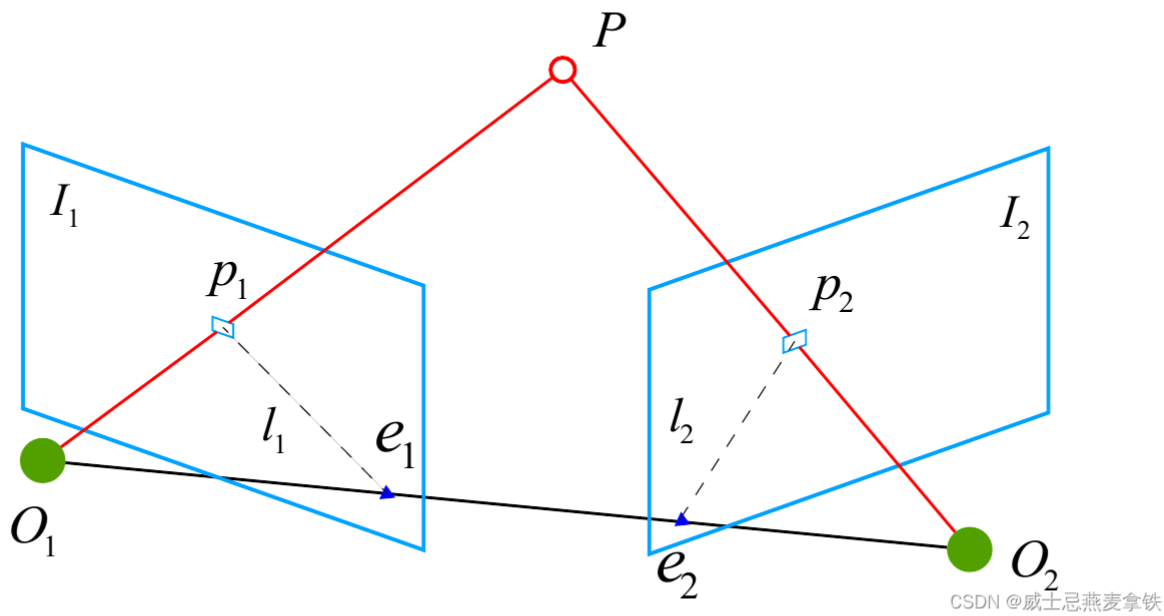
对极几何用来解决"根据两组2D特征匹配点来估计相机运动"的问题。

符号定义

假设我们从两张图像中，得到了一对正确配对的特征点。如下图，设第一帧到第二帧的运动为 R , t R,t R,t（需要求解的变量）。两个相机中心分别为 O 1 O_1 O1 和 O 2 O_2 O2。现在，考虑 I 1 I_1 I1 中有一个特征点 p 1 p_1 p1，它在 I 2 I_2 I2 中对应着特征点 p 2 p_2 p2。这两个特征点通过特征匹配得到。



理想情况下，连线 O 1 p 1 → \overrightarrow{O_{1} p_{1}} O1p1

和连线 O 2 p 2 → \overrightarrow{O_{2} p_{2}} O2p2

在三维空间中会相交于点 P P P 。这时可以用一些术语来表示它们之间的几何关系：

- 极平面（Epipolar plane）： O 1 , O 2 , P O_{1}, O_{2}, P O1, O2, P 三个点构成的平面被称为极平面。
- 极点（Epipoles）： O 1 O 2 O_{1} O_{2} O1O2 连线与成像平面 I 1 ， I 2 I_1， I_2 I1，I2 的交点 e 1 , e 2 e_{1}, e_{2} e1, e2 被称为极点。
- 极线（Epipolar line）：极平面与两个成像平面 I 1 ， I 2 I_1， I_2 I1，I2 之间的交线 l 1 , l 2 l_{1}, l_{2} l1, l2 被称为极线。
- 基线（Baseline）： O 1 O 2 O_{1} O_{2} O1O2 被称为基线。

