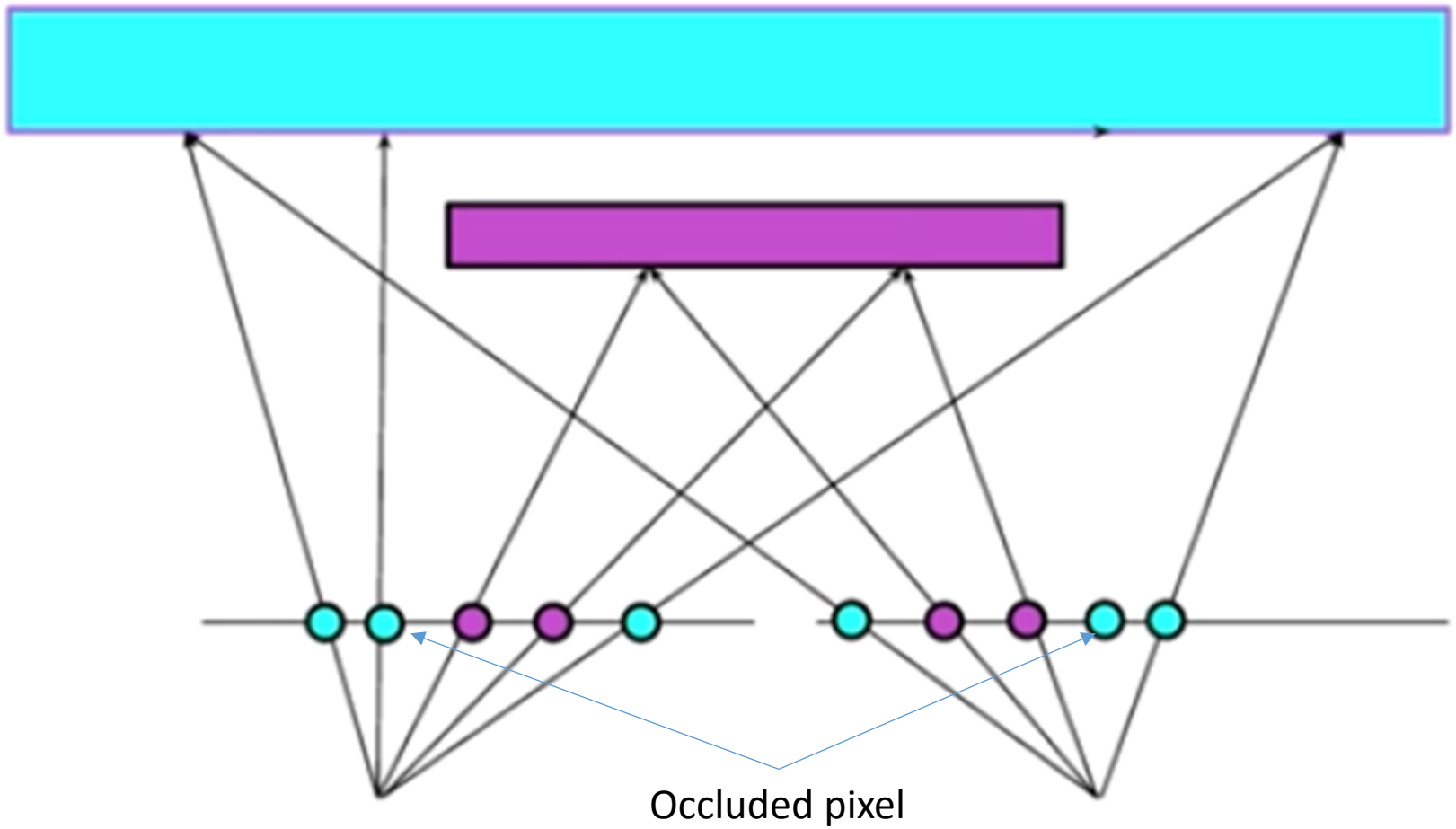
- Beyond the hard constraint of epipolar geometry, there are soft constraints to help identify corresponding points
  - Similarity
  - **Uniqueness**
  - **Ordering**
  - Disparity gradient is limited

