STEREO VISION

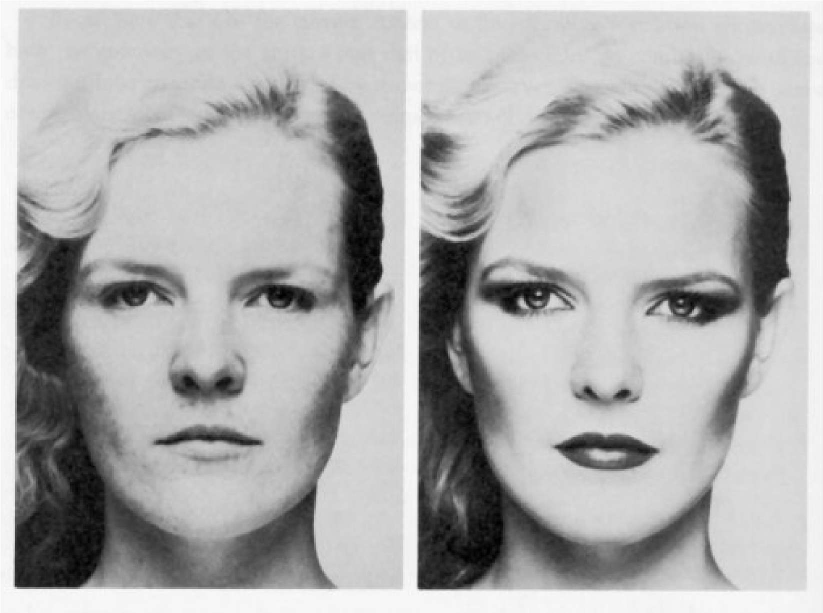
Goal of stereo vision

- The recovery of the 3D structure of a scene using two or more images of the 3D scene, each acquired from a different viewpoint in space.
- The images can be obtained using multiple cameras.
- The term *binocular vision* is used when two cameras are employed.

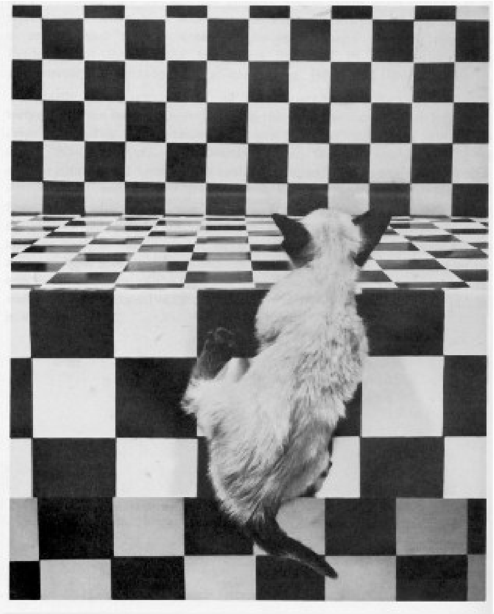
Visual cues for 3D

- When we look at image, what properties indicate the differences in depth or provide hints about object’s shape?