几何关系

在第一帧的坐标系下，设 P P P 的空间位置为： P = [X , Y , Z] T \boldsymbol{P}=[X, Y, Z]^{T} P = [X, Y, Z]T，那么由针孔相机模型可以得到两个像素点 p 1 p_1 p1 和 p 2 p_2 p2 的像素位置为：

s 1 p 1 = K P , s 2 p 2 = K (R P + t) (1) s_{1} \boldsymbol{p}_{1}=\boldsymbol{K} \boldsymbol{P}, \quad s_{2} \boldsymbol{p}_{2}=\boldsymbol{K}(\boldsymbol{R} \boldsymbol{P}+\boldsymbol{t}) \tag{1}
s1p1 = KP, s2p2 = K(RP + t)(1)

其中， K K K 为相机内参矩阵， R , t R,t R,t 为两个坐标系的相机运动。

我们通常会使用齐次坐标表示像素点。在使用其次坐标时，一个向量将等于它自身乘上任意常数，这可以用于表示一个投影关系。比如， s 1 p 1 s_1\boldsymbol{p}_{1} s1p1 和 p 1 \boldsymbol{p}_{1} p1 成投影关系，它们在齐次坐标系的意义是相等的，我们称这种关系为尺度意义下相等，记为：

s 1 p 1 ≃ p 1 (2) s_1\boldsymbol{p}_{1} \simeq \boldsymbol{p}_{1} \tag{2} s1p1 ≃ p1(2)

因此，可以将 (1) 中的投影关系用齐次坐标投影到归一化坐标上，写成：

p 1 ≃ K P , p 2 ≃ K (R P + t) (3) \boldsymbol{p}_{1} \simeq \boldsymbol{K} \boldsymbol{P}, \quad \boldsymbol{p}_{2} \simeq \boldsymbol{K}(\boldsymbol{R} \boldsymbol{P}+\boldsymbol{t}) \tag{3}
p1 ≃ KP, p2 ≃ K(RP + t)(3)

然后取：

x 1 = K − 1 p 1 , x 2 = K − 1 p 2 (4) \boldsymbol{x}_{1}=\boldsymbol{K}^{-1} \boldsymbol{p}_{1}, \quad \boldsymbol{x}_{2}=\boldsymbol{K}^{-1} \boldsymbol{p}_{2} \tag{4} x1 = K−1p1, x2 = K−1p2(4)

这里的 x 1 , x 2 \boldsymbol{x}_{1}, \boldsymbol{x}_{2} x1, x2 是两个像素点在归一化平面上的坐标。

带入 (3) 式，有：

x 2 ≃ R x 1 + t (5) \boldsymbol{x}_{2} \simeq \boldsymbol{R} \boldsymbol{x}_{1}+\boldsymbol{t} \tag{5} x2 ≃ Rx1 + t(5)

两边同时左乘 t ∧ \boldsymbol{t}^{\wedge} t∧（ 相当于两边同时与 t t t 做外积），有：

t ∧ x 2 ≃ t ∧ R x 1 (6) \boldsymbol{t}^{\wedge} \boldsymbol{x}_{2} \simeq \boldsymbol{t}^{\wedge} \boldsymbol{R} \boldsymbol{x}_{1} \tag{6} t∧x2 ≃ t∧Rx1(6)

然后，两侧同时左乘 x 2 T \boldsymbol{x}_{2}^{T} x2T：

x 2 T t ∧ x 2 ≃ x 2 T t ∧ R x 1 (7) \boldsymbol{x}_{2}^{T} \boldsymbol{t}^{\wedge} \boldsymbol{x}_{2} \simeq \boldsymbol{x}_{2}^{T} \boldsymbol{t}^{\wedge} \boldsymbol{R} \boldsymbol{x}_{1} \tag{7}
x2Tt∧x2 ≃ x2Tt∧Rx1(7)

