Computer Vision for HCI

Stereo Vision

1

Ambiguity in Single View

- Structure and depth are inherently ambiguous from single view

This is __________ perspective

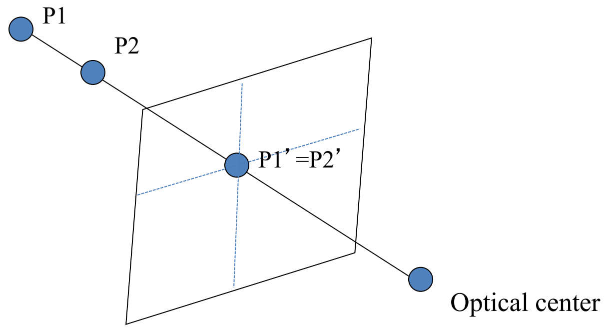
Some slides adapted from Kristen Grauman, James Hays, Oliver Grau

Images from Lana Lazebnik

2

Ambiguity in Single View

- Structure and depth are inherently ambiguous from single view

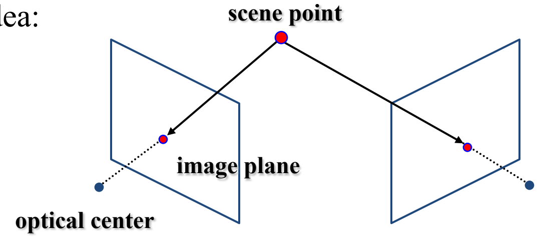


3

Extracting 3-D Information

- What cues help us to perceive 3-D shape and depth?
  - Shading, focus, texture, motion, …
- Stereo:
  - Shape from difference between two views
  - Infer 3-D shape of scene from two (multiple) images from different viewpoints

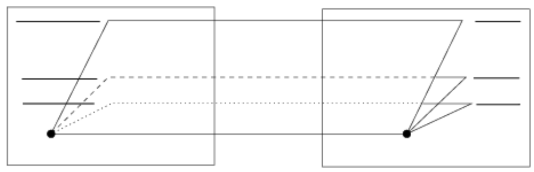
Main idea:



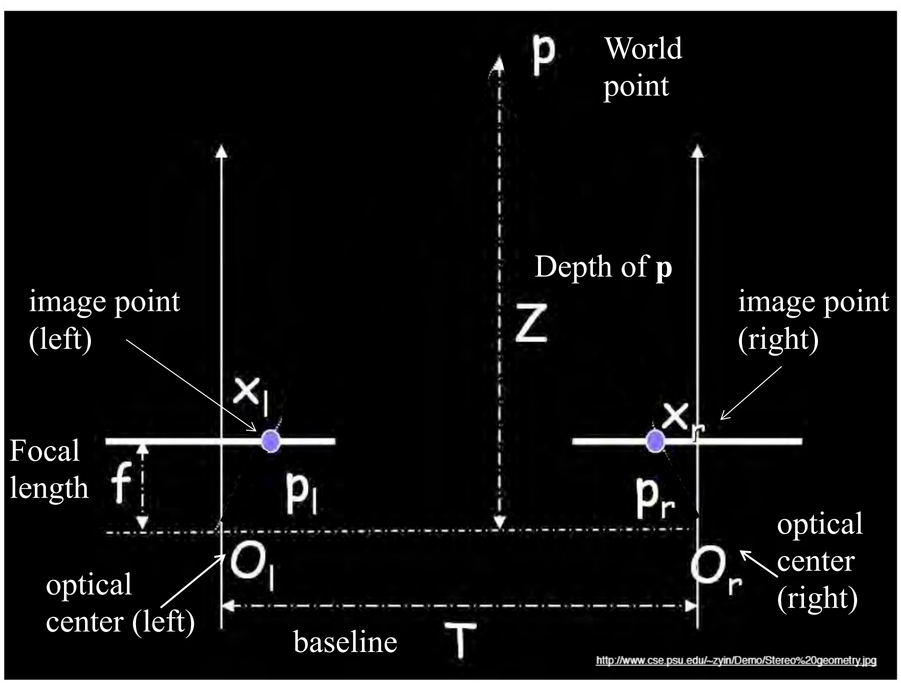
4

Geometry for a Simple Stereo System

- First, assume parallel optical axes, known camera parameters (i.e., calibrated cameras):



