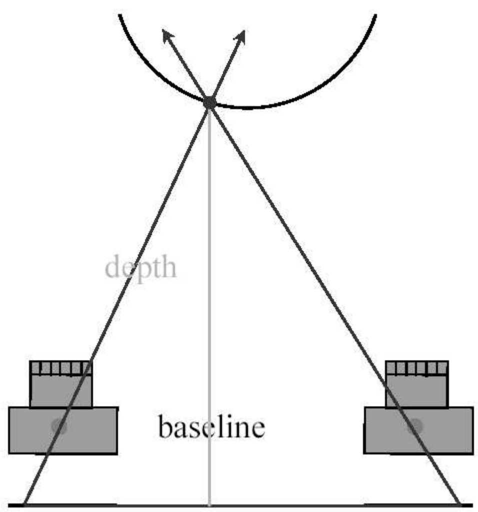
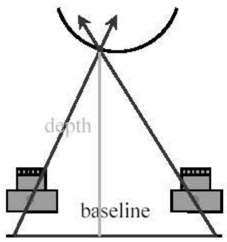
Stereopsis

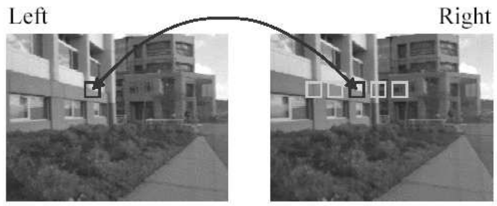


Reconstruction



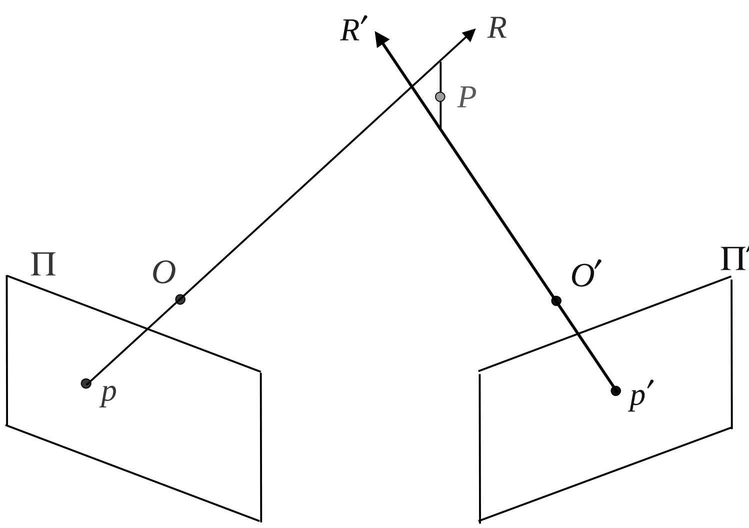
Triangulate on two images of the same point to recover depth.