观察等式左侧， t ∧ x 2 t^{\wedge} x_{2} t∧x2 是一个与 t t t 和 x 2 {x}_{2} x2 都垂直的向量。把它再和 x 2 {x}_{2} x2 做内积时，将得到 0。因此，有：

x 2 T t ∧ R x 1 ≃ 0 (8) \boldsymbol{x}_{2}^{T} \boldsymbol{t}^{\wedge} \boldsymbol{R} \boldsymbol{x}_{1} \simeq 0 \tag{8} x2Tt∧Rx1 ≃ 0(8)

由于等式左侧严格为0，乘以任意常数之后也都为0，于是我们可以把 ≃ \simeq ≃ 写成等号：

x 2 T t ∧ R x 1 = 0 (9) \boldsymbol{x}_{2}^{T} \boldsymbol{t}^{\wedge} \boldsymbol{R} \boldsymbol{x}_{1}=0 \tag{9} x2Tt∧Rx1 = 0(9)

重新代入 p 1 p_1 p1 和 p 2 p_2 p2，有：

p 2 T K − T t ∧ R K − 1 p 1 = 0 (10) \boldsymbol{p}_{2}^{T} \boldsymbol{K}^{-T} \boldsymbol{t}^{\wedge} \boldsymbol{R} \boldsymbol{K}^{-1} \boldsymbol{p}_{1}=0 \tag{10} p2TK−Tt∧RK−1p1 = 0(10)

(8) 和 (9) 这两个式子都称为**对极约束**，它以形式简洁著名。它的几何意义是 O 1 , O 2 , P O_{1}, O_{2}, P O1, O2, P 三点共面。

因为对极约束包含的未知数有旋转 R \boldsymbol{R} R 和平移 t \boldsymbol{t} t，旋转和平移的自由度都为3，由于尺度等价性，去掉一个自由度。因此，如果我们有5对以上匹配点，就可以通过这些二维图像点的对应关系，恢复出在两帧之间摄像机的运动。

本质矩阵和基础矩阵

对极约束中同时包含了平移和旋转。我们把中间部分记作两个矩阵：本质矩阵 E（Essential Matrix）和基础矩阵 F（Fundamental Matrix）。

基础矩阵：

$$\boldsymbol{E}=\boldsymbol{t}^{\wedge} \boldsymbol{R} \tag{11}$$

本质矩阵：

$$\boldsymbol{F}=\boldsymbol{K}^{-T} \boldsymbol{E} \boldsymbol{K}^{-1} \tag{12}$$

因此，对极约束可以进一步简化为：

$$\boldsymbol{x}_{2}^{T} \boldsymbol{E} \boldsymbol{x}_{1}=\boldsymbol{p}_{2}^{T} \boldsymbol{F} \boldsymbol{p}_{1}=0 \tag{13}$$

对极约束简洁地给出了两个匹配点的空间位置关系。于是，相机位姿估计问题变为以下两步:

1. 根据配对点的像素位置，求出 E 或者 F；
2. 根据 E 或者 F，求出 R, t。

由于 E 和 F 只相差了相机内参，而内参在 SLAM 中通常是已知的，所以实践当中往往使用形式更简单的 E。我们以 E 为例，介绍上面两个问题如何求解。

本质矩阵 E

由定义有，本质矩阵 $E=t^{\wedge} R$，有以下性质：

- **尺度等价性**：本质矩阵是由对极约束定义的。由于对极约束是等式为零的约束，所以对 E 乘以任意非零常数后，对极约束依然满足。我们把这件事情称为 E 在不同尺度下是等价的。
- **本质矩阵的内在性质**：根据 $E=t^{\wedge} R$，可以证明，本质矩阵 E 的奇异值必定是 $[\sigma, \sigma, 0]^{T}$ 的形式。
- **本质矩阵的自由度**：因为平移和旋转各有三个自由度，故 $t^{\wedge} R$ 共有六个自由度。但由于尺度等价性,故 E 实际上只有五个自由度。