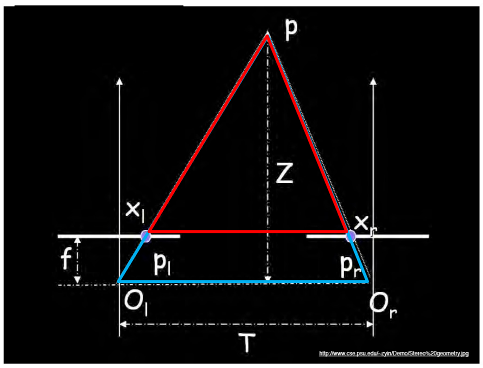
5



6

Geometry for a Simple Stereo System

- Assume parallel optical axes, known camera parameters (i.e., calibrated cameras). **What is expression for Z (depth)?**



Similar triangles (p_l, p, p_r) and (O_l, p, O_r):

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

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

disparity

7

Depth from Disparity

image I(x,y)

Disparity map D(x,y)

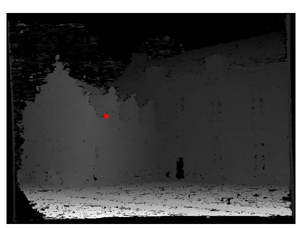


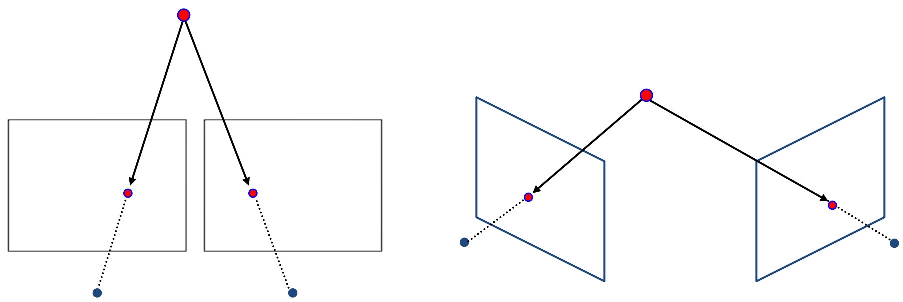
image I´(x´,y´)

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

8

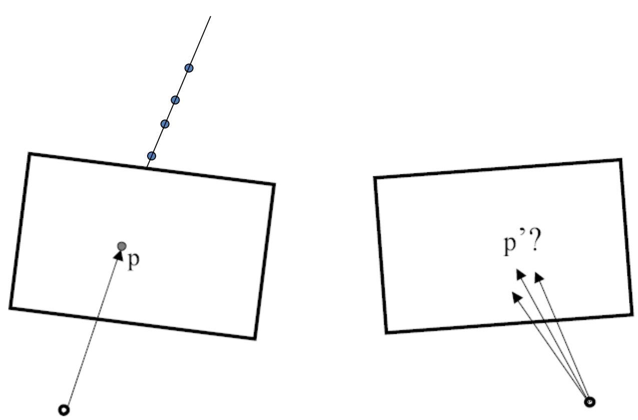
General Case, with Calibrated Cameras

- The two cameras need not have parallel optical axes.



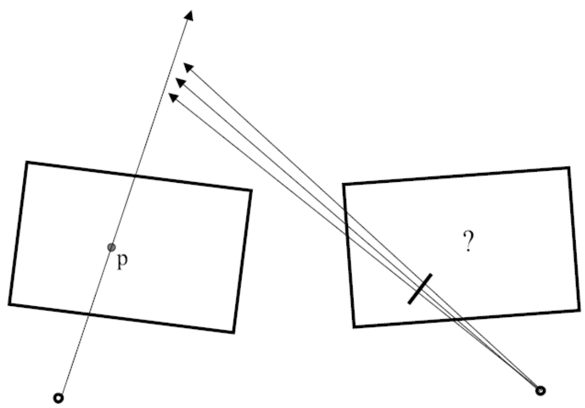
vs.

Stereo Correspondence Constraints



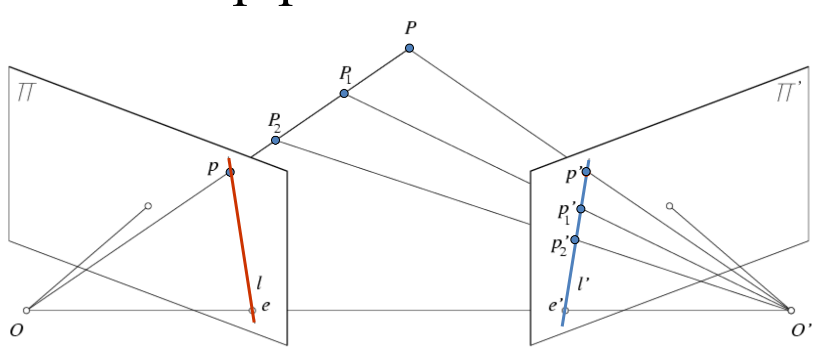
Given p in left image, where can corresponding point p' be?

