Vision par ordinateur pour la robotique

Chapitre 1 – Vision géométrique

Hicham Hadj-Abdelkader

hicham.hadjabdelkader@univ-evry.fr

https://www.ibisc.univ-evry.fr/~hhadjabdelkader/

Année universitaire 2026–2027

Université d'Évry Paris-Saclay – UFR Sciences et Technologies
Master 2 – Parcours RI (robotique intelligente et industrielle)

Organisation de l'enseignement

Volume

- CM (4 h) : ch. 1 *Vision géométrique* (2 h) et ch. 2 *Asservissement visuel* (2 h)
- TD (4 h) : 2 séances de 2 h, qui préparent directement les TP
- TP (8 h) : 2 séances de 4 h – *calibration main–œil et plan de travail*, puis *asservissement visuel d'un bras* (MATLAB)

Évaluation : comptes rendus de TP + examen écrit.

Compléments (hors volume) : version complète des chapitres, CM 3 *commande référencée capteurs*, TD 5, TP reconstruction 3D.

Plan du chapitre 1

1. Introduction : la vision pour la manipulation
2. Rappels : rotations, poses et vitesses
3. Modèle sténopé de la caméra
4. Calibration, main–œil, pose (PnP)
5. Primitives et appariement
6. Estimation robuste (RANSAC)
7. Homographie du plan de travail
8. Géométrie épipolaire (aperçu)

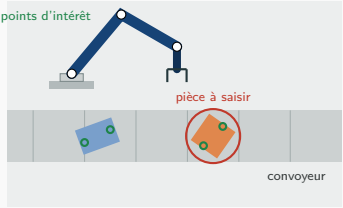
Introduction : la vision pour la robotique de manipulation

Pourquoi la vision pour un robot manipulateur ?

- Capteur sans contact, bon marché, riche : géométrie (où est la pièce ? quelle est sa pose ?) et sémantique (quelle pièce ? est-elle conforme ?).
- Lève la contrainte des cellules « figées » : pièces en vrac, positions variables, petites séries, robots collaboratifs.

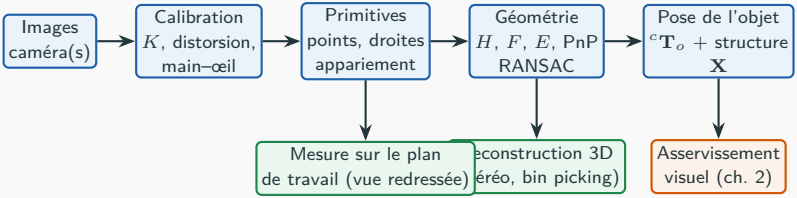
Questions traitées dans ce cours

1. Comment un point 3D se projette-t-il ? (*modèle, calibration, main–œil*)
2. Où est l'objet, quel est le mouvement entre deux vues ? (*PnP, H, géométrie épipolaire*)
3. Quelle est la forme 3D de la scène ? (*triangulation, stéréo*)
4. Comment guider le bras depuis l'image ? (*ch. 2*)



Cellule robotisée : pièces sur un convoyeur, bras équipé d'une caméra sur la pince.

Briques de la vision pour la robotique



Exemple robotique

Fil conducteur du cours et du TP : un **bras 6 axes** équipé d'une caméra sur la pince (*eye-in-hand*) et une **caméra fixe** au-dessus du poste (*eye-to-hand*) : localisation de pièces sur un convoyeur, saisie, positionnement précis par rapport à une pièce.

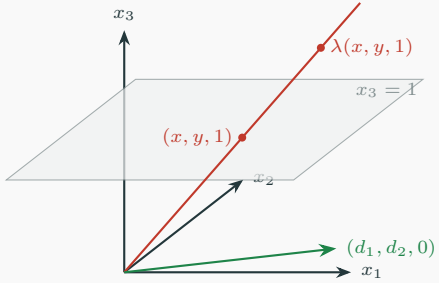
Rappels : géométrie projective et transformations rigides

Coordonnées homogènes dans le plan

Un point $(x, y) \in \mathbb{R}^2$ est représenté par tout vecteur

$$\mathbf{x} = (x_1, x_2, x_3)^\top \neq \mathbf{0}, \qquad (x, y) = \left(\frac{x_1}{x_3}, \frac{x_2}{x_3}\right).$$

- $\mathbf{x}$ et $\lambda\mathbf{x}$ ($\lambda \neq 0$) représentent le même point : $\mathbf{x} \sim \lambda\mathbf{x}$.
- Plan projectif $\mathbb{P}^2$: 2 degrés de liberté (ddl).
- Un point de $\mathbb{P}^2$ = une *droite passant par l'origine* de $\mathbb{R}^3$ (un rayon) ; le plan $x_3 = 1$ en est une carte.
- $x_3 = 0$: points à l'infini (directions).



direction : ne coupe pas le plan $x_3 = 1$

Rotations : le groupe $\mathrm{SO}(3)$

$$\mathrm{SO}(3) = \{R \in \mathbb{R}^{3\times 3} \ : \ R^\top R = I, \ \det R = +1\} \qquad (3 \text{ ddl})$$

- $R^{-1} = R^\top$; colonnes = axes du repère tourné.
- Composition non commutative : $R_1 R_2 \neq R_2 R_1$.
- Rotations élémentaires, p. ex. lacet :

$$R_z(\psi) = \begin{bmatrix} \cos\psi & -\sin\psi & 0 \\ \sin\psi & \cos\psi & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Matrice antisymétrique associée à $\boldsymbol{\omega}$:

$$[\boldsymbol{\omega}]_\times = \begin{bmatrix} 0 & -\omega_3 & \omega_2 \\ \omega_3 & 0 & -\omega_1 \\ -\omega_2 & \omega_1 & 0 \end{bmatrix}, \quad [\boldsymbol{\omega}]_\times \mathbf{v} = \boldsymbol{\omega} \times \mathbf{v}.$$

$[\boldsymbol{\omega}]_\times^\top = -[\boldsymbol{\omega}]_\times$, $[\boldsymbol{\omega}]_\times \boldsymbol{\omega} = \mathbf{0}$, $\mathrm{rang}[\boldsymbol{\omega}]_\times = 2$.

