3-D Reconstruction

COMPSCI 773 S1 T

VISION GUIDED CONTROL

A/P Georgy Gimel'farb

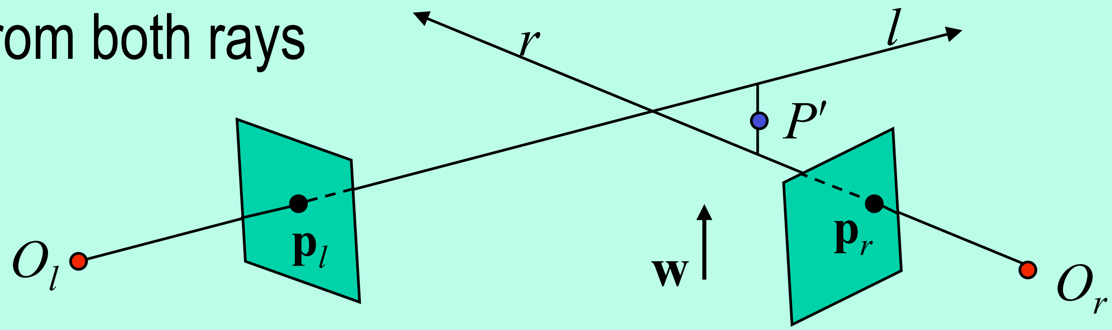
Three Basic Cases

Depending on the amount of a priori knowledge:

1. **Both intrinsic and extrinsic parameters**: the unique reconstruction of a 3-D scene by ***triangulation***
2. **Only the intrinsic parameters**: a 3-D scene is still reconstructed and also the extrinsic parameters are estimated, but **up to an unknown scaling factor**
3. **Only pixel correspondences**: a 3-D scene is still reconstructed, but **up to an unknown, global projective transformation**

Triangulation from Projections

- Point P, projected into the pair of corresponding points $\mathbf{p}_l$ and $\mathbf{p}_r$, lies at the intersection of the two rays from O_l through $\mathbf{p}_l$ and from O_r through $\mathbf{p}_r$, respectively
  - Approximate parameters and image locations: the two rays may not actually intersect in space
  - **Estimate of the intersection**: the point of minimum distance from both rays



Triangulation (in the left reference frame)

- $a\mathbf{p}_l$; $a \in \mathbf{R} = [-\infty, \infty]$ - the ray, l, through O_l and $\mathbf{p}_l$
- $\mathbf{T} + bR^{\mathsf{T}}\mathbf{p}_r$; $b \in \mathbf{R}$ - the ray, r, through O_r and $\mathbf{p}_r$
- $\mathbf{w} = \mathbf{p}_l \times R^{\mathsf{T}}\mathbf{p}_r$ - a vector orthogonal to both l and r
  – P' - the midpoint of the segment joining l and r and parallel to $\mathbf{w}$
  – Endpoints of the segment, $a_0\mathbf{p}_l$ and $\mathbf{T} + b_0R^{\mathsf{T}}\mathbf{p}_r$ - by solving the linear system of equations for a_0, b_0, and c_0:

$$a\mathbf{p}_i + c\left(\mathbf{p}_i \times R^{\mathsf{T}}\mathbf{p}_j\right) = \mathbf{T} + bR^{\mathsf{T}}\mathbf{p}_j$$

$$\Rightarrow a\mathbf{p}_i - bR^{\mathsf{T}}\mathbf{p}_j + c\left(\mathbf{p}_i \times R^{\mathsf{T}}\mathbf{p}_j\right) = \mathbf{T}$$

Reconstruction up to a Scale

- Reconstruction by using the essential matrix
  - Only the intrinsic parameters and n point correspondences, $n \geq 8$, are known
  - Since the baseline is unknown, the true scale of the viewed scene cannot be recovered
- The estimated essential matrix, E, can only be known up to an arbitrary scale factor
  - Convenient normalisation of E - by normalising the length of the translation vector $\mathbf{T}$ to unit

$$\|T\| = \sqrt{T_x^2 + T_y^2 + T_z^2}$$

Normalisation of E

- From the definition of the essential matrix, $E = RS$:

$$E^\top E = S^\top R^\top R S = S^\top S = \begin{bmatrix} 0 & T_z & -T_y \\ -T_z & 0 & T_x \\ T_y & -T_x & 0 \end{bmatrix} \begin{bmatrix} 0 & -T_z & T_y \\ T_z & 0 & -T_x \\ -T_y & T_x & 0 \end{bmatrix}$$

$$= \begin{bmatrix} T_y^2 + T_z^2 & -T_x T_y & -T_x T_z \\ -T_y T_x & T_z^2 + T_x^2 & -T_y T_z \\ -T_z T_x & -T_z T_y & T_x^2 + T_y^2 \end{bmatrix} \Rightarrow Tr\left(E^\top E\right) = 2\|\mathbf{T}\|^2 \Rightarrow \hat{\mathbf{T}} = \frac{\mathbf{T}}{\|\mathbf{T}\|} \equiv \begin{bmatrix} \hat{T}_x \\ \hat{T}_z \\ \hat{T}_z \end{bmatrix}$$