Uniqueness Constraint

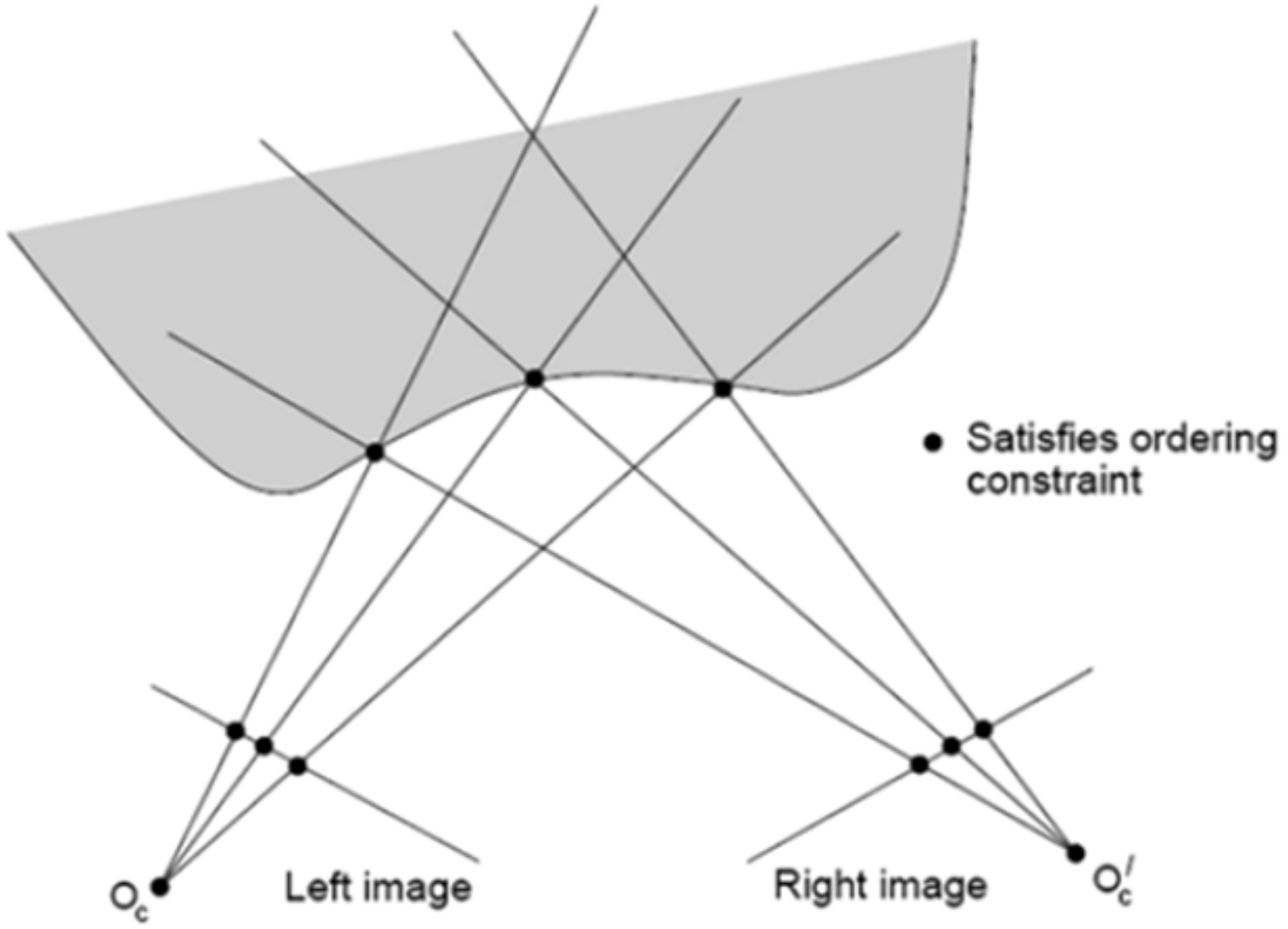
- No more than one match in right image for every point in left image