Représentation axe–angle $\theta\mathbf{u}$

Rotation d'angle θ autour de l'axe unitaire $\mathbf{u}$ (formule de Rodrigues) :

$$\boxed{R = \exp(\theta[\mathbf{u}]_\times) = I + \sin\theta\,[\mathbf{u}]_\times + (1-\cos\theta)\,[\mathbf{u}]_\times^2}$$

Inverse : $\cos\theta = \dfrac{\operatorname{tr} R - 1}{2}, \quad [\mathbf{u}]_\times = \dfrac{R - R^\top}{2\sin\theta} \quad (0 < \theta < \pi).$

- $\theta\mathbf{u} \in \mathbb{R}^3$: paramétrage minimal, singulier seulement en $\theta = \pi$; utilisé en asservissement visuel 3D et 2D 1/2 (ch. 2).
- Petits angles : $R \approx I + [\theta\mathbf{u}]_\times$.

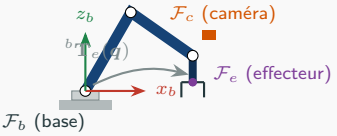
Attention

Angles d'Euler (roulis, tangage, lacet) : intuitifs (consignes RPY des contrôleurs de robots) mais singuliers (*gimbal lock*) ; on préfère $\theta\mathbf{u}$ ou les quaternions pour l'estimation et la commande.

Transformations rigides : le groupe $\mathrm{SE}(3)$

$$ {}^a\mathbf{T}_b = \begin{bmatrix} {}^a\mathbf{R}_b & {}^a\boldsymbol{t}_b \\ \mathbf{0}^\top & 1 \end{bmatrix} \in \mathrm{SE}(3), \qquad {}^a\mathbf{X} = {}^a\mathbf{R}_b\, {}^b\mathbf{X} + {}^a\boldsymbol{t}_b. $$

- Composition : ${}^a\mathbf{T}_c = {}^a\mathbf{T}_b\, {}^b\mathbf{T}_c$.
- Inverse : ${}^a\mathbf{T}_b^{-1} = \begin{bmatrix} R^\top & -R^\top\mathbf{t} \\ \mathbf{0}^\top & 1 \end{bmatrix}$.
- 6 ddl : 3 en rotation, 3 en translation.
- Paramètres extrinsèques de la caméra : $[R\,|\,\mathbf{t}] = {}^c\mathbf{T}_w$; centre optique $\mathbf{C} = -R^\top\mathbf{t}$ (dans $\mathcal{F}_w$).



${}^b\mathbf{T}_c = {}^b\mathbf{T}_e(\boldsymbol{q})\, {}^e\mathbf{T}_c$: modèle géométrique du robot (variable) × montage caméra (fixe, calibration main–œil).

Exemple : caméra montée sur la pince (*eye-in-hand*)

Exemple robotique

Caméra fixée sur la bride, axe optique selon l'axe de l'outil z_e, décalée de 6 cm et reculée de 4 cm :
Repère effecteur : z_e axe de l'outil, x_e, y_e dans le plan de la bride.
Repère caméra : $z_c = z_e$, $x_c = y_e$, $y_c = -x_e$.

$$ {}^{e}\mathbf{R}_c = \begin{bmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \qquad {}^{e}\boldsymbol{t}_c = (0.06,\, 0,\, 0.04)^\top. $$

- Colonnes de ${}^{e}\mathbf{R}_c$: axes x_c, y_c, z_c exprimés dans $\mathcal{F}_e$.
- Un point ${}^{c}\mathbf{X}$ vu par la caméra est dans le repère effecteur en ${}^{e}\mathbf{X} = {}^{e}\mathbf{R}_c\, {}^{c}\mathbf{X} + {}^{e}\boldsymbol{t}_c$.
- Exemple : ${}^{c}\mathbf{X} = (0.02,\, 0.05,\, 0.56)^\top$ $\Rightarrow {}^{e}\mathbf{X} = (0.01,\, 0.02,\, 0.60)^\top$: 60 cm devant la bride.

Transformation homogène : lire et composer les poses

${}^{a}\mathbf{T}_{b}$ = pose du repère $\mathcal{F}_b$ dans le repère $\mathcal{F}_a$:

- colonnes de ${}^{a}\mathbf{R}_{b}$: axes x_b, y_b, z_b exprimés dans $\mathcal{F}_a$;
- ${}^{a}\boldsymbol{t}_{b}$: origine de $\mathcal{F}_b$ exprimée dans $\mathcal{F}_a$;
- elle transporte des coordonnées de b vers a : ${}^{a}\tilde{\mathbf{X}} = {}^{a}\mathbf{T}_{b}\,{}^{b}\tilde{\mathbf{X}}$.

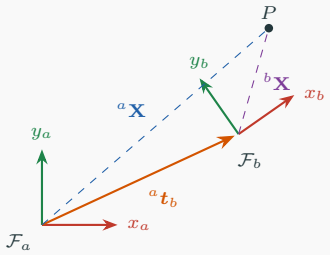
Point ou direction ?

$$\tilde{\mathbf{X}} = \begin{bmatrix} \mathbf{X} \\ 1 \end{bmatrix} \text{ (point)}, \qquad \tilde{\mathbf{d}} = \begin{bmatrix} \mathbf{d} \\ 0 \end{bmatrix} \text{ (direction)}$$

la translation ne s'applique pas aux directions.

Règle des indices : ils se « simplifient » de proche en proche

$${}^{a}\mathbf{T}_{c} = {}^{a}\mathbf{T}_{b}\,{}^{b}\mathbf{T}_{c}, \qquad {}^{b}\mathbf{T}_{a} = \left({}^{a}\mathbf{T}_{b}\right)^{-1}.$$



Attention

Erreurs fréquentes : ordre des produits, confusion entre ${}^{a}\mathbf{T}_{b}$ et son inverse, translation appliquée à une direction.