- Feature matching across views
- Calibrated cameras

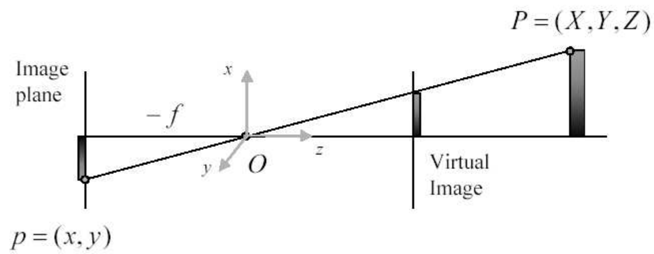


Only need to match features across epipolar lines

Geometric Reconstruction

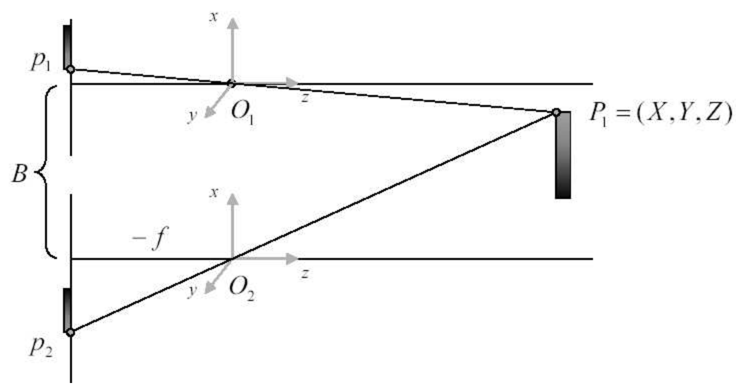


Pinhole Camera Model



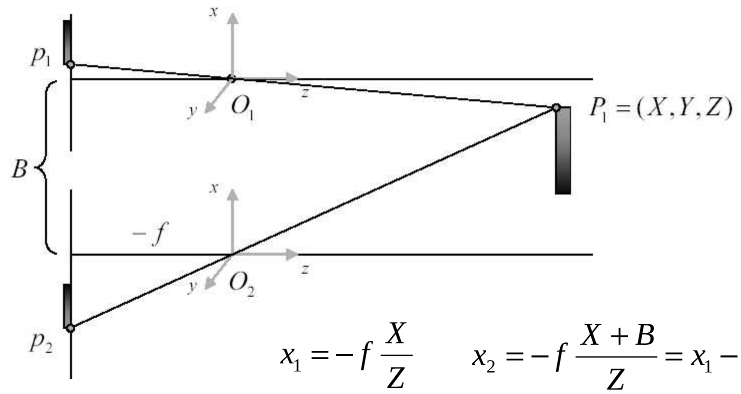
$$x = -f \frac{X}{Z}$$

Basic Stereo Derivations



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

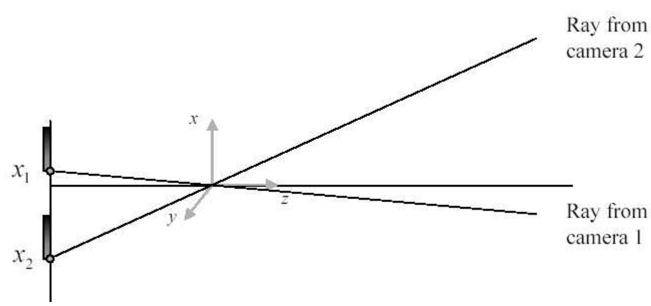
Basic Stereo Derivations



$$x_1 = -f\,\frac{X}{Z} \qquad x_2 = -f\,\frac{X+B}{Z} = x_1 - f\,\frac{B}{Z}$$

$$Z = \frac{fB}{x_1 - x_2}$$

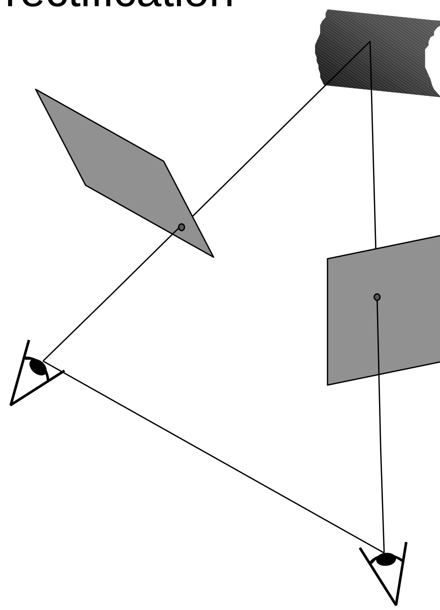
Basic Stereo Derivations



Define the disparity: $d = x_1 - x_2$

$$Z = \frac{fB}{d}$$

Stereo image rectification



Stereo image rectification

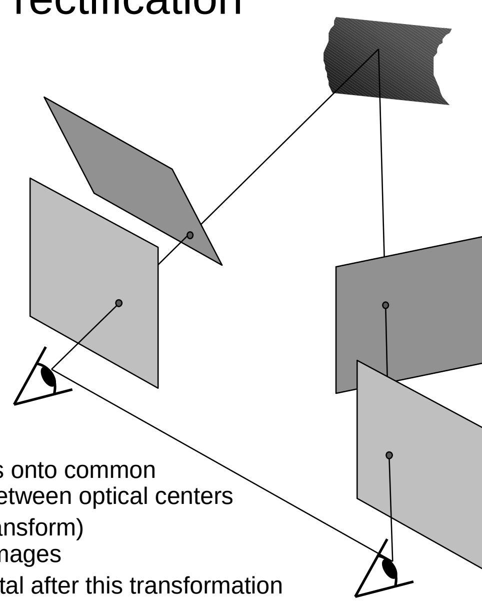


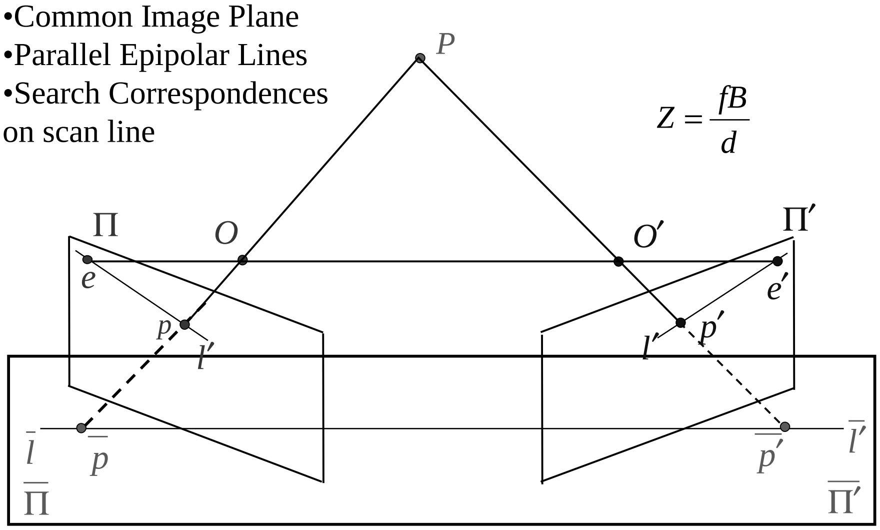
Image Reprojection

- reproject image planes onto common plane parallel to line between optical centers
- a homography (3x3 transform) applied to both input images
- pixel motion is horizontal after this transformation
- C. Loop and Z. Zhang. Computing Rectifying Homographies for Stereo Vision. IEEE Conf. Computer Vision and Pattern Recognition, 1999.