Stereo Correspondence Constraints



11

Epipolar Constraint

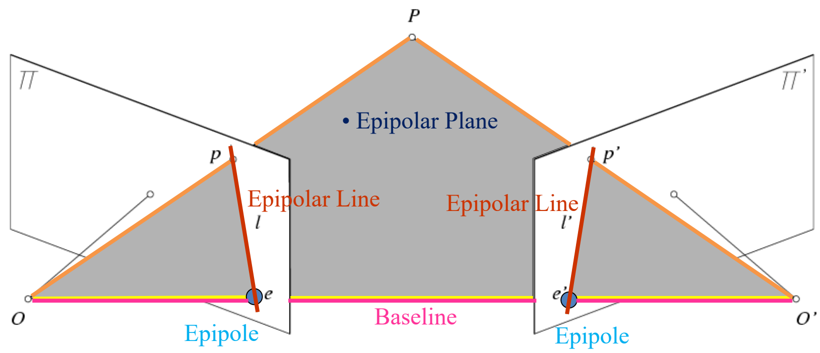


Geometry of two views constrains where the corresponding pixel for some image point in the first view must occur in the second view

- It must be on the line carved out by a plane connecting the world point and optical centers

12

Epipolar Geometry



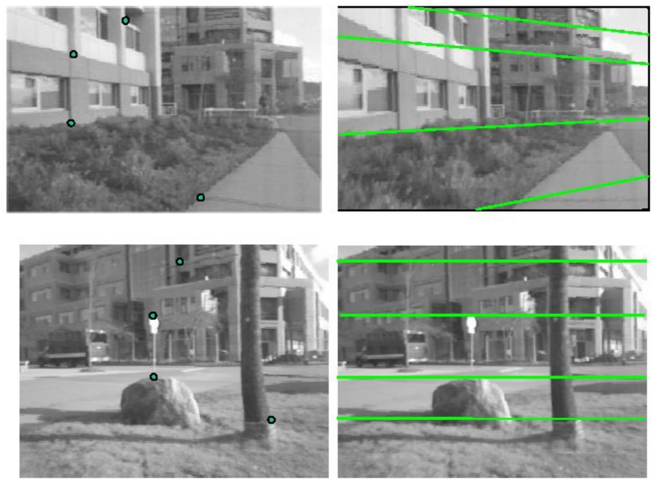
13

Why is the epipolar constraint useful?

14

Epipolar Constraint

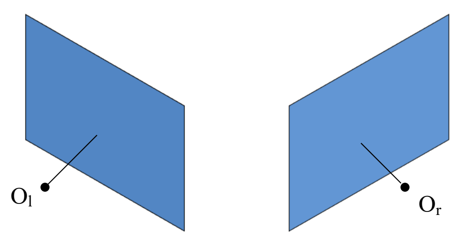
This is useful because it reduces the correspondence problem to a 1D search along an epipolar line.



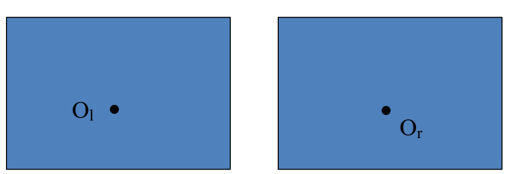
15

What do the Epipolar Lines Look Like?

1.



2.