Vitesse angulaire : dérivée d'une matrice de rotation

${}^a\mathbf{R}_b(t)$ varie dans le temps. En dérivant $RR^\top = I$: $\dot{R}R^\top + (\dot{R}R^\top)^\top = 0$, donc $\dot{R}R^\top$ est antisymétrique :

$$\dot{R} = [\boldsymbol{\omega}]_\times R = R\,[\boldsymbol{\omega}_b]_\times, \qquad \boldsymbol{\omega} = R\,\boldsymbol{\omega}_b$$

- $\boldsymbol{\omega}$: vitesse angulaire de $\mathcal{F}_b$ par rapport à $\mathcal{F}_a$, exprimée dans $\mathcal{F}_a$;
- $\boldsymbol{\omega}_b = R^\top \boldsymbol{\omega}$: la même, exprimée dans $\mathcal{F}_b$ (c'est ce que mesure un gyromètre embarqué) ;
- c'est un vecteur : direction = axe instantané de rotation, norme = vitesse de rotation (rad/s).
- point fixé à $\mathcal{F}_b$, rotation pure : $\dot{\mathbf{X}} = \boldsymbol{\omega} \times \mathbf{X}$;
- exemple : $R = R_z(\psi(t)) \Rightarrow \dot{R}R^\top = \dot{\psi}\,[\mathbf{e}_z]_\times$, $\boldsymbol{\omega} = (0, 0, \dot{\psi})$;
- ce résultat simple n'est vrai que pour une rotation autour d'un axe fixe.

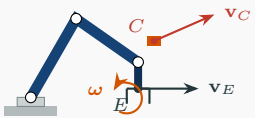
Torseur cinématique d'un solide

Champ des vitesses d'un solide : pour deux points O et P du solide,

$$\mathbf{v}_P = \mathbf{v}_O + \boldsymbol{\omega} \times \overrightarrow{OP}$$

- $\boldsymbol{\omega}$ est la même pour tous les points ; la vitesse linéaire dépend du point ;
- le torseur cinématique $\boldsymbol{v} = (\mathbf{v}_O, \boldsymbol{\omega})$ décrit tout le mouvement, réduit au point O et exprimé dans un repère choisi ;
- forme matricielle (repère du corps) :

$$\mathbf{T}^{-1}\dot{\mathbf{T}} = \begin{bmatrix} [\boldsymbol{\omega}_b]_\times & \mathbf{v}_b \\ \mathbf{0}^\top & 0 \end{bmatrix} = [\boldsymbol{v}_b]^\wedge, \text{ avec } \mathbf{v}_b = R^\top \dot{\mathbf{t}}.$$



l'effecteur tourne : la caméra, décalée, reçoit en plus la vitesse $\boldsymbol{\omega} \times \overrightarrow{EC}$.

Exemple : vitesse de la caméra embarquée

La bride translate à $0{,}1$ m/s selon x_e et tourne autour de l'axe de l'outil à $0{,}5$ rad/s ; torseur réduit au centre de la bride E, dans $\mathcal{F}_e$: $\mathbf{v}_E = (0.1, 0, 0)$, $\boldsymbol{\omega} = (0, 0, 0.5)$. Caméra en C, $\overrightarrow{EC} = (0.06, 0, 0.04)$:

$$\mathbf{v}_C = \mathbf{v}_E + \boldsymbol{\omega} \times \overrightarrow{EC} = (0.1,\ 0.03,\ 0) \text{ m/s} \qquad (\text{dans } \mathcal{F}_e)$$

Exprimé dans $\mathcal{F}_c$ (multiplication par ${}^{e}\mathbf{R}_{c}^{\top}$) :

$$\boldsymbol{v}_c = (0.03,\ -0.1,\ 0), \qquad \boldsymbol{\omega}_c = (0,\ 0,\ 0.5)$$

Une rotation de l'outil autour de son axe n'est pas une rotation pure de la caméra : le centre optique décalé décrit un cercle (3 cm/s ici).

À retenir

Un torseur se transporte (changement de point) et se projette (changement de repère d'expression) : c'est la matrice ${}^{c}\mathbf{V}_e$ du chapitre 2, indispensable pour passer de la vitesse calculée pour la caméra à la consigne de l'effecteur.

Mini-quiz – Poses et vitesses

1. On connaît ${}^{b}\mathbf{T}_{c} = [R \,|\, \mathbf{t}]$. Écrire ${}^{c}\mathbf{T}_{b}$, la matrice utilisée pour projeter.
2. Un outil dont le repère est aligné avec la base tourne autour de l'axe vertical à $0{,}5\,\mathrm{rad/s}$. Que vaut $\boldsymbol{\omega}_e$? Et si l'outil est basculé de $\theta = 90^\circ$?
3. L'outil tourne autour de son axe à $1\,\mathrm{rad/s}$; la caméra est à $6\,\mathrm{cm}$ de cet axe. Quelle est la vitesse linéaire du centre optique ?
4. Pourquoi intégrer $\dot{R} = R\,[\boldsymbol{\omega}_b]_\times$ plutôt que les angles d'Euler ?

Mini-quiz – Poses et vitesses

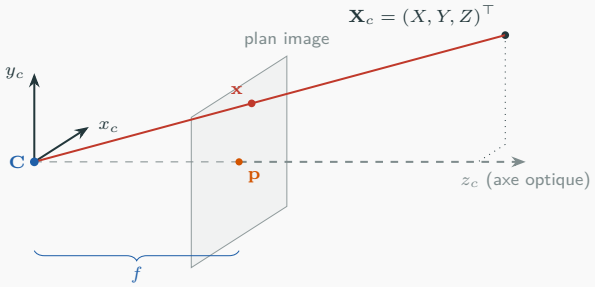
1. On connaît ${}^{b}\mathbf{T}_{c} = [R \,|\, \mathbf{t}]$. Écrire ${}^{c}\mathbf{T}_{b}$, la matrice utilisée pour projeter.
2. Un outil dont le repère est aligné avec la base tourne autour de l'axe vertical à $0{,}5\,\text{rad/s}$. Que vaut $\boldsymbol{\omega}_e$? Et si l'outil est basculé de $\theta = 90°$?
3. L'outil tourne autour de son axe à $1\,\text{rad/s}$; la caméra est à $6\,\text{cm}$ de cet axe. Quelle est la vitesse linéaire du centre optique ?
4. Pourquoi intégrer $\dot{R} = R\,[\boldsymbol{\omega}_b]_\times$ plutôt que les angles d'Euler ?

