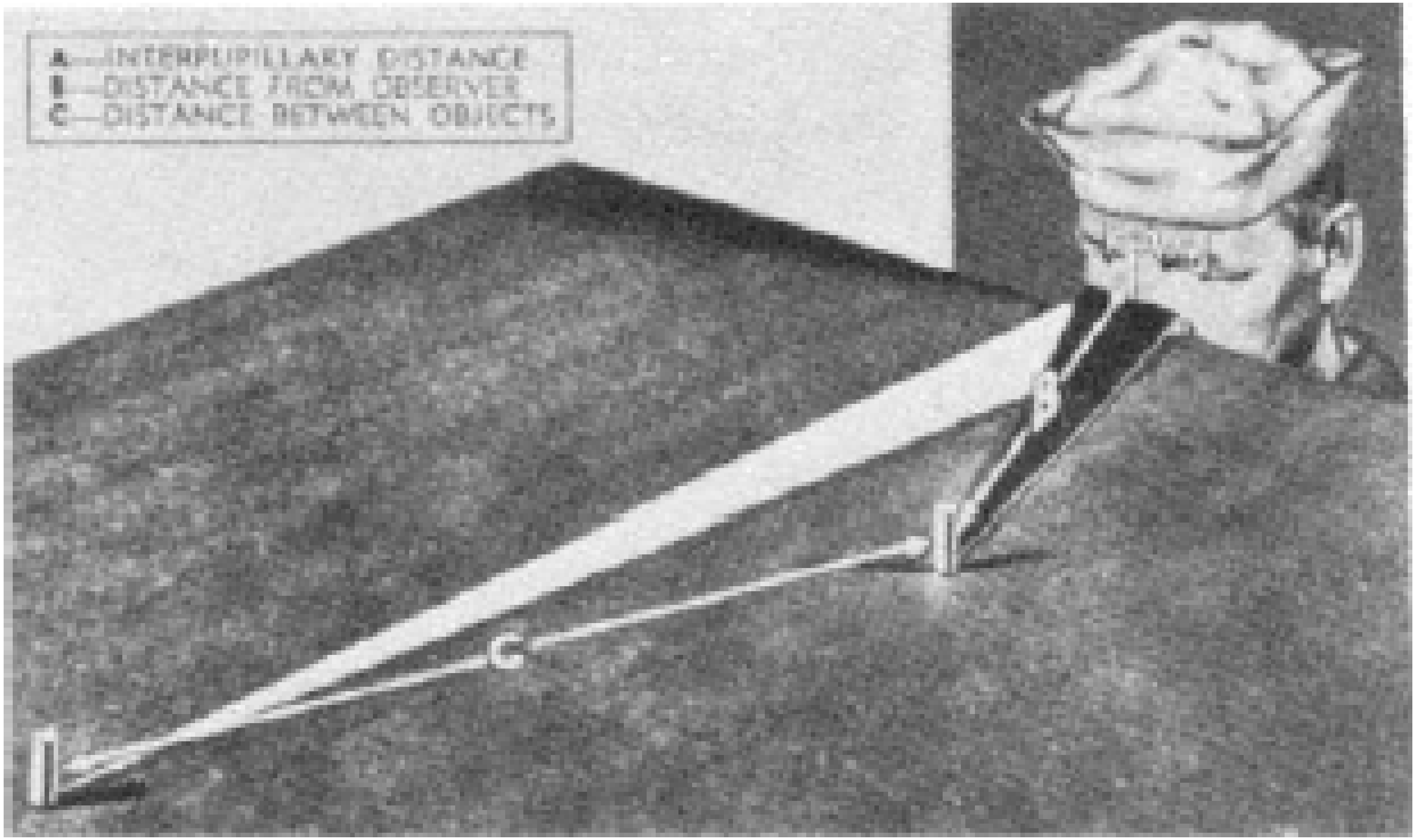
6.801/866

Stereopsis

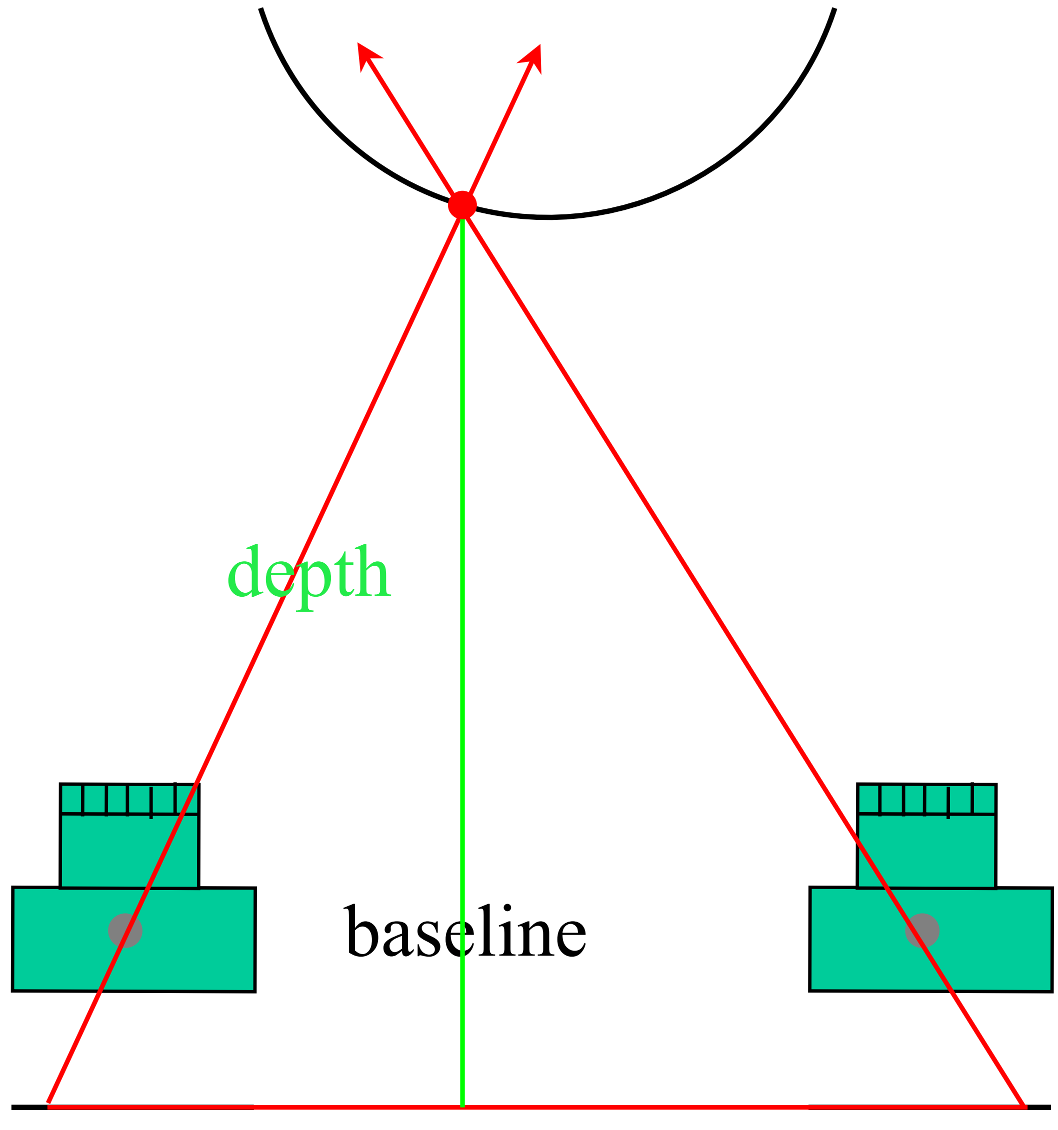
T. Darrell

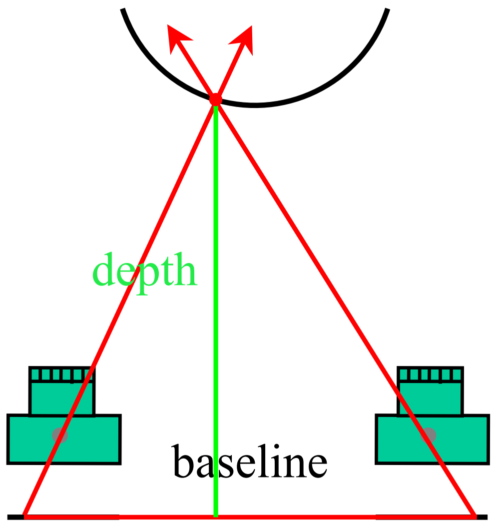

A—INTERPUPILLARY DISTANCE
B—DISTANCE FROM OBSERVER
C—DISTANCE BETWEEN OBJECTS
B
C


IRIA

Outline

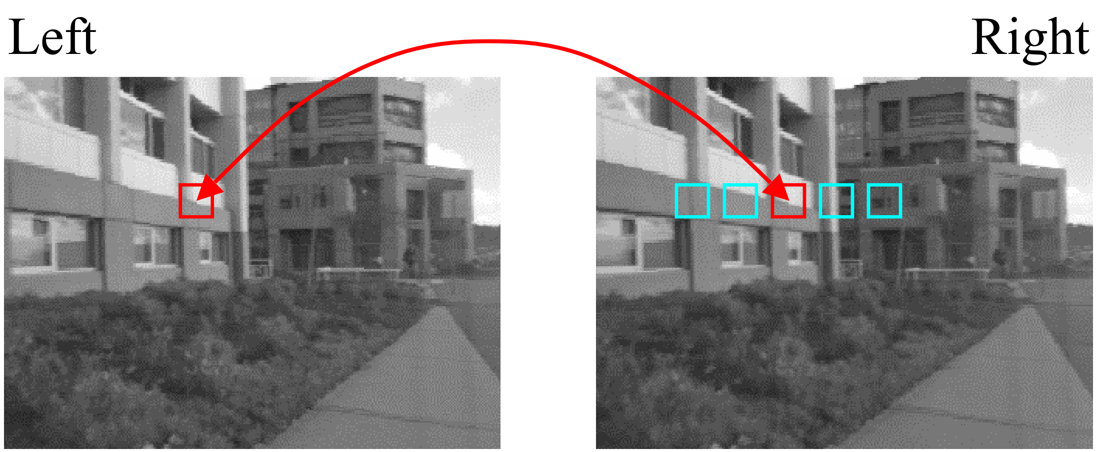
- Reconstruction
- Rectification
- Early vs. Late
- Window Matching
- Edge Matching
- Ordering Constraint
- More views


depth
baseline



Triangulate on two images of the same point to recover depth.

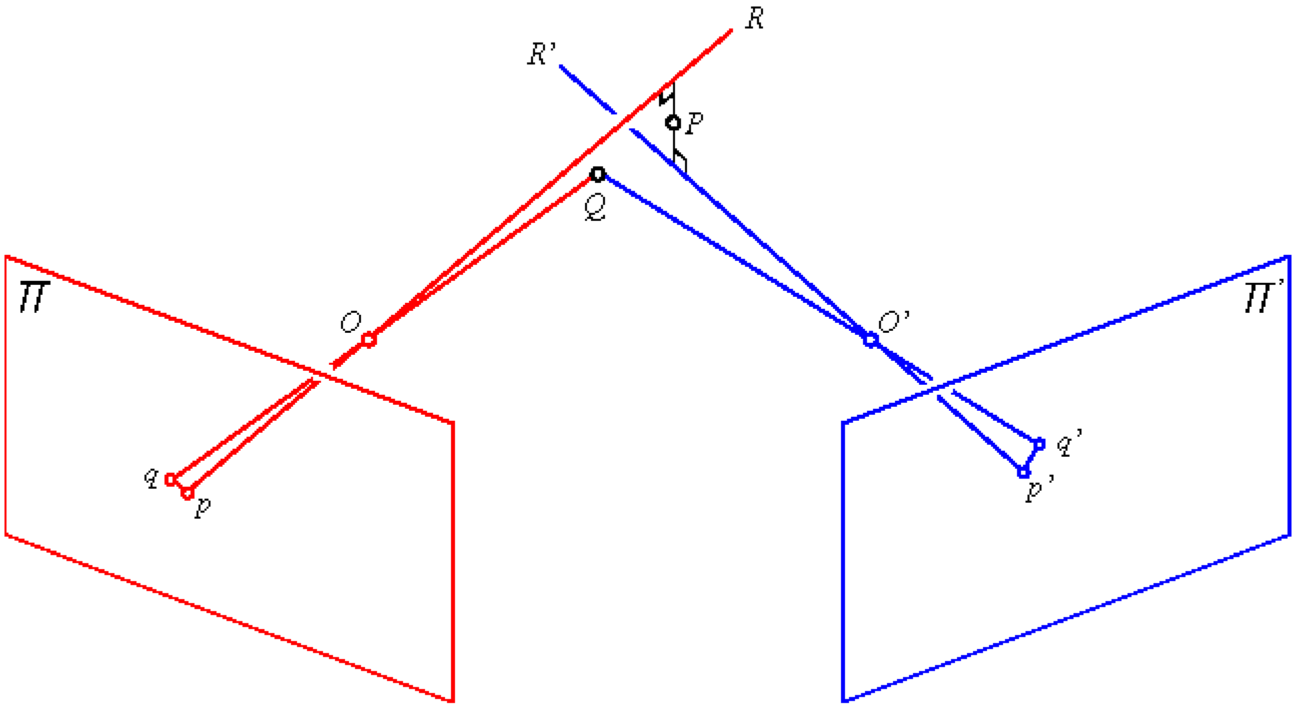
- Feature matching across views
- Calibrated cameras



Last lecture: only need to match features across epipolar lines…

Geometric Reconstruction

Geometric: midpoint of intersection of Op and O'p'



Multiple views?

Algebraic Reconstruction

Point must satisfy both projection equations:

$$z\boldsymbol{p} = \mathcal{M}\boldsymbol{P}$$

$$z'\boldsymbol{p}' = \mathcal{M}\boldsymbol{P}$$

and therefore:

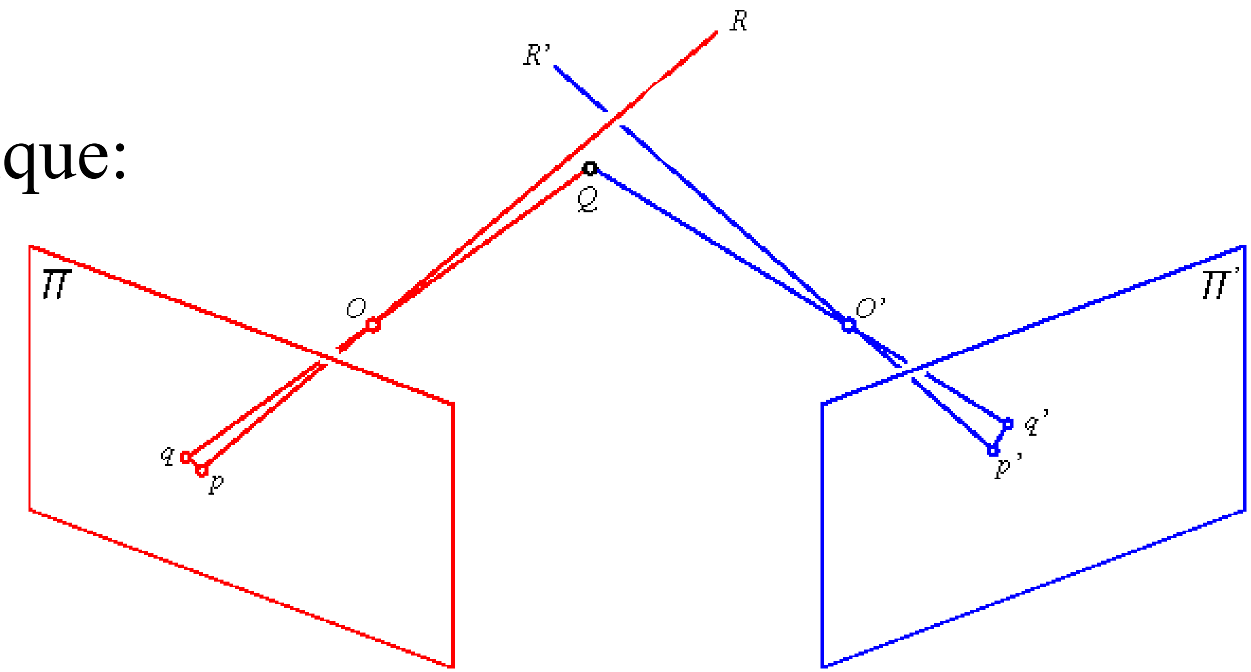
$$\begin{cases} \boldsymbol{p} \times \mathcal{M}\boldsymbol{P} = 0 \\ \boldsymbol{p}' \times \mathcal{M}'\boldsymbol{P} = 0 \end{cases} \iff \begin{pmatrix} [\boldsymbol{p}_{\times}]\mathcal{M} \\ [\boldsymbol{p}'_{\times}]\mathcal{M}' \end{pmatrix} \boldsymbol{P} = 0.$$

solve using least squares.

No obvious geometric interpretation; well-defined for multiple views.

