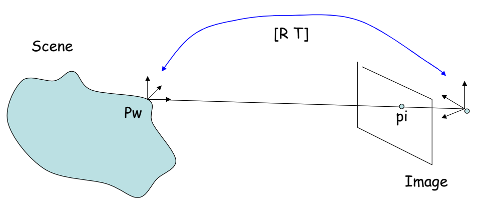
MultiView Geometry and Reconstruction

- Last day: Vision Overview
  - Image formation (light,surfaces,projection)
  - Computing features to improve salience and tractability (edges, corners, segments etc)
  - Adding constraints to understand the world:
    - Simplified camera (pinhole) and surfaces (Lambertian)
    - Model and template based recognition
- Today Multiview Geometry and Reconstruction (Forsyth&Ponce ch10 & 11)

Quiz # 4

- How is Computer Vision different from Image Processing?
  - Image Processing computes (modified) images, Computer Vision infers something about world state.
- What are the important factors in image formation?
  - Light sources, surface properties, sensor properties (camera model) including their relative positions.
- What is visual correspondence?
  - Finding (matching) the projections of the same world point in multiple images.
- Why is feature extraction useful?
  - Reduces the amount of raw data (tractability).
  - Emphasizes task relevant image properties (salience).

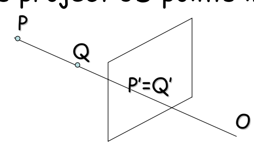
A Word on Coordinate Frames



- Typically we will have a world coordinate frame W and a camera coordinate frame related by Pc=[R T]Pw
- Points Pc are projected to the image frame by pi=KPc
- Points in the image plane can be described by 3D camera frame (normalized) coordinates, or pixel coordinates.

Why MultiView Vision?

- 2D images project 3D points into 2D:



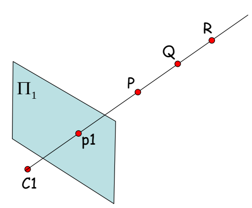
- 3D Points on the same viewing line have the same 2D image:
  - 2D imaging results in depth information loss

(Camps)

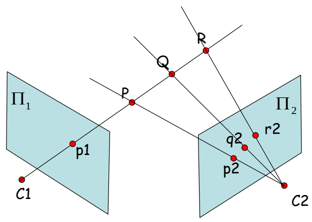
Multi-View Geometry

- Single image doesn't indicate the depth of a scene point along projection ray.
- 2 or more images allow depth measurement via triangulation.
- How do multiple views of the same scene help us infer something about
  - 3D scene structure?
  - camera configuration?
- Add more information add more constraints!

Epipolar Geometry

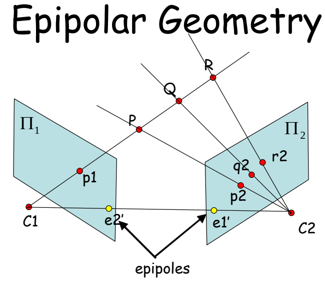


Epipolar Geometry



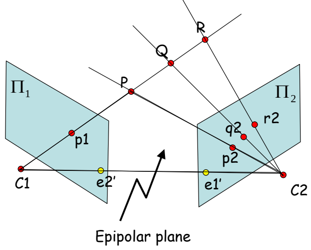
Second View can distinguish points which give rise to p1.
IF we can determine **correspondence** – which point p2,q2,r2 arises from the same world point as p1?

Epipolar Geometry

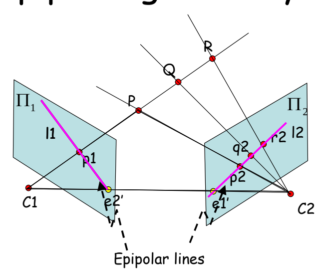


- Epipoles are projections of camera centers from other views.

Epipolar geometry



Epipolar geometry



Epipolar constraint: Correspondence search for point p1 is constrained to *epipolar line* l2. l2 is the projection of the viewing ray C1->P in Π2.

Epipolar Geometry Summary

- **Epipole:** ***e1'*** projection of optical centre C1 of one camera in the image of the other.
- **Epipolar plane** for a point *P* and cameras with optical centres *C1* and *C2*, is defined by intersecting rays *C1P* and *C2P*.
- **Epipolar line** *l2* for *p2* is the intersection of this plane and the image plane *Π2*. It passes through the **epipole** *e2* where the baseline joining *C1* and *C2* intersects *Π2*.
- ***Epipolar constraint***: for projections *p1* and *p2, p2* must lie on the epipolar line associated with *p1*.