Occlusion

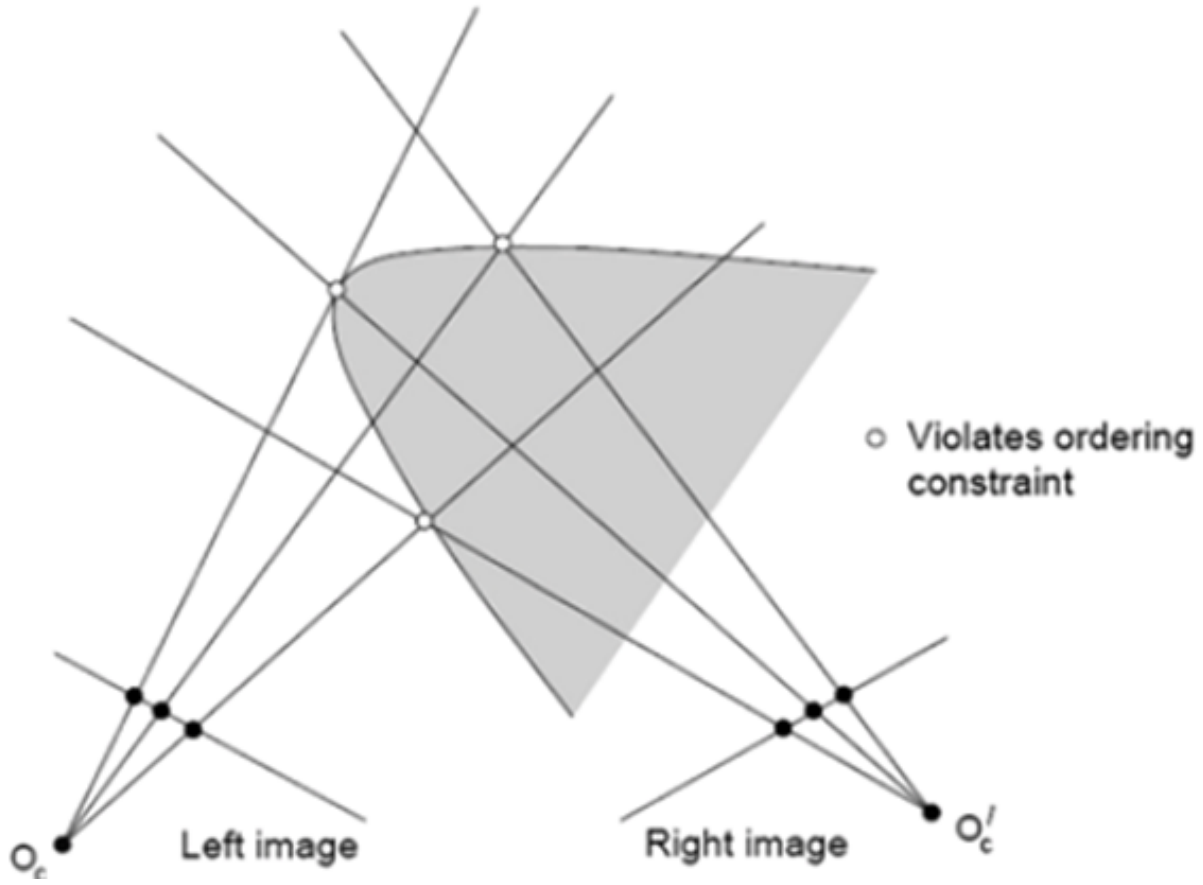


Ordering Constraint



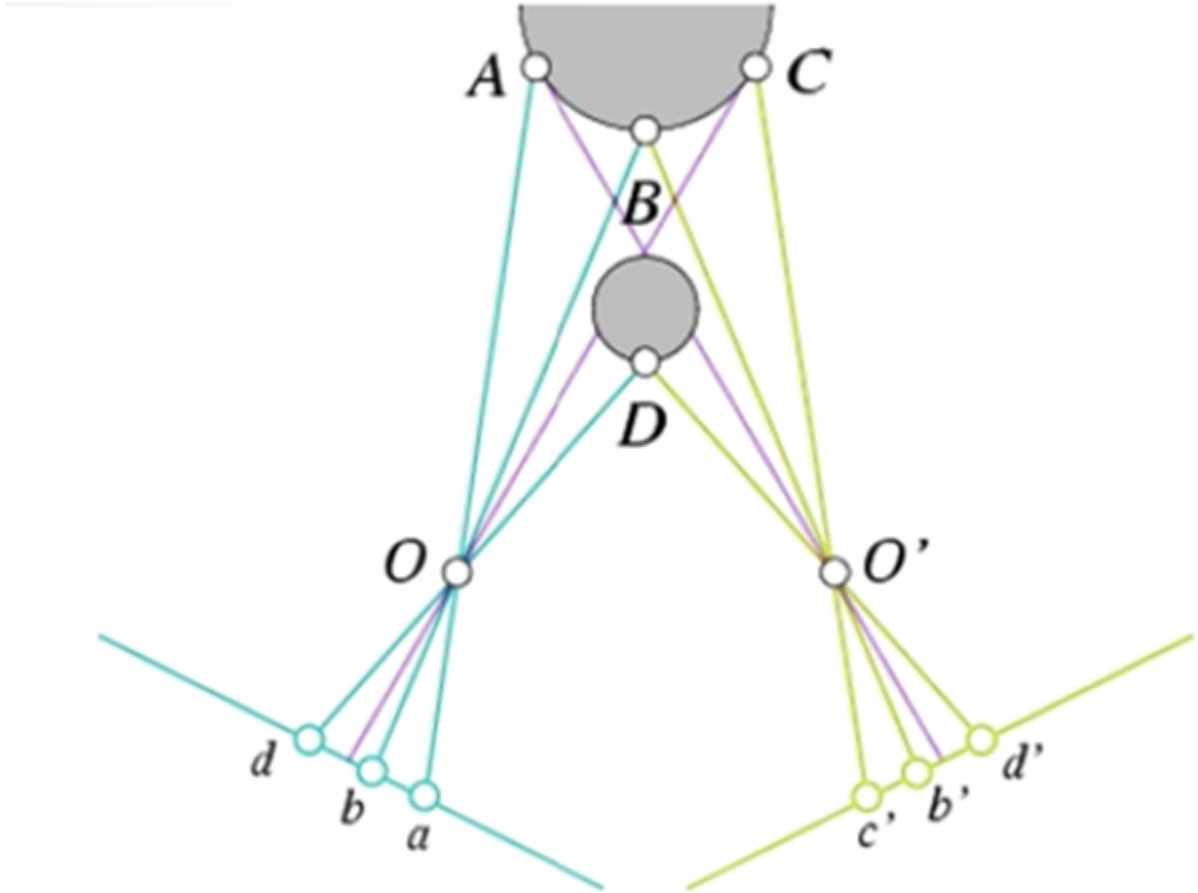
Ordering Constraint

- Will not always hold, e.g. consider transparent object



Ordering Constraint

- A narrow occluding surface