Nonlinear Reconstruction

Nonlinear technique:



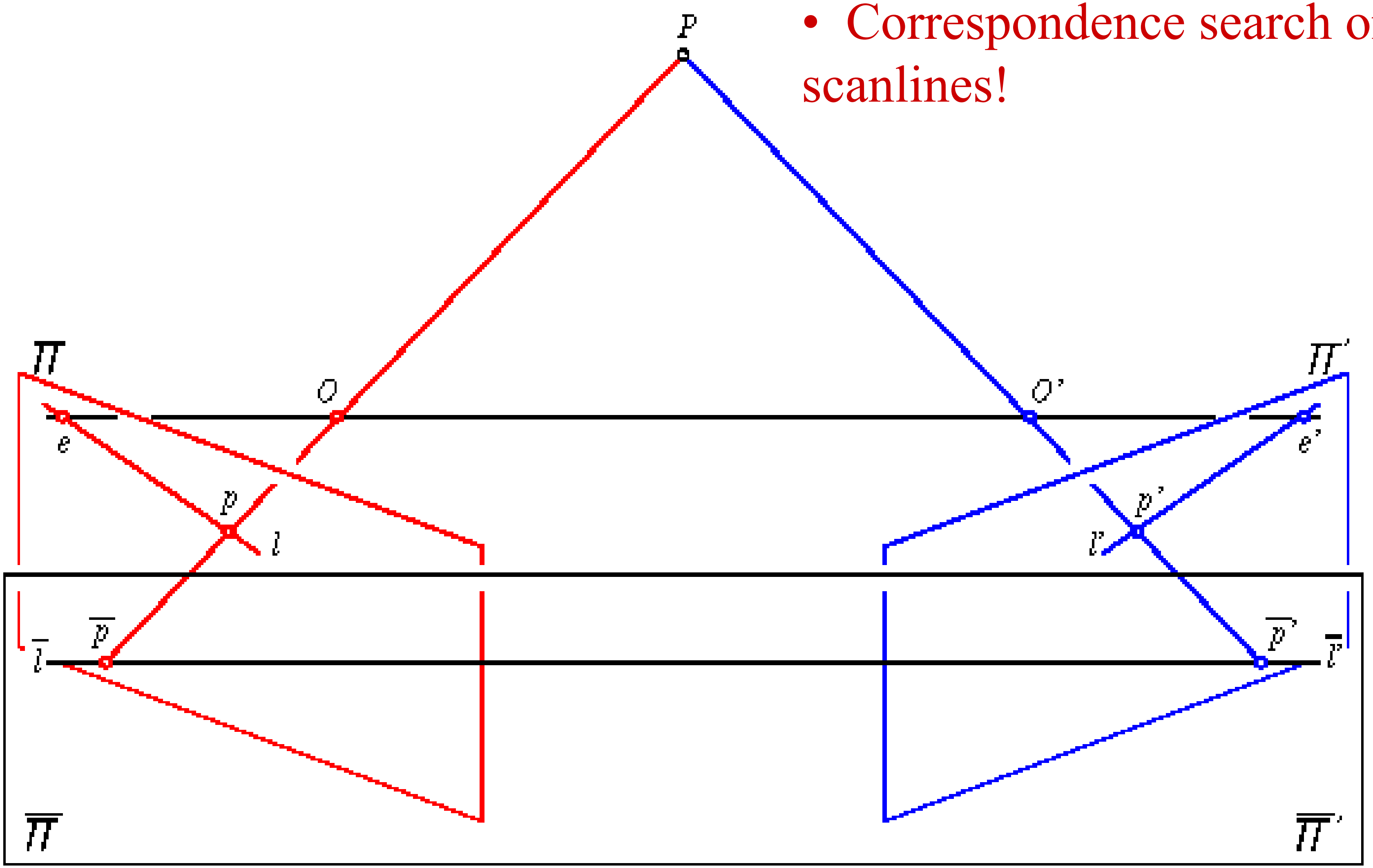
search for point Q with minimal projection error:

$$d^2(p, q) + d^2(p', q')$$

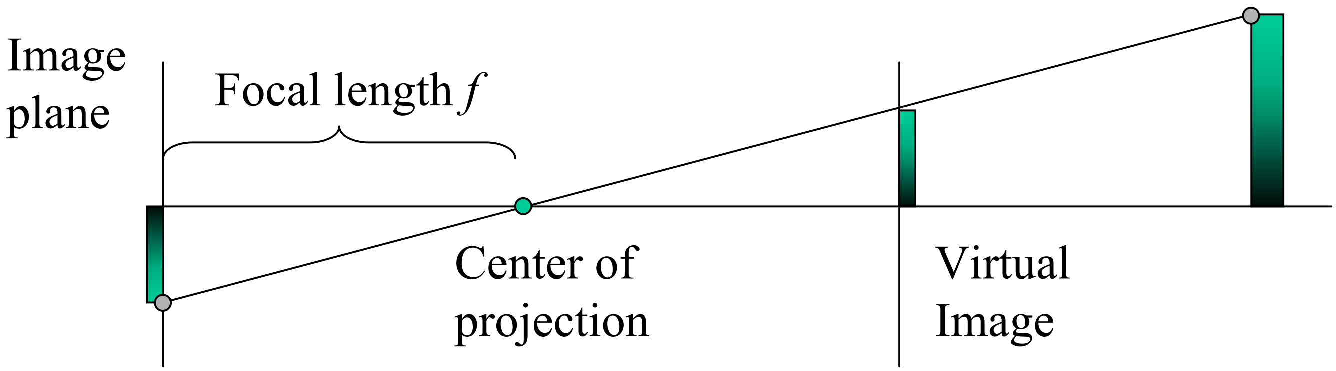
(initialize with one of previous linear methods.)

Rectified image pair

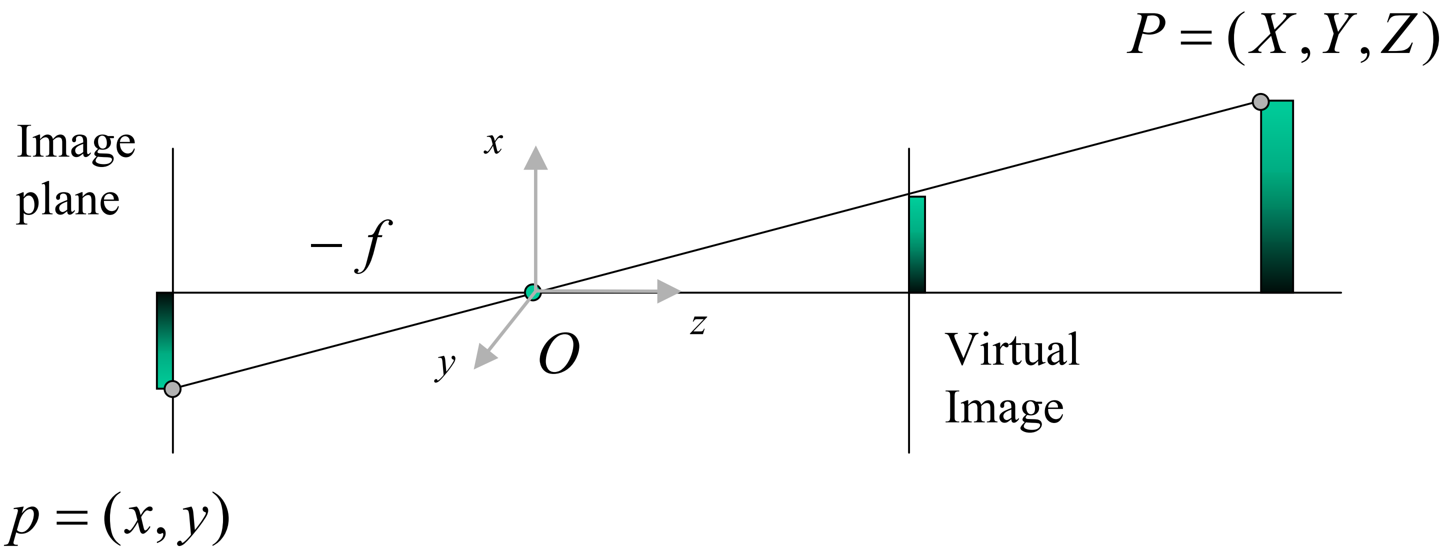
- Choose normalized image locations
- Common image plane
- Horizontal epipolar lines
- Correspondence search on scanlines!



Pinhole Camera Model

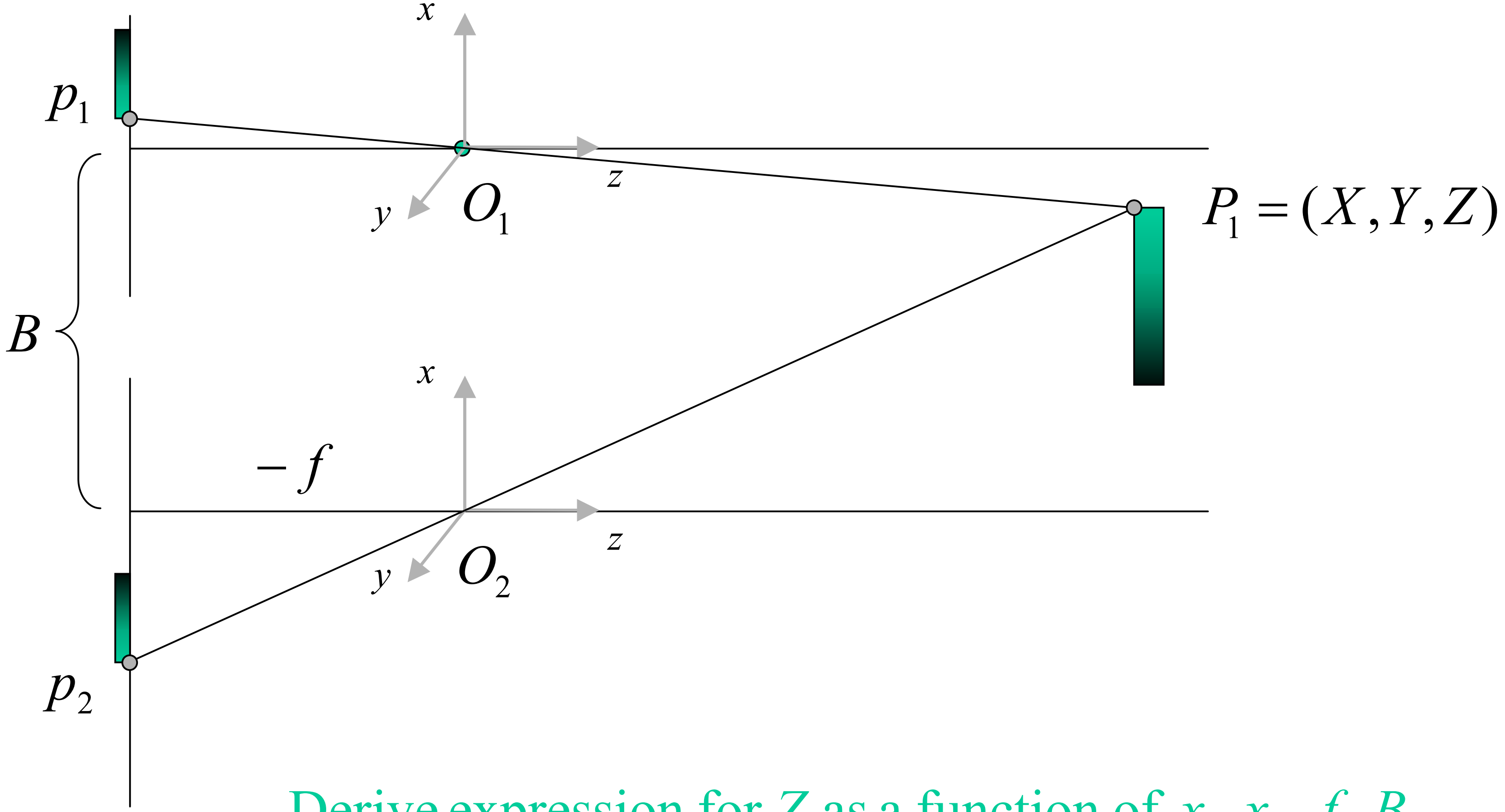


Pinhole Camera Model



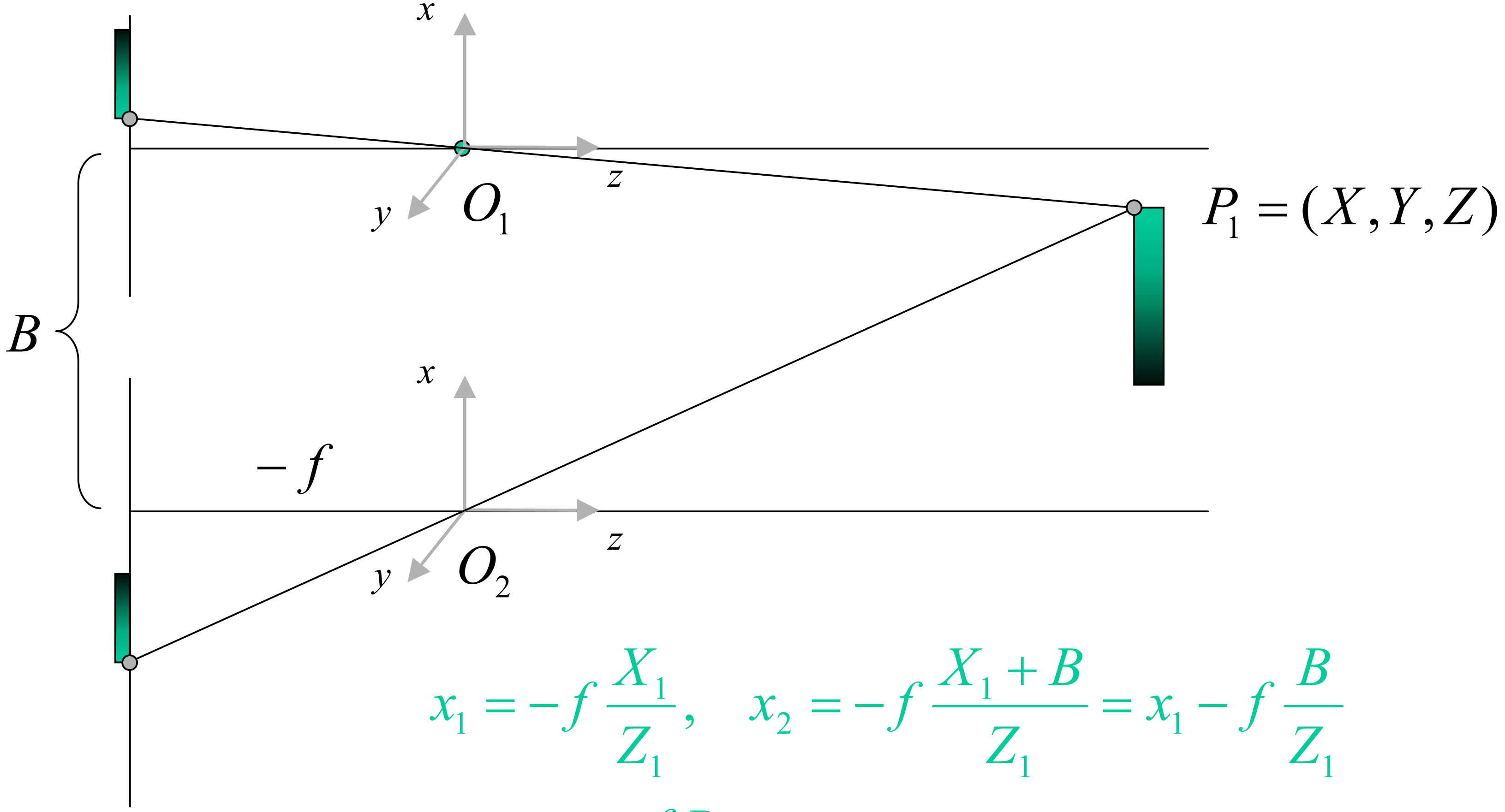
$$x_1 = -f\frac{X_1}{Z_1}$$

Basic Stereo Derivations



Derive expression for Z as a function of x_1, x_2, f, B

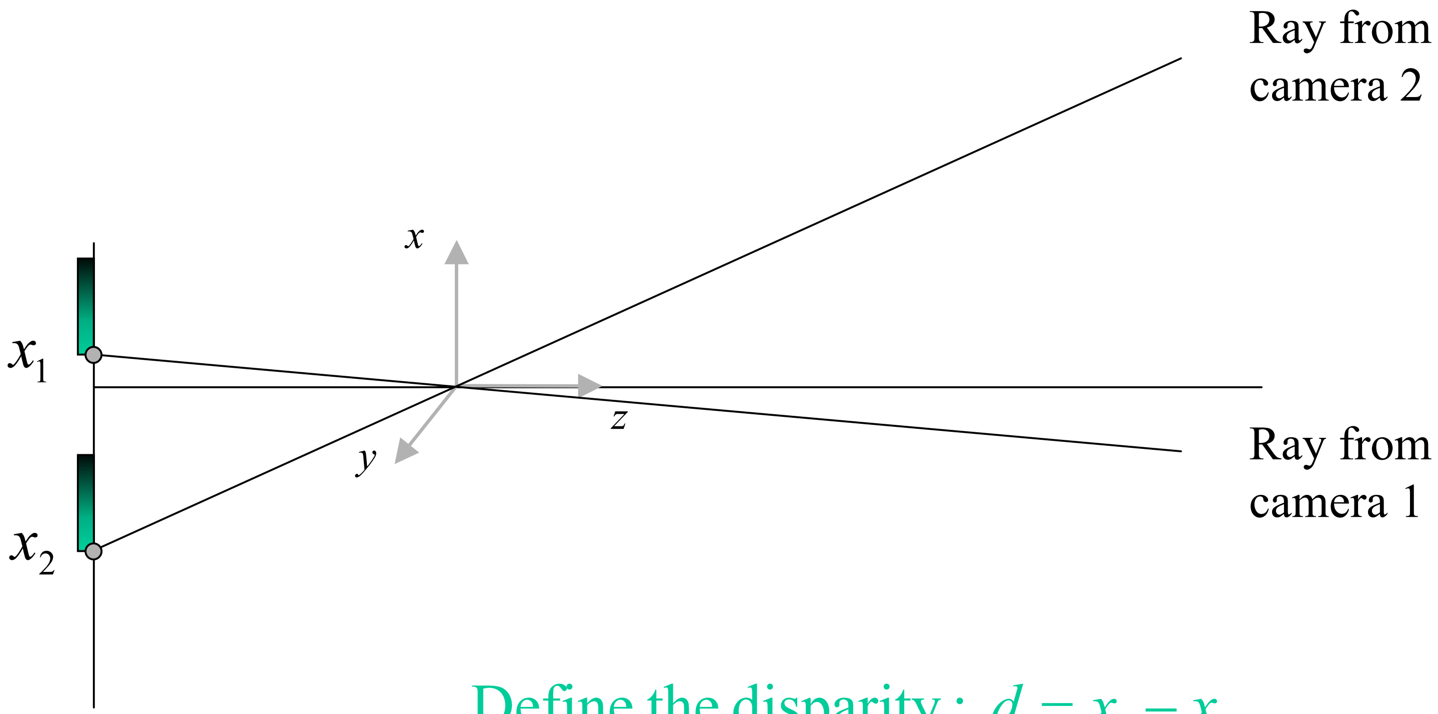
Basic Stereo Derivations



$$x_1 = -f\frac{X_1}{Z_1}, \quad x_2 = -f\frac{X_1 + B}{Z_1} = x_1 - f\frac{B}{Z_1}$$

$$\Rightarrow Z_1 = \frac{f\,B}{x_1 - x_2}$$

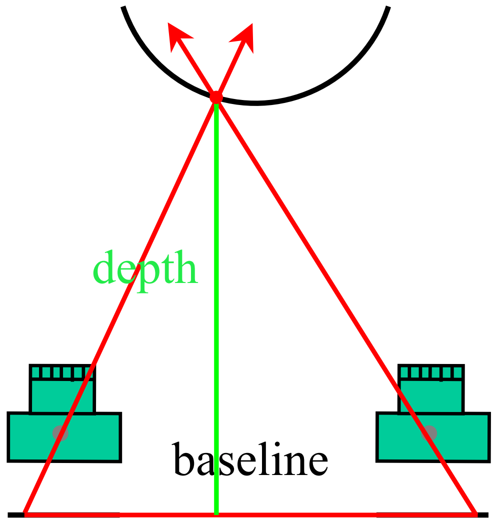
Basic Stereo Derivations



Define the disparity : $d = x_1 - x_2$

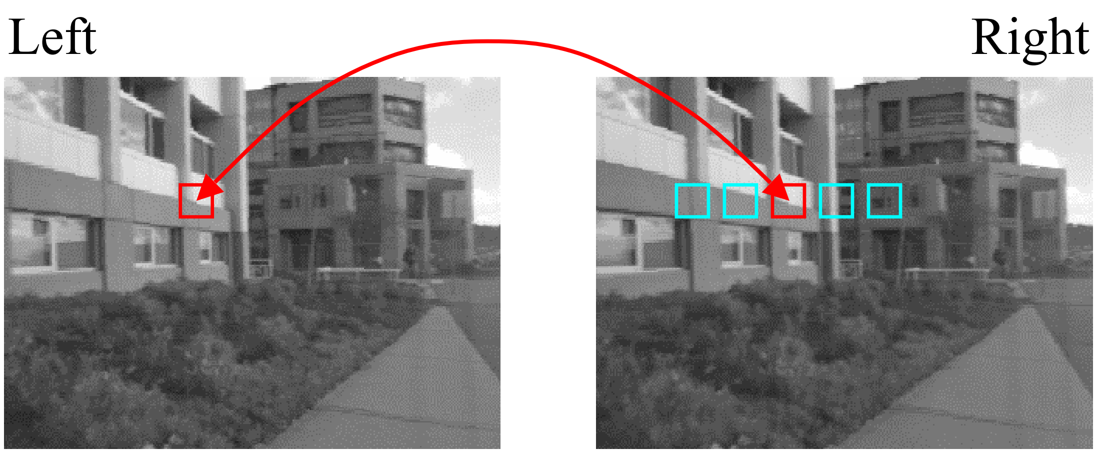
$$Z = \frac{f\,B}{d}$$

Stereo Vision



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

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



Matching across scan lines

How do matching errors affect the depth error?