E 具有五个自由度的事实，表明我们最少可以用五对点来求解 E。但是，E 的内在性质是一种非线性性质，在求解线性方程时会带来麻烦，因此，也可以只考虑它的尺度等价性，使用八对点来估计——这就是经典的八点法(Eight-point-algorithm)。然后对八点法进行奇异值分解，求解出最终值。

八点法求解本质矩阵 E

考虑一对匹配点，它们的归一化坐标为： $\boldsymbol{x}_{1}=\left[u_{1}, v_{1}, 1\right]^{T}, \boldsymbol{x}_{2}=\left[u_{2}, v_{2}, 1\right]^{T}$。根据对极约束，有：

$$\left(u_{1}, v_{1}, 1\right)\left(\begin{array}{ccc}e_{1} & e_{2} & e_{3} \\ e_{4} & e_{5} & e_{6} \\ e_{7} & e_{8} & e_{9}\end{array}\right)\left(\begin{array}{c}u_{2} \\ v_{2} \\ 1\end{array}\right)=0 \tag{14}$$

我们把矩阵 E 展开，写成向量的形式：

$$\boldsymbol{e}=\left[e_{1}, e_{2}, e_{3}, e_{4}, e_{5}, e_{6}, e_{7}, e_{8}, e_{9}\right]^{T} \tag{15}$$

那么对极约束可以写成与 $\boldsymbol{e}$ 有关的线性形式：

$$\left[u_{1} u_{2}, u_{1} v_{2}, u_{1}, v_{1} u_{2}, v_{1} v_{2}, v_{1}, u_{2}, v_{2}, 1\right] \cdot \boldsymbol{e}=0$$

同理，对于其它点对也有相同的表示。我们把所有点都放到一个方程中，变成线性方程组：

$$\left(\begin{array}{ccccccccc}u_{1}^{1} u_{2}^{1} & u_{1}^{1} v_{2}^{1} & u_{1}^{1} & v_{1}^{1} u_{2}^{1} & v_{1}^{1} v_{2}^{1} & v_{1}^{1} & u_{2}^{1} & v_{2}^{1} & 1 \\ u_{1}^{2} u_{2}^{2} & u_{1}^{2} v_{2}^{2} & u_{1}^{2} & v_{1}^{2} u_{2}^{2} & v_{1}^{2} v_{2}^{2} & v_{1}^{2} & u_{2}^{2} & v_{2}^{2} & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ u_{1}^{8} u_{2}^{8} & u_{1}^{8} v_{2}^{8} & u_{1}^{8} & v_{1}^{8} u_{2}^{8} & v_{1}^{8} v_{2}^{8} & v_{1}^{8} & u_{2}^{8} & v_{2}^{8} & 1\end{array}\right)\left(\begin{array}{l}e_{1} \\ e_{2} \\ e_{3} \\ e_{4} \\ e_{5} \\ e_{6} \\ e_{7} \\ e_{8} \\ e_{9}\end{array}\right)=0 \tag{16}$$

上面的方程组可以通过 SVD 分解法解出，具体解法就不阐述了，见SLAM十四讲。

对极几何约束的局限性

尺度不确定性

由于本质矩阵 E 会有尺度等价性，所以它分解得到的 t 和 R 也会有尺度等价性。通常我们会把 t 进行归一化再进行求解。这样就会造成**尺度不确定性**。

初始化中的纯旋转

从 E 分解得到 R 和 t 的过程中，如果发生的是纯旋转，即 $t=0$，那么也就无法退出 R。

因此，在单目视觉的初始化中，必须要有一定程度的平移。

多于8点对的情况

当给定的点对数多于8对时，我们可以计算一个最小二乘解。

当可能出现误匹配的时候，更倾向于使用RANSAC（随机采样一致性）的方法来求解。

参考文献：视觉SLAM十四讲