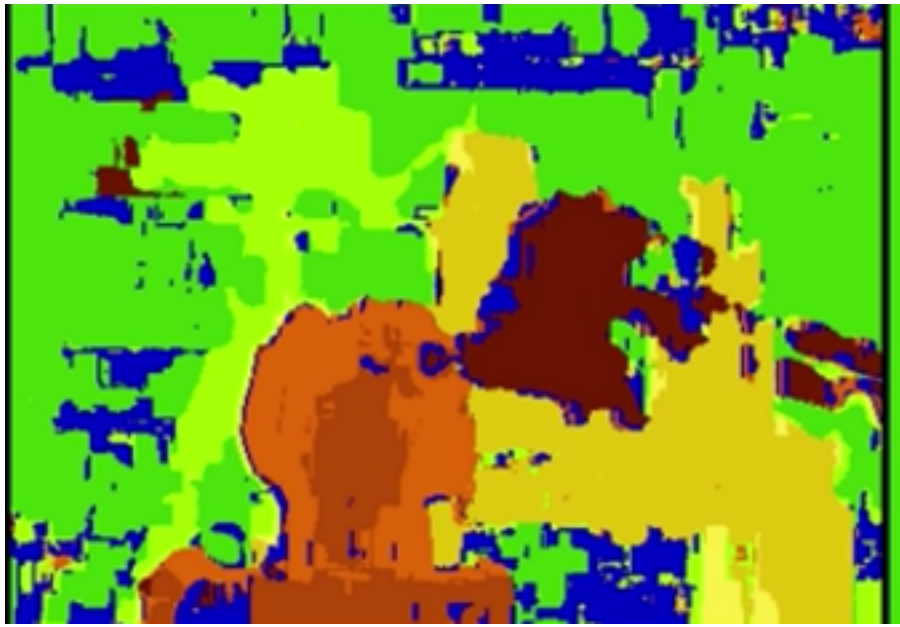
Stereo Results

- Image data from University of Tsukuba

Image

Ground truth depth

Results with Window Search

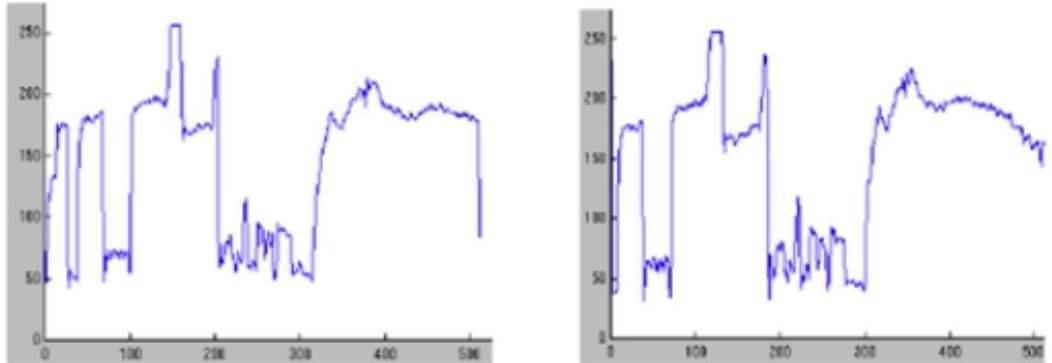


Window-based Matching
(best window size)

