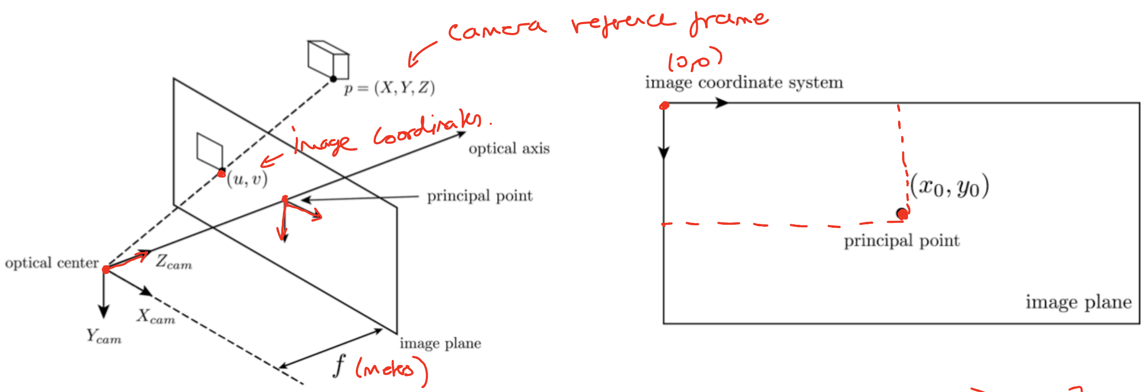
# Discussion 5: computer vision in practice

1. Realistic camera matrices




Recall: $u = f\dfrac{X}{Z}$ , $v = f\dfrac{Y}{Z}$ $\Rightarrow$ $\begin{pmatrix} u \\ v \\ 1 \end{pmatrix} = \dfrac{1}{Z} \underbrace{\begin{bmatrix} f & 0 & 0 \\ 0 & f & 0 \\ 0 & 0 & 1 \end{bmatrix}}_{K} \begin{bmatrix} X \\ Y \\ Z \end{bmatrix}$

K (camera "intrinsic" matrix).

① u,v to be in units of pixels

→ Scale factor s (pixels/m)

(often have rectangular pixels)

→ Need two factors $s_x$ $s_y$

$u = \underbrace{s_x f}_{f_x} \dfrac{X}{Z}$ $\qquad v = \underbrace{s_y f}_{f_y} \dfrac{Y}{Z}$ ← "focal lengths" in pixels.

$u = f_x \dfrac{X}{Z}$ $\qquad v = f_y \dfrac{Y}{Z}$ $\Rightarrow$ $\begin{pmatrix} u \\ v \\ 1 \end{pmatrix} = \dfrac{1}{Z} \underbrace{\begin{bmatrix} f_x & 0 & 0 \\ 0 & f_y & 0 \\ 0 & 0 & 1 \end{bmatrix}}_{K} \begin{bmatrix} X \\ Y \\ Z \end{bmatrix}$

② u,v should not be centered in the center of the image plane.

$u = f_x \dfrac{X}{Z} + x_0$ $\qquad v = f_y \dfrac{Y}{Z} + y_0$

s ← skew

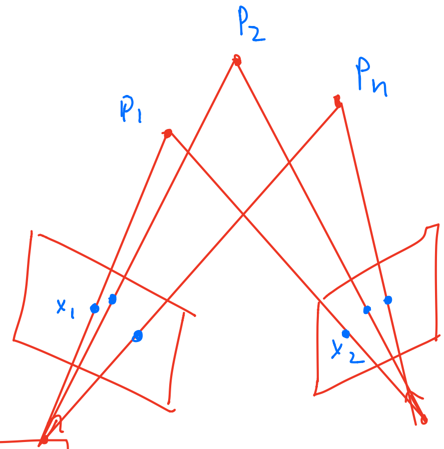
$\begin{pmatrix} u \\ v \\ 1 \end{pmatrix} = \dfrac{1}{Z} \begin{bmatrix} f_x & 0 & x_0 \\ 0 & f_y & y_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \end{bmatrix}$

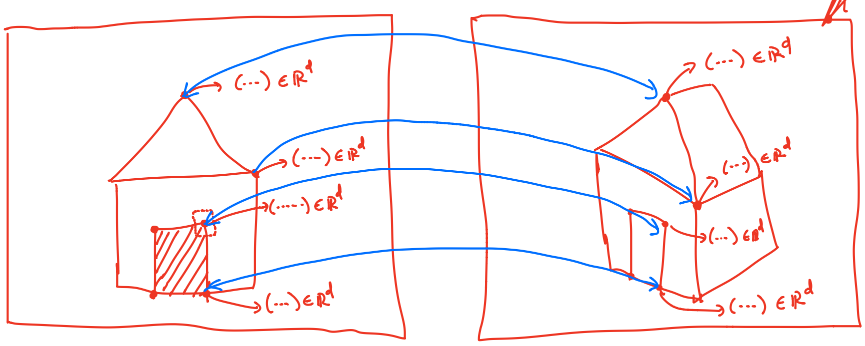
$K \leftarrow$ usually what K looks like.