Réponses

1. ${}^{c}\mathbf{T}_{b} = \begin{bmatrix} R^\top & -R^\top \mathbf{t} \\ \mathbf{0}^\top & 1 \end{bmatrix}$.
2. $\boldsymbol{\omega}_e = (0, 0, 0.5)$; avec $\theta = 90°$: $\boldsymbol{\omega}_e = (-0.5, 0, 0)$, et $\dot{\phi}$, $\dot{\psi}$ ne sont plus distinguables ($\det E = 0$).
3. $\mathbf{v}_C = \boldsymbol{\omega} \times \overrightarrow{EC}$: $6\,\text{cm/s}$; la vitesse angulaire reste $1\,\text{rad/s}$.
4. Pas de singularité, et avec l'exponentielle la matrice reste une rotation.

Modèle sténopé de la caméra

Le modèle sténopé (*pinhole*)



- Tous les rayons passent par le centre optique $\mathbf{C}$.
- Plan image (virtuel, devant $\mathbf{C}$) à la distance focale f.
- Point principal $\mathbf{p}$: intersection de l'axe optique et du plan image.
- Théorème de Thalès :

$$x = f\frac{X}{Z}, \qquad y = f\frac{Y}{Z}.$$

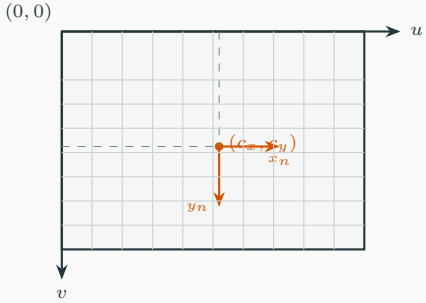
- En homogène, la projection devient linéaire :

$$Z\begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \begin{bmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}.$$

Des coordonnées métriques aux pixels : la matrice K

Coordonnées normalisées : $\mathbf{x}_n = (X/Z,\, Y/Z,\, 1)^\top$. Alors

$$\mathbf{x} = K\mathbf{x}_n, \qquad K = \begin{bmatrix} f_x & s & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix}.$$



- $f_x = f/\rho_u$, $f_y = f/\rho_v$: focale en pixels (ρ_u, ρ_v : taille d'un pixel).
- (c_x, c_y) : point principal (pixels), proche du centre de l'image.
- s : *skew* (axes non orthogonaux), $s \approx 0$ pour les capteurs actuels.
- 5 paramètres intrinsèques.
- En asservissement visuel (ch. 2) : coordonnées normalisées $x = (u - c_x)/f_x$, $y = (v - c_y)/f_y$.

Matrice de projection complète

Avec la pose (paramètres extrinsèques) $\mathbf{X}_c = R\mathbf{X}_w + \mathbf{t}$:

$$\lambda\,\mathbf{x} = K\,[R\,|\,\mathbf{t}]\,\mathbf{X} = P\mathbf{X}, \qquad P \in \mathbb{R}^{3\times 4}$$

- P est définie à un facteur près : 11 ddl = 5 (intrinsèques) + 6 (pose).
- Profondeur : $\lambda = Z = \mathbf{r}_3^\top \mathbf{X}_w + t_3$; le point est devant la caméra ssi $\lambda > 0$.
- Notation fonctionnelle : $\mathbf{x} = \pi(P\mathbf{X})$ avec $\pi(a,b,c) = (a/c,\ b/c)$.

Exemple numérique

$K = \begin{bmatrix} 800 & 0 & 320 \\ 0 & 800 & 240 \\ 0 & 0 & 1 \end{bmatrix}$, $R = I$, $\mathbf{t} = (0.1,\ 0,\ 2)^\top$ m, $\mathbf{X}_w = (0.4,\ -0.2,\ 2)^\top$ m.

1. $\mathbf{X}_c = (0.5,\ -0.2,\ 4)^\top$, profondeur $4\,\text{m} > 0$; normalisées $\mathbf{x}_n = (0.125,\ -0.05,\ 1)^\top$.
2. Pixels : $u = 800 \times 0.125 + 320 = \mathbf{420}$, $v = 800 \times (-0.05) + 240 = \mathbf{200}$.

Sensibilité : 1 pixel à $4\,\text{m}$ correspond à $Z/f_x = 5\,\text{mm}$ latéralement ; à $40\,\text{m}$, $5\,\text{cm}$.

Distorsion des objectifs

Les objectifs réels (surtout les courtes focales, fréquentes sur une caméra embarquée) s'écartent du sténopé. Modèle de Brown–Conrady sur les coordonnées normalisées, $r^2 = x^2 + y^2$:

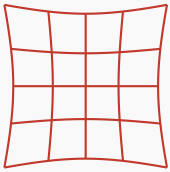
$$x_d = x\,(1 + k_1 r^2 + k_2 r^4 + k_3 r^6) + 2p_1 xy + p_2(r^2 + 2x^2)$$
$$y_d = y\,(1 + k_1 r^2 + k_2 r^4 + k_3 r^6) + p_1(r^2 + 2y^2) + 2p_2 xy$$

puis $\mathbf{x} = K(x_d, y_d, 1)^\top$.

- k_i : distorsion radiale ($k_1 < 0$: barillet ; $k_1 > 0$: coussinet) ;
- p_1, p_2 : distorsion tangentielle (décentrement).
- Fisheye / caméras omnidirectionnelles : modèles spécifiques (Kannala–Brandt, sphère unitaire).



barillet ($k_1 < 0$)



coussinet ($k_1 > 0$)

Exemple

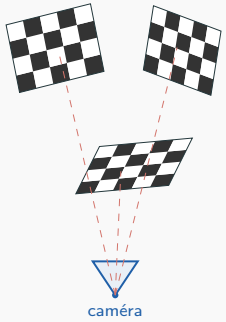
$k_1 = -0.2$, point normalisé à $r = 0.5$: facteur $1 - 0.2 \times 0.25 = 0.95$; avec $f_x = 800$, un point à 400 px du centre est déplacé d'environ 20 px : une ligne de marquage droite apparaît courbée – il faut corriger la distorsion avant toute géométrie.