16

Example: Converging Cameras

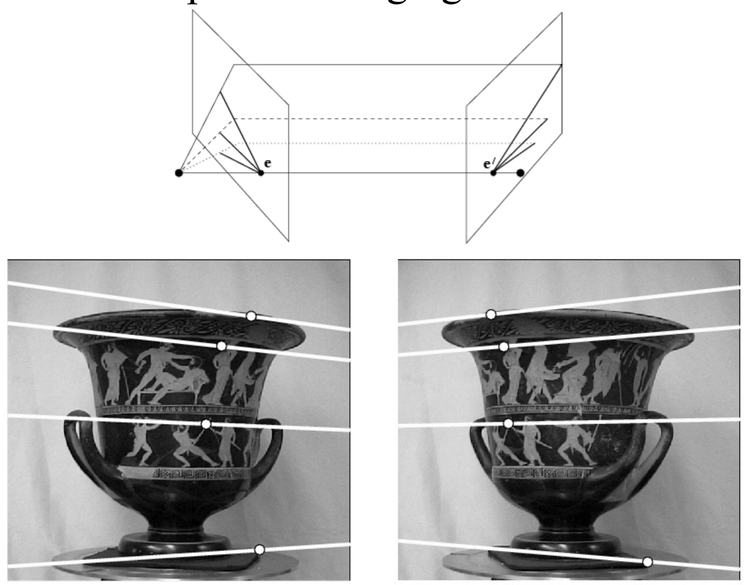
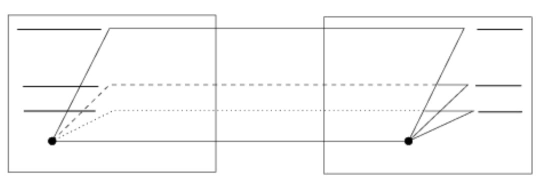


Figure from Hartley & Zisserman

17

Example: Parallel Cameras



(epipoles at infinity)

Figure from Hartley & Zisserman

18

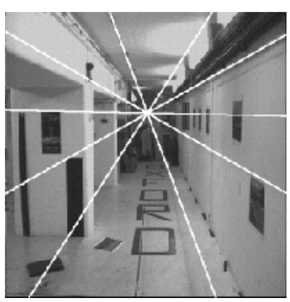
Camera Motion

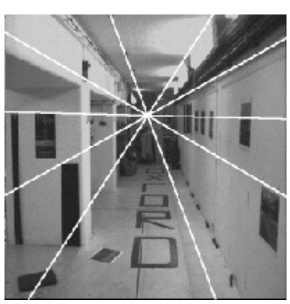
Consider two images from a moving camera (rather than two distinct cameras)

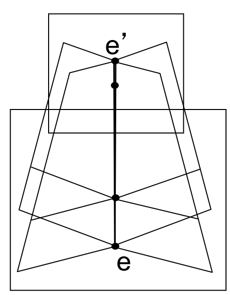
What would the epipolar lines look like if the camera moves directly forward?

19

Example: Forward Motion



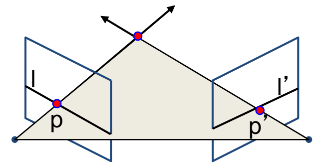




Epipole has same coordinates in both images.
Points move along lines radiating from e:
"Focus of expansion"

20

How are Two Image Planes Related?



- Let p be a point in left image, p' in right image
  - Homogeneous coordinates: [x, y, 1], [x', y', 1]
- p and p' are related by a rotation and translation of the cameras
  - Yields epipolar constraint with formula $p'^T F p = 0$
    - Longuet-Higgins 1981
- Epipolar relation
  - p maps to epipolar line l'
  - p' maps to epipolar line l
- Equation for line: ax + by + c = 0
  - Defined by parameter column vector $r = [a, b, c]^T$
  - For any point $p = [x,y,1]^T$ on the line defined by r we have $r^T p = 0$
- Recall $p'^T F p = 0$, hence the epipolar lines are

$$0 = p'^T F p = (p'^T F) p = (F^T p')^T p = l^T p \quad \rightarrow \quad l = F^T p'$$

Similarly, $l' = Fp$ Defines parameters of line (l) through point p

21

Fundamental Matrix

- This matrix F is called
  - The "**Fundamental Matrix**"
    - General, uncalibrated camera case
  - The "Essential Matrix"
    - When image intrinsic parameters are known
- Can solve for F from point correspondences
  - Each (p, p') pair gives one linear equation involving elements of F

$$p'^T F p = 0$$

22

Computing Fundamental Matrix

- Set up system of equations

$$\begin{bmatrix} x' & y' & 1 \end{bmatrix} \begin{bmatrix} f_{11} & f_{12} & f_{13} \\ f_{21} & f_{22} & f_{23} \\ f_{31} & f_{32} & f_{33} \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = 0$$

$$x'xf_{11} + x'yf_{12} + x'f_{13} + y'xf_{21} + y'yf_{22} + y'f_{23} + xf_{31} + yf_{32} + f_{33} = 0$$

$$\begin{bmatrix} x'x & x'y & x' & y'x & y'y & y' & x & y & 1 \end{bmatrix} \mathbf{f} = 0$$