Image Rectification

- Common Image Plane
- Parallel Epipolar Lines
- Search Correspondences on scan line

$$Z = \frac{fB}{d}$$



Reconstruction

$$p = f\,\frac{P}{Z}$$

$$p' = f'\,\frac{P'}{Z'}$$

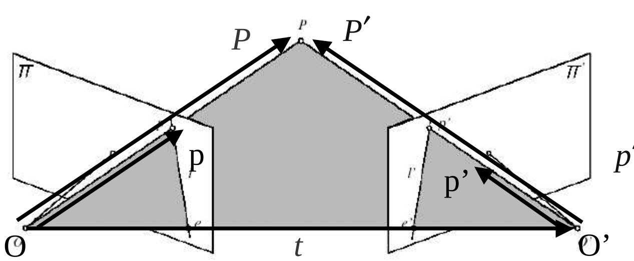


FIGURE 11.1: Epipolar geometry: the point P, the optical centers O and O' of the two cameras, and the two images p and p' of P all lie in the same plane.

$$P = RP' + t$$

$$P' = R^{-1}(P - t) = R^{T}(P - t)$$

Reconstruction

$$p' = f' \frac{P'}{Z'}$$

$$P' = R^T(P - t) = R'(P - t) \qquad R' = \begin{bmatrix} R_1'^T \\ R_2'^T \\ R_3'^T \end{bmatrix}$$

$$p' = f' \frac{R'(P - t)}{R_3'^T (P - t)}$$

$$x' = f' \frac{R_1'^T (P - t)}{R_3'^T (P - t)} \qquad \text{Equation 1}$$

$$p = f \frac{P}{Z} \Rightarrow P = \frac{pZ}{f} \qquad \text{Equation 2}$$

$$Z = f \frac{(x'R_3' - f'R_1')^T t}{(x'R_3' - f'R_1')^T p}$$

(From equations 1 and 2)

Reconstruction up to a Scale Factor

- Assume that intrinsic parameters of both cameras are known
- Essential Matrix is known up to a scale factor (for example, estimated from the 8 point algorithm).

