CIS 580 Spring 2012 - Lecture 20

April 2, 2012

Notes and figures by Matthieu Lecce.

Review:

Single-view geometry

- We showed how focal length and projection center of a camera could be recovered from the projections of three orthogonal vanishing points on the screen.
- We showed how the camera pose can be recovered from a set of 2D-3D matchings, when K is unknown (*uncalibrated* pose estimation)
- This time we will cover *calibrated* pose estimation.

Single-view geometry (continued)

Pose from a single view, using 3 points

Here we assume the camera is *calibrated*, i.e. K is known.

Given A, B, C in world (object) coordinate system, their projections a, b, c in calibrated coordinates ($a \sim K^{-1} a_{\text{pixels}}$). We want to recover the camera extrinsics (R, T). This problem is very similar to the problem of GPS localization (triangulation).

See figure 1. From the law of cosines, we have:

$$\begin{aligned} BC^2 &= d_B^2 + d_C^2 - 2 d_B d_C \cos \delta_{BC} \\ AC^2 &= d_A^2 + d_C^2 - 2 d_A d_C \cos \delta_{AC} \\ AB^2 &= d_A^2 + d_B^2 - 2 d_A d_B \cos \delta_{AB} \end{aligned}$$

If we introduce u, v such that $d_B = u d_A$ and $d_C = v d_A$, we have:

$$\begin{aligned} BC^2 &= d_A^2 (u^2 + v^2 - 2uv \cos \delta_{BC}) \\ AC^2 &= d_A^2 (u^2 + v^2 - 2uv \cos \delta_{AC}) \\ AB^2 &= d_A^2 (u^2 + v^2 - 2uv \cos \delta_{AB}) \end{aligned}$$

Therefore we have the following:

$$\begin{aligned} \frac{BC^2}{(u^2 + v^2 - 2uv \cos \delta_{BC})} &= \frac{AC^2}{(u^2 + v^2 - 2uv \cos \delta_{AC})} \\ \frac{AB^2}{(u^2 + v^2 - 2uv \cos \delta_{AB})} &= \frac{AC^2}{(u^2 + v^2 - 2uv \cos \delta_{AC})} \end{aligned}$$

These two equations of second order in u, v yield one equation of fourth order: we find four solutions for u^2, which correspond to eight pairs (u, v).

Each of the pairs (u, v) yields a set of d_A, d_B, d_C values, and finally:

$$\begin{aligned} OA &= d_A a = RA + T \\ OB &= d_B b = RB + T \\ OC &= d_C c = RC + T \end{aligned}$$

(where a, b, c have been normalized to unit vectors)

Absolute pose or absolute orientation

Let's consider two sets of 3D points (as measured by a Kinect sensor for instance), and assume we want to estimate the transformation that maps one set to the other. See figure 2. $\boxed{P_1^i = RP_2^i + T}$ $(R \in SO(3), T \in \mathbb{R}^3)$.

It can be easily shown that the **least-squares** translation estimate between the two point clouds is given by the following formula:

$$T = \underbrace{\frac{1}{n}\sum P_1^i}_{\overline{P_1}} - R\,\underbrace{\frac{1}{n}\sum P_2^i}_{\overline{P_2}} = \overline{P_1} - R\overline{P_2}$$

We now want to find R verifying the following equation:

$$P_1^i = RP_2^i + \overline{P_1} - R\overline{P_2}$$

We can rewrite the above equation in the simple form $P = RQ$. As usual, this equation cannot generally be solved exactly, so perform a least-squares estimation. The squared Frobenius norm of $P - RQ$ can be expressed as follows:

$P = P_1 - \overline{P_1} + R\overline{P_2}$, $Q = P_2$

$$\begin{aligned} \|P - RQ\|_F^2 &= \text{tr}\left((P - RQ)^T (P - RQ)\right) \\ &= \text{tr}\left(P^T P + Q^T \underbrace{R^T R}_{I} Q - P^T RQ - Q^R T\right) \\ &= -\text{tr}\left(P^T RQ + Q^T RP\right) \\ &= -2\text{tr}\left(R\, \underbrace{QP^T}_{H}\right) \end{aligned}$$

We use the fact that tr $(A^T) = $ tr (A) and the trace is invariant to circular permutation of the matrix product

To summarize we are trying to find $\max_R \text{tr}\left(R \underset{3\times3}{H}\right)$

THEOREM: If H is symmetric and positive definite, and R is orthogonal, then:

$$\text{tr}\,(H) \geq \text{tr}\,(RH)$$

To maximize tr (RH), we want to find an R such that tr (RH) is positive definite. We can achieve this by performing an SVD decomposition of H: $H = USV^T$, set $R = VU^T$ then $RH = VU^TUSV^T = VSV^T$ which is symmetric, positive and definite.

Therefore we can use the following recipe to estimate R:

- take the SVD of $QP^T = H = USV^T$,
- return optimal $R = VU^T$.

Is it that simple? **No**, we forgot to take into account that we want $R \in \mathcal{SO}(3)$. If we set $R = VU^T$ with $V^TV = U^TU = I$, we have $R^TR = UV^TVU^T = I$, and $\det R = 1$ or -1. Therefore instead of $R = VU^T$, we should return the following *normalized* estimate:

$$R = V\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & \det(UV^T) \end{bmatrix}U^T$$

.

Two views uncalibrated

- 1966 soccer game, goal incorrectly awarded because of perspective error.
- *Goal-directed Video Metrology*, I. Reid and A. Zisserman

Two views synced Let's assume that we know the projective transformation of the ground $x_1 \sim Px_2$, $b_1 \nsim Pb_2$

$l_1 \sim b_1 \times v_1$, $l_2 \sim b_2 \times v_2$, f_1, f_2 are the back-projections of l_1, l_2 on ground plane, f_1, f_2 have to intersect at F because both planes (l_1, l_2) and

If $x_1 \sim Px_2$ then $l_1 \sim P^{-T}l_1'$, l_1' is the projection of f_1 on the second image plane. Then $f \sim l_1 \times l_2$

We have done all this without any knowledge of the focal length.

Two views of a plane

See figure 4. $X_1 = RX_2 + T$, $x_1 = \begin{bmatrix} x_1' \\ y_1' \\ 1 \end{bmatrix}$, $x_2 = \begin{bmatrix} x_2' \\ y_2' \\ 1 \end{bmatrix}$, $Z_1x_1 = RZ_2x_2 + T$.

Plane equation w.r.t second camera $N^TX_2 = d$, or equivalently $\frac{N^TX_2}{d} = 1$