- Eight degrees of freedom in F (solution is good up to scale)
  - Need at least **8 (x, y) coordinates** to solve for F

$$A\mathbf{f} = \begin{bmatrix} x_1'x_1 & x_1'y_1 & x_1' & y_1'x_1 & y_1'y_1 & y_1' & x_1 & y_1 & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ x_n'x_n & x_n'y_n & x_n' & y_n'x_n & y_n'y_n & y_n' & x_n & y_n & 1 \end{bmatrix} \mathbf{f} = \mathbf{0}$$

23

Computing Fundamental Matrix
Normalized 8-Point Algorithm

- Solve for F using analogous approach as when solving for homography H
- For robustness, compute similarity transformation matrices T^a and T^b that make p_i and p_i' be:
  - Zero mean
  - Have an average distance to the origin of $\sqrt{2}$
- Set **f** equal to Eigenvector of $(A^T A)$ corresponding to smallest Eigenvalue
  - Ensure **f** is unit norm
- Unrasterize **f** to give fundamental matrix $\widetilde{F}$ (of the normalized points)
  - Noise causes rank($\widetilde{F}$) = 3 [rank is # independent cols (or rows)]
- **Due to epipolar geometry, need to enforce singularity constraint (rank($\widetilde{F}$) = 2)**
  - Take SVD of $\widetilde{F}$ ($\widetilde{F} = UDV^T$)
    - D = diag($[r,s,t]$), where singular values $r \geq s \geq t$
  - Set $\widetilde{F} = U\text{diag}([r,s,\mathbf{0}])V^T$
- Finally, remove normalization by setting $F = (T^b)^T \tilde{F} T^a$

$p'^T F p = 0$
$p'^T[(T^b)^T \tilde{F} T^a]p = 0$
$(p'^T (T^b)^T)\, \tilde{F} (T^a p) = 0$
$(T^b p')^T \tilde{F}\, (T^a p) = 0$

***Note, slightly different process (not inverse of T^b) for removing normalization than with homography!

24

Building Depth Maps

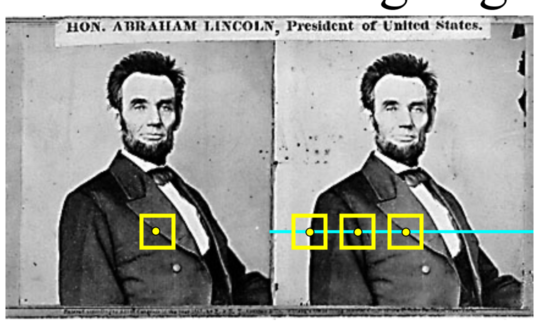
- With <u>parallel optical axes</u>, one can build depth maps
- Depth maps are inversely related to disparity maps

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

- Need to find point correspondences to find disparities
  - Epipolar geometry constrains search for point correspondences
    - Finding correspondences is still a difficult problem

25

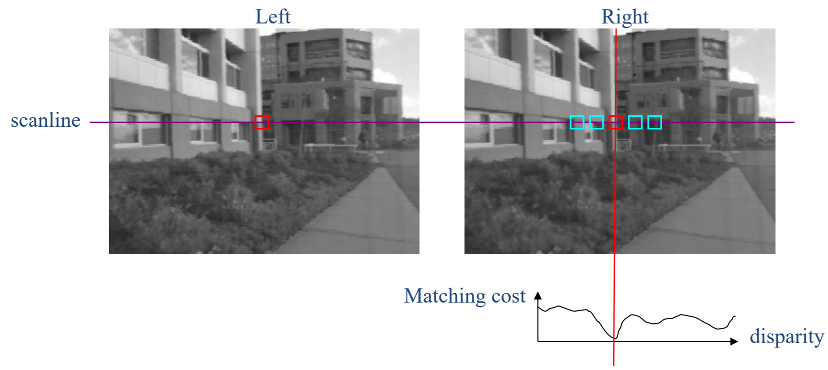
Basic Stereo Matching Algorithm



- (If necessary, rectify the two stereo images to transform epipolar lines into horizontal scanlines)
- For each pixel x in the first image
  - Find corresponding epipolar scanline in the right image
  - Examine pixels on the scanline and pick the best match x'
  - Compute disparity x-x' and set Depth(x) = $fT/(x-x')$
    - Or use relative depth as RelDepth(x) = $1/(x-x')$

