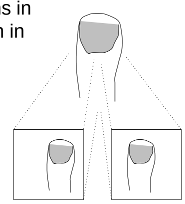
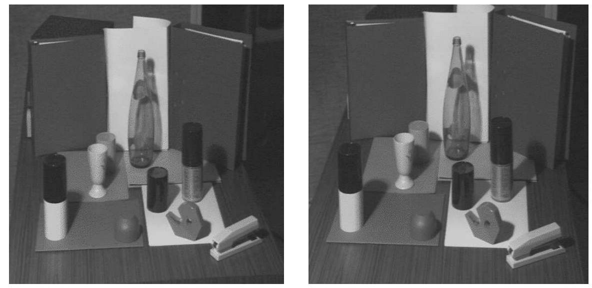
Stereo Vision

- Stereo vision - inferring 3-D structure from two images taken from different viewpoints.
- Object appears in different positions in each image depending on its depth in the scene.
- Depth $\propto$ position difference
- 3-D structure from stereo images.



Stereo Pair

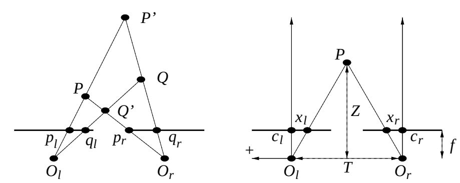


The Two Problems of Stereo

- To estimate depth from a pair of stereo images we need to solve two main problems .
- **Correspondence problem**:
  - for all items in the left image, find their corresponding item in the right image
  - 'items' - pixels, features (edges, etc), regions, objects, etc
- **Reconstruction problem**:
  - using the estimated disparities between items, reconstruct the 3-D structure of the scene
  - needs additional information about the cameras and assumptions about the scene

Triangulation

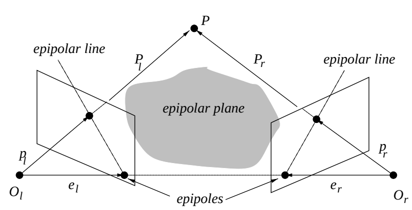
- A simple stereo system (parallel optical axes):



- Using similar triangles: $\frac{T+x_l-x_r}{Z-f} = \frac{T}{Z} \quad \rightarrow \quad Z = \frac{fT}{x_r - x_l}$
- Need to know: $d = x_r - x_l$, f, T, c_l and c_r to compute Z

Epipolar Geometry

- The general stereo problem:



Stereo Extrinsic Parameters

- Vectors P_l and P_r refer to same 3-D point P with respect to left and right camera frames respectively.
- Relationship between P_l and P_r given by rotation matrix R and translation T:
$$\mathrm{P}_r = R(\mathrm{P}_l - \mathrm{T}) \qquad (1)$$
- Defines the *extrinsic parameters* of the stereo system.
- Image points p_l and p_r (defined wrt camera frames) related to 3-D points by perspective equations:
$$\mathrm{p}_l = f_l \mathrm{P}_l / Z_l \qquad \mathrm{p}_r = f_r \mathrm{P}_r / Z_r \qquad (2)$$

Epipolar Constraint

- Each 3-D point defines an epipolar plane intersecting each image along the epipolar line.
- Given a point p_l in left image, correct match in right image MUST lie along corresponding epipolar line.
- Epipolar plane therefore constrains locations of possible matches for points in either image.
- Known as the *epipolar constraint*.
- Can be used to help in solving correspondence problem - 2-D search becomes a 1-D search (more later).

Finding the Epipolar Line

- Knowing the epipolar line for a point helps to find its corresponding point in the other image.
- Therefore need way of determining equation for the epipolar line.
- Also needed to solve opposite problem - determining extrinsic parameters of stereo system given set of corresponding points, ie calibration.
- Need to compute two matrices:
  - the *essential matrix*, defining relationship between an image point defined wrt to camera coordinates and the epipolar line;
  - the *fundamental matrix*, defining relationship between an image point defined wrt to pixel coordinates and the epipolar line.

The Essential Matrix

- The 3 vectors P_l, T, and $(\mathrm{P}_l - \mathrm{T})$ all lie in the epipolar plane.
- Equation of the plane is therefore (using eqn (1)):

$$(\mathrm{P}_l - \mathrm{T})^T(\mathrm{T} \times \mathrm{P}_l) = 0 \quad \rightarrow \quad (R^T\mathrm{P}_r)^T(\mathrm{T} \times \mathrm{P}_l) = 0 \qquad (3)$$

- NB: the *cross product* $\mathrm{T} \times \mathrm{P}_l$ is a vector perpendicular to the plane containing T and P_l, ie the epipolar plane, and since $\mathrm{P}_l - \mathrm{T}$ is also in the plane, the *dot product* $(\mathrm{P}_l - \mathrm{T})^T(\mathrm{T} \times \mathrm{P}_l)$ is zero.
- Cross product can be written as:

$$\mathrm{T} \times \mathrm{P}_l = S\mathrm{P}_l \quad \rightarrow \quad S = \begin{bmatrix} 0 & -T_z & T_y \\ T_z & 0 & -T_x \\ -T_y & T_x & 0 \end{bmatrix} \qquad (4)$$

