X points for VSLAM

Tuesday, March 24, 2020 3:33 PM

This doc provides a summary of X-point method used in VSLAM

① 2 points: P2P

This is often used in 2-point RANSAC, e.g. MSCKF.

$su = K(RP + t)$

P: 3D point

u: projection of P on the image

K: intrinsics

We know some 3D-2D pairs, $\{P_i, u_i\}_{i=1,2,\dots}$ Solve R, t.

Usually PnP needs 3 pairs to solve for R, P, but if the R is known, e.g. via integration of gyro measurements, we only need two points to solve for translation. ⇒ 2 point method.

Since R, P are known, let $P' = RP$. move K to left side, let $u' = K^{-1}u$, u' is the normalized coordinates.

we can simplify the eq. as

$$su' = P' + t$$

$$\begin{bmatrix} su'_x \\ su'_y \\ s \end{bmatrix} = \begin{bmatrix} P'_x \\ P'_y \\ P'_z \end{bmatrix} + \begin{bmatrix} t_x \\ t_y \\ t_z \end{bmatrix}$$

use the third line $s = P'_z + t_z$. for the first two eqs:

$$(P'_z + t_z)u'_x = P'_x + t_x \Rightarrow t_x + u'_x t_z = P'_x - P'_z u'_x$$

$$(P'_z + t_z)u'_y = P'_y + t_y \Rightarrow t_y + u'_y t_z = P'_y - P'_z u'_y$$

unknowns are t_x, t_y, t_z, and each 3D-2D pair can only get two eqs. So we need at least 2 points (4 eqs) to solve for t.

② 3 point method

a. 3D-3D match, ICP.

Known N pairs 3D-3D $\{p_i, q_i\}_{i=1,2,\dots,N}$, solve for R, t

$$R, t = \underset{R,t}{\operatorname{argmin}} \frac{1}{2}\sum_{i=1}^{N} \| p_i - (Rq_i + t) \|^2$$

we need at least 3 non colinear 3D pairs.

Step 1: centralize points (subtract off the mean)

$$p'_i = p_i - \mu_p, \quad q'_i = q_i - \mu_q$$

$$\mu_p = \frac{1}{N}\sum_{i=1}^{N} p_i, \quad \mu_q = \frac{1}{N}\sum_{i=1}^{N} q_i$$

Step 2: Solve for R:

use SVD: $W = \sum_{i=1}^{N} p'_i q'^T_i = U\Sigma V^T$

$$R = UV^T$$

Step 3: Solve for t: $t = \mu_p - R\mu_q$.

$$\min \frac{1}{2}\sum_{i=1}^{N} \| (p'_i + \mu_p) - (R(q'_i + \mu_q) + t) \|^2$$

$$= \min \frac{1}{2}\sum_{i=1}^{N} \| (p'_i - Rq'_i) + (\mu_p - (R\mu_q + t)) \|^2 \quad \text{(expand out)}$$

$$= \min \frac{1}{2}\sum_{i=1}^{N} \| p'_i - Rq'_i \|^2 + \frac{1}{2}\sum_{i=1}^{N} \| \mu_p - (R\mu_q + t) \|^2 + \sum_{i=1}^{N} (p'_i - Rq'_i)^T(\mu_p - (R\mu_q + t))$$

$$= \min \frac{1}{2}\sum_{i=1}^{N} \| p'_i - Rq'_i \|^2 + \frac{N}{2}\| \mu_p - (R\mu_q + t) \|^2 + (\mu_p - (R\mu_q + t))^T \Big(\underbrace{\sum_{i=1}^{N} p'_i}_{0} - R\underbrace{\sum_{i=1}^{N} q'_i}_{0}\Big)$$

$$= \min \frac{1}{2}\sum_{i=1}^{N} \| p'_i - Rq'_i \|^2 + \frac{N}{2}\| \mu_p - (R\mu_q + t) \|^2$$