26

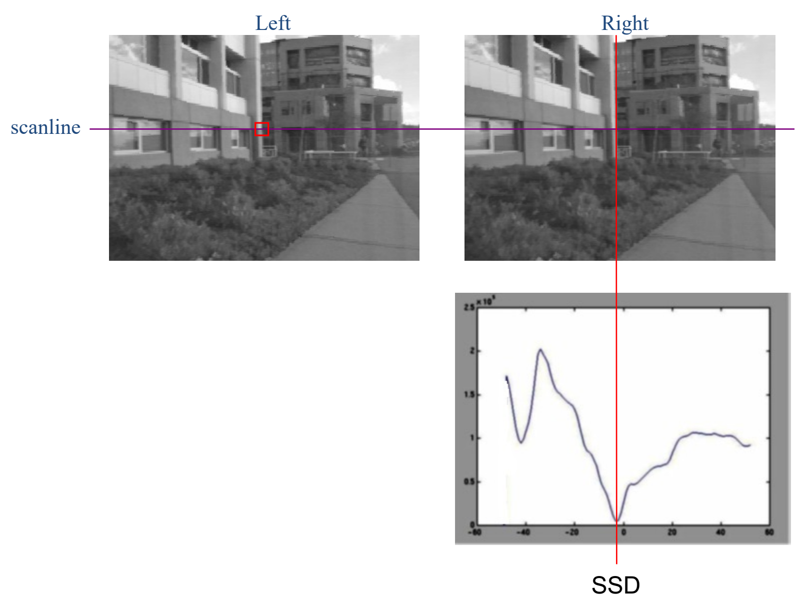
Correspondence Search



- Slide a window along the right scanline and compare contents of that window with the reference window in the left image
- Matching cost: SAD, SSD, or **normalized cross-correlation**

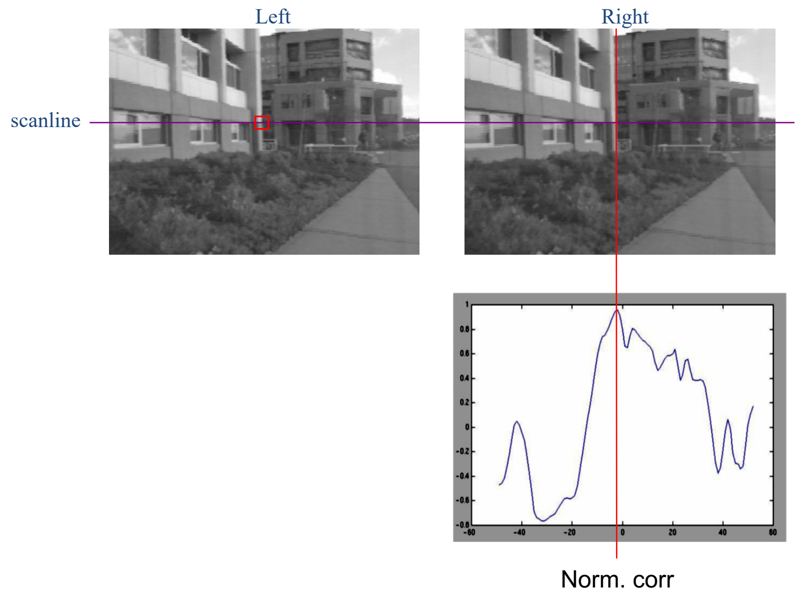
27

Correspondence Search



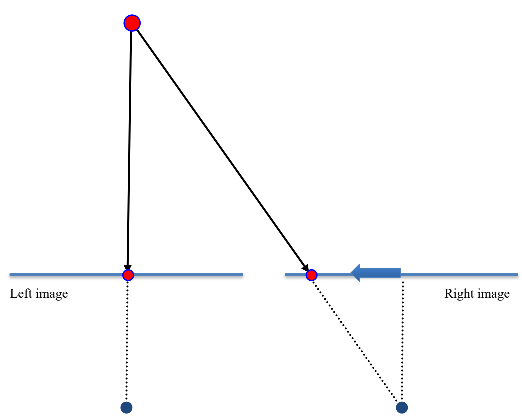
28

Correspondence Search



29

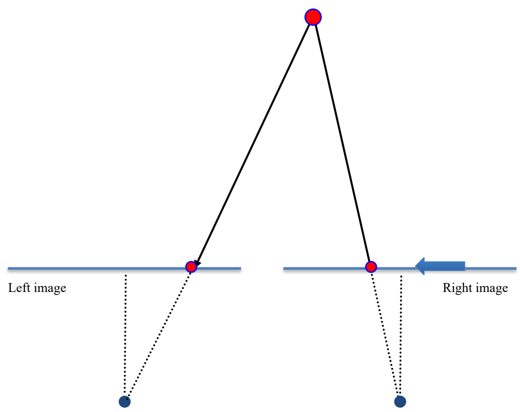
Limited Search – Leftward Only



Given the geometry, need only search "to the left" in the right image from starting point taken from the left image