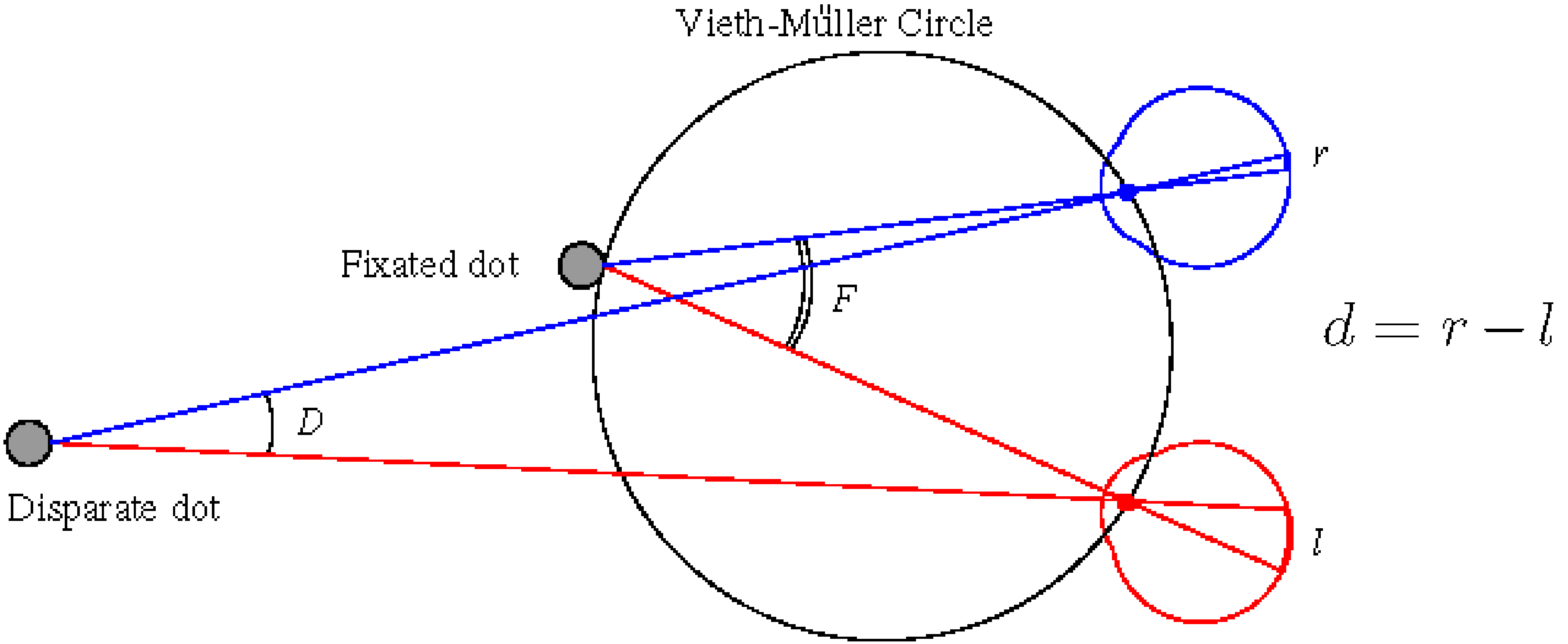
Outline

- ✓ Reconstruction
- ✓ Rectification
- Early vs. Late
- Window Matching
- Edge Matching
- Ordering Constraint
- More views

Consider human vision

Differences?

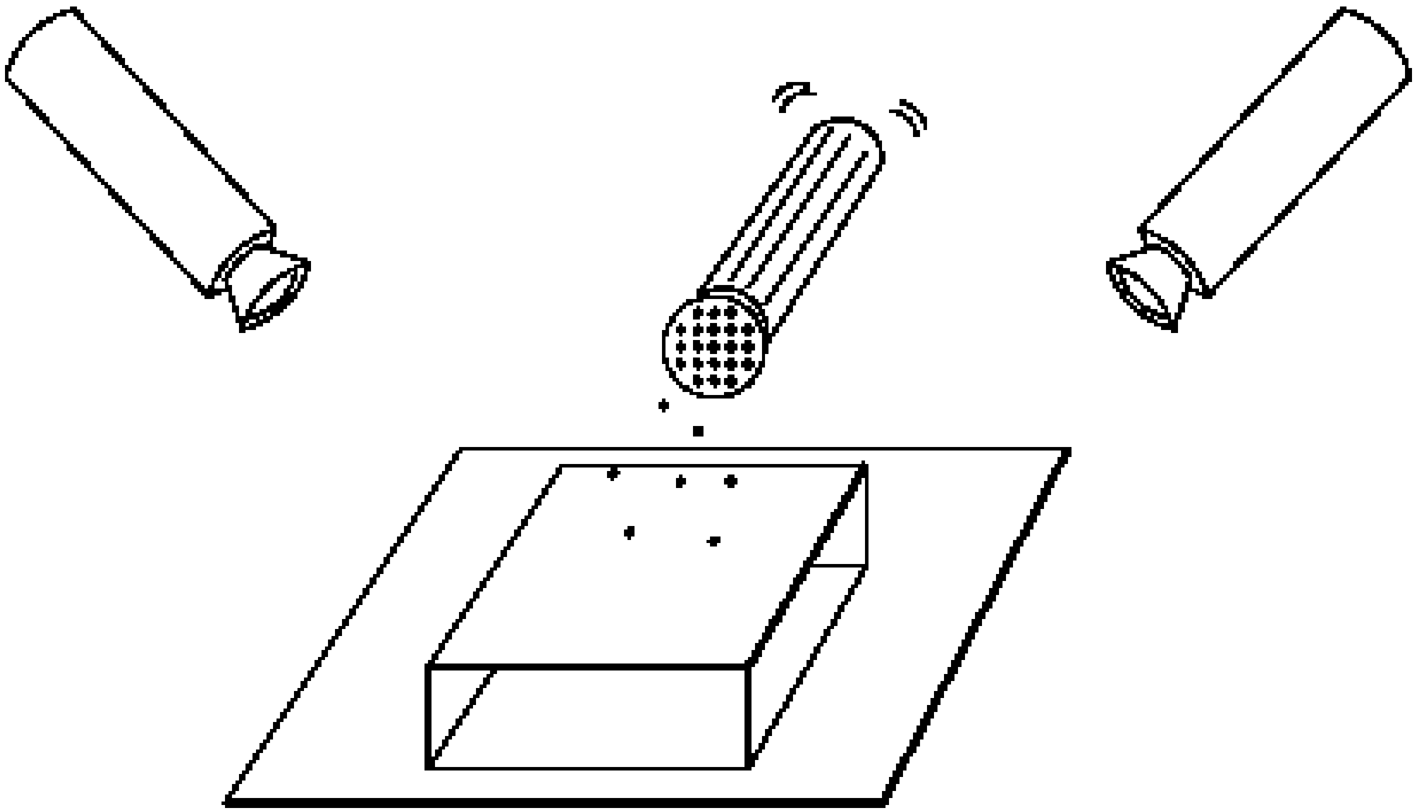
- active foveation
- image plane
- object recognition



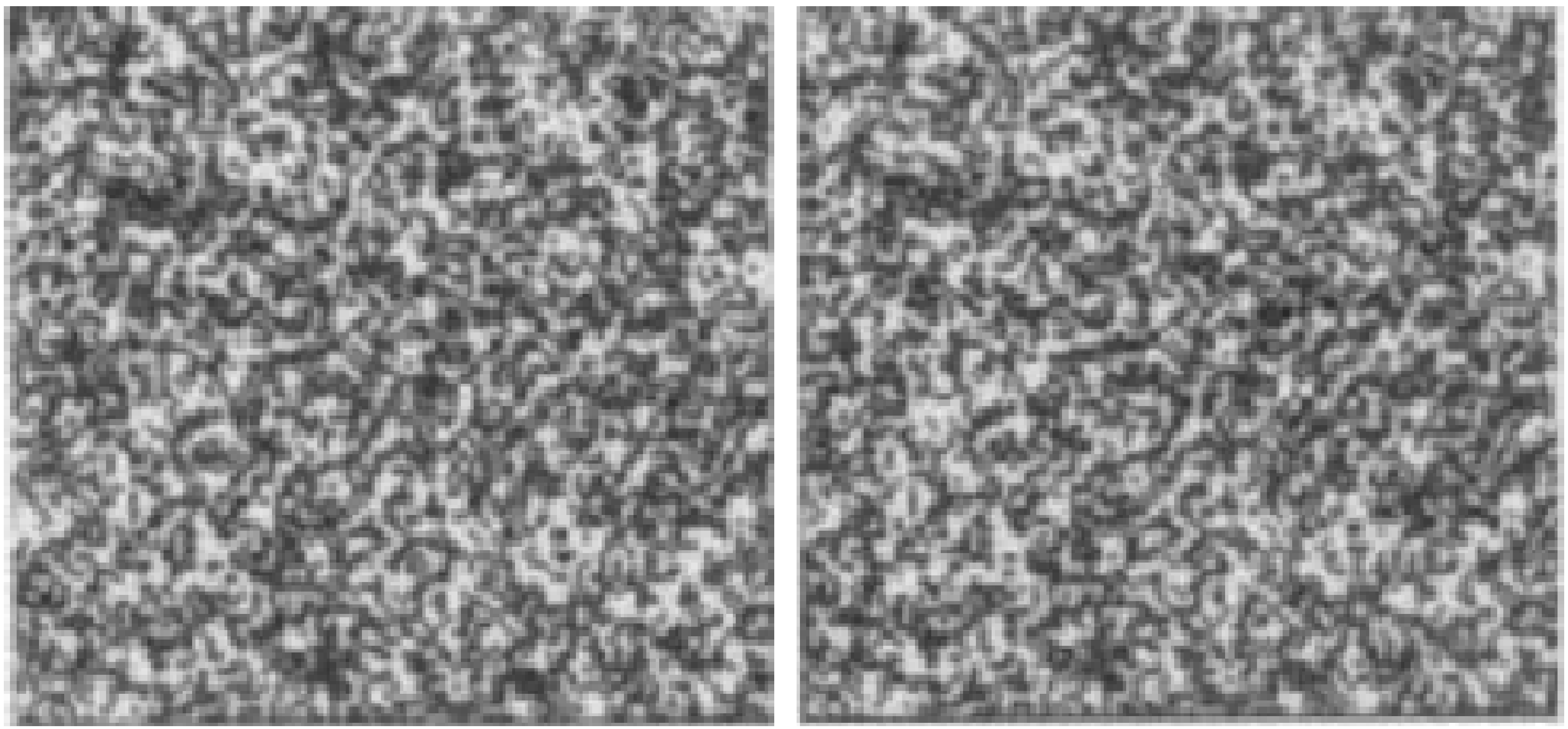
Human vision is very accurate at relative depth judgements.

Julez (1960): which happens first, recognition or fusion?

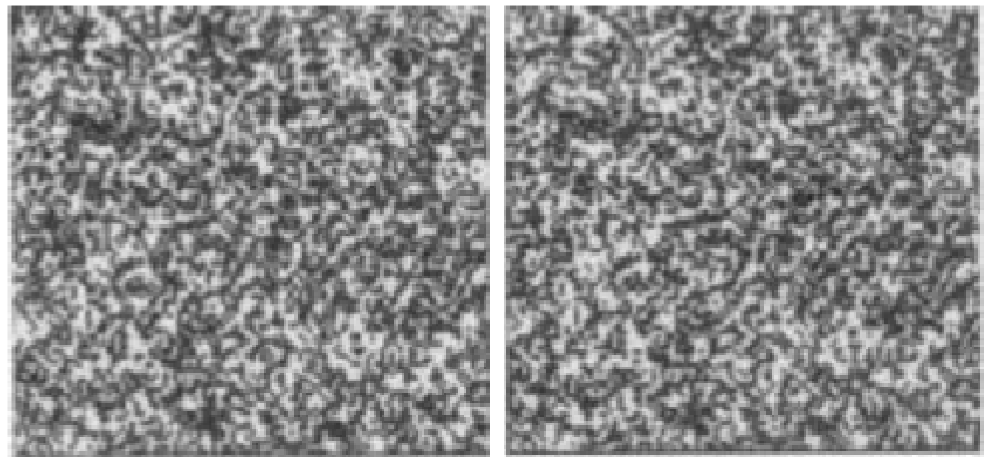
Random dot stereograms

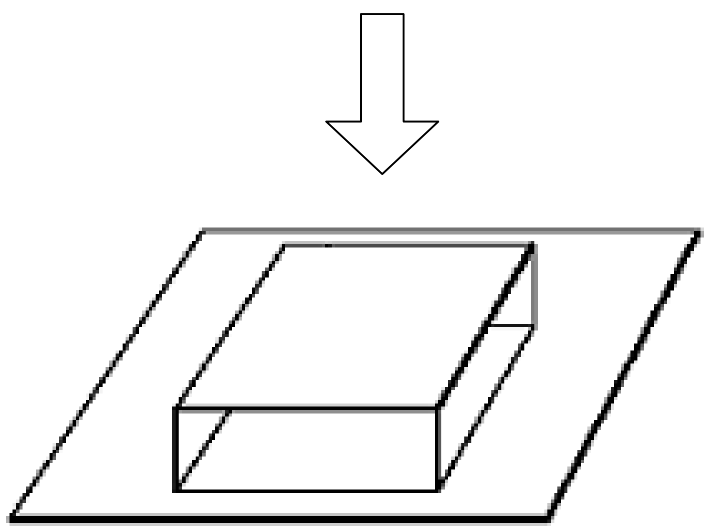


Random dot stereograms

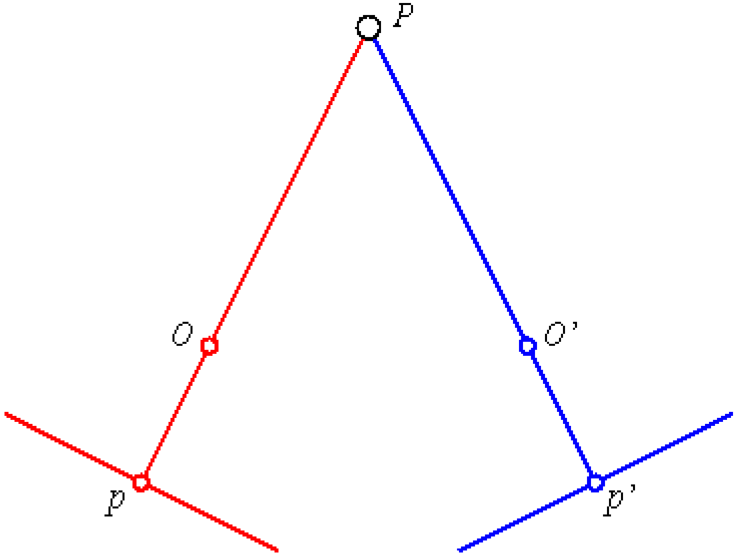


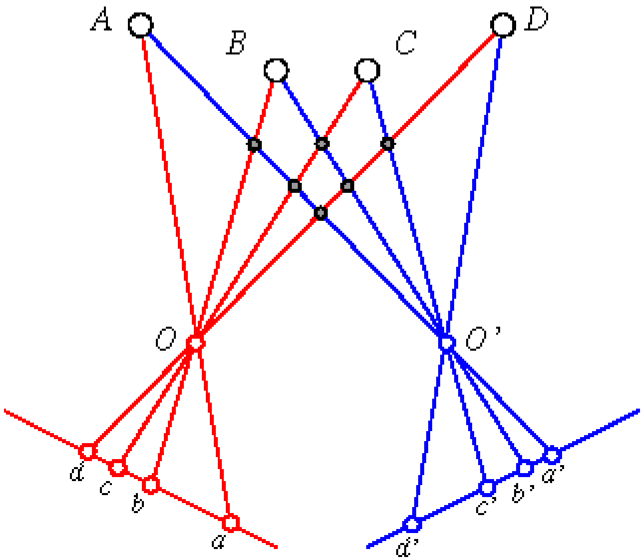
Random dot stereograms



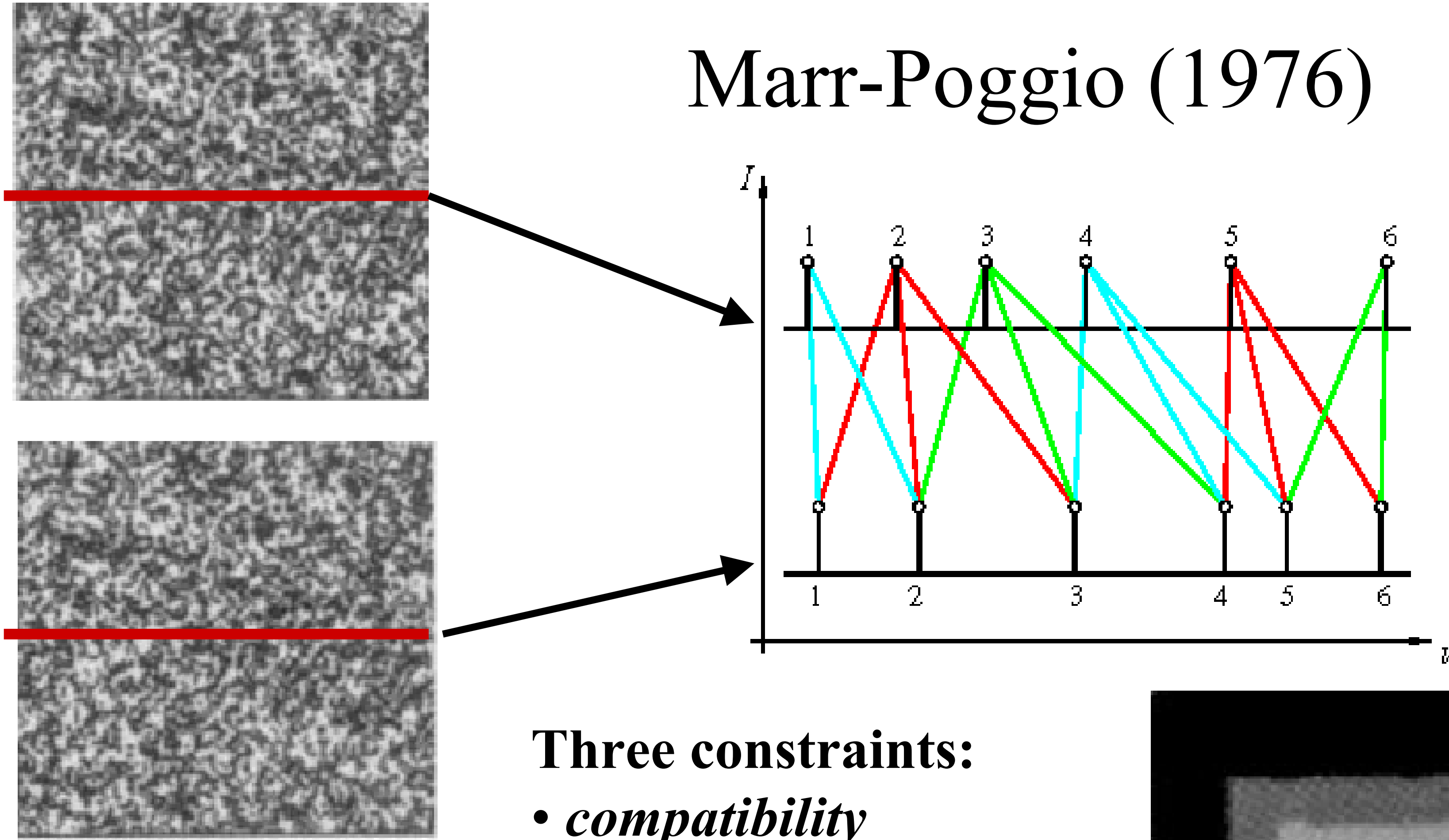


Correspondence is ambiguous



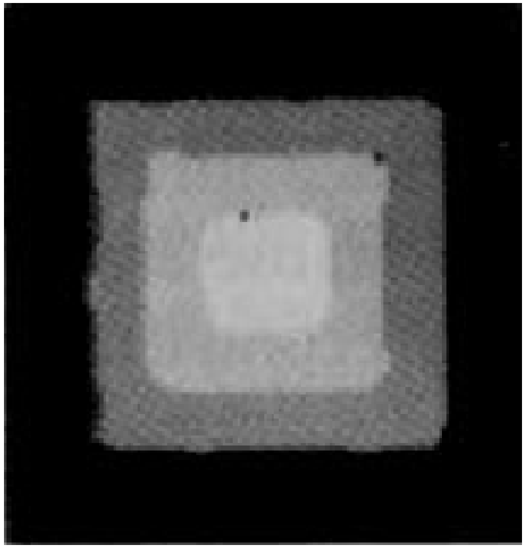


Marr-Poggio (1976)



