VIO from scratch 7-2, 7-3

Friday, April 10, 2020   11:17 AM

Triangulation:

Landmark y is seen in keyframes $k=1,\dots,n$.

$y\in R^4$ in homogeneous coordinates, measurements are $x_k=[u_k,v_k,1]^T$, on the normalized plane.

Projection matrix is $P_k=[R_k,t_k]\in R^{3\times4}$, from world frame to camera frame.

$\forall k,\ \lambda_k x_k = P_k y$. We know x_k, P_k, want to get y.

λ_k is unknown depth. don't care about λ_k.

$\lambda_k = P_{k,3}^T y$, $P_{k,3}^T$ is 3rd row of P_k.

$$\lambda_k \begin{bmatrix}u_k\\v_k\\1\end{bmatrix} = [R_k|t_k]\begin{bmatrix}y_1\\y_2\\y_3\\1\end{bmatrix} \qquad P_k=\begin{bmatrix}P_{k,1}^T\\P_{k,2}^T\\P_{k,3}^T\end{bmatrix}$$

$\therefore \lambda_k = P_{k,3}^T y$

Substitute λ_k into $\lambda_k x_k = P_k y$:

$$u_k P_{k,3}^T y = P_{k,1}^T y$$
$$v_k P_{k,3}^T y = P_{k,2}^T y$$

$\therefore$ Stack all measurements:

$$\begin{bmatrix}u_1P_{1,3}^T - P_{1,1}^T\\ v_1P_{1,3}^T - P_{1,2}^T\\ \vdots\\ u_nP_{n,3}^T - P_{n,1}^T\\ v_nP_{n,3}^T - P_{n,2}^T\end{bmatrix} y = 0 \rightarrow Dy=0$$

$\because D\in R^{2n\times4}$, when $n\geq2$. D is full rank and doesn't have a nullspace.

$\therefore \min_y \|Dy\|_2^2$, s.t. $\|y\|=1$.

$A_{n,m}$, $r(A)=n$, $r(B)=k$, $B\approx A$.

use SVD for D^TD:

$$D^TD=\sum_{i=1}^{4}\sigma_i^2 u_i u_i^T$$

σ_i is singular value and from large to small

$\therefore y=u_4$. $\|Dy\|_2^2=\sigma_4^2$.

if $\sigma_4 \ll \sigma_3$, then y is in the nullspace, triangulation is successful.

otherwise, triangulation fails.

E.g. In autonomous driving, we see far away points, then the triangulation often fails, since the baseline is too small.

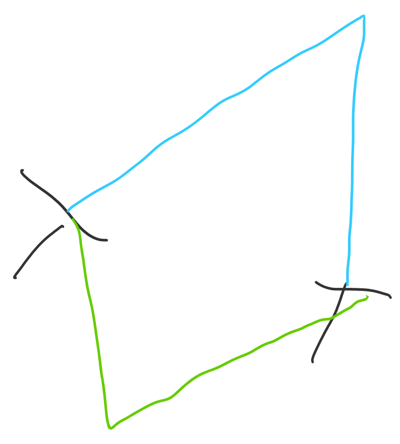
If we use GPS for pose, then the values are not numerically stable. Since GPS values are too large, and position changes are too small.

$\therefore$ We rescale D:

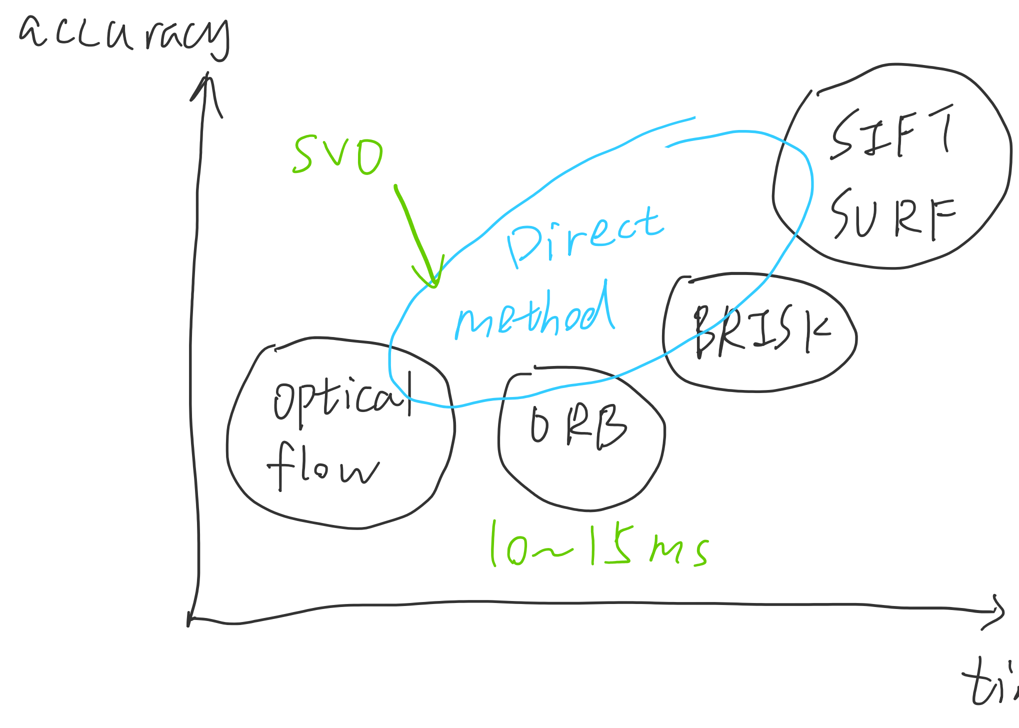
$$Dy = \underbrace{DS}_{\tilde{D}}\ \underbrace{S^{-1}y}_{\tilde{y}} = 0$$

S is diag matrix, use D's largest element's inverse.

Also need to check the sign of y.



Points go to the back of cameras due to noise.



Good accuracy but needs 200ms+

Direct method in general is better than feature based, since textureless scene is very common.

Harris Corner:

$$M=\sum_{x,y}\begin{bmatrix}I_x^2 & I_{xy}\\ I_xI_y & I_y^2\end{bmatrix},\quad \det(M)=\lambda_1\lambda_2,\quad tr(M)=\lambda_1+\lambda_2$$

$$S_{Harris}=\det(M)-k\,tr(M)^2,\quad k=0.04\sim0.06.$$

$$S_{Shi-Tomasi}=\min(\lambda_1,\lambda_2)$$

Optical flow:

$$I(x+dx,y+dy,t+dt)=I(x,y,t)$$
$$I(x+dx,y+dy,t+dt)\approx I(x,y,t)+\underbrace{\frac{\partial I}{\partial x}dx+\frac{\partial I}{\partial y}dy+\frac{\partial I}{\partial t}dt}_{0}$$
$$\therefore \frac{\partial I}{\partial x}\frac{dx}{dt}+\frac{\partial I}{\partial y}\frac{dy}{dt}=-\frac{\partial I}{\partial t}$$

warp function: to handle the intensity change.

$$I(x,y,t)=I(W(x+dx,y+dy),t+dt)$$
$$W(x+dx,y+dy)=\begin{bmatrix}1+P_1 & P_3 & P_5\\ P_2 & 1+P_4 & P_6\end{bmatrix}\begin{bmatrix}x+dx\\y+dy\\1\end{bmatrix}$$

$P_1\sim P_6$ are parameters.

$\therefore$ In addition to allowing translation, we also allow affine transformation to satisfy brightness consistency condition.