Ground truth depth

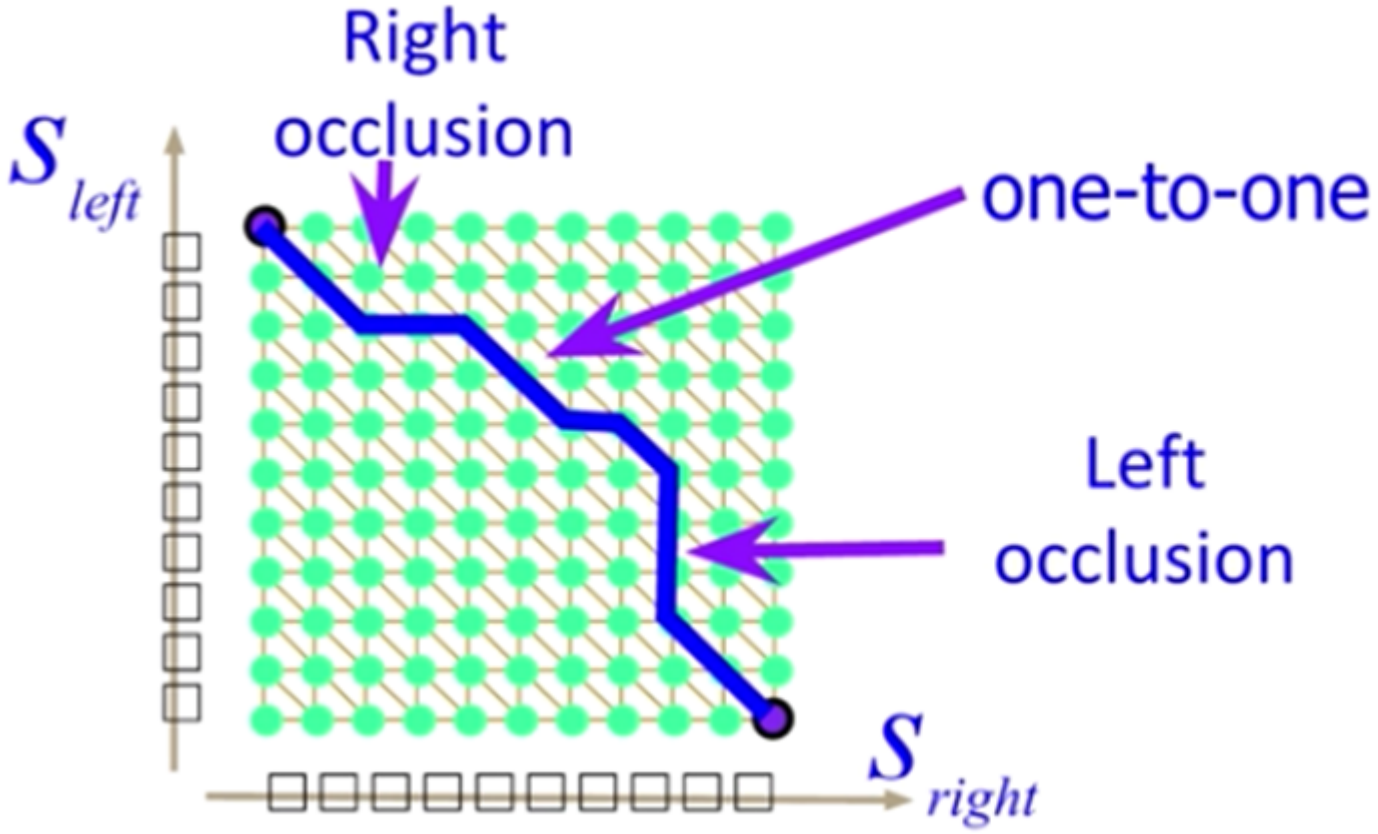
Better Solution

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