Calibration de la caméra

Objectif de la calibration

Estimer K (et la distorsion), éventuellement la pose, à partir de correspondances $\mathbf{X}_j \leftrightarrow \mathbf{x}_j$ entre les points d'une mire connue et leurs images.

- Mire 3D (deux plans) : DLT directe sur P.
- Mire plane (damier) vue sous plusieurs poses : méthode de Zhang (OpenCV, MATLAB `cameraCalibrator`).
- Dans tous les cas : raffinement non linéaire de l'erreur de reprojection.



Exemple robotique

Pour un robot, on calibre aussi les **extrinsèques** : caméra / effecteur ${}^{e}\mathbf{T}_{c}$ (calibration main–œil, caméra embarquée) ou caméra / base ${}^{b}\mathbf{T}_{c}$ (caméra fixe). Le robot déplace lui-même la mire ou la caméra.

Méthode de Zhang (2000) : mire plane

Repère monde attaché à la mire : $Z = 0$. Alors

$$\lambda \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = K[\mathbf{r}_1\ \mathbf{r}_2\ \mathbf{r}_3\ \mathbf{t}] \begin{bmatrix} X \\ Y \\ 0 \\ 1 \end{bmatrix} = \underbrace{K[\mathbf{r}_1\ \mathbf{r}_2\ \mathbf{t}]}_{H=[\mathbf{h}_1\ \mathbf{h}_2\ \mathbf{h}_3]} \begin{bmatrix} X \\ Y \\ 1 \end{bmatrix}.$$

Chaque vue donne une homographie H. Comme $\mathbf{r}_1 \perp \mathbf{r}_2$ et $\|\mathbf{r}_1\| = \|\mathbf{r}_2\|$, avec $\mathbf{r}_i = \mu K^{-1}\mathbf{h}_i$:

$$\boxed{\mathbf{h}_1^\top\, \omega\, \mathbf{h}_2 = 0, \qquad \mathbf{h}_1^\top\, \omega\, \mathbf{h}_1 = \mathbf{h}_2^\top\, \omega\, \mathbf{h}_2}, \qquad \omega = K^{-\top}K^{-1} \text{ (symétrique, 6 coefficients).}$$

- Chaque image : 2 équations linéaires en $\mathbf{b} = (\omega_{11}, \omega_{12}, \omega_{22}, \omega_{13}, \omega_{23}, \omega_{33})^\top$.
- $n \geq 3$ images en position générale ; SVD de $V\mathbf{b} = \mathbf{0}$.
- $\omega = LL^\top$ (Cholesky) $\Rightarrow K^{-1} = L^\top$; puis pour chaque vue $\mu = 1/\|K^{-1}\mathbf{h}_1\|$, $\mathbf{r}_1 = \mu K^{-1}\mathbf{h}_1$, $\mathbf{r}_2 = \mu K^{-1}\mathbf{h}_2$, $\mathbf{r}_3 = \mathbf{r}_1 \times \mathbf{r}_2$, $\mathbf{t} = \mu K^{-1}\mathbf{h}_3$.
- Orthogonalisation : $Q = [\mathbf{r}_1\, \mathbf{r}_2\, \mathbf{r}_3] = U\Sigma V^\top \Rightarrow R = UV^\top$.

Raffinement non linéaire et bonnes pratiques

$$\min_{K,\mathbf{k},\{R_i,\mathbf{t}_i\}} \sum_{i=1}^{n}\sum_{j=1}^{N} \left\| \mathbf{x}_{ij} - \hat{\mathbf{x}}(K,\mathbf{k},R_i,\mathbf{t}_i,\mathbf{X}_j) \right\|^2 \quad \text{(Levenberg–Marquardt)}$$

- Estimateur du maximum de vraisemblance pour un bruit image gaussien.
- **Erreur RMS de reprojection** : bonne calibration 0.1 à 0.5 px.

- 15 à 30 images, mire inclinée de ± 30–$45°$ (les vues fronto-parallèles seules sont dégénérées).
- Couvrir tout le champ, en particulier les bords (distorsion).
- Mise au point et zoom fixes ; mire rigide et plane.

Exemple

15 images d'un damier 9×6 coins : 810 observations, 1620 équations pour $4 + 5 + 15 \times 6 = 99$ inconnues.

À retenir

Erreur *algébrique* pour initialiser (linéaire, rapide) → erreur *géométrique* (reprojection) pour raffiner. Ce schéma revient dans tout le cours.

Calibration main–œil : le problème $AX = XB$

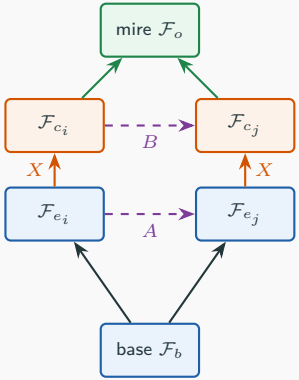
Caméra embarquée : on cherche $X = {}^{e}\mathbf{T}_{c}$. Le robot amène la caméra devant une mire fixe sous plusieurs poses i :

- ${}^{b}\mathbf{T}_{e_i}$: donnée par le modèle géométrique du robot ;
- ${}^{c_i}\mathbf{T}_{o}$: pose de la mire, donnée par la calibration (Zhang) ou un PnP.

La mire est fixe : ${}^{b}\mathbf{T}_{e_i}\, X\, {}^{c_i}\mathbf{T}_{o} = {}^{b}\mathbf{T}_{o}$ pour tout i. Entre deux poses i et j :

$$\underbrace{{}^{b}\mathbf{T}_{e_i}^{-1}\,{}^{b}\mathbf{T}_{e_j}}_{A}\, X = X\, \underbrace{{}^{c_i}\mathbf{T}_{o}\,{}^{c_j}\mathbf{T}_{o}^{-1}}_{B}$$

A : déplacement de l'effecteur, B : déplacement de la caméra, exprimés chacun dans son repère.