30

Limited Search – Leftward Only



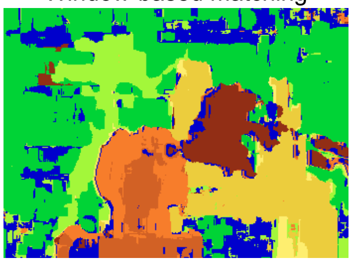
Given the geometry, need only search "to the left" in the right image from starting point taken from the left image

31

Results with Window Search

Scene

Window-based matching



Ground truth

32

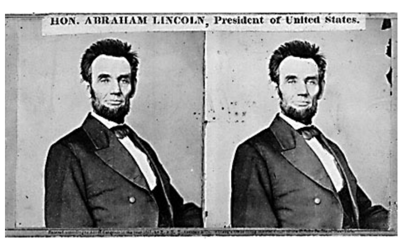
Effect of Window Size

W = 3 W = 20

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

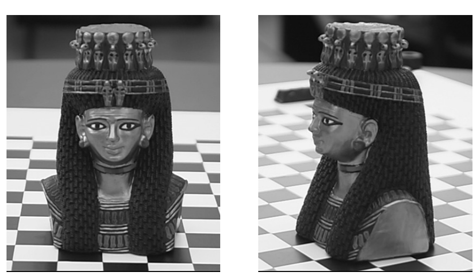
33

Difficulties of Correspondence Search



Textureless surfaces

Occlusions



Specularities (Non-Lambertian surfaces)

34

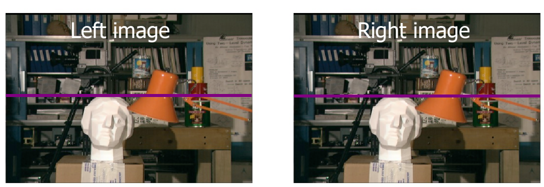
Priors and Constraints

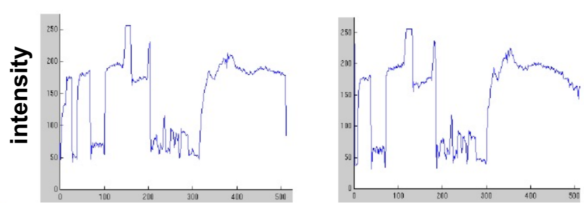
- Uniqueness
  - For any point in one image, there should be at most one matching point in the other image
- Ordering
  - Corresponding points should be in the same order in both views
- Smoothness
  - We expect disparity values to change slowly (for the most part)

35

Scanline Stereo

- Try to coherently match pixels on the entire <u>scanline</u>
- Different scanlines are still optimized independently

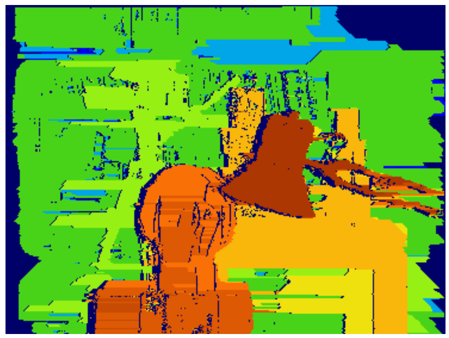




36

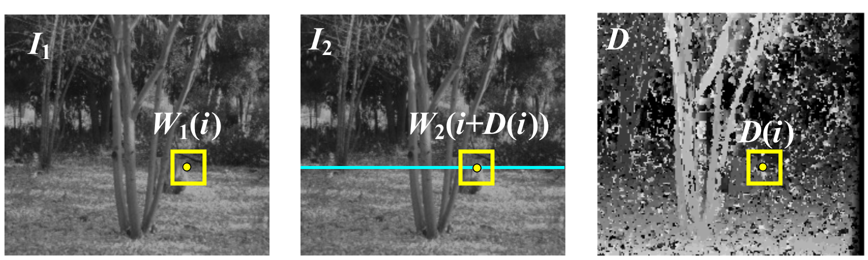
Coherent Stereo on 2-D Grid

- Scanline stereo can generate *streaking* artifacts
  - Dynamic programming approaches Ohta & Kanade '85, Cox et al. '96



