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 ∝ position difference
• 3-D structure from stereo images.

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):

\[ \frac{x_r - x_l}{Z-f} = \frac{T_z}{Z} \rightarrow Z = \frac{fT_z}{x_r - x_l} \]

• Need to know: \( d = x_r - x_l \), \( f \), \( T \), \( c_l \) and \( c_r \) to compute \( Z \)

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 \( T \):

\[ P_r = R(P_l - T) \] (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:

\[ P_l = f_lP_l/Z_l \quad P_r = f_rP_r/Z_r \] (2)

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 $(P_l - T)$ all lie in the epipolar plane.
- Equation of the plane is therefore (using eqn (1)):
  $$(P_l - T)^T(T \times P_l) = 0 \quad \rightarrow \quad (R^T P_r)^T(T \times P_l) = 0 \quad (3)$$
- NB: the cross product $T \times P_l$ is a vector perpendicular to the plane containing $T$ and $P_l$, i.e., the epipolar plane, and since $P_l - T$ is also in the plane, the dot product $(P_l - T)^T(T \times P_l)$ is zero.
- Cross product can be written as:
  $$T \times P_l = SP_l \quad \rightarrow \quad S = \begin{bmatrix} 0 & -T_z & T_y \\ T_z & 0 & -T_x \\ -T_y & T_x & 0 \end{bmatrix} \quad (4)$$

The Fundamental Matrix

- If $p_l$ represents a point in the left image in pixel coordinates, then from eqn (3) in Lecture 1:
  $$p_l = M_l^{-1} \tilde{p}_l \quad M_l = \begin{bmatrix} 1/s_x & 0 & a_x/f \\ 0 & 1/s_y & a_y/f \\ 0 & 0 & 1 \end{bmatrix}$$
- Given a corresponding point $p_r$, also in pixel coordinates, then
  $$p_l^T F p_l = 0 \quad \rightarrow \quad F = (M_l^{-1})^T E M_l^{-1}$$
- Matrix $F$ is known as the fundamental matrix - defines epipolar line in pixel coordinates, i.e., if $\vec{u}_l = F \tilde{p}_l$ then epipolar line given by:
  $$\bar{x} \bar{u}_x + \bar{y} \bar{u}_y + \bar{f} \bar{u}_z = 0$$

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 $p_l^T F p_l = 0$ in form
  $$\sum_{i,j} a_{ij} F_{ij} = 0$$
- For $n$ correspondences we have $n$ such equations, giving homogeneous linear system
  $$A \vec{F} = 0$$
  where $A$ is $n \times 9$ and $\vec{F}$ 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.

Reconstruction by Triangulation

- Given corresponding points $p_l$ and $p_r$, need to determine where rays $a p_l$ and $b p_r$ intersect, i.e., need to find $a$ and $b$.
- WRT left camera frame rays given by $a p_l$ and $T + b R^T p_r$
- In general, rays will not intersect:
  $$O_l \quad O_r \quad P_l \quad P_r$$
- In practice - find closest point to both rays.

Finding the Best 3-D Point

- Denote rays $a p_l$ and $T + b R^T p_r$, 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 $a p_l$ and $T + b R^T p_r$, by solving the following system of 3 linear equations:
  $$a p_l - b R^T p_r - T + c (p_l \times R^T p_r) = 0$$
- Choosing $a$, $b$, and $c$ to make their difference $= 0$ therefore gives the required segment and hence the mid-point $p'$.