Three constraints:

- ***compatibility***
- ***uniqueness***
- ***continuity***



Works well on RDS….but not so well on natural images…

Outline

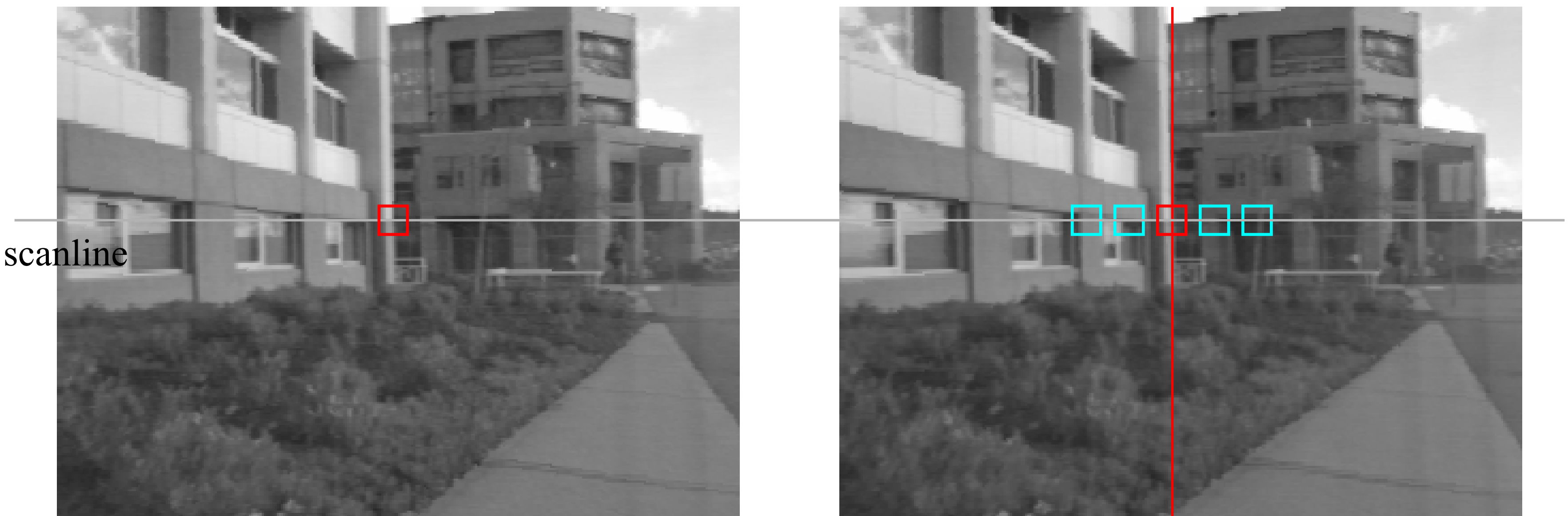
- ✓ Reconstruction
- ✓ Rectification
- ✓ Early vs. Late
- Window Matching
- Edge Matching
- Ordering Constraint
- More views

Correspondence using window matching

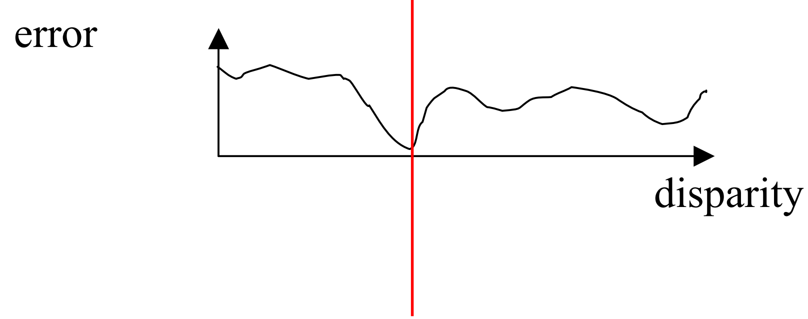
Points are highly individually ambiguous…

More unique matches are possible with small regions of image.

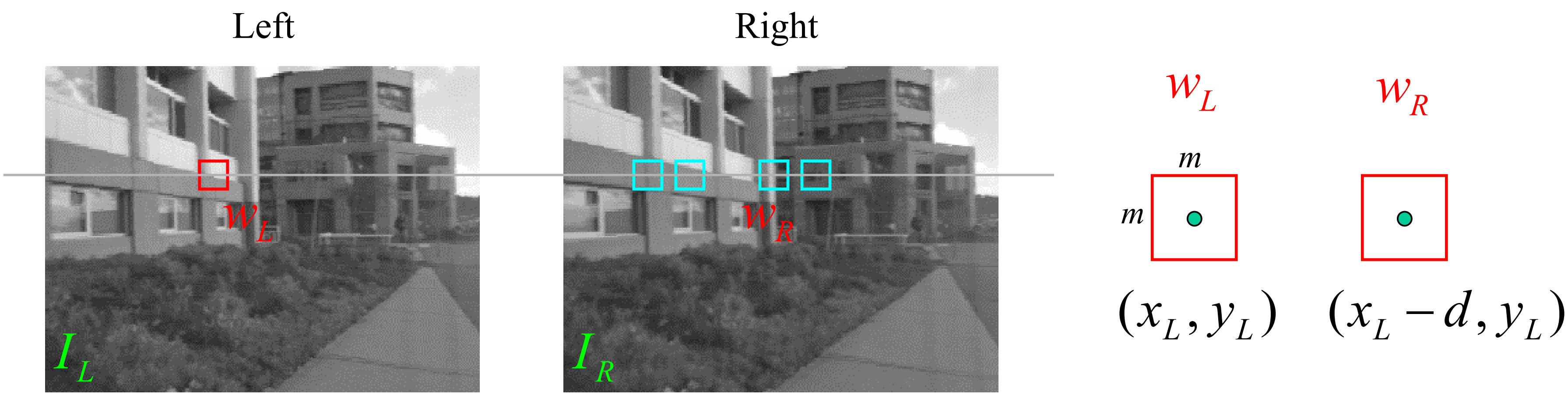
Correspondence using window matching



Criterion function:



Sum of Squared (Pixel) Differences



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

We define the window function :

$$W_m(x, y) = \{u, v \mid x - \tfrac{m}{2} \le u \le x + \tfrac{m}{2}, y - \tfrac{m}{2} \le v \le y + \tfrac{m}{2}\}$$

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

$$C_r(x, y, d) = \sum_{(u,v) \in W_m(x,y)} [I_L(u, v) - I_R(u - d, v)]^2$$

Image Normalization

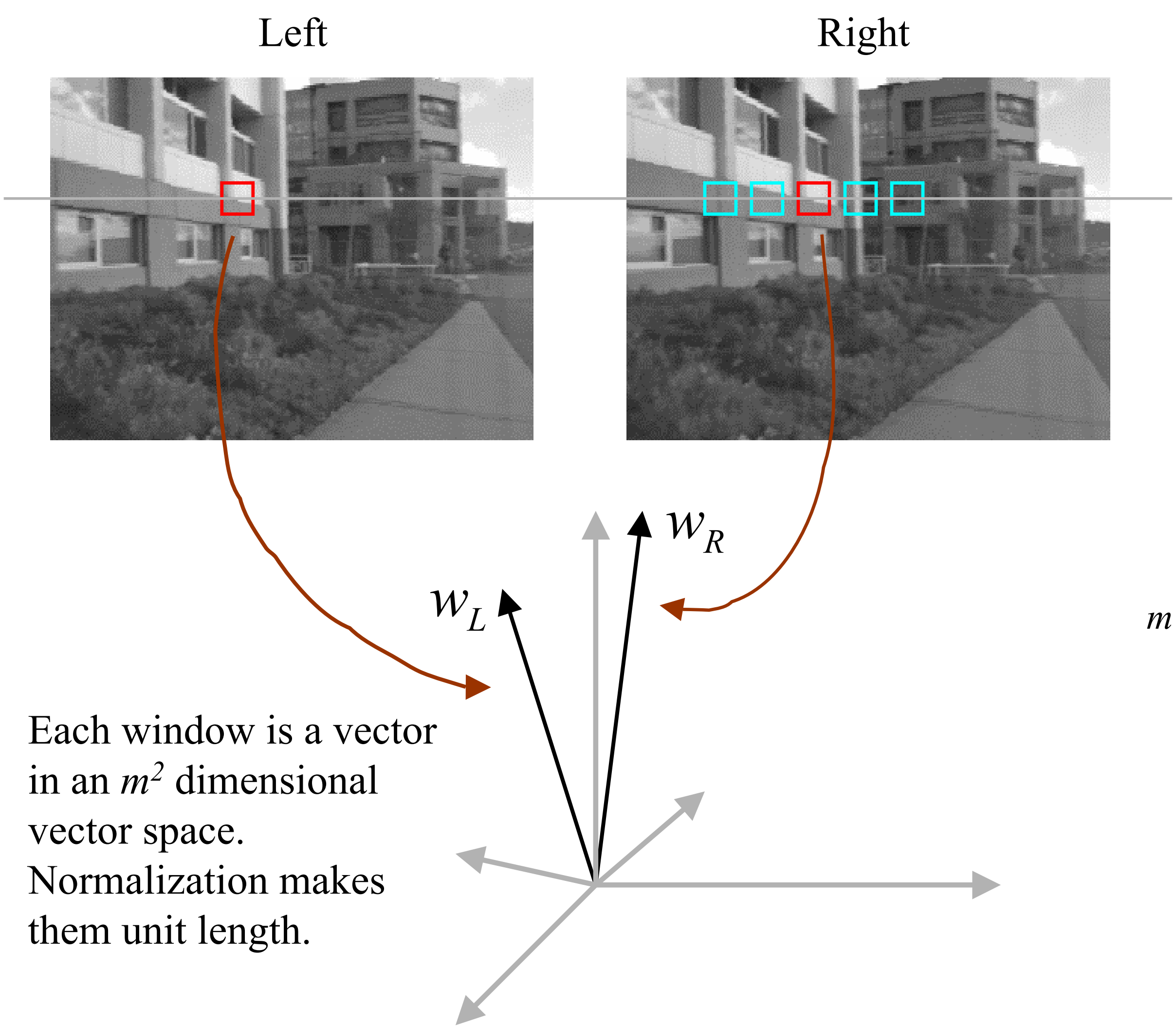
- Even when the cameras are identical models, there can be differences in gain and sensitivity.
- The cameras do not see exactly the same surfaces, so their overall light levels can differ.
- For these reasons and more, it is a good idea to normalize the pixels in each window:

$$\bar{I} = \frac{1}{|W_m(x,y)|} \sum_{(u,v)\in W_m(x,y)} I(u,v)$$ Average pixel

$$\|I\|_{W_m(x,y)} = \sqrt{\sum_{(u,v)\in W_m(x,y)} [I(u,v)]^2}$$ Window magnitude

$$\hat{I}(x,y) = \frac{I(x,y) - \bar{I}}{\|I - \bar{I}\|_{W_m(x,y)}}$$ Normalized pixel

Images as Vectors



Each window is a vector in an m^2 dimensional vector space. Normalization makes them unit length.

"Unwrap" image to form vector, using raster scan order

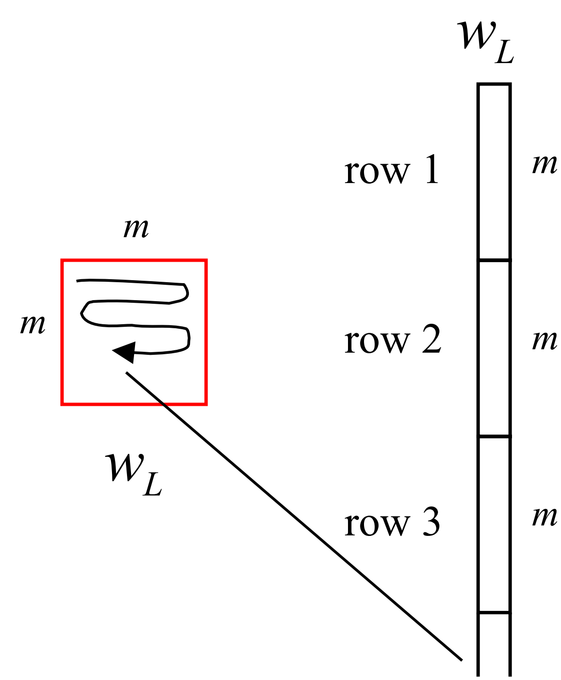
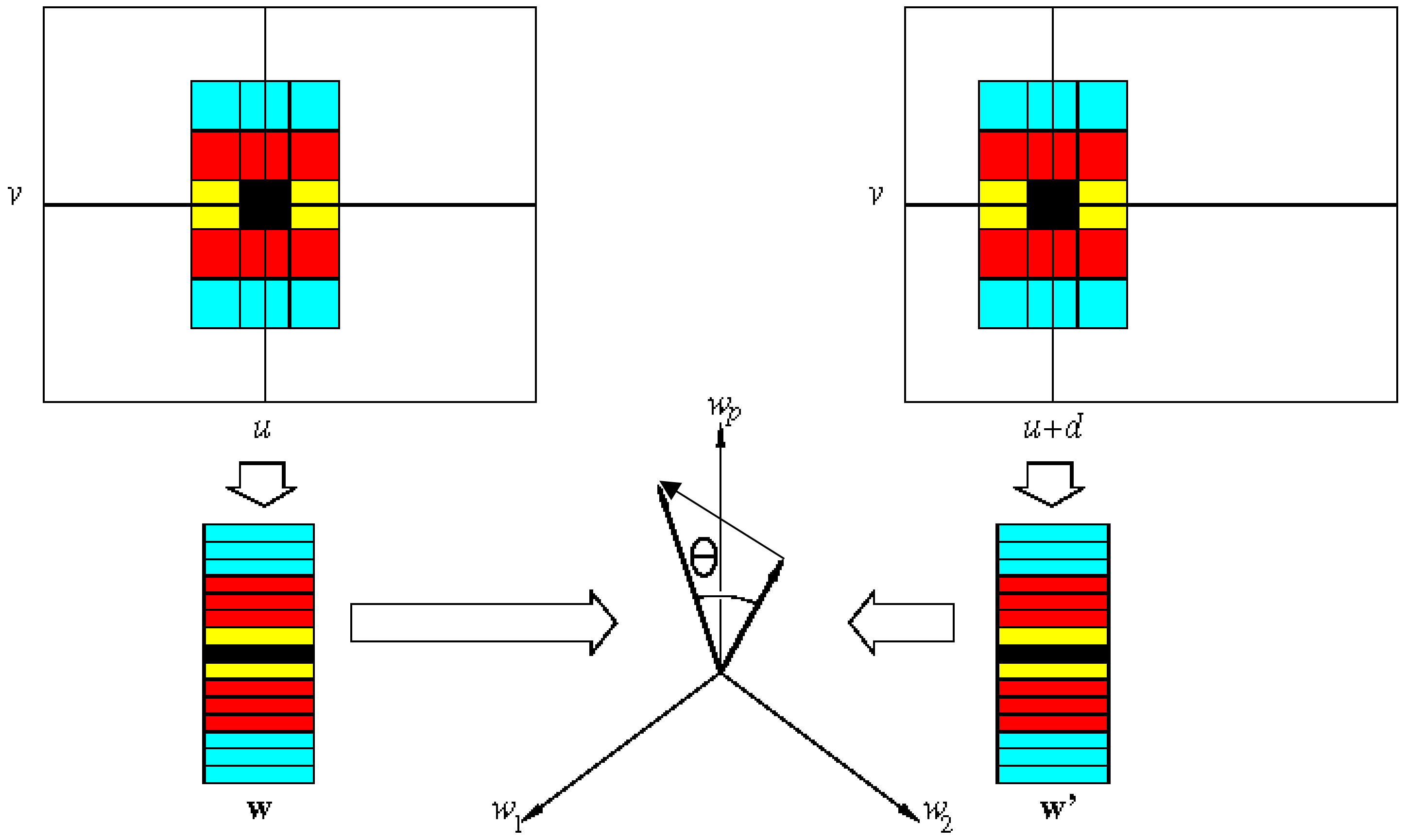