Résolution de $AX = XB$ (Tsai–Lenz, Park–Martin)

En séparant rotation et translation :

$$R_A R_X = R_X R_B, \qquad (R_A - I)\,\mathbf{t}_X = R_X \mathbf{t}_B - \mathbf{t}_A.$$

1) Rotation. R_A et R_B sont semblables : mêmes angles, axes liés par

$$\theta \mathbf{u}_A = R_X\, \theta \mathbf{u}_B.$$

Avec $n \geq 2$ mouvements d'axes non parallèles : R_X par la méthode de Procrustes (SVD de $\sum_i \mathbf{a}_i \mathbf{b}_i^\top$).

2) Translation. On empile les équations $(R_{A_i} - I)\,\mathbf{t}_X = R_X \mathbf{t}_{B_i} - \mathbf{t}_{A_i}$: moindres carrés.

Bonnes pratiques

- 10 à 20 poses, avec de grandes rotations autour d'axes variés (une translation pure ne contraint pas R_X) ;
- la précision dépend de celle du robot (modèle géométrique) et de la pose de mire ;
- raffinement final : minimiser l'erreur de reprojection de la mire sur toutes les vues.

Caméra fixe (*eye-to-hand*)

La mire est portée par l'effecteur ; on cherche ${}^b\mathbf{T}_c$: même équation en échangeant les rôles.

Localiser un objet connu : estimation de pose (PnP)

Modèle 3D de l'objet (points $\mathbf{X}_j$ dans $\mathcal{F}_o$) + images $\mathbf{x}_j$ + K connue $\Rightarrow$ pose ${}^c\mathbf{T}_o = [R \,|\, \mathbf{t}]$: problème Perspective-n-Point.

- $n = 3$: P3P, jusqu'à 4 solutions ; avec RANSAC pour rejeter les faux appariements ;
- $n \geq 4$: EPnP (linéaire, rapide), DLT ($n \geq 6$) ;
- objet plan : homographie $H = K[\mathbf{r}_1\ \mathbf{r}_2\ \mathbf{t}]$, puis $\mathbf{r}_1, \mathbf{r}_2, \mathbf{t}$ comme pour Zhang ;
- toujours suivi d'un raffinement de l'erreur de reprojection (Levenberg–Marquardt).

De la pose de l'objet à la saisie

Pose de saisie dans la base, par composition :

$$ {}^b\mathbf{T}_g = \underbrace{{}^b\mathbf{T}_e(\boldsymbol{q})}_{\text{robot}}\ \underbrace{{}^e\mathbf{T}_c}_{\text{main–œil}}\ \underbrace{{}^c\mathbf{T}_o}_{\text{PnP}}\ \underbrace{{}^o\mathbf{T}_g}_{\text{prise}} $$

${}^o\mathbf{T}_g$: prise définie une fois sur le modèle CAO. Toute erreur de la chaîne se retrouve dans la saisie : c'est l'intérêt de l'asservissement visuel (ch. 2).

Primitives visuelles et leur détection

Quelles primitives pour la manipulation ?

Toute la géométrie multi-vues repose sur des correspondances $\mathbf{x}_j \leftrightarrow \mathbf{x}'_j$.

Points

- coins (Harris, FAST)
- blobs multi-échelles (SIFT, SURF)
- usage : objets texturés, homographie, PnP

Droites / contours

- gradients, Canny
- transformée de Hough
- usage : arêtes de pièces, joints de soudure, bords de convoyeur

Régions / mires

- marqueurs (ArUco, AprilTag)
- moments d'une région
- usage : centrage et orientation d'une pièce, prise

Détecteur : *où* ?

- répétable (changement de point de vue, d'échelle, d'éclairage) ;
- bien localisé (sous-pixel) ; nombreux et bien répartis.

Descripteur : *quoi* ?

- discriminant ;
- invariant (rotation, échelle, éclairage affine) ;
- compact et rapide à comparer.

Invariance à l'échelle : SIFT et ORB

SIFT (Lowe, 2004)

- Extrema de la différence de gaussiennes $D(\mathbf{x}, \sigma) = L(\mathbf{x}, k\sigma) - L(\mathbf{x}, \sigma)$ dans l'espace-échelle (26 voisins).
- Localisation sous-pixel, rejet des faibles contrastes et des contours.
- Orientation dominante (histogramme 36 classes).
- Descripteur : 4×4 cellules $\times$ 8 orientations $= 128$ dimensions.

ORB (Rublee *et al.*, 2011)

- FAST (cercle de 16 pixels) + pyramide d'échelles ;
- orientation par centroïde d'intensité ;
- BRIEF orienté : 256 comparaisons binaires ;
- distance de Hamming (XOR + popcount).

	SIFT	ORB
Taille	128 flottants	256 bits
Distance	L_2	Hamming
Échelle	oui (DoG)	pyramide
Vitesse	lente	~ 10–$100\times$ plus rapide
Usage	SfM hors ligne	SLAM temps réel

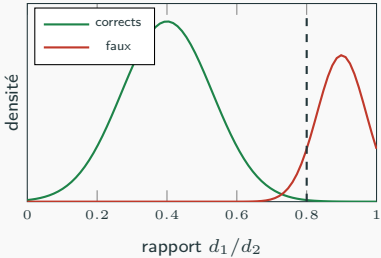
Exemple robotique

Reconnaissance d'objets texturés (cartons, étiquettes) : ORB ou SIFT + RANSAC. Pièces non texturées : appariement de gabarit ou de contours (*chamfer matching*).

Appariement et suivi

Appariement par descripteurs. Pour $\mathbf{d}$ de l'image 1, chercher les deux plus proches voisins $\mathbf{d}'_1, \mathbf{d}'_2$ dans l'image 2.

- **Test du rapport de Lowe** : accepter si $\|\mathbf{d} - \mathbf{d}'_1\| / \|\mathbf{d} - \mathbf{d}'_2\| < \tau \approx 0.8$.
- **Vérification croisée** : $i \to j$ et $j \to i$ doivent se choisir mutuellement.
- Recherche approchée (kd-tree, FLANN) pour de grands ensembles.