No matter what is R, we can always choose $t = \mu_p - R\mu_q$ to make the 2nd term zero. ∴ We only need to optimize R over the 1st term:

$$\min \frac{1}{2}\sum_{i=1}^{N} \| p'_i - Rq'_i \|^2 \quad \text{(expand)}$$

$$= \min \frac{1}{2}\sum_{i=1}^{N} (p'^T_i p'_i + q'^T_i R^T R q'_i - 2p'^T_i R q'_i)$$

$$= \max \sum_{i=1}^{N} p'^T_i R q'_i \qquad R^TR = I$$

$$= \max \sum_{i=1}^{N} \operatorname{tr}(R q'_i p'^T_i)$$

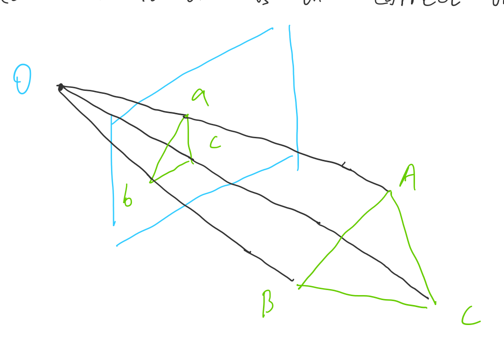
$$= \max \operatorname{tr}\Big(R\sum_{i=1}^{N} q'_i p'^T_i\Big)$$

$$= \max \operatorname{tr}(RW^T)$$

Note that we need at least 3 non co-linear points to make sure W is invertible and solve for R.

b. P3P.

P3P is the minimal algorithm for solving 6 DoF. We need at least 3 pairs to solve for R, t, but we need 1 extra pair to validate which solution is the correct one. ∴ We still need 4 pairs.



a, b, c are the 2D measurements, A, B, C are unknown 3D points in the camera frame.

from the cosine theorem:

$$OA^2 + OB^2 - 2OA \cdot OB \cdot \cos\langle a, b\rangle = AB^2$$

$$OB^2 + OC^2 - 2OB \cdot OC \cdot \cos\langle b, c\rangle = BC^2$$

$$OA^2 + OC^2 - 2OA \cdot OC \cdot \cos\langle a, c\rangle = AC^2$$

divide all terms above by OC^2, let

$$x = \frac{OA}{OC}, \quad y = \frac{OB}{OC}, \quad v = \frac{AB^2}{OC^2}, \quad uv = \frac{BC^2}{OC^2}, \quad wv = \frac{AC^2}{OC^2}, \quad u = \frac{BC^2}{AB^2}, \quad w = \frac{AC^2}{AB^2}$$

$$x^2 + y^2 - 2xy\cos\langle a, b\rangle = v \quad ①$$

$$y^2 + 1 - 2y\cos\langle b, c\rangle = uv \quad ②$$

$$x^2 + 1 - 2x\cos\langle a, c\rangle = wv \quad ③$$

Substitute ① into ②, ③, cancel v, then

$$\begin{cases} (1-u)y^2 - ux^2 - \cos\langle b, c\rangle y + 2u\cos\langle a, b\rangle xy + 1 = 0 \\ (1-w)x^2 - wy^2 - \cos\langle a, c\rangle x + 2w\cos\langle a, b\rangle xy + 1 = 0 \end{cases}$$

This is about unknown x, y, and cosines can be computed from 2D measurements, u, w can be computed from known 3D points in the world frame.

There are 4 solutions, and we need an extra pair $\langle P, d\rangle$ to validate the solutions.

Once we know x, y, we can get 3D points A, B, C in camera frame. then using 3D-3D match we can solve for R, t.

③ 4 point

a. Homography.

Homography is trying to solve 3D to 2D transformation, its assumption is that 3D points are on the same plane. so we can simplify this to plane to plane transformation.