Image windows as vectors



Possible metrics

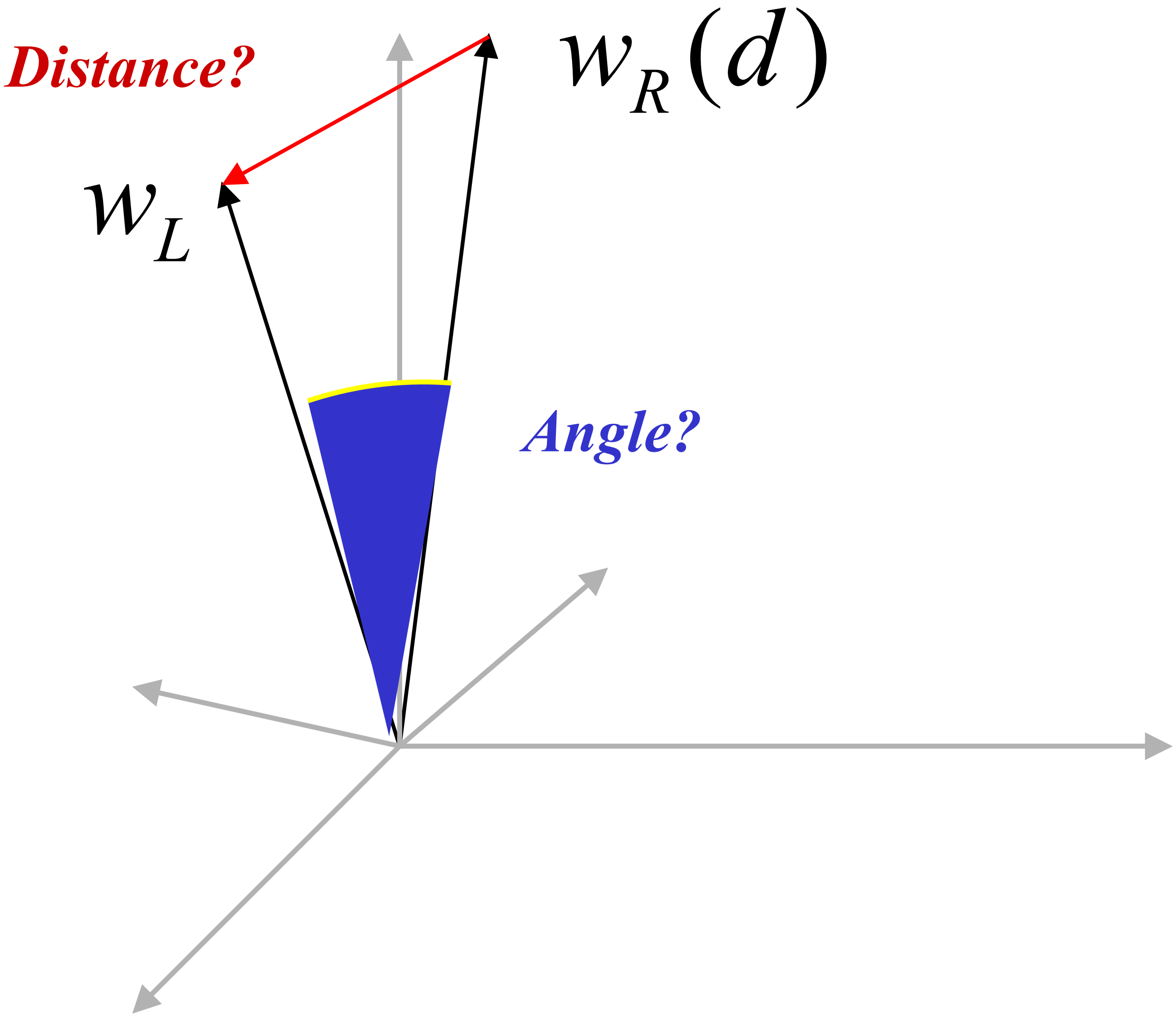
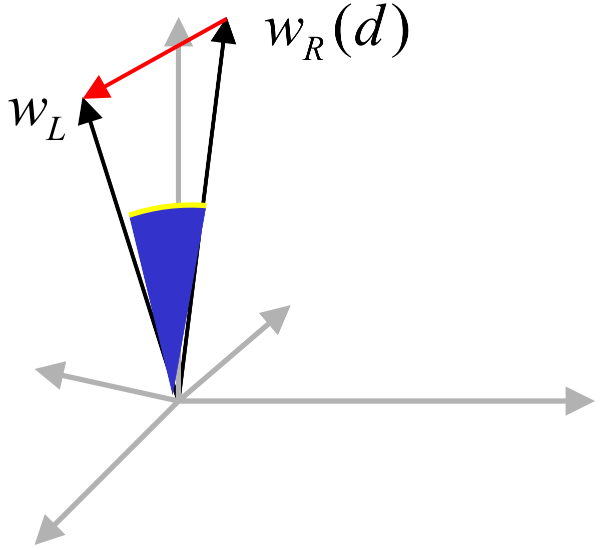


Image Metrics



(Normalized) Sum of Squared Differences

$$C_{\mathrm{SSD}}(d) = \sum_{(u,v)\in W_m(x,y)} [\hat{I}_L(u,v) - \hat{I}_R(u-d,v)]^2$$

$$= \left\| w_L - w_R(d) \right\|^2$$

Normalized Correlation

$$C_{\mathrm{NC}}(d) = \sum_{(u,v)\in W_m(x,y)} \hat{I}_L(u,v)\hat{I}_R(u-d,v)$$

$$= w_L \cdot w_R(d) = \cos\theta$$

$$d^* = \arg\min_d \left\| w_L - w_R(d) \right\|^2 = \arg\max_d w_L \cdot w_R(d)$$

Correspondence Using Correlation

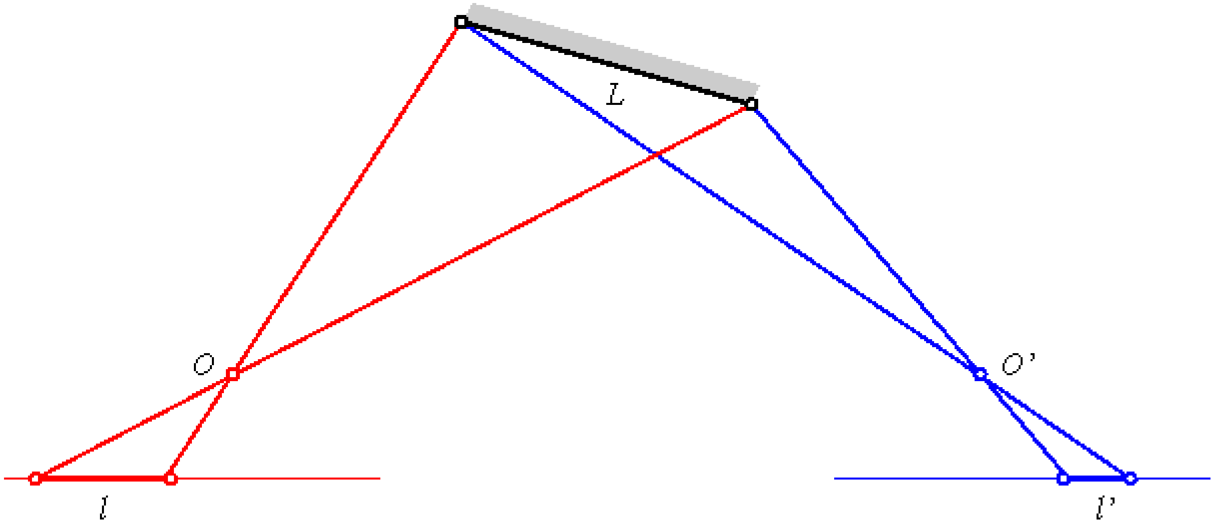
Left

Disparity Map

Images courtesy of Point Grey Research

Foreshortening

Window methods assume fronto-parallel surface at 3-D point.



Initial estimates of the disparity can be used to warp the correlation windows to compensate for unequal amounts of foreshortening in the two pictures [Kass, 1987; Devernay and Faugeras, 1994].

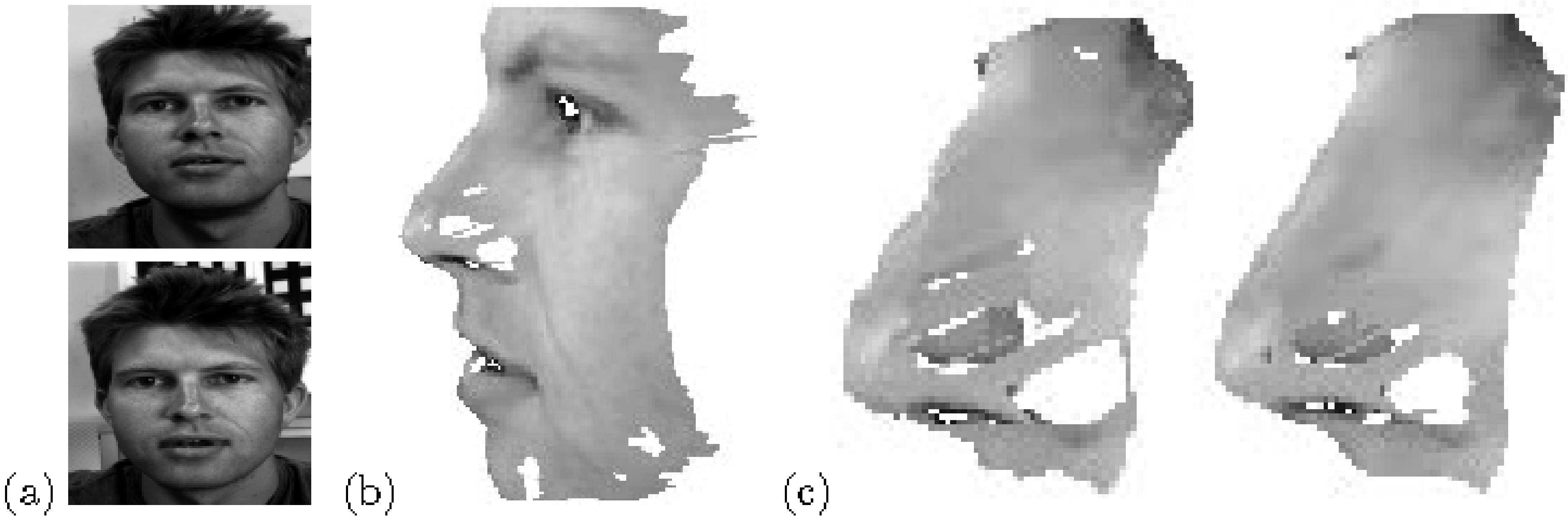


FIGURE 12.13: Correlation-based stereo matching: (a) a pair of stereo pictures; (b) a texture-mapped view of the reconstructed surface; (c) comparison of the regular (left) and refined (right) correlation methods in the nose region. Reprinted from [Devernay and Faugeras, 1994], Figures 5, 8 and 9.

Outline

- ✓ Reconstruction
- ✓ Rectification
- ✓ Early vs. Late
- ✓ Window Matching
- Edge Matching
- Ordering Constraint
- More views

Problems with window methods

Patch too small?

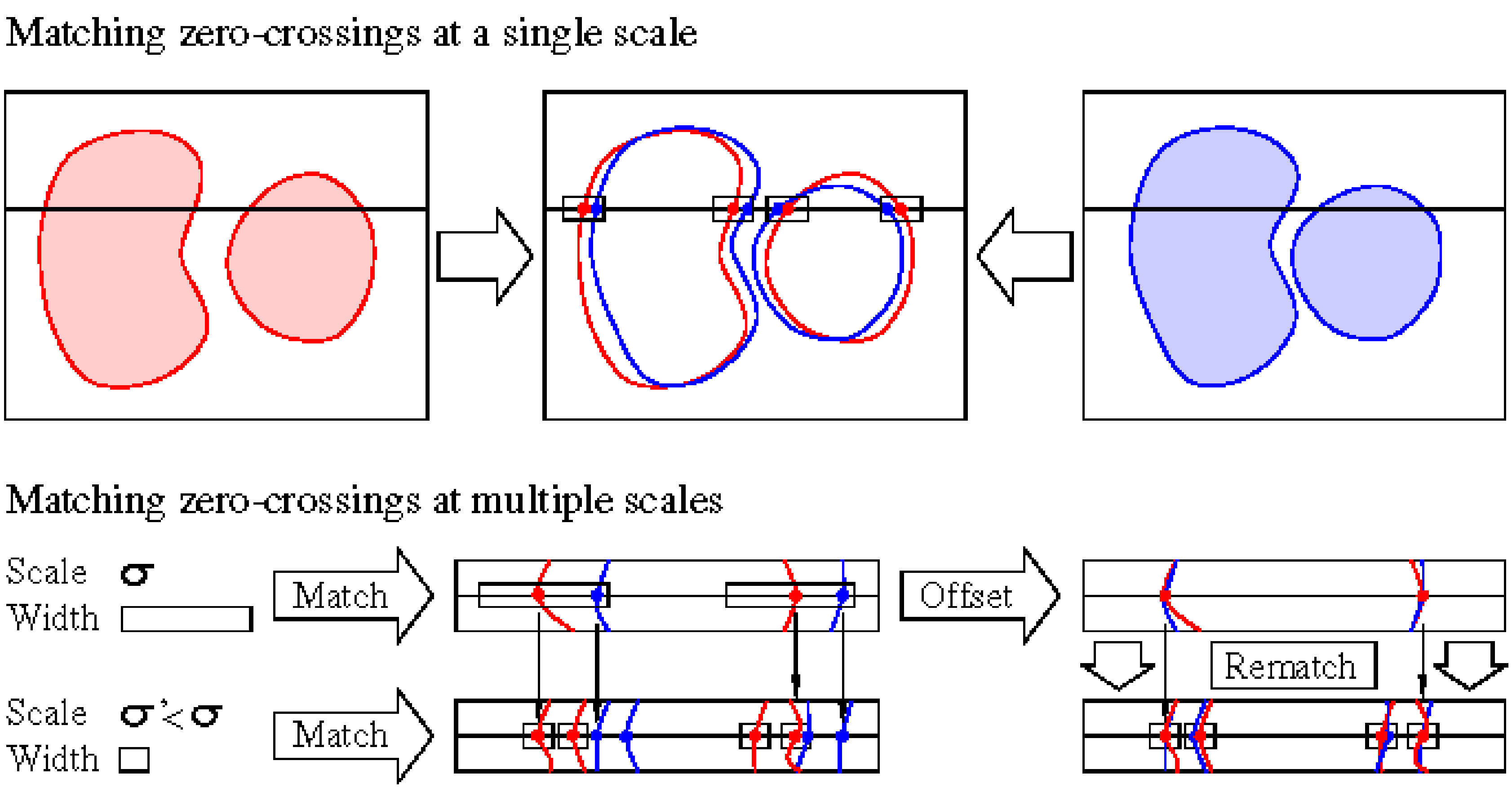
Patch too large?

Can try variable patch size [Okutomi and Kanade], or arbitrary window shapes [Veksler and Zabih]

Should match between physically meaningful quanties, and at multiple scales [Marr]…

Marr-Poggio algorithm

Search for edges, a.k.a. "zero crossings": (more during edge detection lectures…)

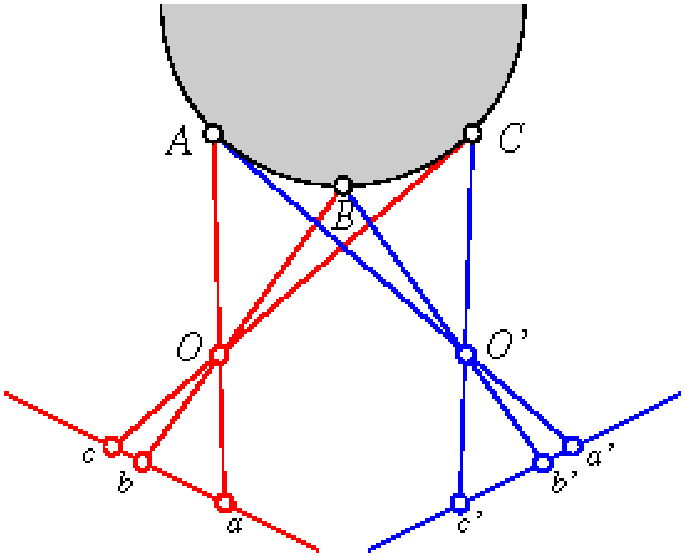


Ordering constraint

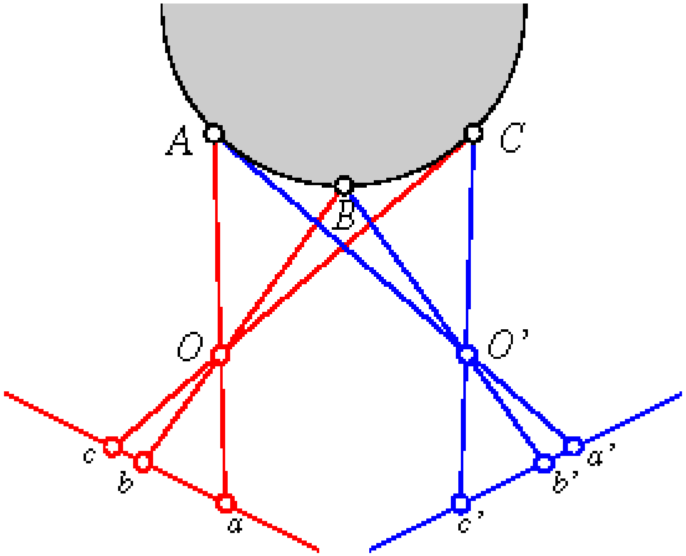
"It is reasonable to assume that the order of matching image features along a pair of epipolar lines is the inverse of the order of the corresponding surface attributes along the curve where the epipolar plane intersects the observed object's boundary."

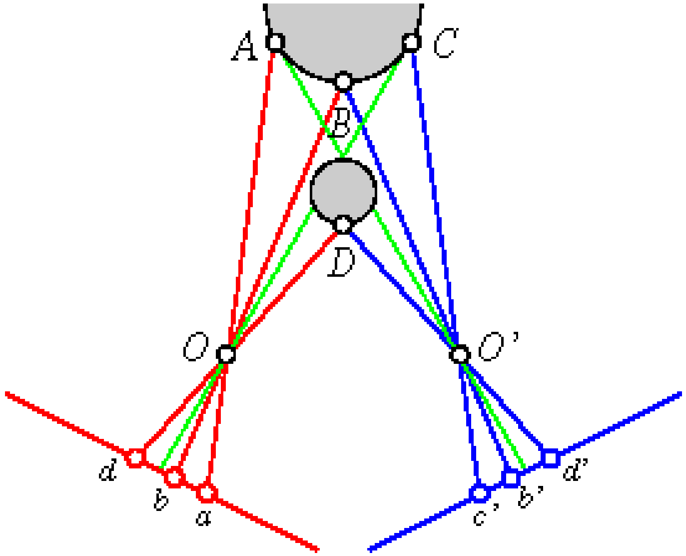
This is the so-called *ordering constraint* introduced by [Baker and Binford, 1981; Ohta and Kanade, 1985].

Ordering constraint



Ordering constraint



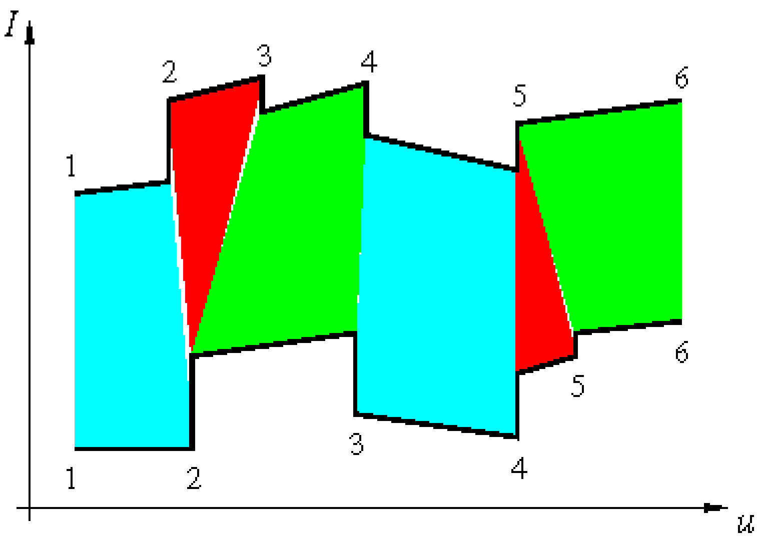


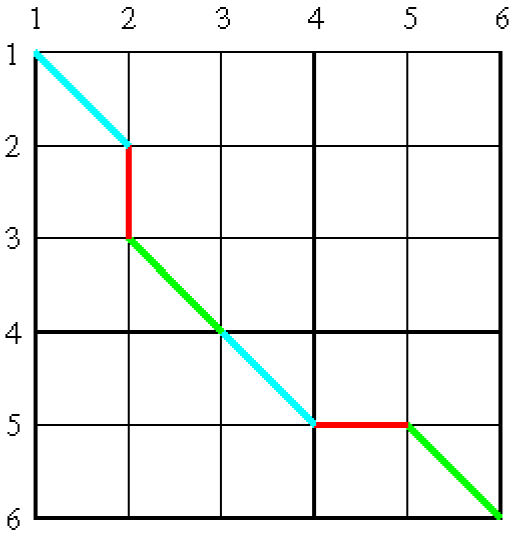
Oops!