Shading



Textures



Focus

Motion

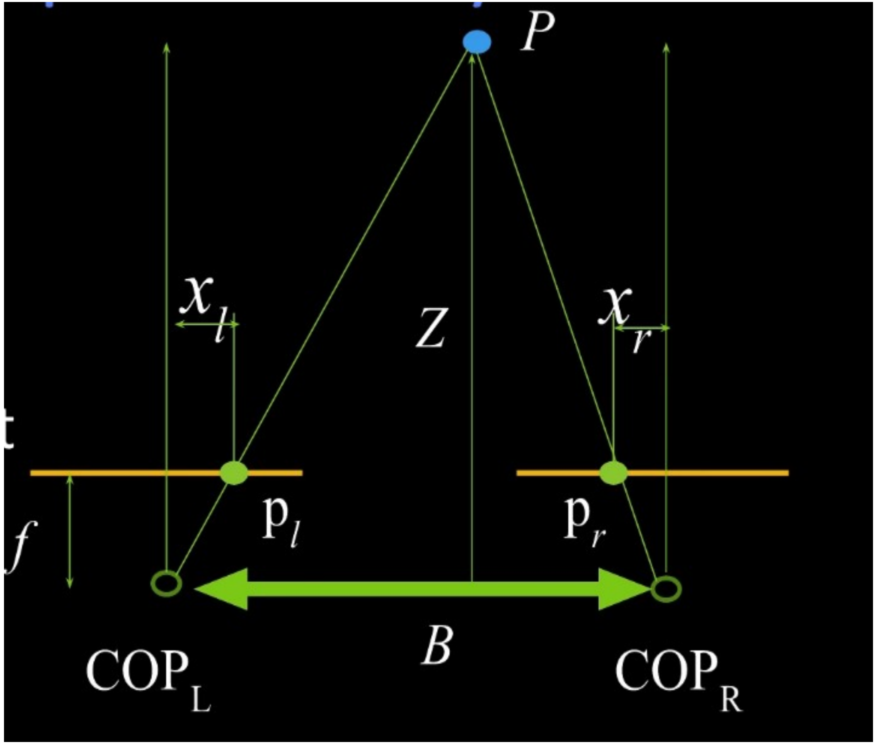
Shape from X

- X = shading, texture, focus, motion, ...
- We’ll focus on the motion cue

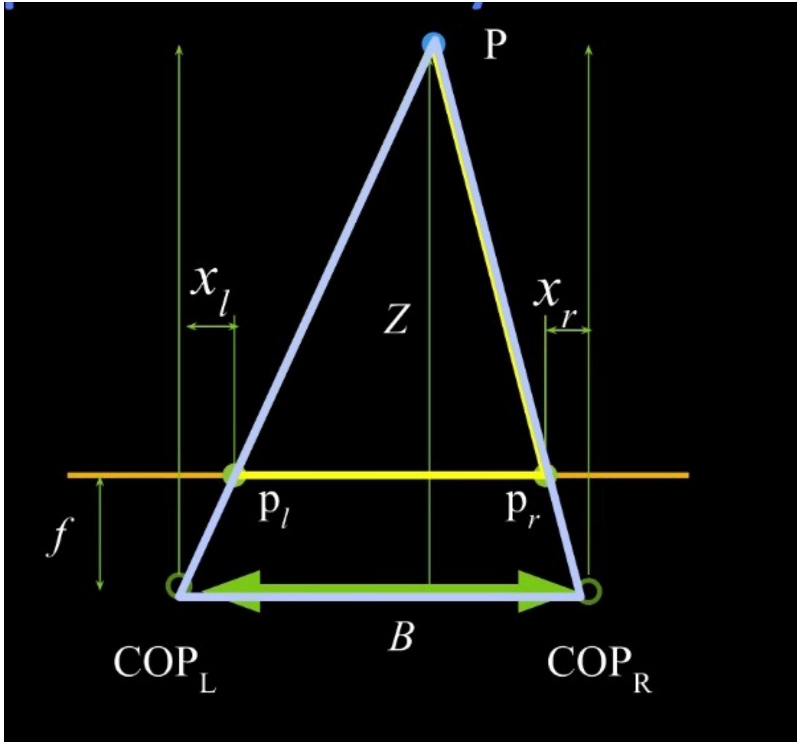
Estimating depth with stereo

- **Stereo :** Shape from motion between the two views
- We need to consider
  - Reconstruction
  - Correspondence

Geometry for a stereo system



- Two Image planes are coplanar
- Image planes at the front of image camera
- Origins are at the center of the planes
- B is baseline width,distance between two cameras
- Z is the depth to be obtained from the two formed images.
- f is the focal length.



