Stereo

CS 510

May 8th, 2013

Colorado State University

Where are we?

- We are done! (essentially)
- We covered image matching
  - Correlation & Correlation Filters
  - Fourier Analysis
  - PCA
- We covered feature-based matching
  - Bag of Features approach
  - Constellation approach
  - Including:
    - Feature Extraction
    - Feature Description
    - Classification
- Not a bad introduction to vision

But we skipped a few topics…

- Like 3D vision
  - 3D sensors (e.g. Kinect)
  - Stereo (i.e. multiple overlapping cameras)
  - Structure from Motion
- Like Video
  - Object tracking
  - Motion segmentation
  - Activity recognition

Let's fix one omission: Stereo

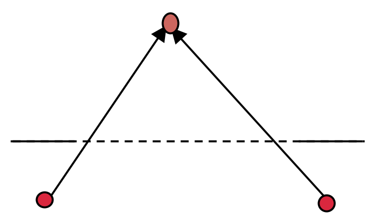
- The ability to infer 3D structure and distance from two or more overlapping images taken simultaneously from different viewpoints

Are these stereo images? Describe the viewpoints

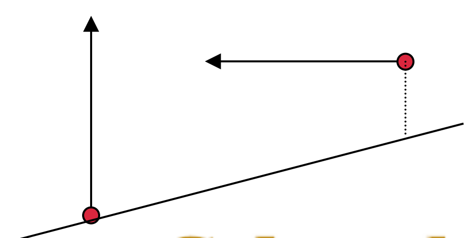
Scenarios

- Most common: perpendicular optical axes

- Also common: converging optical axes (e.g. eyes)



- More common than you might think: arbitrary axes



Two SubProblems:

- Image Matching (correspondence)
  - identifying which points in image #1 match which points in image #2
  - note: not all points in image #1 match anything in image #2. *Why not?*
  - Note: not all matching points can be found.

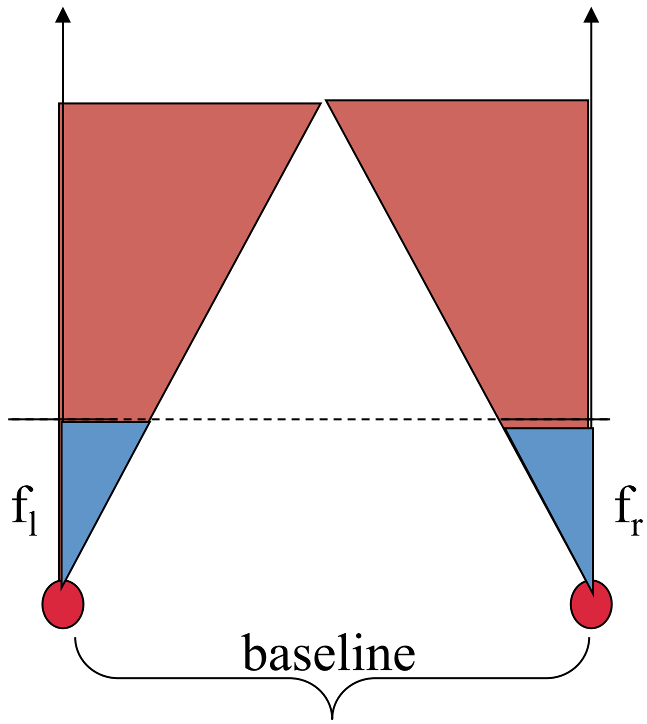
- Reconstruction
  - Given point matches, determine their 3D position
  - Requires triangulation (implicit or explicit)

Image Matching

- Find common scene points in two images
  - Occlusion
  - Incomplete overlap of visual fields
  - Potentially strong perspective effects
- General Methods:
  - Correlation based
    - Cross-correlate every pixel in left image to right image
    - Epipolar geometry can constrain this search…
  - Feature based
    - Extract points, edges, lines, etc., and match them across image

Reconstruction as Triangulation

- Assume that the positions and baselines of the cameras are known:



$$P = t_l\left[x_l, y_l, f_l\right]$$

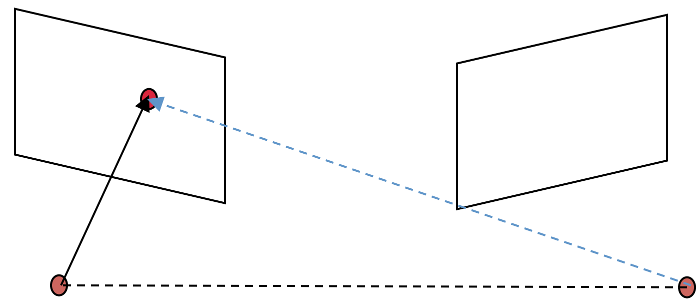
$$P = \left[b_x, b_y, b_z\right] + t_r\left[x_r, y_r, f_r\right]$$

Solve for t's, compute coordinate of point.

Q: Isn't this overconstrained?