Suppose all 3D points are on $z=0$ plane, then the projection is:

$$s\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = K[r_1\ r_2\ r_3\ t]\begin{bmatrix} X \\ Y \\ 0 \\ 1 \end{bmatrix}$$

$$= K[r_1\ r_2\ t]\begin{bmatrix} X \\ Y \\ 1 \end{bmatrix}$$

$$= H\begin{bmatrix} X \\ Y \\ 1 \end{bmatrix}$$

$H = K[r_1\ r_2\ t]$ is the homography.

$$H = \begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{bmatrix} \Rightarrow \begin{aligned} su &= h_{11}X + h_{12}Y + h_{13} \quad ① \\ sv &= h_{21}X + h_{22}Y + h_{23} \quad ② \\ s &= h_{31}X + h_{32}Y + h_{33} \quad ③ \end{aligned}$$

Substitute ③ into ①, ② and get

$$(h_{31}X + h_{32}Y + h_{33})u = h_{11}X + h_{12}Y + h_{13}$$

$$(h_{31}X + h_{32}Y + h_{33})v = h_{21}X + h_{22}Y + h_{23}$$

let $h = [h_{11}, h_{12}, h_{13}, h_{21}, h_{22}, h_{23}, h_{31}, h_{32}, h_{33}]^T$, then

$$\begin{bmatrix} -x & -y & -1 & 0 & 0 & 0 & xx_1 & yx_1 & x_1 \\ 0 & 0 & 0 & -x & -y & -1 & xy_1 & yy_1 & y_1 \end{bmatrix} h = 0$$

there's 9 unknowns, so we need 5 points to solve those. but consider that H is equivalent to kH, we let $h_{33} = 1$, then we only need 4 points to solve for H.

Note that we have to make sure the element is nonzero to set it to 1. from definition:

$$h_{33} = [0\ 0\ 1][t_x\ t_y\ t_z]^T = t_z.$$

$$H = K[r_1\ r_2\ t], \quad K = \begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix}$$

$$\therefore H_{33} = [0\ 0\ 1][t_x\ t_y\ t_z]^T = t_z.$$

Since camera extrinsic t is non-zero?? i.e. t_z is nonzero, we can set h_{33} to 1.

b. EPnP.

EPnP selects 4 control points in the world frame, then express all the points in the world frame as a weighted sum of the control points:

$$P_i^w = \sum_{j=1}^{4} \alpha_{ij} c_j^w, \quad \text{with} \sum_{j=1}^{4} \alpha_{ij} = 1 \quad (*)$$

c_j^w is the control point, $\alpha_{i1}, \alpha_{i2}, \alpha_{i3}, \alpha_{i4}$ are weights, as soon as we fix c_j^w, P_j^w is only dependent on α. there is only one solution of α, since we have five equations.

In the camera frame, we also have

$$P_i^c = \sum_{j=1}^{4} \alpha_{ij} c_j^c, \quad \text{with} \sum_{j=1}^{4} \alpha_{ij} = 1$$

let $P_i^c = {}_cT_w P_i^w$, $C_i^c = {}_cT_w C_i^w$, multiply ${}_cT_w$ on both sides of $(*)$ we get:

$${}_cT_w P_i^w = {}_cT_w \sum_{j=1}^{4} \alpha_{ij} c_j^w$$

$$\therefore P_i^c = \sum_{j=1}^{4} \alpha_{ij} c_j^c$$

$\because s_i u_i = K P_i^c$, substitute $P_i^c = \sum_{j=1}^{4} \alpha_{ij} c_j^c$

$$s_i\begin{bmatrix} u_i \\ v_i \\ 1 \end{bmatrix} = \begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix} \sum_{j=1}^{4} \alpha_{ij} \begin{bmatrix} x_j^c \\ y_j^c \\ z_j^c \end{bmatrix} \quad \begin{matrix} ① \\ ② \\ ③ \end{matrix}$$