37

Stereo Matching as Energy Minimization



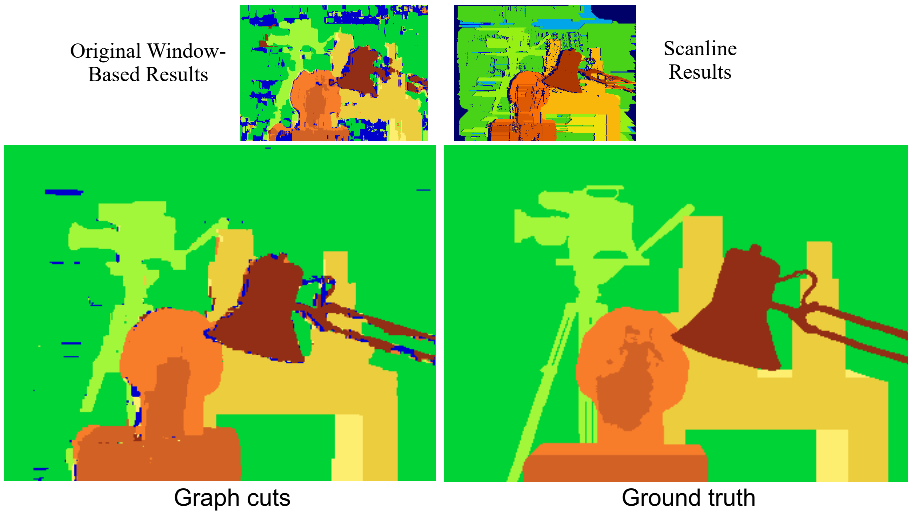
$$E(D) = \underbrace{\sum_i \left(W_1(i) - W_2(i + D(i))\right)^2}_{\text{data term}} + \lambda \underbrace{\sum_{\text{neighbors } i,j} \rho\left(D(i) - D(j)\right)}_{\text{smoothness term}}$$

- Energy functions of this form can be minimized using *graph cuts*

Y. Boykov, O. Veksler, and R. Zabih, Fast Approximate Energy Minimization via Graph Cuts, PAMI 2001

38

Graph Cuts Results



Y. Boykov, O. Veksler, and R. Zabih, [Fast Approximate Energy Minimization via Graph Cuts]

39

Depth Estimation using Deep Learning With a Single Image

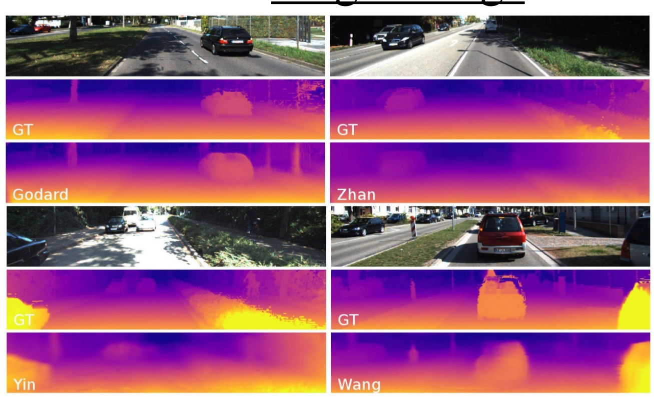


Fig. 1. An example of monocular depth estimation. The depth maps are predicted from deep neural networks proposed by Godard *et al.* [13], Zhan *et al.* [14], Yin *et al.* [15], and Wang *et al.* [16]. The results are taken from [17]. As shown in these figures, the 3D structures of objects, like trees, street and cars, can be effectively perceived from single images by deep depth networks.

40

Kinect: Structured Infrared Light

http://www.youtube.com/watch?v=dTKlNGSH9Po

41

Summary

- Epipolar geometry
  - Epipoles are intersection of baseline with image planes
  - Matching point in second image is on a line passing through its epipole
  - Fundamental matrix maps from a point in one image to a line (its epipolar line) in the other
  - Can solve for F given corresponding points (e.g., interest points)
- Stereo depth estimation
  - Estimate disparity by finding corresponding points
  - Depth is inversely related to disparity

42