DP-based search

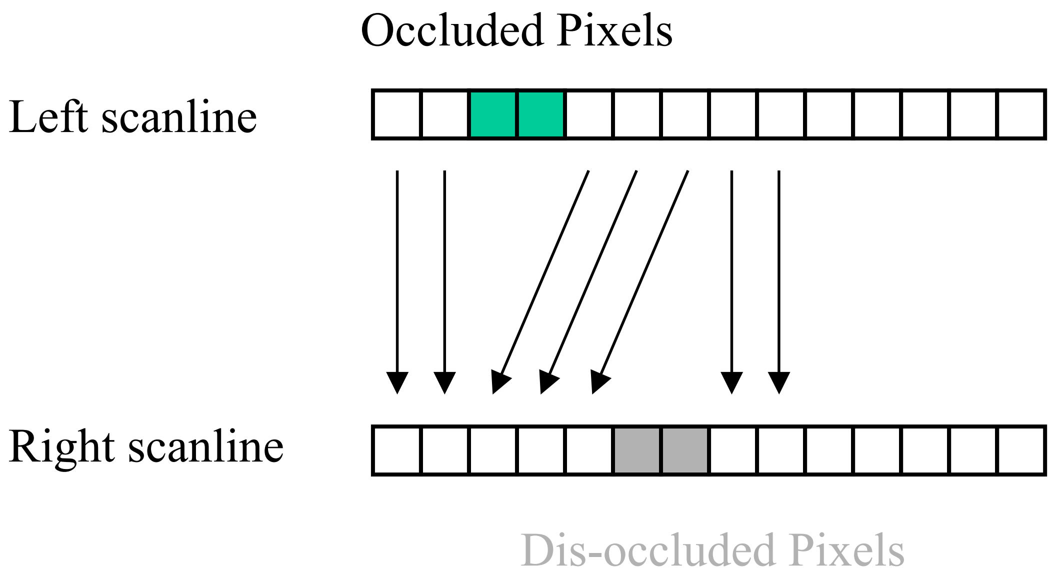
Ordering constraint == smooth path through match graph.

Consider path's cost over a graph whose nodes correspond to pairs of left and right image features, and arcs represent matches between left and right intensity profile intervals bounded by the features of the corresponding nodes





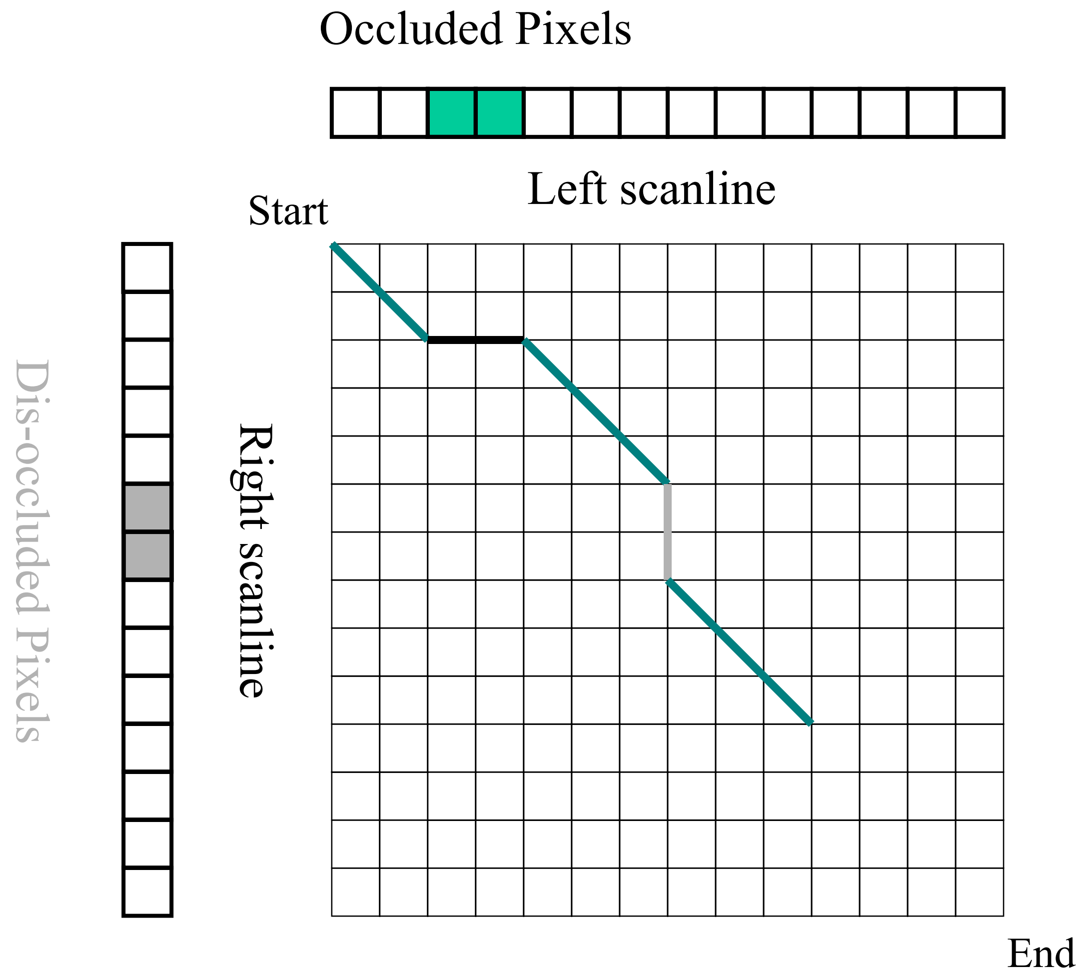
Search Over Correspondences



Three cases:

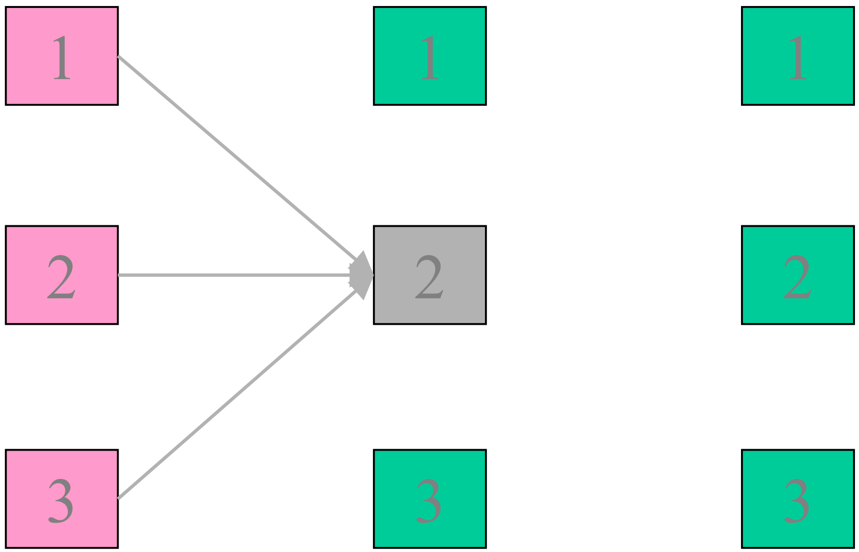
- Sequential – cost of match
- Occluded – cost of no match
- Disoccluded – cost of no match

Dynamic Programming



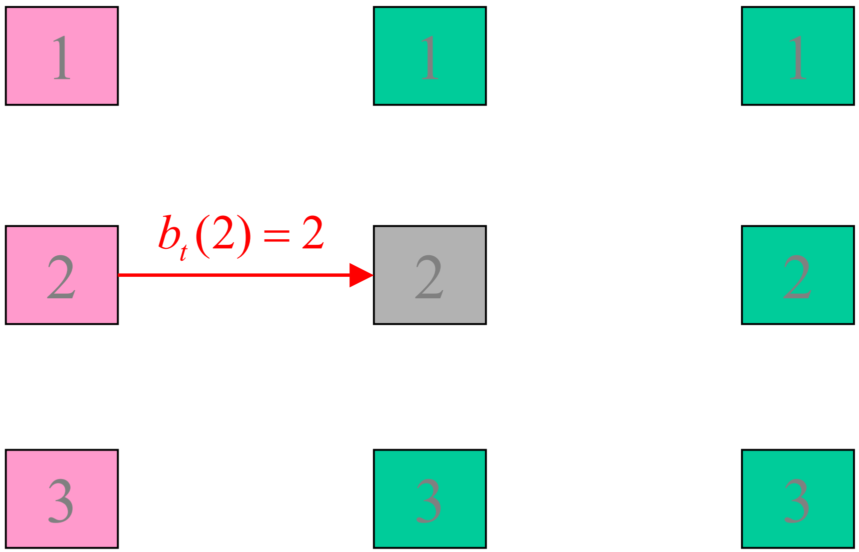
Dynamic programming yields the optimal path through grid. This is the best set of matches that satisfy the ordering constraint

Dynamic Programming



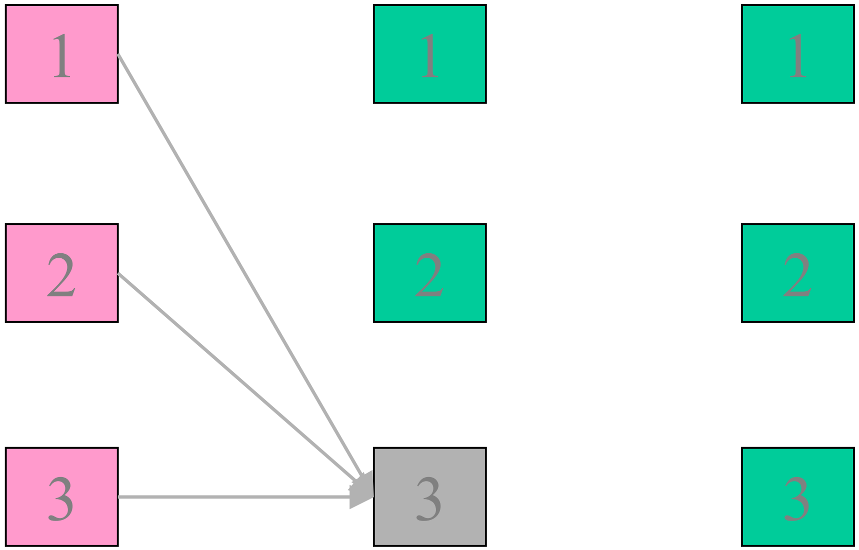
Principle of Optimality for an n-stage assignment problem:

Dynamic Programming



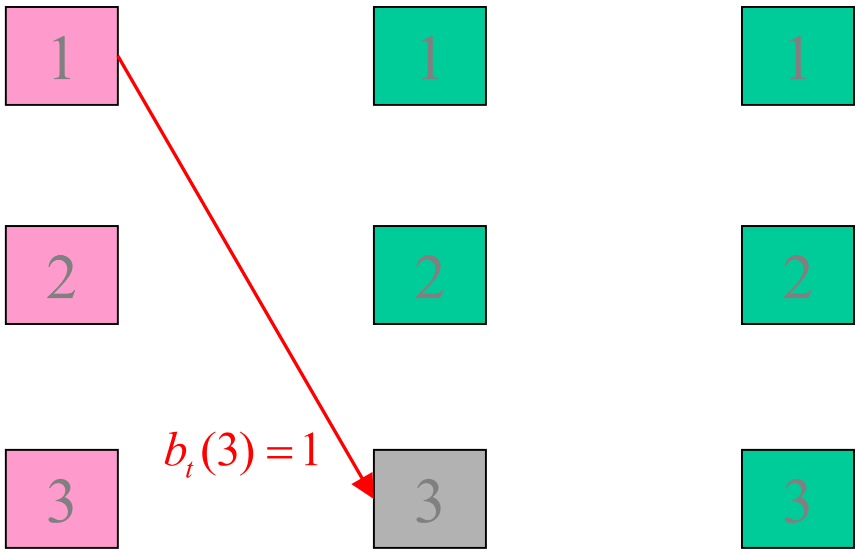
Principle of Optimality for an n-stage assignment problem:

Dynamic Programming



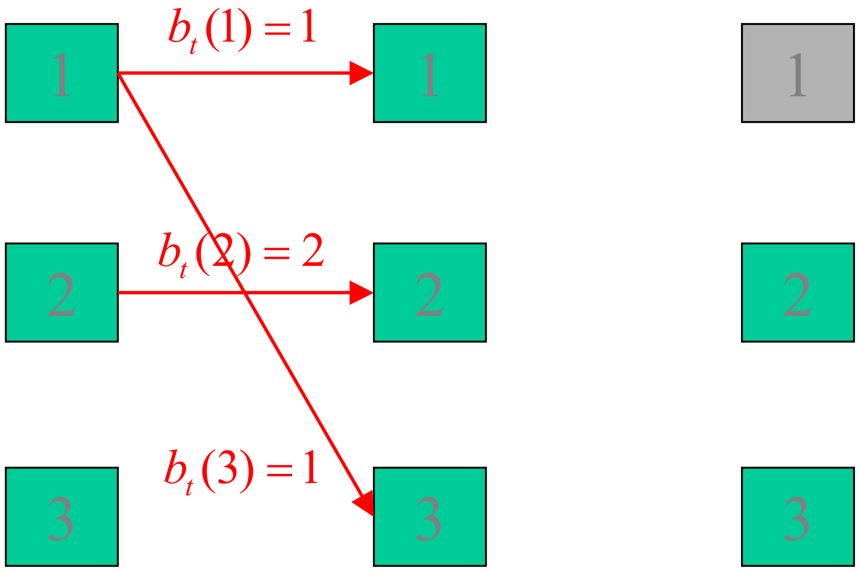
Principle of Optimality for an n-stage assignment problem:

Dynamic Programming



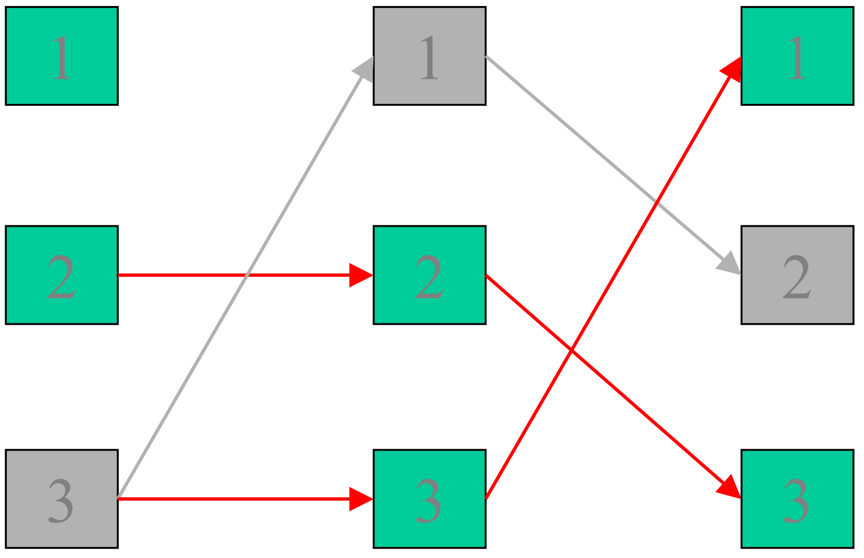
Principle of Optimality for an n-stage assignment problem:

Dynamic Programming



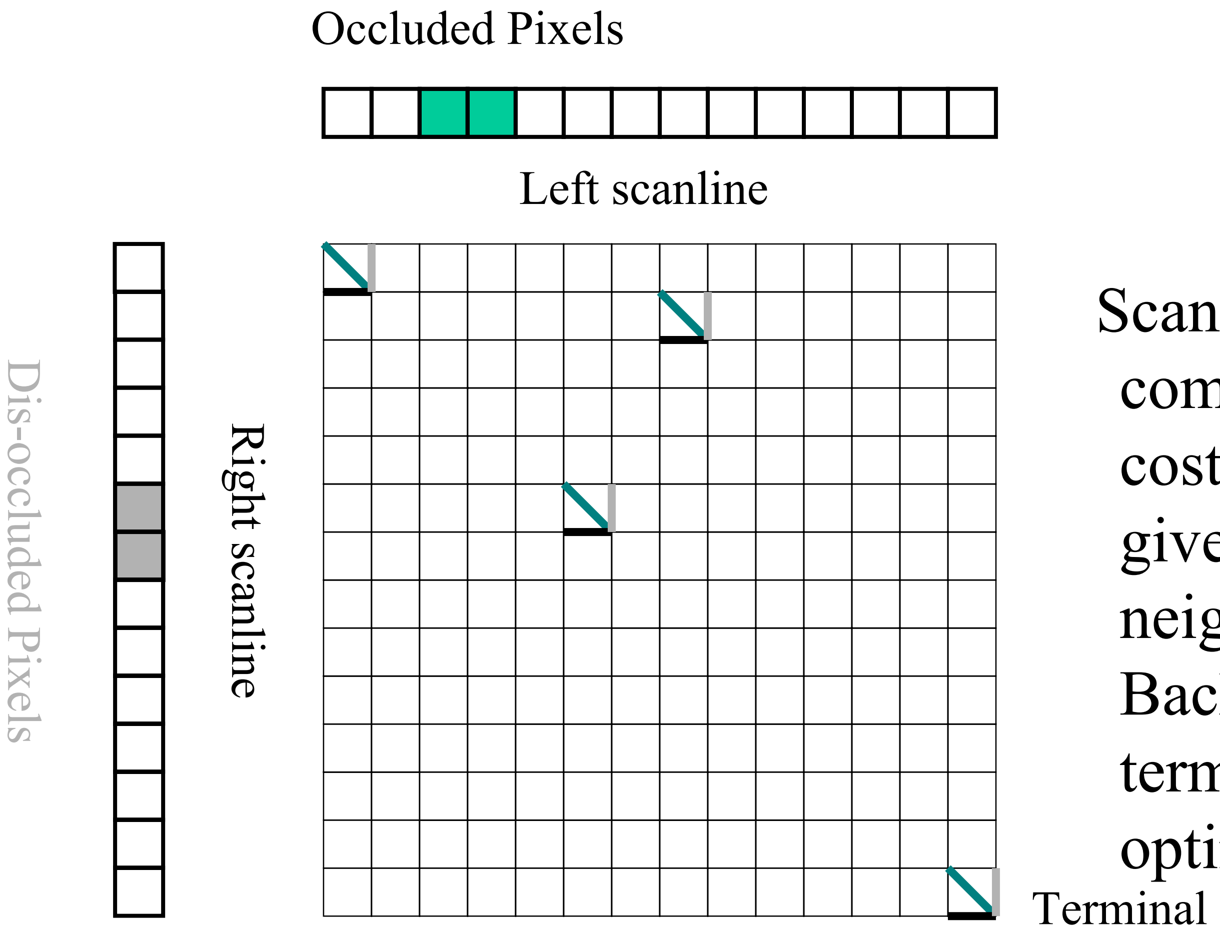
Principle of Optimality for an n-stage assignment problem:

Dynamic Programming



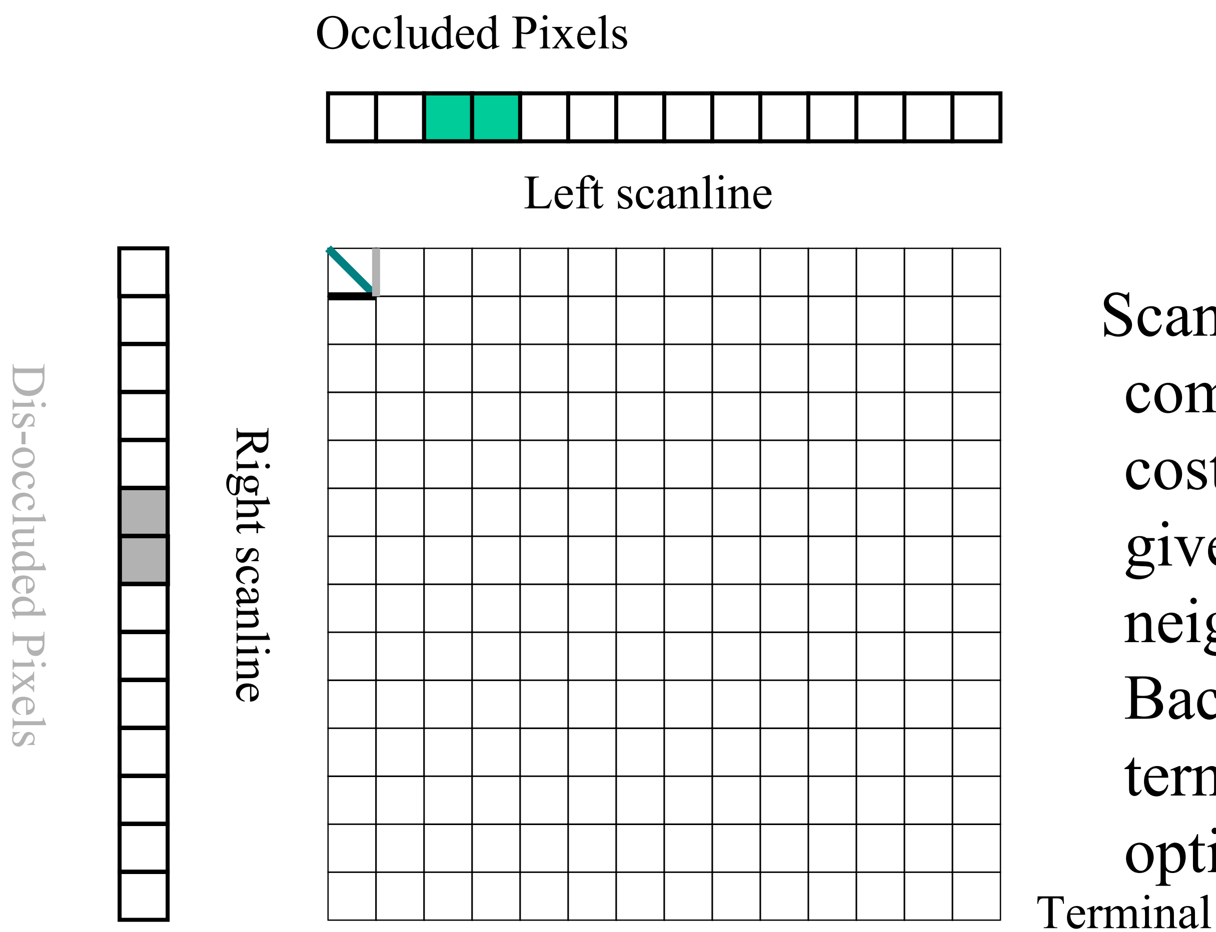
Back-chaining recovers the optimal path and its cost:

Stereo Matching with Dynamic Programming



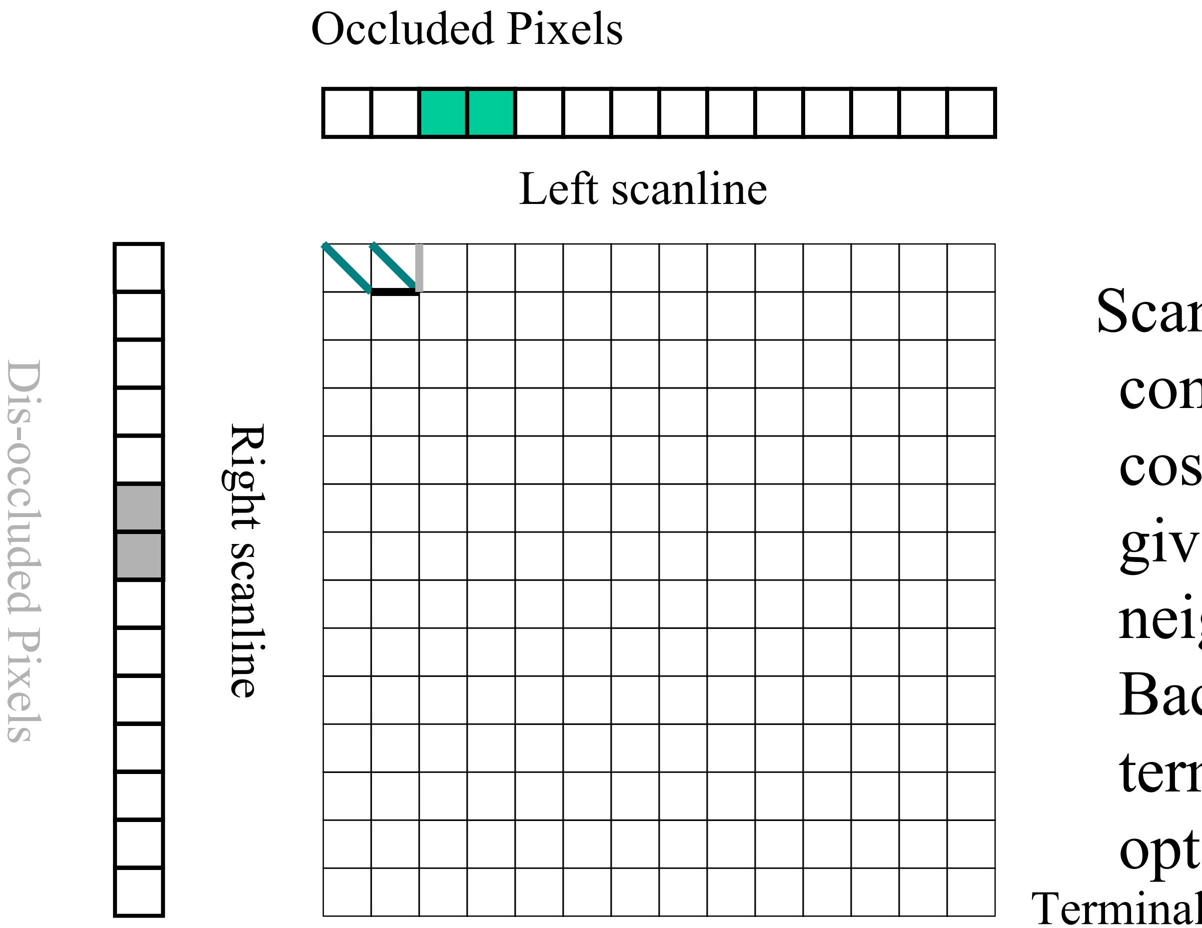
Scan across grid computing optimal cost for each node given its upper-left neighbors. Backtrack from the terminal to get the optimal path.

Stereo Matching with Dynamic Programming



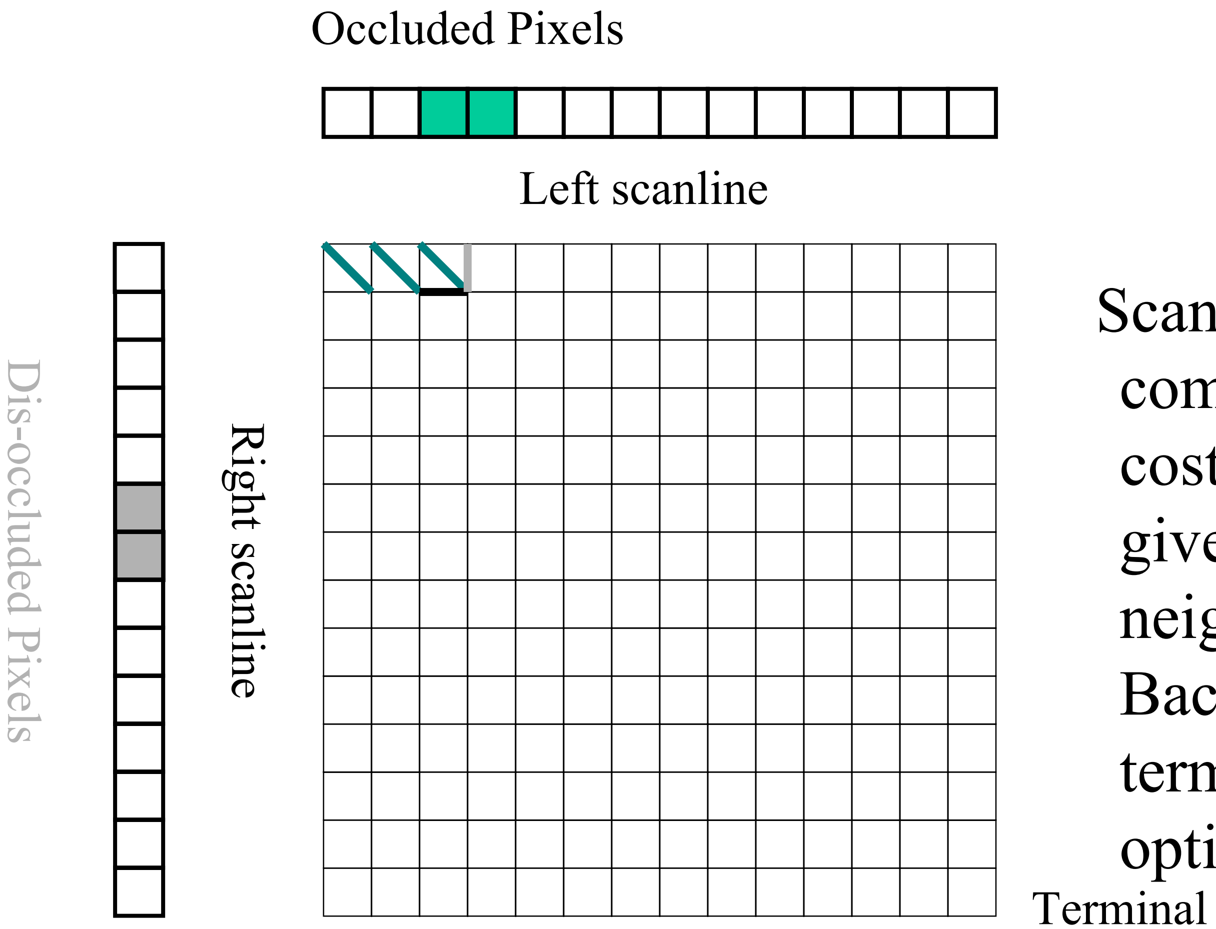
Scan across grid computing optimal cost for each node given its upper-left neighbors. Backtrack from the terminal to get the optimal path.

Stereo Matching with Dynamic Programming



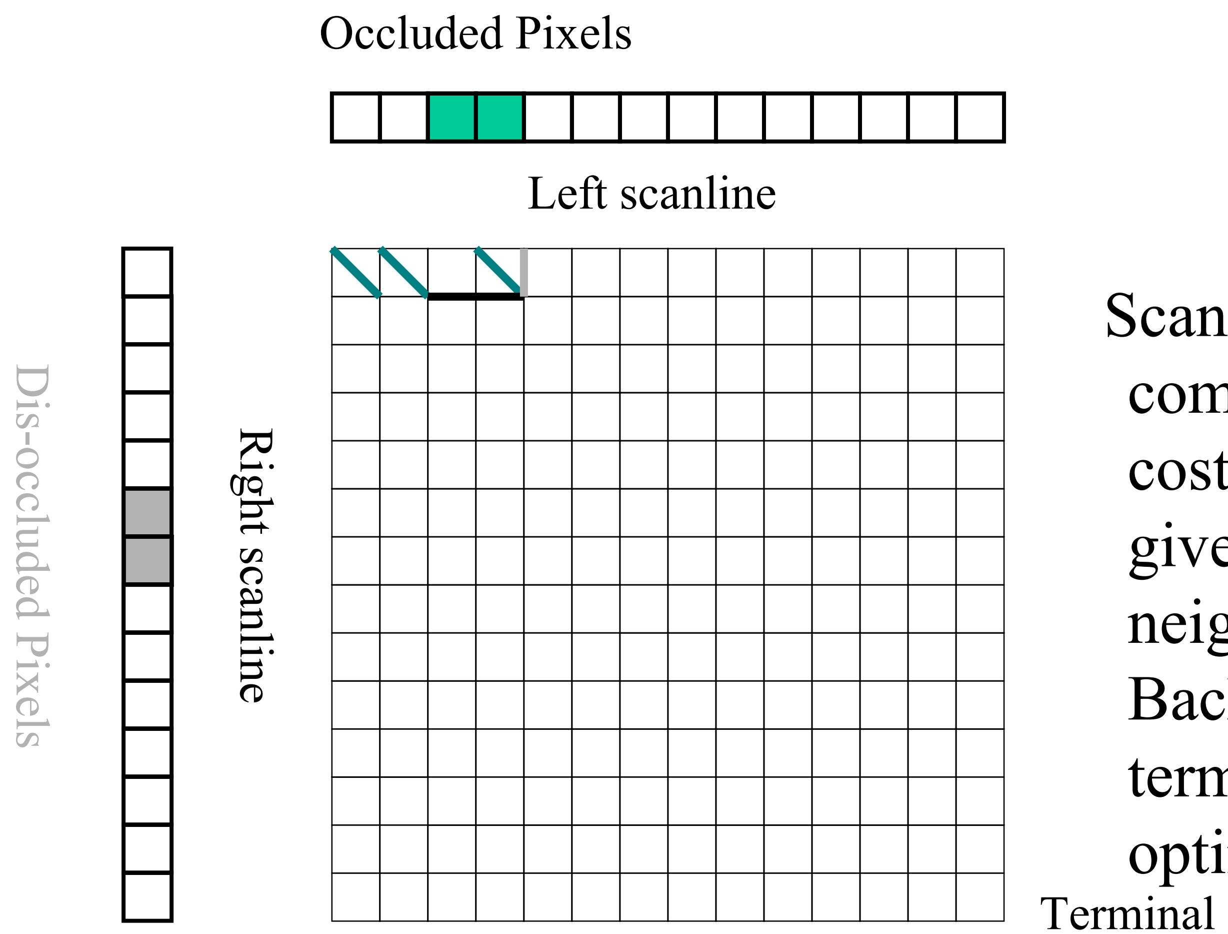
Scan across grid computing optimal cost for each node given its upper-left neighbors. Backtrack from the terminal to get the optimal path.

Stereo Matching with Dynamic Programming



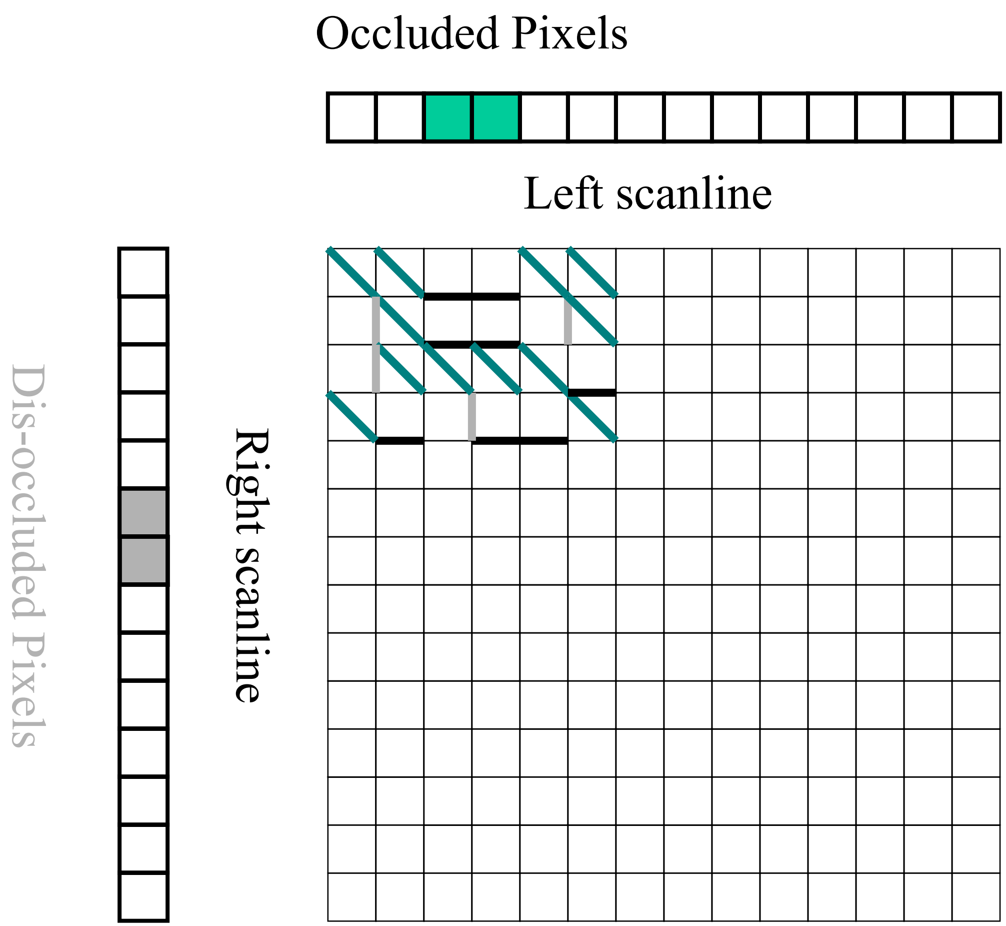
Scan across grid computing optimal cost for each node given its upper-left neighbors. Backtrack from the terminal to get the optimal path.

Stereo Matching with Dynamic Programming



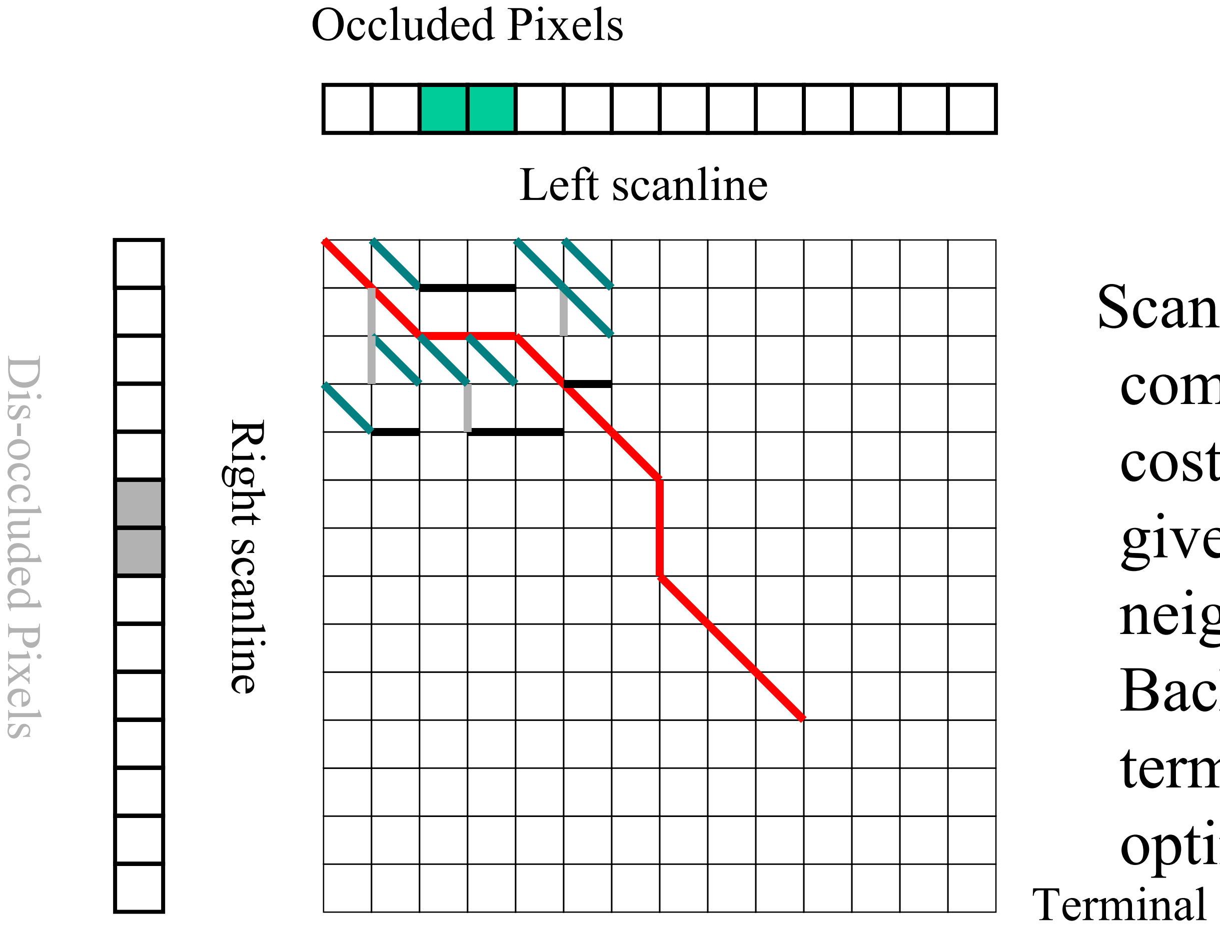
Scan across grid computing optimal cost for each node given its upper-left neighbors. Backtrack from the terminal to get the optimal path.

