Estimating Camera Pose

COSC450

Lecture 3

Perspective-n-Point Pose (PnP)

Input:

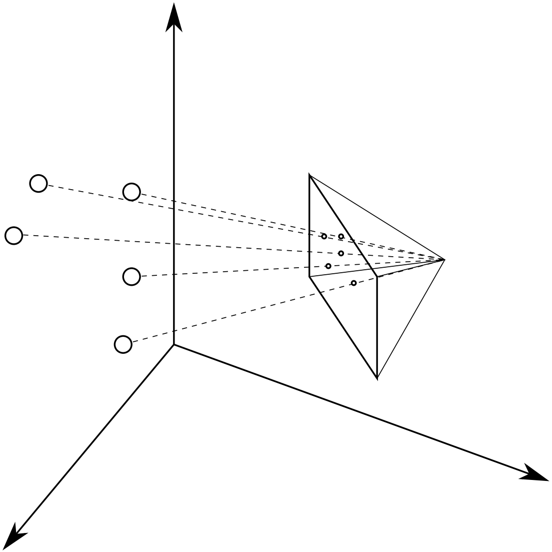
- n 3D point locations, $\mathbf{x}_i$
- Corresponding 2D image locations, $\mathbf{u}_i$
- Camera calibration information, K

Output:

- Camera rotation, R, and translation, $\mathbf{t}$

$$\mathbf{u}_i \equiv \mathrm{K} \begin{bmatrix} \mathrm{R} & \mathbf{t} \end{bmatrix} \mathbf{x}_i$$

$$\begin{bmatrix} u_i \\ v_i \\ 1 \end{bmatrix} \equiv \begin{bmatrix} f_u & s & c_u \\ 0 & f_v & c_v \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} | & | & | & | \\ \mathbf{r}_1 & \mathbf{r}_2 & \mathbf{r}_3 & \mathbf{t} \\ | & | & | & | \end{bmatrix} \begin{bmatrix} x_i \\ y_i \\ z_i \\ 1 \end{bmatrix}$$



Perspective-n-Point Pose

Minimal case is $n = 3$

- Each point gives 2 constraints
- Six unknowns (what are they?)
- Points in *general position*
- In this case: not co-linear

More than minimal case:

- Can lead to simpler algorithms
- Least-squares fit
- Removing bad matches (outliers)

We've already seen this problem

- Estimating Rs and $\mathbf{t}$s in calibration

Checkerboard is an easy case:

- $\mathbf{u}_i$s easy to find and match to $\mathbf{x}_i$s
- $\mathbf{x}_i$s aligned to X-Y plane
- Simple solution because all $z_i = 0$

$$\begin{bmatrix} u_i \\ v_i \\ 1 \end{bmatrix} \equiv \begin{bmatrix} f_u & s & c_u \\ 0 & f_v & c_v \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} | & | & | & | \\ \mathbf{r}_1 & \mathbf{r}_2 & \mathbf{r}_3 & \mathbf{t} \\ | & | & | & | \end{bmatrix} \begin{bmatrix} x_i \\ y_i \\ 0 \\ 1 \end{bmatrix}$$

$$\mathbf{u}_i \equiv \mathrm{K} \begin{bmatrix} \mathbf{r}_1 & \mathbf{r}_2 & \mathbf{t} \end{bmatrix} \begin{bmatrix} x_i & y_i & 1 \end{bmatrix}^\mathsf{T}$$

Checkerboard Pose

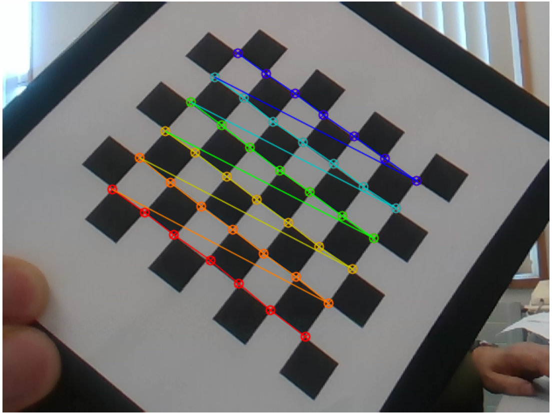
As before, this is a homography $\mathbf{u}_i = \lambda \mathrm{H} \mathbf{x}_i$