The Essential Matrix ...`cont..`

- From eqns (3) and (4) we get: $\mathrm{P}_r^T E \mathrm{P}_l = 0 \quad \rightarrow \quad E = RS$
- Matrix E is known as the *essential matrix* - defines epipolar constraint in terms of system extrinsic parameters.
- Also, using eqn (2) and dividing by $Z_r Z_l / f_r f_l$ gives

$$\mathrm{p}_r^T E \mathrm{p}_l = 0$$

- Hence, for image point p_l, corresponding point p_r must satisy:

$$\mathrm{p}_r^T \mathrm{u}_r = 0 \quad \rightarrow \quad \mathrm{u}_r = E\mathrm{p}_l$$

- Since $\mathrm{p}_r = [x_r, y_r, f]^T$, this gives the equation of the epipolar line in the right image (f is a constant):

$$x_r u_{rx} + y_r u_{ry} + f u_{rz} = 0$$

The Fundamental Matrix

- If $\bar{\mathrm{p}}_l$ represents a point in the left image in pixel coordinates, then from eqn (3) in Lecture 1:

$$\mathrm{p}_l = M_l^{-1}\bar{\mathrm{p}}_l \qquad M_l = \begin{bmatrix} 1/s_x & 0 & o_x/f \\ 0 & 1/s_y & o_y/f \\ 0 & 0 & 1 \end{bmatrix}$$

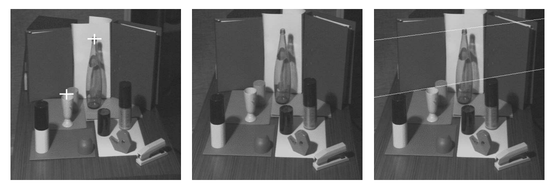
- Given a corresponding point $\bar{\mathrm{p}}_r$, also in pixel coordinates, then

$$\bar{\mathrm{p}}_r^T F \bar{\mathrm{p}}_l = 0 \quad \rightarrow \quad F = (M_r^{-1})^T E M_l^{-1}$$

- Matrix F is known as the *fundamental matrix* - defines epipolar line in pixel coordinates, ie if $\bar{\mathrm{u}}_r = F\bar{\mathrm{p}}_l$ then epipolar line given by:

$$\bar{x}_r \bar{u}_{rx} + \bar{y}_r \bar{u}_{ry} + f \bar{u}_{rz} = 0$$

Epipolar Lines - Example



Determining F from Correspondences

- If we know corresponding points, then can determine F (or E)
- Enables epipolar lines to be found without need to calibrate camera
- We can rewrite $\bar{\mathrm{p}}_r^T F \bar{\mathrm{p}}_l = 0$ in form

$$\sum_i \sum_j a_{ij} F_{ij} = 0$$

- For n correspondences we have n such equations, giving homogeneous linear system

$$A\bar{\mathrm{f}} = 0$$

where A is $n \times 9$ and $\bar{\mathrm{f}}$ is vector containing elements of F.

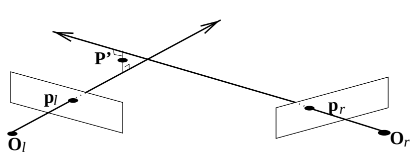
- Solution up to scale factor can be obtained using *singular value decomposition* (SVD) if $n \geq 8 \rightarrow$ 8-point algorithm.

3-D Reconstruction

- Given a set of corresponding points now need to compute 3-D coordinates.
- Complexity of problem depends on whether stereo system is calibrated.
- Also - reconstruction will always result in sparse set of 3-D points - need to fill in the gaps, eg using surface fitting.
- Requires model of 3-D structure, eg piecewise planar, B-splines, etc.
- Relatively easy case - determining 3-D points using triangulation for calibrated system.

Reconstruction by Triangulation

- Given corresponding points p_l and p_r, need to determine where rays $a\mathrm{p}_l$ and $b\mathrm{p}_r$ intersect, ie need to find a and b.
- WRT left camera frame rays given by $a\mathrm{p}_l$ and $\mathrm{T} + bR^T\mathrm{p}_r$
- In general, rays will not intersect:



- In practice - find closest point to both rays.

Finding the Best 3-D Point

- Denote rays $a\mathrm{p}_l$ and $\mathrm{T} + bR^T\mathrm{p}_r$ by l and r.
- Required point P' is then the mid-point of segment which is perpendicular to l and r AND joins l and r.
- We can find the endpoints of this segment, say $a_o\mathrm{p}_l$ and $\mathrm{T} + b_oR^T\mathrm{p}_r$, by solving the following system of 3 linear equations:

$$a\mathrm{p}_l - bR^T\mathrm{p}_r - \mathrm{T} + c(\mathrm{p}_l \times R^T\mathrm{p}_r) = 0$$

- $a\mathrm{p}_l - bR^T\mathrm{p}_r - \mathrm{T}$ is a segment joining l and r
- $c(\mathrm{p}_l \times R^T\mathrm{p}_r)$ is perpendicular to l and r
- Choosing a, b and c to make their difference $= 0$ therefore gives the required segment and hence the mid-point P'.