substitute ③ into ①, ②:

$$\sum_{j=1}^{4} [\alpha_{ij} f_x x_j^c + \alpha_{ij}(c_x - u_i) z_j^c] = 0$$

$$\sum_{j=1}^{4} [\alpha_{ij} f_y y_j^c + \alpha_{ij}(c_y - v_i) z_j^c] = 0$$

the unknowns are 12 coordinates of control points. Let $X = [x_1^c\ y_1^c\ z_1^c\ x_2^c\ y_2^c\ z_2^c\ x_3^c\ y_3^c\ z_3^c\ x_4^c\ y_4^c\ z_4^c]^T$, we have

$$MX = 0,$$

where M is 8×12 since we have 4 points.

We have 12 unknowns but only 8 eqs so the solution is more than 1. We also need the rank of the nullspace of M^TM, which can only be 1, 2, 3, 4, and solve for the control points in 4 conditions, and select the one that min. the reprojection error.

When we have the control points, we can solve for R, t using 3D-3D match.

recall that PnP is the problem of estimating the pose of the camera, given a set of 3D points in the world and their 2D projections on the image.

The problem of EPnP is:

given features, we know the coordinates in the world frame:

$$P_i^w = \begin{bmatrix} x_i^w \\ y_i^w \\ z_i^w \end{bmatrix}$$

its projection on the image:

$$P_i = \begin{bmatrix} u_i \\ v_i \end{bmatrix}$$

we want to solve

$${}_wT_c = \begin{bmatrix} {}_wR_c & t \\ 0 & 1 \end{bmatrix}$$

EPnP becomes:

Step 1: compute control points in camera frame

Step 2: compute feature positions in camera frame

Step 3: compute R, t.

④. 5 points.

There exists an Essential matrix, for matched pairs on the normalized plane, we have

$$x'^T_n E x_n = 0.$$

it's relationship with Fundamental matrix is:

$$x'^T_n E x_n = (K_2^{-1}x')^T E (K_1^{-1}x)$$

$$= x'^T K_2^{-T} E K_1^{-1} x$$

$$= 0$$

$$\therefore F = K_2^{-T} E K_1^{-1} \Rightarrow E = K_2^T F K_1$$

E and R, t are associated by

$$E = \hat{t}R,$$

$\operatorname{rank}(E) = 2$, DoF $= 5 = 3$ DoF(R) + 3 DoF(t) − 1 DoF(scale)

Nister 5 point algorithm:

$$x'^T_n E x_n = 0 \Rightarrow [u'\ v'\ 1]\begin{bmatrix} e_{11} & e_{12} & e_{13} \\ e_{21} & e_{22} & e_{23} \\ e_{31} & e_{32} & e_{33} \end{bmatrix}\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = 0$$

$$\Rightarrow u'ue_{11} + u've_{12} + u'e_{13} + uv'e_{21} + vv'e_{22} + v'e_{23} + ue_{31} + ve_{32} + e_{33} = 0$$

let $e = [e_{11}\ e_{12} \cdots e_{33}]$, we have

$$A_{5\times 9}\, e = 0$$

∴ Solution has 4 DoF, $e = a_x x + a_y y + a_z z + a_w w$, where x, y, z, w are the solutions to $Ae = 0$, in matrix form

$$E = a_x X + a_y Y + a_z Z + a_w W$$

any a_x, a_y, a_z, a_w satisfy $x'^T_n E x_n = 0$. So we need to consider that

$$x'^T_n E x_n = 0 \Rightarrow x'^T_n (sE) x_n = 0$$

WLOG, let $a_w = 1$, only a_x, a_y, a_z are unknowns, Nister introduced the constraint:

$$EE^TE - \frac{1}{2}\operatorname{tr}(EE^T)E = 0.$$

substitute $E = a_x X + a_y Y + a_z Z + W$, we get 3rd order eqs for a_x, a_y, a_z, there are $3 \times 3 = 9$ eqns. we get only 1 solution for E.