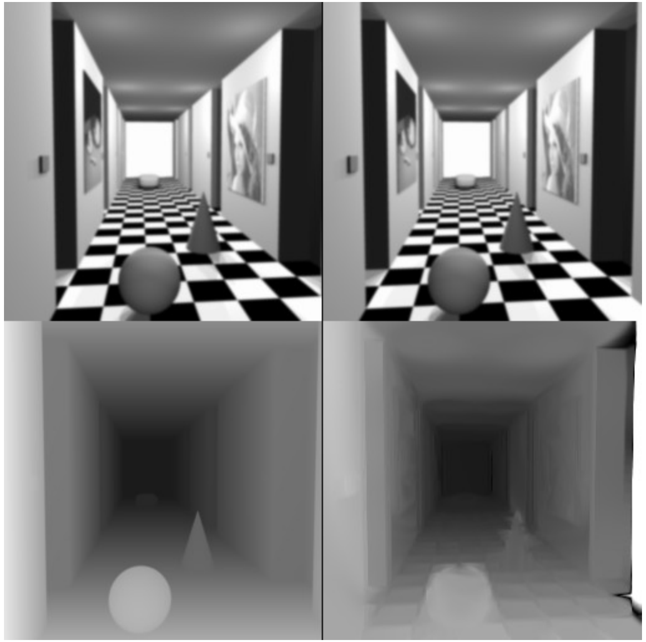
- Similar triangles (COP_L, P,COP_R) and (p_l,P,p_r)

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

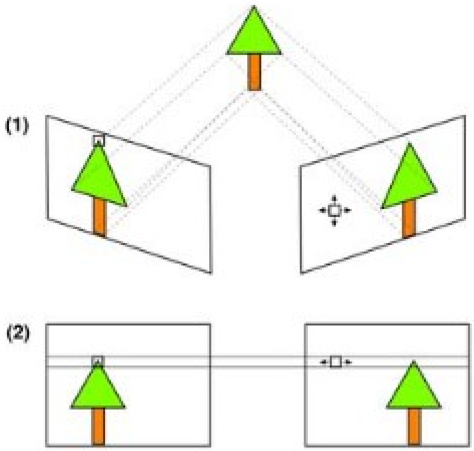
$$\boxed{Z = f\,B / x_l - x_r}$$

- x_l - x_r is called as disparity
- Depth is inversely proportional to Disparity

Disparity map



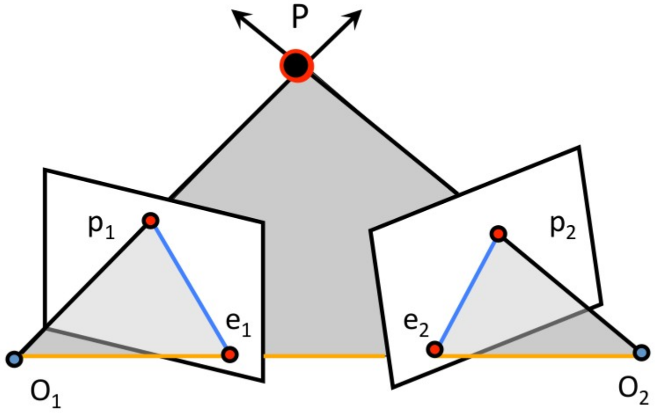
Stereo Correspondence



Search problem: Given an element in the left image, we search for the element in the right image.This involves two decisions:

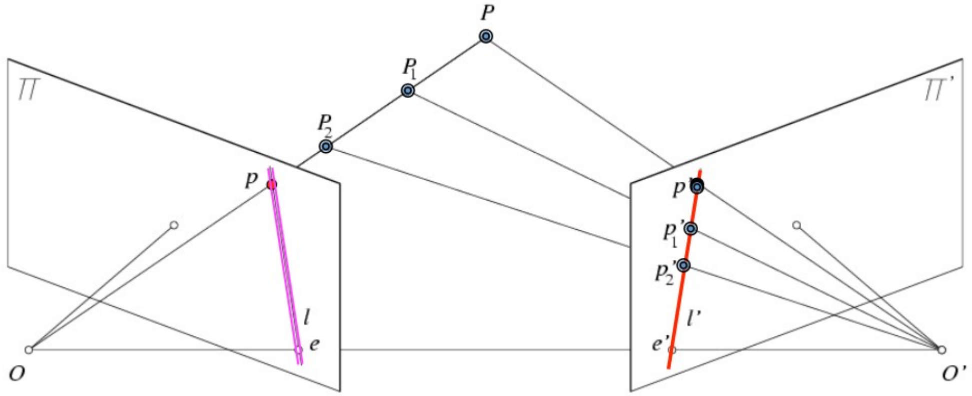
1. Which image element to match and
2. Which similarity measure to adopt