- Four (or more) points, find H
- Gives solution up to scale λ

$$
\begin{aligned}
\mathbf{r}_1 &= \lambda \mathrm{K}^{-1} \mathbf{h}_1 \\
\mathbf{r}_2 &= \lambda \mathrm{K}^{-1} \mathbf{h}_2 \\
\mathbf{r}_3 &= \mathbf{r}_1 \times \mathbf{r}_2 \\
\mathbf{t} &= \lambda \mathrm{K}^{-1} \mathbf{h}_3
\end{aligned}
$$

- $\mathbf{r}_i$s are unit vectors so

$$
\lambda = \frac{1}{\|\mathrm{K}^{-1} \mathbf{h}_1\|}
$$



Refining Checkerboard Pose

Generally R is not quite a rotation

- The SVD can help us

$$\mathrm{R} = \mathrm{U}\Sigma\mathrm{V}^{\mathsf{T}}$$

- U and V are rotations
- Σ is a scaling
- Rotations don't scale so $\Sigma \approx \mathrm{I}$
- Setting $\Sigma = \mathrm{I}$ gives a rotation
- This is the nearest true rotation to R

Minimise the reprojection error

- As before, the homography estimation doesn't use a meaningful error term
- Minimise the reprojection error:

$$\sum_i \|\mathbf{u_i} - \tilde{\mathbf{u}}_i\|^2$$

where

$$\tilde{\mathbf{u}}_i \equiv \mathrm{K}\begin{bmatrix}\mathrm{R} & \mathbf{t}\end{bmatrix}\mathbf{x}_i$$

- Non-linear in R, $\mathbf{t}$
- How to represent R?

Perspective-3-Point Pose (P3P, a sketch)

In the general case:

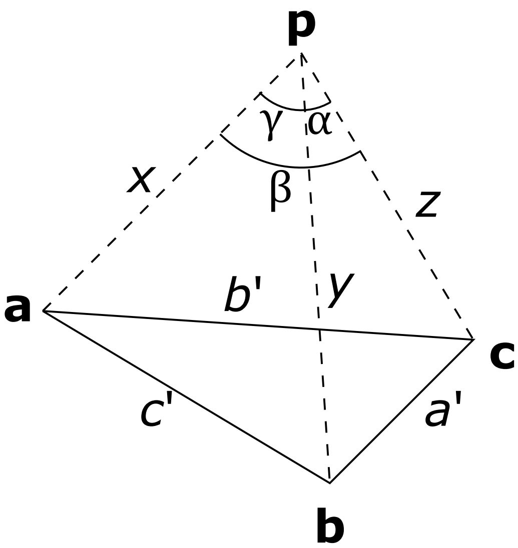
- We have 3D points $\mathbf{a}, \mathbf{b}, \mathbf{c}$
- We want to find the camera centre $\mathbf{p}$
- Our image points give us angles α, β, γ
- Define

$$x = \|\mathbf{p} - \mathbf{a}\| \quad y = \|\mathbf{p} - \mathbf{b}\| \quad z = \|\mathbf{p} - \mathbf{c}\|$$
$$a' = \|\mathbf{b} - \mathbf{c}\| \quad b' = \|\mathbf{a} - \mathbf{c}\| \quad c' = \|\mathbf{b} - \mathbf{a}\|$$

- For a triangle with sides length a, b, c

$$c^2 = a^2 + b^2 - 2ab\cos\theta$$

where θ is angle opposite c



P3P (A sketch)

Applying this to triangles **apb**, **apc**, **bpc**:

$$a'^2 = y^2 + z^2 - 2yz\cos\alpha$$
$$b'^2 = x^2 + z^2 - 2xz\cos\beta$$
$$c'^2 = x^2 + y^2 - 2xy\cos\gamma$$

- 3 equations in 3 unknowns
- Non-linear equations – tricky
- Can reduce to 2 quadratics
- Gives up to 4 solutions
- Disambiguate with extra point(s)