⑤. 6 points:

DLT needs at least 6 points to solve for R, t.

$$\begin{bmatrix} su_x \\ su_y \\ s \end{bmatrix} = \begin{bmatrix} r_1 & r_2 & r_3 \\ r_4 & r_5 & r_6 \\ r_7 & r_8 & r_9 \end{bmatrix}\begin{bmatrix} P_x \\ P_y \\ P_z \end{bmatrix} + \begin{bmatrix} t_x \\ t_y \\ t_z \end{bmatrix} \quad \begin{matrix} ① \\ ② \\ ③ \end{matrix}$$

Substitute ③ into ①, ②,

$$(r_7P_x + r_8P_y + r_9P_z + t_z)u_x = r_1P_x + r_2P_y + r_3P_z + t_x$$

$$(r_7P_x + r_8P_y + r_9P_z + t_z)u_y = r_4P_x + r_5P_y + r_6P_z + t_y$$

there are 12 unknowns $r_1, r_2 \dots r_9, t_x, t_y, t_z$, every matched pair provides 2 eqns.

∴ We need at least 6 points

Note that R is treated as a normal matrix in DLT, we need to do QR decomposition for R to have an approx. rotation matrix.

⑥ 7 point

DoF of fundamental matrix is 7, so we need at least 7 points to solve for F. Similar to 5 point algorithm, we have $Ax = 0$, A has size 7×9. Solution space is $x = v_1 + \lambda v_2$. where v_1, v_2 are solution vectors. in matrix form:

$$F = F_1 + \lambda F_2$$

$\because \operatorname{rank}(F) = 2$, we add the constraint

$$\det(F_1 + \lambda F_2) = a_3\lambda^3 + a_2\lambda^2 + a_1\lambda + a_0 = 0$$

Solve for λ and get F. λ can have 1 or 3 solutions, so we may need extra points to validate the solution.

⑦. 8 point

for fundamental matrix: $u_2^T F u_1 = 0$, u_1, u_2 are pixels.

Properties: $\operatorname{rank}(F) = 2$, $\det(F) = 0$.
if $x'^T F x = 0$, then $x'^T kF x = 0$, $k \neq 0$. } DoF$(F) = 9 - 2 = 7$.

$$F = K^{-T} E K^{-1} = K^{-T}[t_\times] R K^{-1}.$$

this can be seen as $RK^{-1}u_1$, $K^{-1}u_2$, t are coplanar

$$(K^{-1}u_2) \cdot (t \times (RK^{-1}u_1)) = 0$$

$$(K^{-1}u_2)^T [t_\times] R K^{-1} = u_2^T K^{-T} [t_\times] R K^{-1} u_1 = 0$$

$$\because x'^T F x = 0 \Rightarrow [u'\ v'\ 1]\begin{bmatrix} f_{11} & f_{12} & f_{13} \\ f_{21} & f_{22} & f_{23} \\ f_{31} & f_{32} & f_{33} \end{bmatrix}\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = 0$$

$$\Rightarrow u'uf_{11} + u'vf_{12} + u'f_{13} + uv'f_{21} + vv'f_{22} + v'f_{23} + uf_{31} + vf_{32} + f_{33} = 0$$

∵ there are 9 unknowns, and each pair provide 1 eq, we need 8 matched pairs. to make $\operatorname{rank}(A) < n$, n is no. of eqns.

Summary:

2 point	P2P	3D-2D, 4 DoF
3 point	ICP	3D-3D, 6 DoF
3 point	P3P	3D-2D, 6 DoF
4 point	Homography	2D-2D
4 point	EPnP	3D-2D, 6 DoF
5 point	Essential matrix	2D-2D
6 point	PnP DLT	3D-2D, 6 DoF
7 point	Fundamental matrix	2D-2D
8 point	Fundamental matrix	2D-2D