Epipolar Geometry



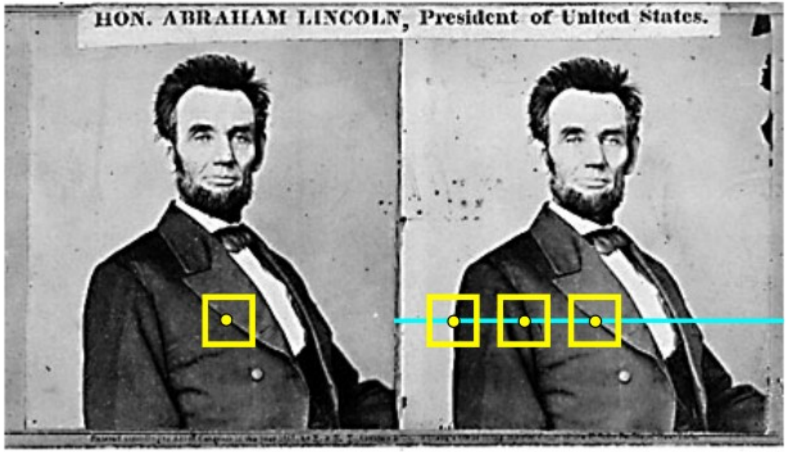
- The standard epipolar geometry setup involves two cameras observing the same 3D point P, whose projection in each of the image planes is located at p1 and p2.
- e1 and e2 are called epipoles.
- Epipolar lines:
  line joining p_1 and e_1
  line joining p_2 and e_2

Epipolar Constraint



- Potential matches for *p* have to lie on the corresponding epipolar line *l'*.
- Potential matches for *p'* have to lie on the corresponding epipolar line *l*.

Why is the epipolar constraint useful?



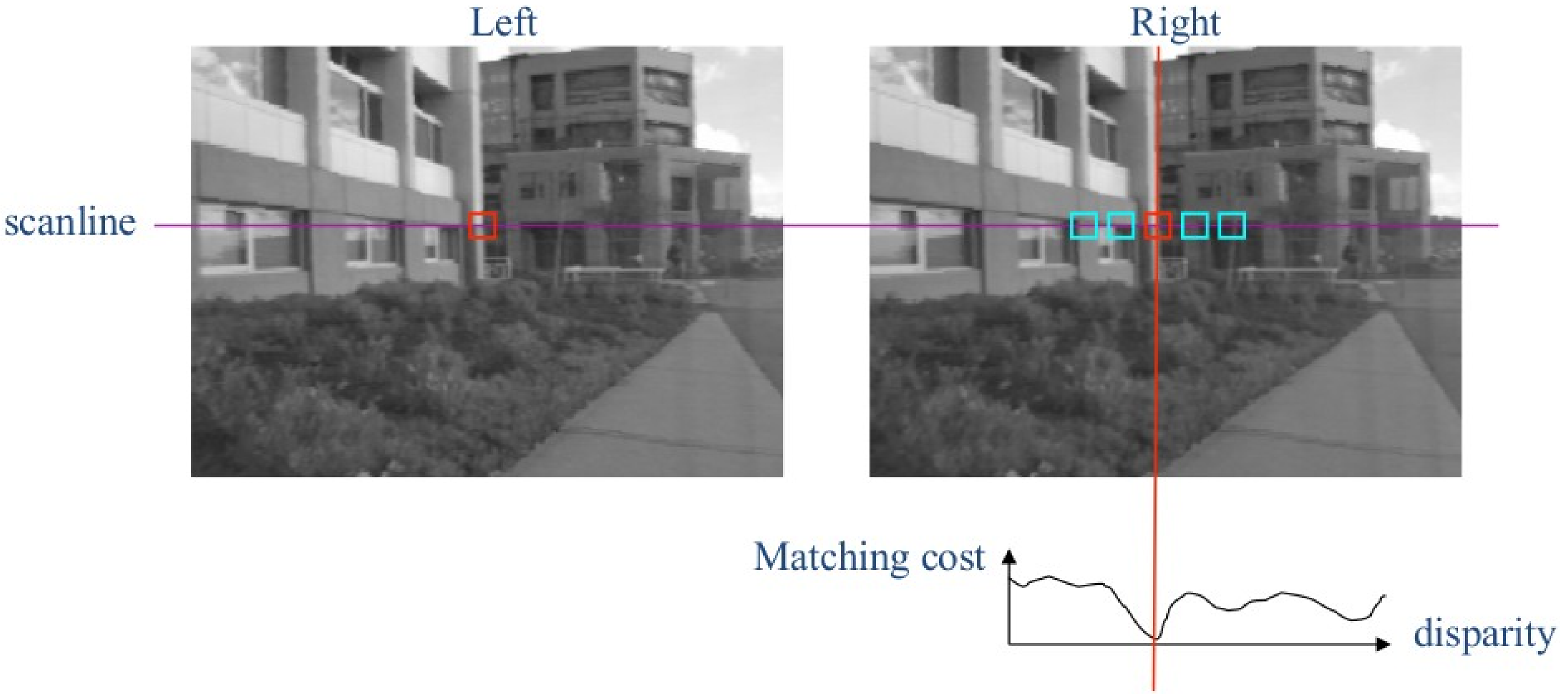
- The epipolar constraint reduces the correspondence problem to a 1D search along the epipolar line.