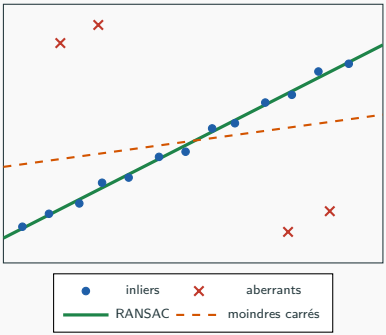
allure schématique (cf. Lowe 2004)

À retenir

Quel que soit le descripteur, une fraction importante des appariements reste fausse (10 à 50 %), auxquels s'ajoutent les points sur les objets mobiles (pièce déplacée, opérateur, reflets) : l'estimation doit être **robuste**.

Estimation robuste : RANSAC

Le problème des données aberrantes



RANSAC (Fischler & Bolles, 1981)

1. Répéter N fois :
   - 1.1 tirer un échantillon minimal de s données ($s = 2$ droite, 4 pour H, 8 pour F, 5 pour E) ;
   - 1.2 estimer le modèle ;
   - 1.3 compter les données cohérentes ($e_j < t$).
2. Garder le modèle de plus grand consensus.
3. Réestimer sur tous ses inliers.

Combien d'itérations ? Quel seuil ?

w : taux d'inliers, p : probabilité voulue de tirer *au moins* un échantillon propre.

$$(1 - w^s)^N = 1 - p \quad \Longrightarrow \quad \boxed{N = \frac{\log(1-p)}{\log(1-w^s)}}$$

$p = 0.99$	$s = 2$	$s = 4$	$s = 5$	$s = 7$	$s = 8$
$w = 0.7$	7	17	26	54	78
$w = 0.5$	17	72	146	588	1 177
$w = 0.3$	49	567	1 893	21 055	70 188

Seuil : bruit gaussien d'écart-type σ (px) ; erreur de codimension m : $t^2 = F^{-1}_{\chi^2_m}(0.95)\,\sigma^2$

- point–droite (distance à la droite épipolaire) : $m = 1$, $t^2 = 3.84\,\sigma^2$;
- point–point (H, reprojection) : $m = 2$, $t^2 = 5.99\,\sigma^2$.

Exemple robotique

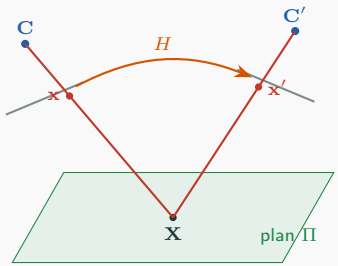
Dans un bac de pièces en vrac, les points d'une autre pièce ou du bac ne suivent pas le modèle de la pièce visée : RANSAC les classe en aberrants. En recommençant sur les points restants, on détecte les autres pièces (RANSAC séquentiel).

À retenir

N croît exponentiellement avec s : d'où l'intérêt des solveurs minimaux (5 points pour E, 3 points pour la pose d'un objet connu (P3P), 2 points pour une rotation pure).

Homographie

Quand deux images sont-elles liées par une homographie ?



Cas 1 – scène plane. Plan Π : $\mathbf{n}^\top \mathbf{X} = d$ (repère caméra 1), $\mathbf{X}' = R\mathbf{X} + \mathbf{t}$. Pour $\mathbf{X} \in \Pi$: $1 = \mathbf{n}^\top \mathbf{X}/d$, d'où

$$\mathbf{X}' = \left(R + \frac{\mathbf{t}\mathbf{n}^\top}{d}\right)\mathbf{X} \;\Rightarrow\; \boxed{H = K\left(R + \frac{\mathbf{t}\,\mathbf{n}^\top}{d}\right)K^{-1}}$$

Cas 2 – rotation pure ($\mathbf{t} = \mathbf{0}$) : $\boxed{H = KRK^{-1}}$, quelle que soit la scène (pas de parallaxe : panoramas, stabilisation).

Exemple robotique

La table, le convoyeur et beaucoup de pièces (tôles, cartons, circuits imprimés) sont plans : deux images d'une pièce plane prises par la caméra embarquée sont liées par une homographie.

Estimation de H : DLT normalisée

Pour chaque correspondance $\mathbf{x}_j = (x, y, 1)^\top \leftrightarrow \mathbf{x}'_j = (x', y', 1)^\top$: $\mathbf{x}'_j \times H\mathbf{x}_j = \mathbf{0}$, soit

$$\begin{bmatrix} \mathbf{0}^\top & -\mathbf{x}_j^\top & y'\mathbf{x}_j^\top \\ \mathbf{x}_j^\top & \mathbf{0}^\top & -x'\mathbf{x}_j^\top \end{bmatrix} \mathbf{h} = \mathbf{0}, \qquad \mathbf{h} = (h_{11}, h_{12}, \ldots, h_{33})^\top \in \mathbb{R}^9.$$

- 8 ddl, 2 équations par point ⇒ 4 correspondances (3 jamais alignées).
- $n > 4$: SVD de $A \in \mathbb{R}^{2n \times 9}$.

DLT normalisée (Hartley)

1. $\tilde{\mathbf{x}}_j = T\mathbf{x}_j$, $\tilde{\mathbf{x}}'_j = T'\mathbf{x}'_j$;
2. DLT ⇒ $\tilde{H}$;
3. $H = T'^{-1}\tilde{H}\,T$.

Normalisation : similitude T telle que le centroïde soit à l'origine et la distance moyenne à l'origine vaille $\sqrt{2}$:

$$T = \begin{bmatrix} s & 0 & -s\bar{x} \\ 0 & s & -s\bar{y} \\ 0 & 0 & 1 \end{bmatrix}, \quad s = \frac{\sqrt{2}}{\frac{1}{n}\sum_j \|\mathbf{x}_j - \bar{\mathbf{x}}\|}.$$

Sans elle, les colonnes de A vont de 1 à 10^5 (pixels) : $A^\top A$ est mal conditionnée.

À retenir

DLT normalisée + RANSAC (4 points) → initialisation ; Levenberg–Marquardt sur l'erreur de transfert symétrique $\sum_j d(\mathbf{x}'_j, H\mathbf{x}_j)^2 + d(\mathbf{x}_j, H^{-1}\mathbf{x}'_j)^2$ → raffinement.