Camera Parameters

- Extrinsic parameters: map world coordinates to camera coordinates [R t]
- Intrinsic parameters relate normalized coordinates X/Z, Y/Z to image plane:

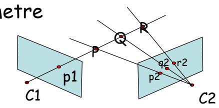
$$u_p = \frac{fX^c}{s_x Z^c} + u_o, \quad v_p = \frac{fY^c}{s_y Z^c} + v_o$$

  - Focal length, f
  - Principal point (centre of image plane) (Uo, Vo)
  - Scale parameters sx, sy – based on sensor pixel
  - Skew parameters α and β
  - Encoded as K matrix

$$K = \begin{bmatrix} f/s_x & \alpha & u_0 \\ 0 & \beta f/s_y & v_0 \\ 0 & 0 & 1 \end{bmatrix}$$

Stereo

- Stereopsis: human ability to fuse images from 2 eyes to give a strong sense of depth.
  - Measures eye rotation required to fixate a point
  - Involves CNS and combined left and right image stimuli
  - Only effective to about 1 metre
- Stereo Applications
  - Robot navigation
  - Aerial reconnaissance
  - Automated 3D modeling



Stereo

- **Stereo:** computing scene depth (shape) from two or more views.
- 2 Problems:
  - **Correspondence:** Which parts of the left and right images are projections of the same scene element?
  - **Reconstruction**: Given correspondences, and possibly camera geometry, what can we say about 3D position and structure of the observed scene?

Terms

- **Disparity**: for scene point P projected to p1 in image 1 and p2 in image 2,
  d=p1-p2
- **Disparity map**: disparity at each pixel/feature (sometimes 2 ½ D sketch)
- **Correspondence**: for image point p1 determining the point p2 which arises from the same scene point P.

Correspondence

- Assumptions:
  - Most scene points are visible from both viewpoints.
  - Corresponding image regions have similar appearance in all views (Lambertian)
  - Fixation distance much larger than baseline
- ***Correspondence as a search problem***: given an element in the left image, find the corresponding element in the right image.
  - Similarity measure.
  - Correlation or feature based

Image Rectification

- Exploits the epipolar constraint
  - Emulates case where optical axes are parallel, epipoles at infinity
- Determines transform (warp) of each image such that pairs of conjugate epipolar lines become collinear and parallel to one of the image axes.
- Correspondence reduced to 1-D search along scanline (d=u1-u2).
- **Problem:** compute transform making conjugate epipolar lines collinear and parallel to horizontal axis.
- Map images onto a common plane parallel to baseline
  - only focal point of camera really matters

Rectification

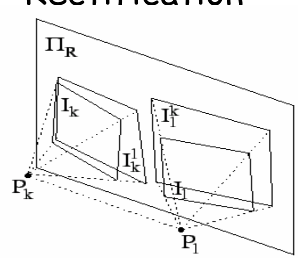
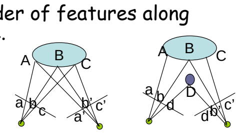


Figure 7.4: Planar rectification: $(\mathbf{I}_k^l, \mathbf{I}_l^k)$ are the rectified images for the pair $(\mathbf{I}_k, \mathbf{I}_l)$ (the plane Π_R should be parallel to the baseline $(\mathbf{P}_k, \mathbf{P}_l)$).

Correspondence

- Computing similarity and finding optimal match.
- Additional assumptions:
  - ***Epipolar constraint:*** search in 1D
  - ***Uniqueness constraint:*** a point in one image should have at most one corresponding point in the other image.
  - ***Continuity Constraint:*** disparity tends to vary slowly across a surface, prefer disparity similar to neighbors
  - ***Ordering constraint:*** order of features along epipolar lines is the same.



Similarity

- Searching for image points/features/regions arising from the same world point.
  - Need to measure similarity!
- Comparing individual pixels is not robust
  - Not statistically meaningful
  - Use extended windows to measure similarity!
  - Use features (edges, corners, patches)

Correlation-Based Correspondence

- Elements matched are fixed size image windows
  - Assumes appearance is constant across window (frontoparallel, Lambertian)
- Similarity measures are correlation scores between windows in the two images:
  - Sum of Absolute Differences (SAD)

    $$C(d) = \sum_W \left| w_{i,j} - w'_{i,j} \right|$$

  - Sum of Squared Differences (SSD)

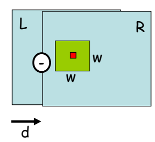
    $$C(d) = \sum_W \left(w_{i,j} - w'_{i,j} \right)^2$$

  - Normalized Cross Correlation (NCC)

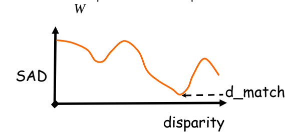
    $$C(d) = \frac{(w - \overline{w}) \cdot (w' - \overline{w}')}{\left| w - \overline{w} \right| \cdot \left| w' - \overline{w}' \right|}$$

- Correspondence is determined by optimal correlation score within the search region (fixed disparity range).