Correspondence problem

How to find corresponding points ?

- Similarity
  - Image patch from the left should match with the right
- Uniqueness
  - There is no more than one match for the pixel in right image
- Ordering
  - If pixels go a,b,c in left, they go a,b,c in right
- Disparity Gradient is limited
  - Depth doesn’t change too quickly

Correspondence search with similarity constraint



Similarity/Dissimilarity measures

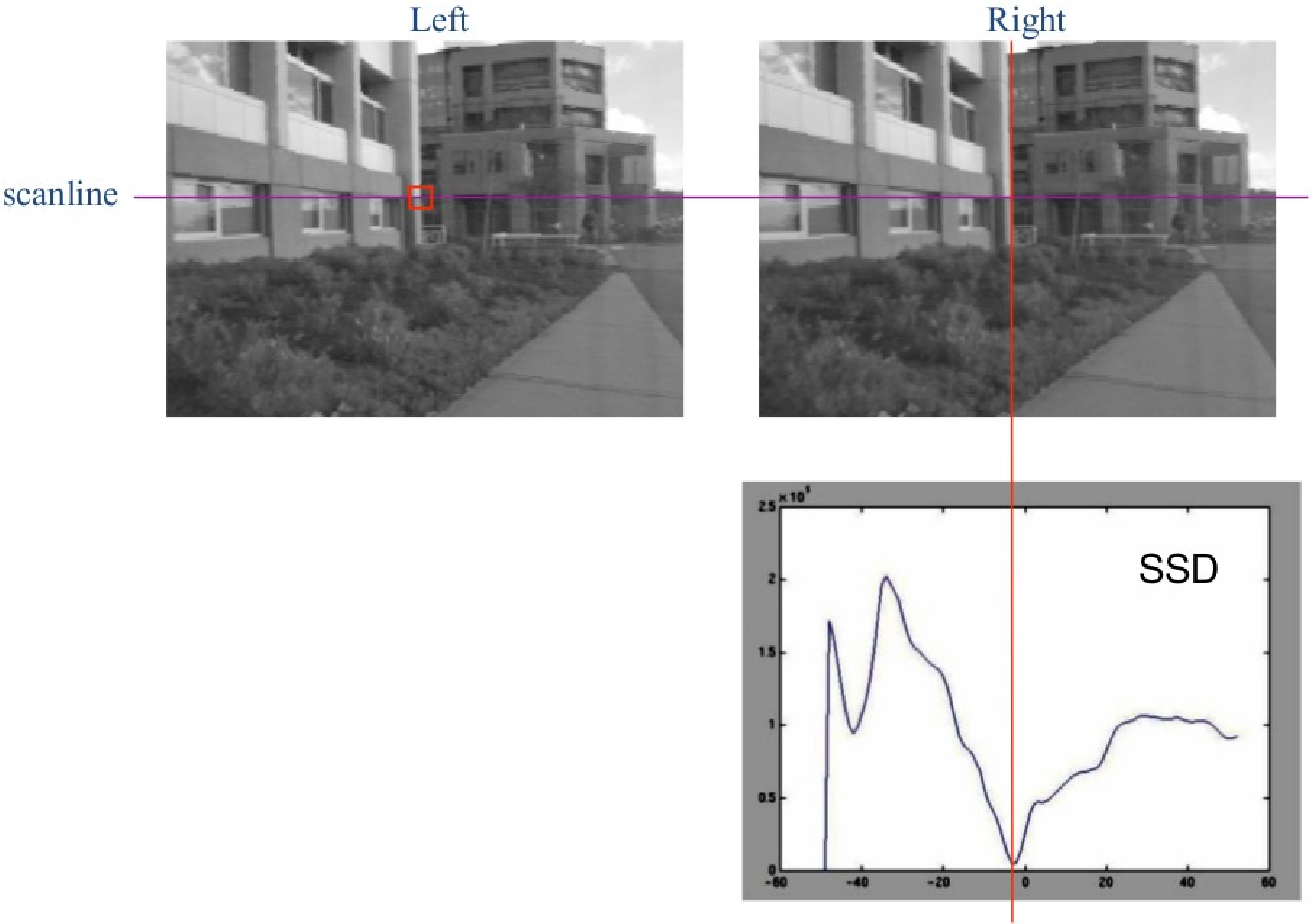
- Sum of Squares difference(SSD)
- Normalized Correlation