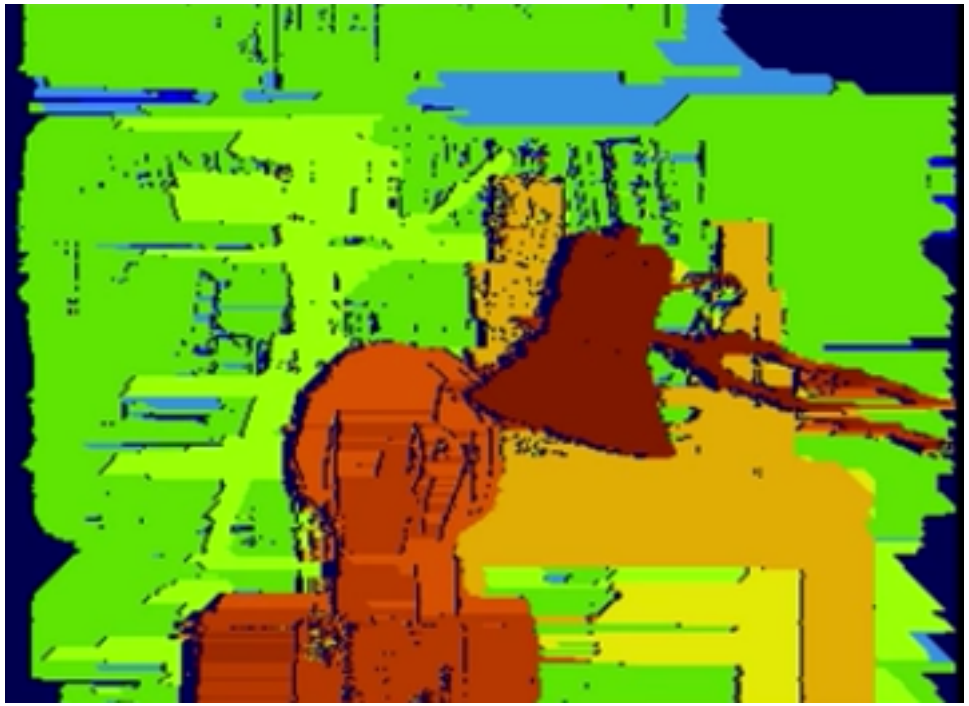
Dynamic Programming Formulation

- Find the best possible path (smallest cost)