Epipolar Geometry

- For any point in image #1, there is a line of points in image #2 such that its match (if one exists) must lie on that line.



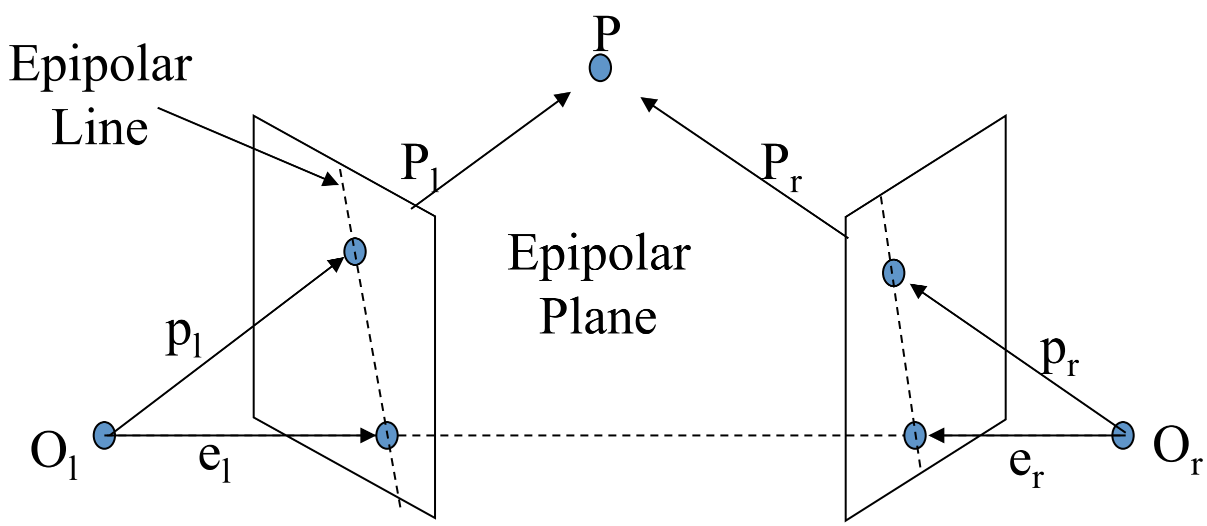
  - There is a plane defined by the two focal points and the 2D point in image #1. The 3D point must lie in this plane.
  - Also, the matching point in image #2 must lie in this plane.

Epipolar (cont.)

- Since the intersection of two planes is a line, there is a line in image #2 on which the matching point must lie. This is called the *epipolar line*.
- If you know the vrp and prp of both cameras, you can compute the epipolar line for any point in image #1.
  - If axes are parallel and B_z=0, then the epipolar lines are scan lines.
- The Essential Matrix (E) allows you to compute epipolar geometry without knowing the camera parameters *a priori*

Getting Formal about Stereo

Do not panic about the next N slides; my goal is just to expose you to terms & concepts in case you go to a vision conference…



Basic Equations

$$P_r = R(P_l - T)$$ 1:Relation between 3D views of point P

$$T \times P_l$$ 2:Normal to epipolar plane

$$(P_l - T)^T \cdot (T \times P_l) = 0$$ 3:Planarity constraint

$$(R^T P_r)^T \cdot (T \times P_l) = 0$$ 4:Rewrite of #3, using #1

Colorado State University

A Clever Equation

You can rewrite a cross product as dot product, so

$$T \times P_l = SP_l$$

where

$$S = \begin{bmatrix} 0 & -T_z & T_y \\ T_z & 0 & -T_x \\ -T_y & T_x & 0 \end{bmatrix}$$

More Equations

$$\left(R^T P_r\right)^T S P_l = 0$$

5: Substitute dot for cross in #4

$$P_R^T R S P_l = 0$$

6: Apply transpose equivalency

$$P_R^T E P_l = 0$$

7: Let RS = E

E is called the Essential Matrix. It is rank 2 (because of S), and shows a linear relationship between the projections of points in two images

Or in 2D….

$$p_l = \frac{f_l}{Z_l} P_l$$ 8: Definition of perspective

$$P_l = \frac{Z_l}{f_l} p_l$$ 9: same

$$\left(\frac{Z_r}{f_r} p_r\right)_R^T E\left(\frac{Z_l}{f_l} p_l\right) = 0$$ 10: rewrite of #7, with #8

$$p_R^T E p_l = 0$$ 11: drop non-zero constants

Colorado State University

Back to Epipolar…

- So E is a linear relation between p_l and p_r
- $u_r = Ep_l$, where u_r is the line of points in R that might match point p_l
- If you know E
  - For every image point p_l:
    - calculate the line u_r
    - only cross-correlate along that line
- E can be calculated from 8 image correspondences
  - Why 8? (How many DOF? How many constraints per correspondence?)

Stereo Practicum

- The larger the baseline, the more the perspective distortion
  - The harder it is to match points
- The smaller the baseline, the smaller the angle between P_l and P_r, the higher the reconstruction error.
  - Errors always highest in Z…