- The normalised essential matrix:

$$\hat{E} = \frac{E}{\sqrt{Tr(E^\top E)/2}} \Rightarrow \hat{E}^\top \hat{E} = \begin{bmatrix} 1 - \hat{T}_x^2 & -\hat{T}_x \hat{T}_y & -\hat{T}_x \hat{T}_z \\ -\hat{T}_y \hat{T}_x & 1 - \hat{T}_y^2 & -\hat{T}_y \hat{T}_z \\ -\hat{T}_z \hat{T}_x & -\hat{T}_z \hat{T}_y & 1 - \hat{T}_z^2 \end{bmatrix}$$

Recovering the Pair $(\hat{\mathbf{T}}, R)$

- The components of $\hat{\mathbf{T}}$ - from any row or column of $\hat{E}^\top \hat{E}$
  - The estimates may differ from the true components by a global sign change (due to quadratic entries of $\hat{E}^\top \hat{E}$)
  - Rotation matrix - from $\hat{E}$ and $\hat{\mathbf{T}}$:

$$\hat{E} = \begin{bmatrix} \hat{\mathbf{E}}_1^\top \\ \hat{\mathbf{E}}_2^\top \\ \hat{\mathbf{E}}_3^\top \end{bmatrix} \quad \Rightarrow \quad \underset{i=1,2,3}{\mathbf{w}_i = \hat{\mathbf{E}}_i \times \hat{\mathbf{T}}} \quad \Rightarrow \quad R = \begin{bmatrix} \mathbf{R}_1^\top = \left(\mathbf{w}_1 + \mathbf{w}_2 \times \mathbf{w}_3\right)^\top \\ \mathbf{R}_2^\top = \left(\mathbf{w}_2 + \mathbf{w}_3 \times \mathbf{w}_1\right)^\top \\ \mathbf{R}_3^\top = \left(\mathbf{w}_3 + \mathbf{w}_1 \times \mathbf{w}_2\right)^\top \end{bmatrix}$$

  - Due to the twofold ambiguity in the sign of $\hat{E}$ and $\hat{\mathbf{T}}$, there are four different estimates for the pair $(\hat{\mathbf{T}}, R)$

Recovering the Pair $(\hat{\mathbf{T}}, R)$

- The 3-D reconstruction of the viewed points resolves the ambiguity and finds the only correct estimate
  - The third component of each point in the left reference frame is computed for each of the four pairs $(\hat{\mathbf{T}}, R)$:

$$\mathbf{P}_r = R\left(\mathbf{P}_l - \hat{\mathbf{T}}\right);\ \mathbf{p}_r = \frac{f_r}{Z_r}\mathbf{P}_r;\ Z_r = \mathbf{R}_3^\top\left(\mathbf{P}_l - \hat{\mathbf{T}}\right) \Rightarrow \mathbf{p}_r = \frac{f_r R\left(\mathbf{P}_l - \hat{\mathbf{T}}\right)}{\mathbf{R}_3^\top\left(\mathbf{P}_l - \hat{\mathbf{T}}\right)}$$

$$\Rightarrow x_r = \frac{f_r \mathbf{R}_1^\top\left(\mathbf{P}_l - \hat{\mathbf{T}}\right)}{\mathbf{R}_3^\top\left(\mathbf{P}_l - \hat{\mathbf{T}}\right)};\ \mathbf{p}_l = \frac{f_l}{Z_l}\mathbf{P}_l \Rightarrow Z_l = f_l \frac{\left(f_r\mathbf{R}_1 - x_r\mathbf{R}_3\right)^\top \hat{\mathbf{T}}}{\left(f_r\mathbf{R}_1 - x_r\mathbf{R}_3\right)^\top \mathbf{p}_l}$$

$$\Rightarrow \mathbf{P}_l = \frac{\left(f_r\mathbf{R}_1 - x_r\mathbf{R}_3\right)^\top \hat{\mathbf{T}}}{\left(f_r\mathbf{R}_1 - x_r\mathbf{R}_3\right)^\top \mathbf{p}_l}\mathbf{p}_l;\quad \mathbf{P}_r = R\left(\mathbf{P}_l - \hat{\mathbf{T}}\right)$$

Reconstruction Algorithm

Input: a set of corresponding points and an estimate $\hat{E}$

1. Recover the normalised translation vector $\hat{\mathbf{T}}$
2. Recover the rotation matrix R
3. Reconstruct the coordinates Z_l and Z_r of each point
4. If the signs of Z_l and Z_r of the reconstructed points are:
   (a) both negative for some point, change the sign of $\hat{\mathbf{T}}$ and go to 3
   (b) one negative, one positive for some point, change the sign of each entry of $\hat{E}$ and go to 2
   (c) both positive for all points, exit