Stereo Matching with Dynamic Programming



Scan across grid computing optimal cost for each node given its upper-left neighbors. Backtrack from the terminal to get the optimal path.

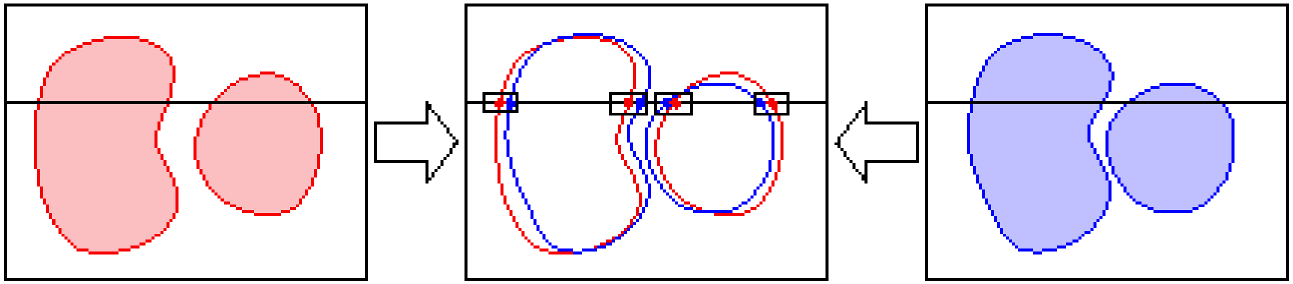
Stereo Matching with Dynamic Programming



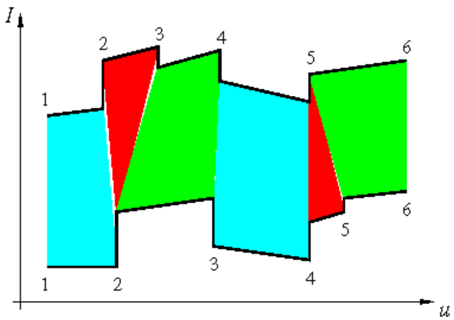
Scan across grid computing optimal cost for each node given its upper-left neighbors.
Backtrack from the terminal to get the optimal path.

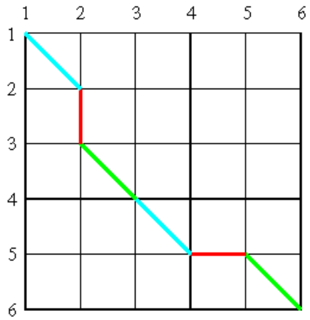
DP vs Edges

Edges:



DP:

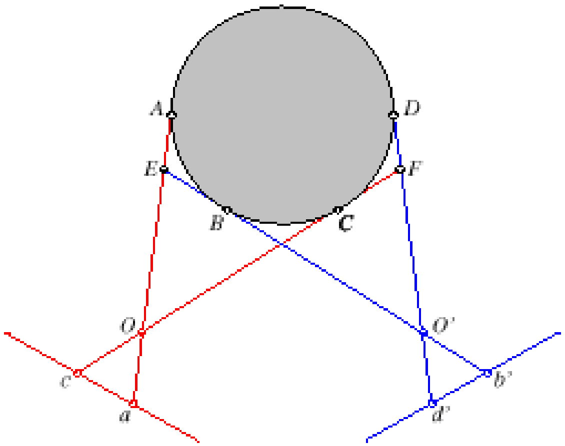




- Which method is better?
  - Edges are more “meaningful” [Marr]…but hard to find!
  - Edges tend to fail in dense texture (outdoors)
  - Correlation tends to fail in smooth featureless areas

Computing Correspondences

Both methods fail for smooth surfaces

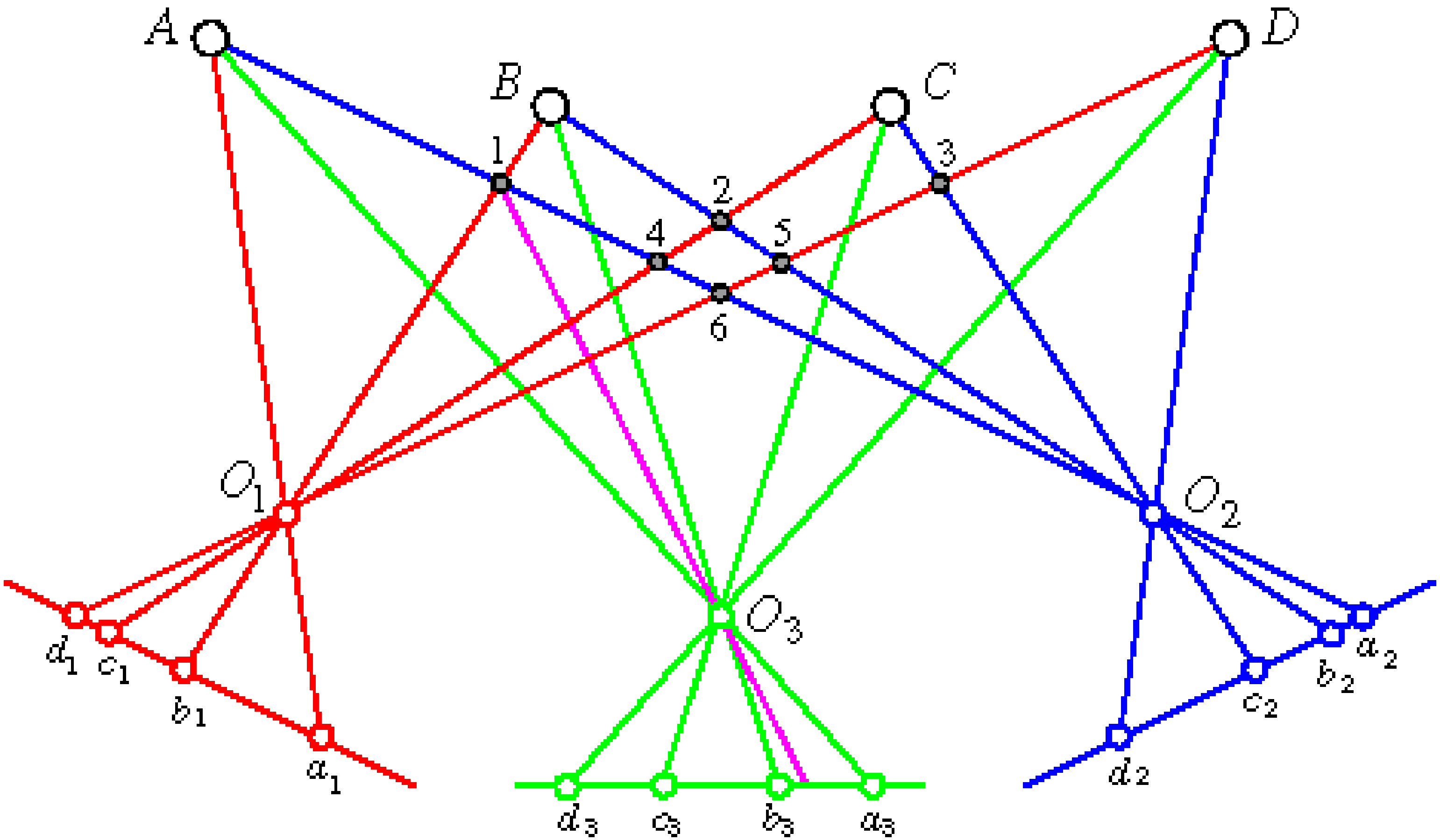


There is currently no good solution to the correspondence problem

Outline

- ✓ Reconstruction
- ✓ Rectification
- ✓ Early vs. Late
- ✓ Window Matching
- ✓ Edge Matching
- ✓ Ordering Constraint
- More views

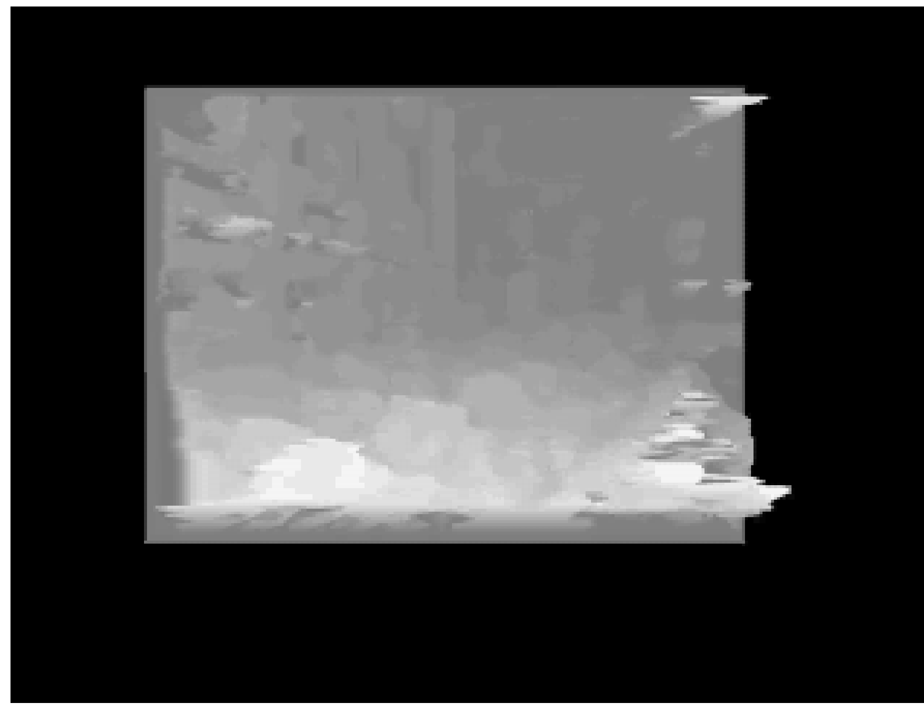
Three (calibrated) views



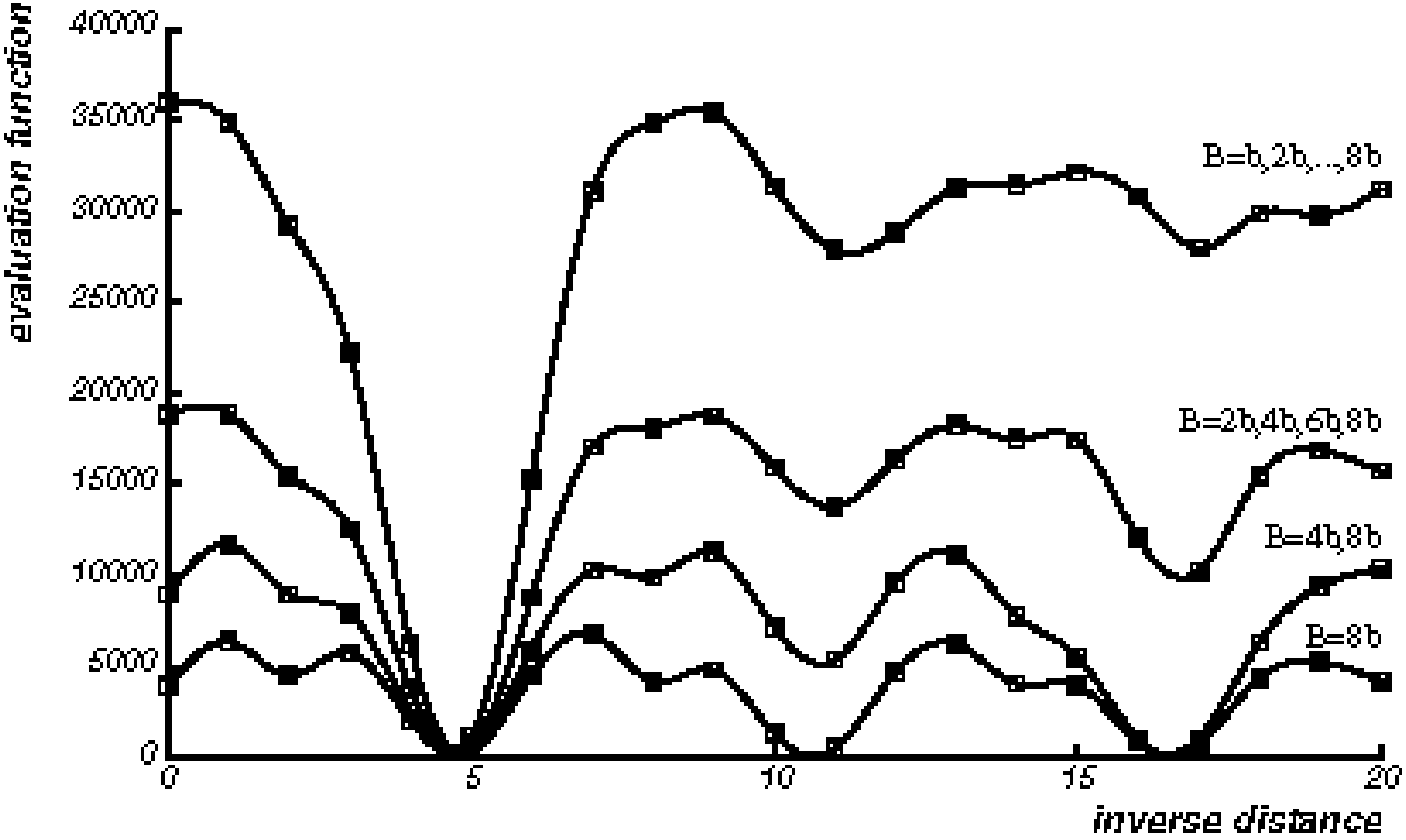
Why this camera arrangement?

Trinocular Stereo Results

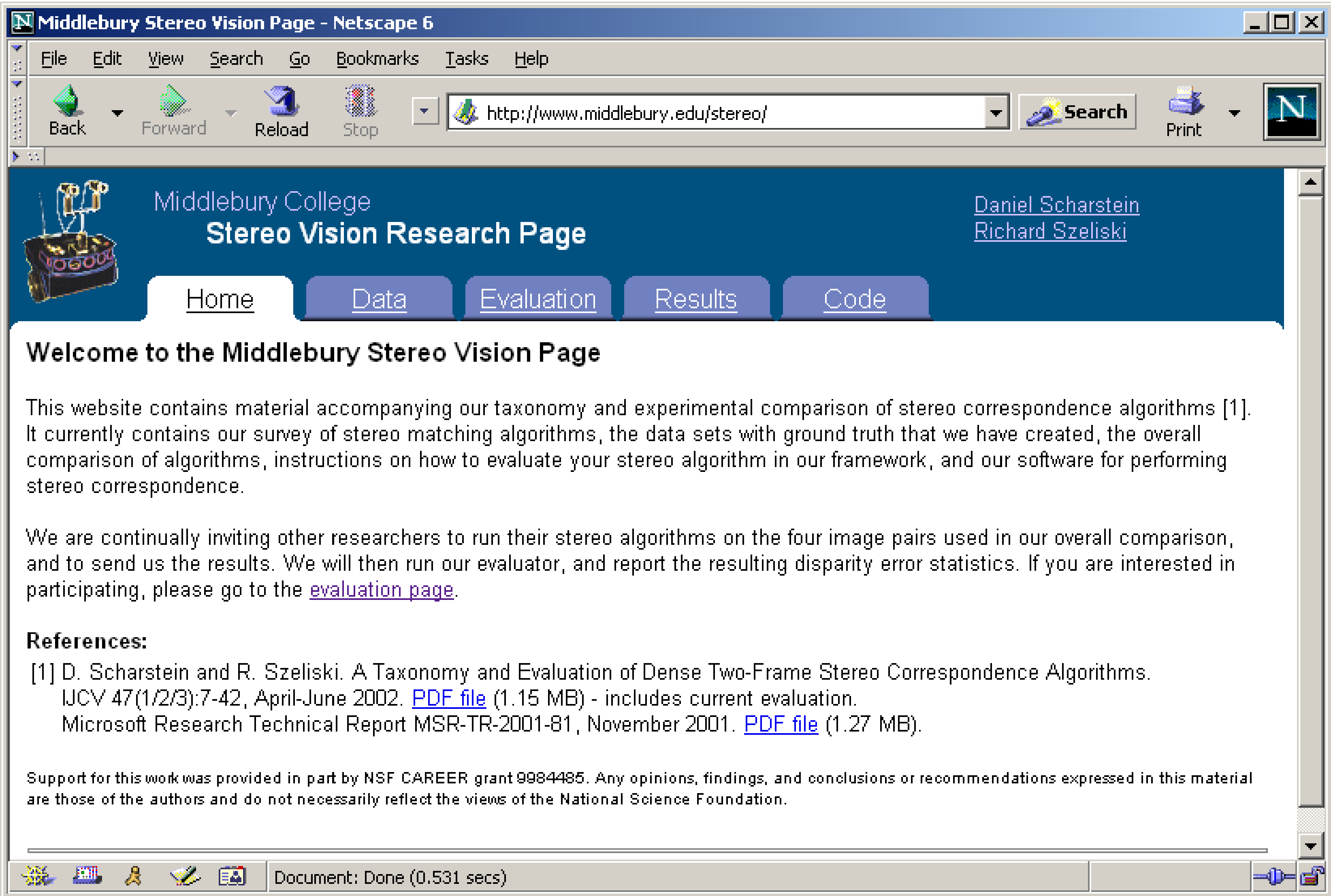
Trinocular stereo system available from *Point Gray Research* for $5K (circa ’97)



More views reduce ambiguity



Middlebury stereo page



http://www.middlebury.edu/stereo/

Outline

- ✓ Reconstruction
- ✓ Rectification
- ✓ Early vs. Late
- ✓ Window Matching
- ✓ Edge Matching
- ✓ Ordering Constraint
- ✓ More views

[Most figures adapted from Forsythe and Ponce]