SAD Correlation Matching



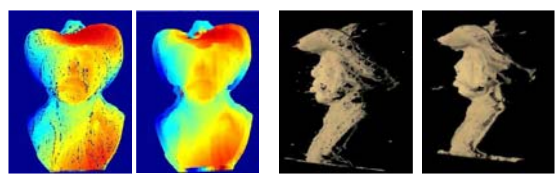
$$SAD(d) = \sum_W \left| w_{i,j} - w'_{i,j} \right|$$



Problems with Correlation

- Pixels in regions with little variation all match equally well
  - Larger window improves statistics, but degrades localization
- Repetitive patterns: (stripes etc) each repetition matches equally well.
- Boundary overreach: at occluding contours continuity and smoothness are violated, but edge gives strong matching feature.
- Half occlusions: regions visible in one image but occluded in the other
  - No valid match exists
  - **Left-right check-** compute matches left to right then right to left, inconsistent disparities imply occlusion

Correlation Window Size



- Window dimension 7 vs 32
- Boundary overreach
- Larger window more variation better match support but more distortion

Other Computational Approaches

- Feature-based Correspondence
  - Restrict correspondence search to a sparse set of features (points, lines, corners).
  - Yields sparse depth maps
- Hierarchical Approaches
  - Fine-to-fine approaches
  - coarse-to-fine approaches

- Optimization: Use a search technique to optimize a cost (energy) function which encodes desired matching constraints
  - **Dynamic Programming**
  - **Graph Cuts**

Stereo results

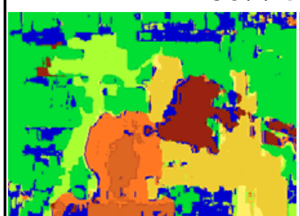
- Data from University of Tsukuba

Scene

Ground truth

(Seitz)

Results with window correlation



Window-based matching
(best window size)

Ground truth

(Seitz)

Results with better method

State of the art method

Ground truth

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

(Seitz)

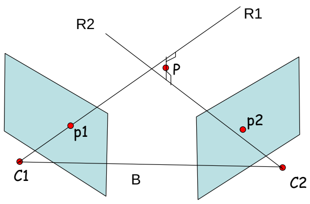
Disparity Refinement

- Matching Metric
  - choose high score matches
- Left-Right Check
  - compute correspondence from left to right, then right to left and compare result
- Suppress Homogeneous Regions
  - ignore regions with low gradients
- Neigbourhood filtering
  - enforce similarity to neighbouring pixels
- Peak properties:
  - curvature of metric, peak ratio (subpixel disparity)

3D Reconstruction

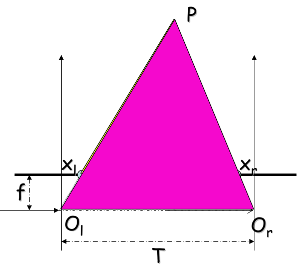
- Given correspondences there are 3 cases depending on prior knowledge:
  - Known intrinsics and extrinsics: unambiguous reconstruction by triangulation
  - Only intrinsics known: reconstruct and estimate extrinsics up to scale factor.
  - Only pixel correspondences known: reconstruction only up to an unknown global projective transformation.

General Reconstruction



- R1 and R2 do not in practice intersect: find midpoint of shortest perpendicular.
- For rectified system z=-B/d, P=-(B/d)p, for normalized point p=(u,v,1)

Reconstruction with rectified geometry (much easier).



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

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

Disparity: $d = x_l - x_r$

$$Z = f\frac{T}{d}$$

Then given Z, we can compute X and Y.

T is the stereo baseline (known for calibrated cameras)
d measures the difference in retinal position between corresponding points

(Camps)

Middlebury Stereo Page

- http://cat.middlebury.edu/stereo/
- Metric comparison of correspondence algorithms
- Top 3:
  - J. Sun, Y. Li, S. B. Kang, and H.-Y. Shum. Symmetric stereo matching for occlusion handling. CVPR 2005.
  - A symmetric patch-based correspondence model for occlusion handling. Y. Deng, Q. Yang, X. Lin, and X. Tang. ICCV 2005.
  - L. Hong and G. Chen. Segment-based stereo matching using graph cuts. CVPR 2004

Quiz #5

- What do additional (more than 1) images give us?
- What is Stereo disparity?
- Why is the epipolar constraint useful in calculating correspondence?
- How is disparity related to depth (Z)?