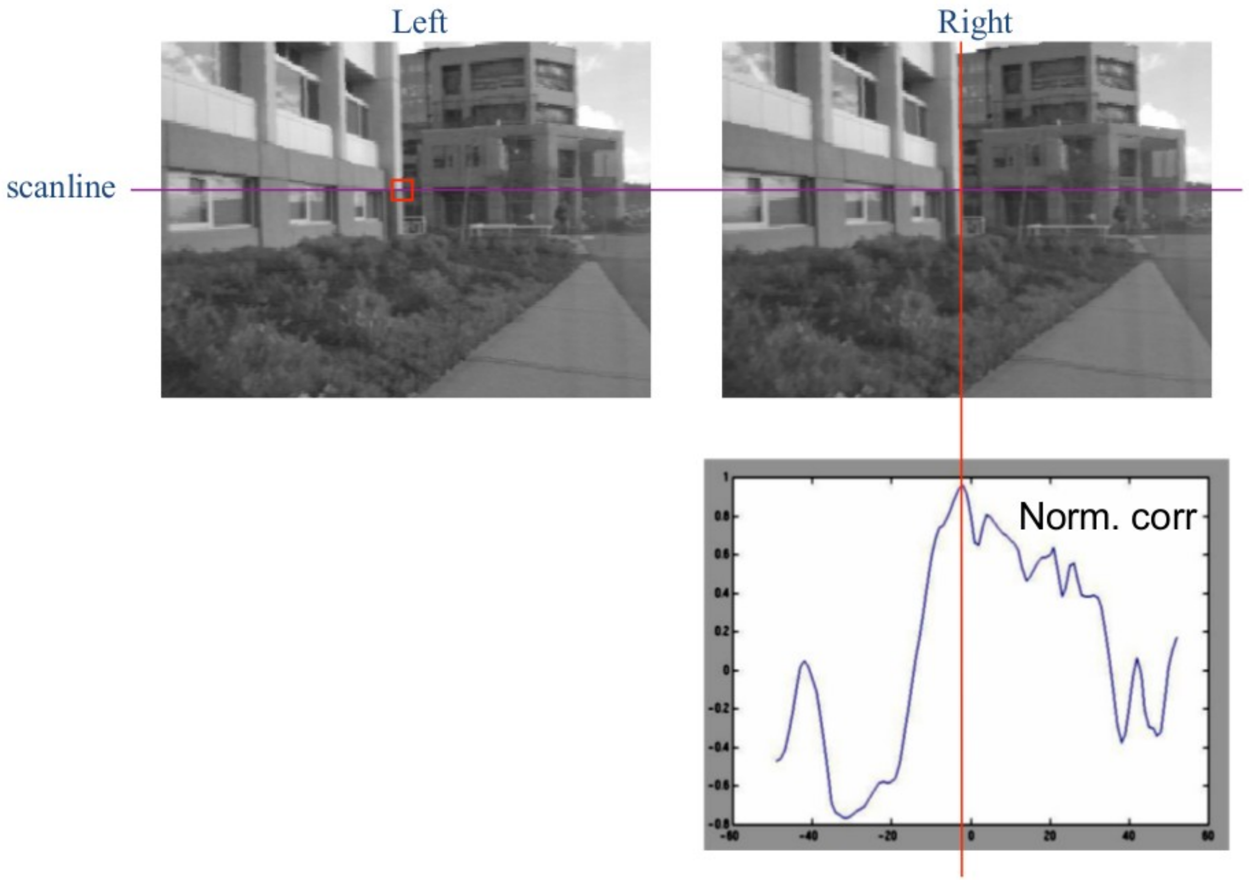
$$r_{ij} = \frac{\sum_m \sum_n [f(m+i, n+j) - \bar{f}][g(m,n) - \bar{g}]}{\sqrt{\sum_m \sum_n [f(m,n) - \bar{f}]^2 \sum_m \sum_n [g(m,n) - \bar{g}]^2}}.$$