Reconstruction up to a Scale Factor

$$\mathcal{E} = k[t_\times]R$$

$$\mathcal{E}\mathcal{E}^T = k^2[t_\times]RR^T[t_\times]^T = k^2[t_\times][t_\times]^T = \begin{bmatrix} k^2(T_Y^2+T_Z^2) & -k^2T_XT_Y & -k^2T_XT_Z \\ -k^2T_XT_Y & k^2(T_X^2+T_Z^2) & -k^2T_YT_Z \\ -k^2T_XT_Z & -k^2T_YT_Z & k^2(T_X^2+T_Y^2) \end{bmatrix}$$

$$Trace[\mathcal{E}\mathcal{E}^T] = 2k^2(T_X^2+T_Y^2+T_Z^2) = 2k^2\|t\|^2$$

$$\frac{\mathcal{E}}{|k|\|t\|} = \mathrm{sgn}(k)\frac{[t_\times]}{\|t\|}R = \mathrm{sgn}(k)\left[\left(\frac{t}{\|t\|}\right)_\times\right]R = \mathrm{sgn}(k)[\hat{t}_\times]R = \hat{E}$$

$$\hat{E}\hat{E}^T = [\hat{t}_\times][\hat{t}_\times]^T = \begin{bmatrix} 1-\hat{T}_X^2 & -\hat{T}_X\hat{T}_Y & -\hat{T}_X\hat{T}_Z \\ -\hat{T}_X\hat{T}_Y & 1-\hat{T}_Y^2 & -\hat{T}_Y\hat{T}_Z \\ -\hat{T}_X\hat{T}_Z & -\hat{T}_Y\hat{T}_Z & 1-\hat{T}_Z^2 \end{bmatrix}$$

Reconstruction up to a Scale Factor

$$\hat{E} = \begin{bmatrix} \hat{E}_1^T \\ \hat{E}_2^T \\ \hat{E}_3^T \end{bmatrix} \qquad R = \begin{bmatrix} R_1^T \\ R_2^T \\ R_3^T \end{bmatrix}$$

Let $w_i = \hat{E}_i \times \hat{t},\ i \in \{1,2,3\}$

It can be proved that

$$R_1 = w_1 + w_2 \times w_3$$
$$R_2 = w_2 + w_3 \times w_1$$
$$R_3 = w_3 + w_1 \times w_2$$

Reconstruction up to a Scale Factor

We have two choices of **t**, (**t**$^+$ and **t**$^-$**)** because of sign ambiguity and two choices of **E,** (E^+ and E^-).

This gives us four pairs of translation vectors and rotation matrices.

Reconstruction up to a Scale Factor

Given $\hat{E}$ and $\hat{t}$

1. Construct the vectors **w**, and compute R
2. Reconstruct the Z and Z' for each point
3. If the signs of Z and Z' of the reconstructed points are
   a) both negative for some point, change the sign of $\hat{t}$ and go to step 2.
   b) different for some point, change the sign of each entry of $\hat{E}$ and go to step 1.
   c) both positive for all points, exit.

$$Z = f \frac{(x'R'_3 - f'R'_1)^T t}{(x'R'_3 - f'R'_1)^T p}$$

$$Z' = -f' \frac{(xR_3 - fR_1)^T (t)}{(xR_3 - fR_1)^T p'}$$

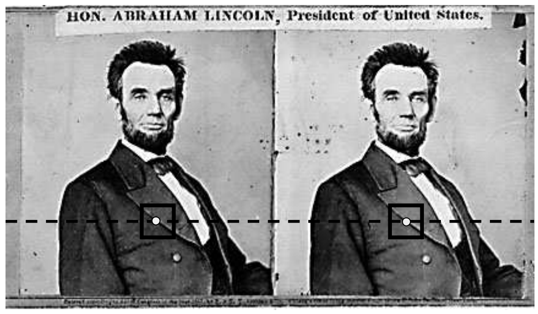
Finding Correspondences

Stereo matching algorithms

Match Pixels in Conjugate Epipolar Lines

- Assume brightness constancy
- This is a tough problem
- Numerous approaches
  - dynamic programming [Baker 81,Ohta 85]
  - smoothness functionals
  - more images (trinocular, N-ocular) [Okutomi 93]
  - graph cuts [Boykov 00]
- A good survey and evaluation: http://www.middlebury.edu/stereo/