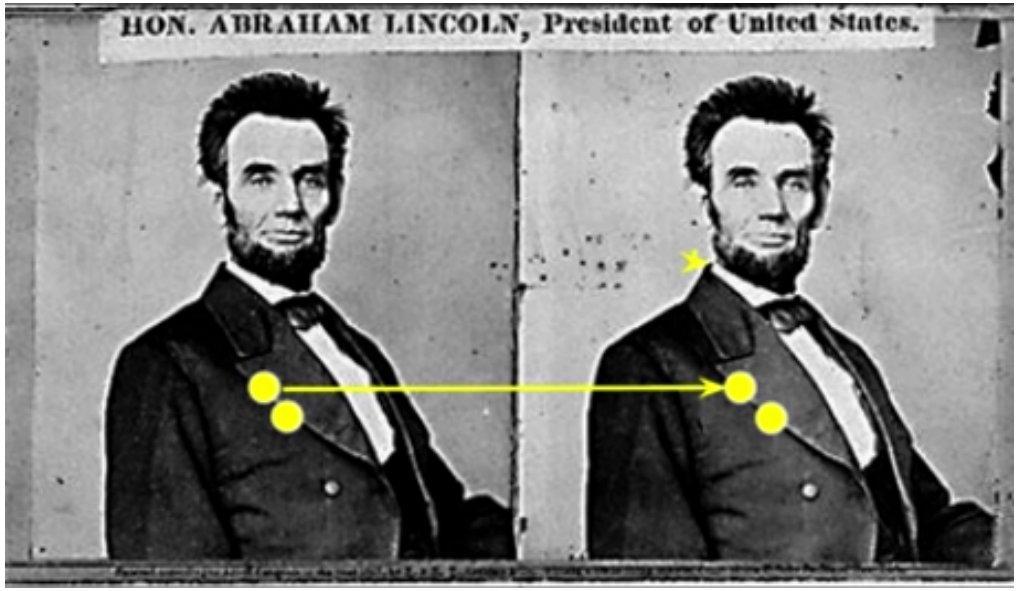
Coherent Stereo on 2D grid -DP

- Scanline stereo generates streaking artifices
- Cannot use dynamic programming to find spatially coherent disparities/correspondences on a 2D grid

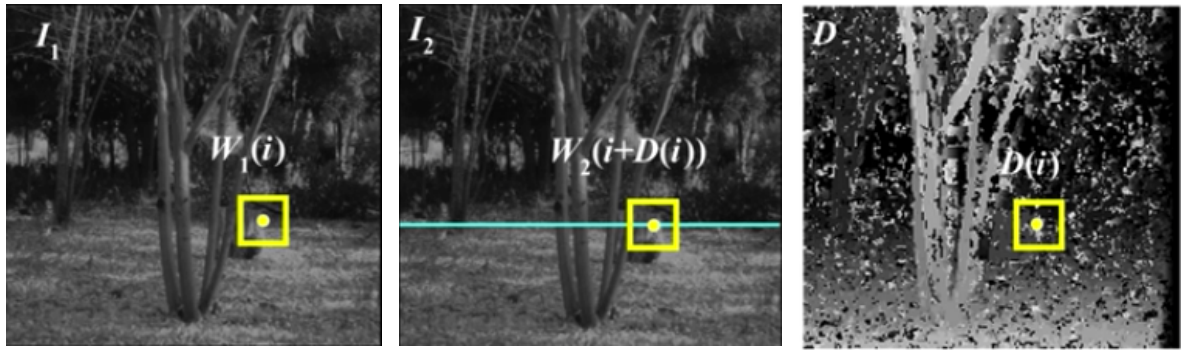


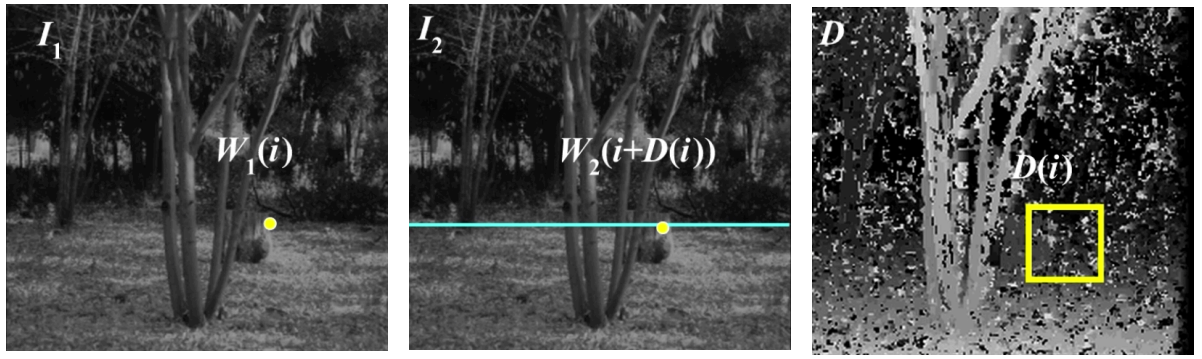
Definition of Good Correspondence

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



Stereo Matching as Energy Minimization





$$Data\ term: E_{data} = \sum_i \left(W_{1(i)} - W_{2(i+D(i))}\right)^2$$

$$Smoothness\ term: E_{Smoothness} = \sum_{neightbors\ i,j} \rho(D(i) - D(j))$$

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

Better Results

- Energy function of this form can be minimized using graph cuts

State of the art

Ground truth

Challenges

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