Uncalibrated Reconstruction

- No information on the intrinsic and extrinsic parameters
- Only n point correspondences, n > 8, are given
  - Thus the location of the epipoles is known
  - The accuracy of the reconstruction is affected by that of the algorithms computing the disparities, not by calibration
- The reconstruction is unique only up to an unknown projective transformation of the world
  - From 5 arbitrary scene points and the epipoles, the projection matrix of each camera is recovered up to this transformation; then the 3-D location of any point is found by triangulation

Uncalibrated Reconstruction

- Homogeneous 3-D / 2-D coordinates: $[X,Y,Z]^\mathsf{T} \Rightarrow [X,Y,Z,1]^\mathsf{T}$ and $[x,y]^\mathsf{T} \Rightarrow [x,y,1]^\mathsf{T}$

- Five points $\mathbf{P}_1,\ldots,\mathbf{P}_5$ to be recovered from their left and right images, $\mathbf{p}_1,\ldots,\mathbf{p}_5$ and $\mathbf{p}'_1,\ldots,\mathbf{p}'_5$
  - No three of them should be collinear and no four coplanar
- <u>Spatial projective transformation</u> is fixed if the destiny of 5 points is known: $M\mathbf{p}_i = \rho_i\mathbf{p}_i$ $(\rho_i \neq 0;\ M$ - the projection matrix$)$
  - Without losing generality, a projective transformation is set up to send these five points to $\mathbf{P}_1 = [1,0,0,0]^\mathsf{T}$, $\mathbf{P}_2 = [0,1,0,0]^\mathsf{T}$, $\mathbf{P}_3 = [0,0,1,0]^\mathsf{T}$, $\mathbf{P}_4 = [0,0,0,1]^\mathsf{T}$, $\mathbf{P}_5 = [1,1,1,1]^\mathsf{T}$

Uncalibrated Reconstruction

- Planar projective transformation is fixed if the destiny of 4 points is known
  - A projective transformation is set up to send the first four $\mathbf{p}_i$ to $\mathbf{p}_1 = [1,0,0]^\mathsf{T}$, $\mathbf{p}_2 = [0,1,0]^\mathsf{T}$, $\mathbf{p}_3 = [0,0,1]^\mathsf{T}$, $\mathbf{p}_4 = [1,1,1]^\mathsf{T}$
    - $\mathbf{p}_5 = [\alpha, \beta, \gamma]^\mathsf{T}$ in this standard projective basis
  - The setups simplify the expression of the projection matrix M:

$$\left.\begin{array}{l} \underset{i=1,2,3,4}{M\mathbf{P}_i = \rho_i \mathbf{p}_i} \;\Rightarrow\; M = \begin{bmatrix} \rho_1 & 0 & 0 & \rho_4 \\ 0 & \rho_2 & 0 & \rho_4 \\ 0 & 0 & \rho_3 & \rho_4 \end{bmatrix} \\ M\mathbf{P}_5 = \rho_5 \mathbf{p}_5 \end{array}\right\} \Rightarrow M = \begin{bmatrix} \alpha\rho_5 - \rho_4 & 0 & 0 & \rho_4 \\ 0 & \beta\rho_5 - \rho_4 & 0 & \rho_4 \\ 0 & 0 & \gamma\rho_5 - \rho_4 & \rho_4 \end{bmatrix}$$

Uncalibrated Reconstruction

- Projection matrices of the left and right camera are found up to the unknown parameters x, x':

$$M = \begin{bmatrix} \alpha x - 1 & 0 & 0 & 1 \\ 0 & \beta x - 1 & 0 & 1 \\ 0 & 0 & \gamma x - 1 & 1 \end{bmatrix}; \; x = \frac{\rho_5}{\rho_4}; \quad M' = \begin{bmatrix} \alpha' x' - 1 & 0 & 0 & 1 \\ 0 & \beta' x' - 1 & 0 & 1 \\ 0 & 0 & \gamma' x' - 1 & 1 \end{bmatrix}; \; x' = \frac{\rho'_5}{\rho'_4}$$

- These parameters are computed using the projection centres $\mathbf{O}, \mathbf{O}'$ found as the null spaces of M, M' ($M\mathbf{O} = 0$; $M'\mathbf{O}' = 0$) and the known location of the epipoles $\mathbf{e}, \mathbf{e}'$ (as $M\mathbf{O}' = \sigma \mathbf{e}$; $M'\mathbf{O} = \sigma' \mathbf{e}'$ with $\sigma \neq 0$ and $\sigma' \neq 0$)

- Then any 3-D point is reconstructed using the projective rays through $\mathbf{O}$, $\mathbf{O}'$