Your basic stereo algorithm



For each epipolar line

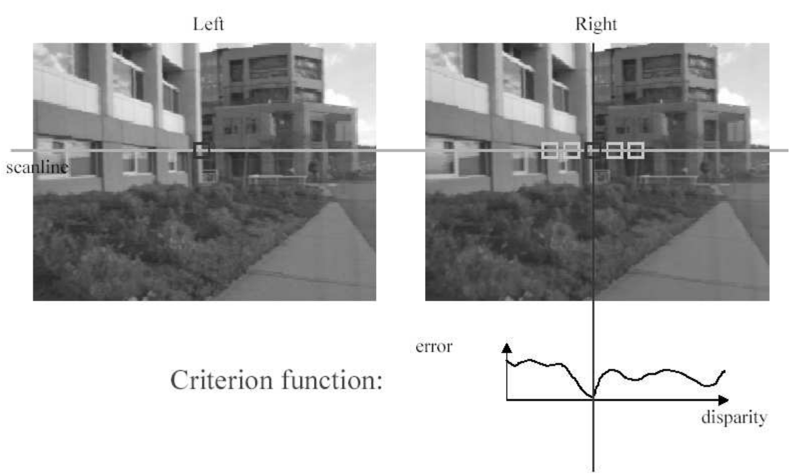
For each pixel in the left image

- compare with every pixel on same epipolar line in right image
- pick pixel with minimum match cost

Improvement: match ***windows***

- This should look familar...
- Can use Lukas-Kanade or discrete search (latter more common)

Correspondence using Discrete Search



Sum of Squared Differences (SSD)

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:

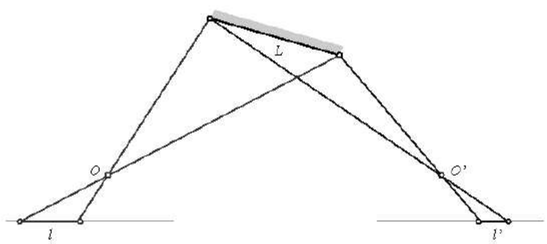
$$\bar{I} = \frac{1}{|W_m(x,y)|} \sum_{(u,v)\in W_m(x,y)} I(u,v)$$ Average pixel

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

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

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].

Problems with window matching

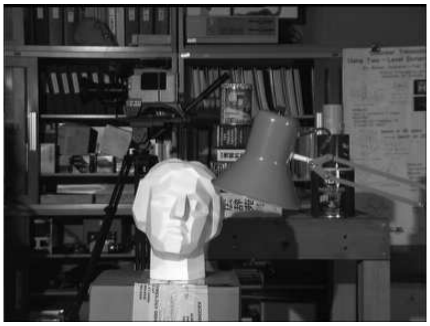
Patch too small?

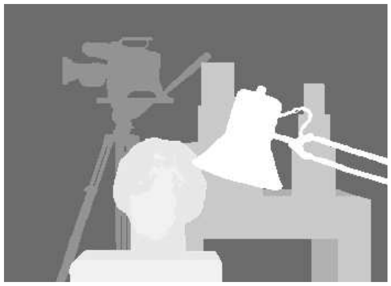
Patch too large?

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

Stereo results

- Data from University of Tsukuba
- Similar results on other images without ground truth

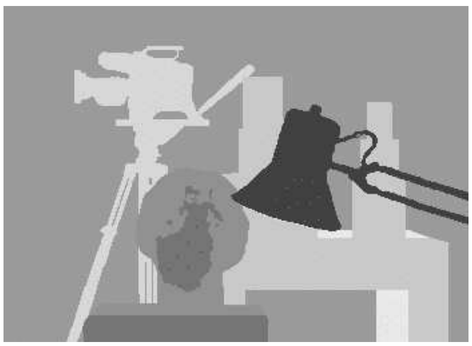




Scene

Ground truth

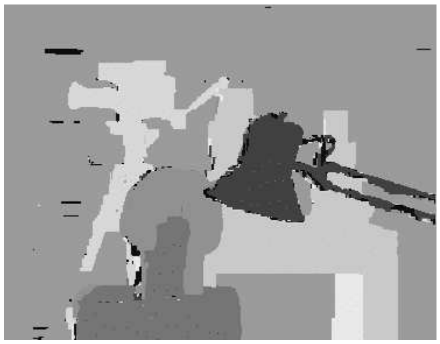
Results with window correlation

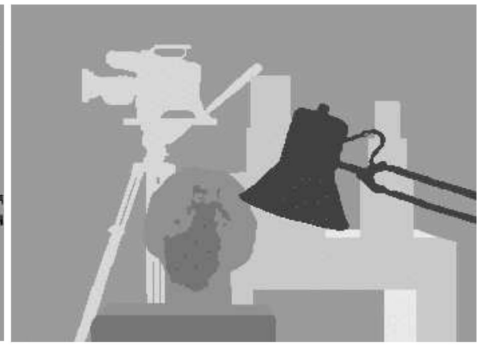


Window-based matching
(best window size)

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

Final Exam

Thursday, April 24, 2003
19:00-21:45