2. Realistic Structure from Motion

Yesterday: $\boxed{\{(x_1^{(i)}, x_2^{(i)})\}_{i=1}^{n}}$

assumed these were given.




feature extraction + Matching

$\hookrightarrow$ SURF, SIFT, ORB, BRISK ...

(1) After matching, we need to reject outliers.

(A) RANSAC (Random Sampling & consensus).

$\rightarrow$ Given only 8 points, we can compute essential matrix $E$.

given an $E$: a match $(x_1, x_2)$ is an "inlier" if $|x_2^T E x| < \varepsilon$

otherwise, if $|x_2^T E x| >> 0$, "outlier".

RANSAC Algorithm:

Input: $S = \{(x_1^{(i)}, x_2^{(i)})\}_{i=1}^{n}$

Output: A "good" $E$ and a set $\bar{S}$ of "good" feature matches.

(a) Sample 8 points from $S$ at random.

(b) compute $E$ from above.

(c) count "inliers" for $E \leftarrow S'$

(d) If $S'$ is the largest set of inliers found so far,
$\bar{S} \leftarrow S'$

(e) repeat until termination criteria.

return $E, \bar{S}$.

## B LMEDS (Least Median of squares)

Let's define an error function

$$e_i(E) = \left| x_2^{(i)T} E \, x_1^{(i)} \right|^2$$

$$\min_E \underbrace{\sum_{i=1}^{n}}_{\text{replace with a median}} e_i(E)$$

$$\min_E \ \operatorname{med}_i \ e_i(E)$$

median will lie in the inliers.

$$\underbrace{\overbrace{(x_1^{(1)}, x_2^{(1)}), \dots, (x_1^{(n-k)}, x_2^{(n-k)})}^{\text{small } e_i}}_{\text{inliers}}, \underbrace{\overbrace{(x_1^{(n-k+1)}, x_2^{(n-k+1)}), \dots, (x_1^{(n)}, x_2^{(n)})}^{\text{big } e_i}}_{\text{outliers}}$$

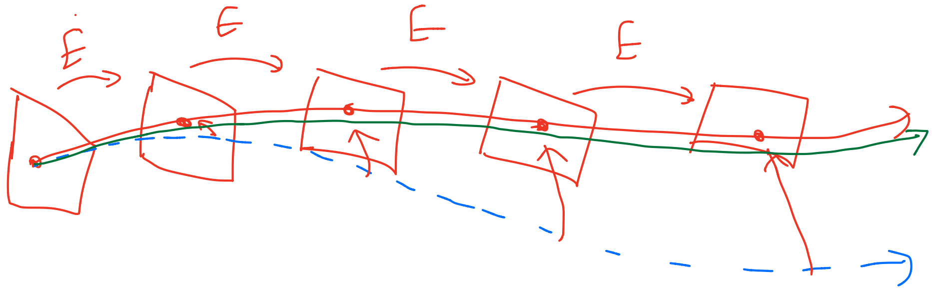
$k$ is small

Solve this optimization problem by sampling.

## (2) Motion Reconstruction:

Yesterday: given $S = \{(x_1^{(i)}, x_2^{(i)})\}$, find $E$

$$T \leftarrow E \rightarrow R$$




Def: "projection function" $\pi_1(P_i, R, T)$ = image coordinates of $P_i$ in frame $\underline{1}$.

point

$\pi_2(P_i, R, T)$ = " " in frame $\underline{2}$.

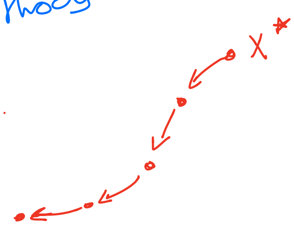
$$f(P, R, T) = \sum_{i=1}^{n} \underbrace{\| x_1^{(i)} - \pi_1(P_i, R, T) \|_2^2}_{\text{reprojection}} + \| x_2^{(i)} - \pi_2(P_i, R, T) \|_2^2$$

$\ell_2$ norm $\|x\|_2 = \sqrt{x_1^2 + \dots x_n^2}$

$\min_{P,R,T} f(P, R, T)$ ← can't just do this globally.

→ start with initial guess $P^*, R^*, T^*$ ← Use not-very-robust methods to compute guess.

→ Use gradient methods to refine further.




$$x_{k+1} \leftarrow x_k - \alpha \nabla f(x_k)$$

This is called BUNDLE ADJUSTMENT