Similarity/Dissimilarity measures

- Mutual Information

$$\mathrm{I}(X;Y) = \int_{\mathcal{Y}} \int_{\mathcal{X}} p_{(X,Y)}(x,y) \log\left(\frac{p_{(X,Y)}(x,y)}{p_X(x)\, p_Y(y)}\right) dx\, dy,$$

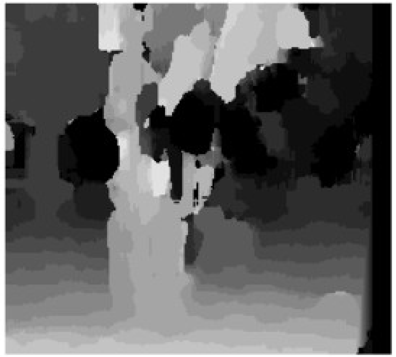

Left
Right
scanline
SSD
2.5
2
1.5
1
0.5
0
-60
-40
-20
0
20
40
60



Left
Right
scanline
Norm. corr

Effect of Window Size

W = 3

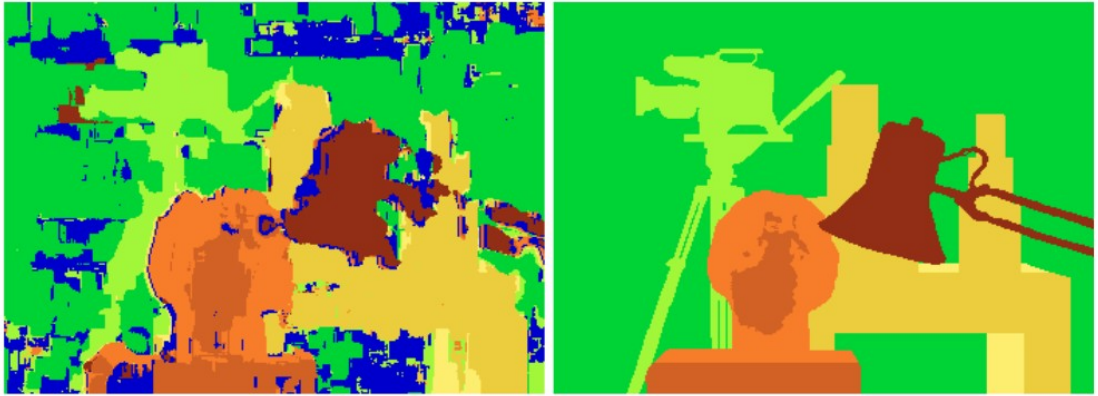


W = 20

- Smaller window
  - \+ More detail
  - • More noise
- Larger window
  - \+ Smoother disparity maps
  - • Less detail

Stereo Correspondence results

Results with window correlation



Window-based matching
(best window size)

Ground truth

Results with better method

State of the art method

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

Ground truth

Challenges in Stereo Vision

- Low-contrast; Texture less image regions
- Occlusions
- Large base lines
- Camera calibration errors etc.,

Thank you!