Application : mesurer sur le plan de travail (vue redressée)

La table (ou le convoyeur) est un plan $Z_t = 0$:

$$\mathbf{x} \sim K[\mathbf{r}_1\ \mathbf{r}_2\ \mathbf{t}] \begin{bmatrix} X_t \\ Y_t \\ 1 \end{bmatrix} = H_t \begin{bmatrix} X_t \\ Y_t \\ 1 \end{bmatrix}.$$

- H_t^{-1} transforme l'image en une carte métrique de la table : position, orientation et dimensions des pièces en mm.
- H_t se calcule à partir de K et de la pose de la caméra, ou s'estime avec 4 points connus (marqueurs collés sur la table).
- Avec ${}^b\mathbf{T}_t$ (pose de la table dans la base), on obtient directement les points de saisie dans la base.
- Attention : seuls les points *du plan* sont mesurés correctement ; le dessus d'une pièce épaisse subit une erreur de parallaxe.

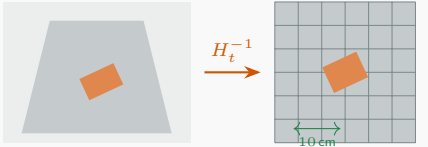


image (caméra inclinée) — vue de dessus (mm)

Calcul de l'homographie table–image (caméra fixe)

Caméra fixe à la hauteur h au-dessus de la table, axe optique vertical (z_c vers le bas, $x_c = x_t$, $y_c = -y_t$) ; point de la table $(X_t, Y_t, 0)$, origine à la verticale du centre optique :

$$\mathbf{X}_c = \begin{bmatrix} X_t \\ -Y_t \\ h \end{bmatrix} = \underbrace{\begin{bmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & h \end{bmatrix}}_{[\mathbf{r}_1\ \mathbf{r}_2\ \mathbf{t}]} \begin{bmatrix} X_t \\ Y_t \\ 1 \end{bmatrix} \quad \Longrightarrow \quad H_t = K[\mathbf{r}_1\ \mathbf{r}_2\ \mathbf{t}] = \begin{bmatrix} f & 0 & c_x h \\ 0 & -f & c_y h \\ 0 & 0 & h \end{bmatrix}.$$

$$u = c_x + f\frac{X_t}{h}, \qquad v = c_y - f\frac{Y_t}{h} \qquad \text{(simple changement d'échelle : 1 px} = h/f \text{ sur la table).}$$

Exemple robotique

$f = 800$, $(c_x, c_y) = (320, 240)$, $h = 1\,\text{m}$: le coin de pièce $(X_t, Y_t) = (0.10, 0.05)\,\text{m}$ se projette en $(u, v) = (400, 200)$; un pixel correspond à $1{,}25\,\text{mm}$. Pour une caméra inclinée, H_t n'est plus diagonale (perspective) mais reste une homographie.

Attention

Dessus d'une pièce de $3\,\text{cm}$ d'épaisseur : il est à $h - 0.03$ de la caméra, l'échelle est fausse de 3 % (3 mm d'erreur à $10\,\text{cm}$ du centre). Utiliser le plan du dessus de la pièce, ou la stéréo.

Décomposition de l'homographie : retrouver le mouvement

Homographie *euclidienne* (calibrée) : $H_e = K^{-1}HK \sim R + \dfrac{\mathbf{t}}{d}\mathbf{n}^\top$.

1. **Échelle** : la deuxième valeur singulière de $R + \mathbf{t}\mathbf{n}^\top/d$ vaut 1, donc $H_e \leftarrow H_e/\sigma_2(H_e)$ (signe choisi pour que $\mathbf{x}'^\top H_e \mathbf{x} > 0$).
2. **Décomposition** (Faugeras 1988, Malis–Vargas 2007) : en général 4 solutions $(R, \mathbf{t}/d, \mathbf{n})$, dont 2 avec le plan devant la caméra.
3. **Désambiguïsation** : normale connue (sol : $\mathbf{n} \approx$ vertical), deuxième plan, ou image suivante.

À retenir

On obtient R exactement, mais seulement $\mathbf{t}/d$: la translation est connue au facteur d près (distance au plan).

Exemple robotique

Caméra embarquée au-dessus d'une pièce plane : d = distance caméra–plan (télémètre, ou table connue via la cinématique) $\Rightarrow$ $\mathbf{t}$ métrique. C'est la base de l'asservissement 2D 1/2 (ch. 2).

Rapport de profondeurs (utile au ch. 2) : pour $\mathbf{x} \leftrightarrow \mathbf{x}'$, $\dfrac{Z'}{Z}\,\mathbf{x}'_n = H_e\,\mathbf{x}_n$, d'où $\rho = Z'/Z = (H_e\mathbf{x}_n)_3$.

Mini-quiz – Homographie

1. Nombre minimal de correspondances pour estimer H ? Dans quels deux cas deux vues sont-elles exactement liées par une homographie ?
2. Après normalisation on obtient $\tilde{H}$. Écrire H.
3. Le robot fait tourner la caméra embarquée autour de son axe optique (poignet), sans la déplacer. Peut-on trianguler la scène ? Quelle est l'homographie entre deux images ?
4. Caméra fixe verticale, $f = 800$, $h = 1\,\text{m}$: quelle taille sur la table correspond à un pixel ? Et à $0{,}8\,\text{m}$?

Mini-quiz – Homographie

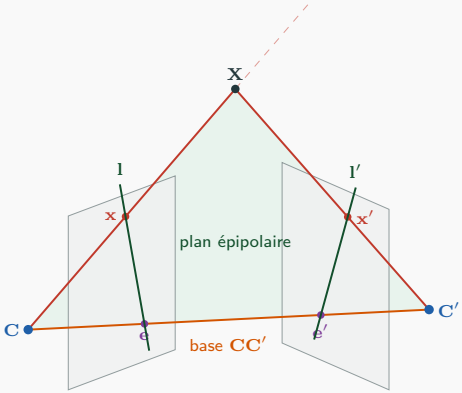
1. Nombre minimal de correspondances pour estimer H ? Dans quels deux cas deux vues sont-elles exactement liées par une homographie ?
2. Après normalisation on obtient $\tilde{H}$. Écrire H.
3. Le robot fait tourner la caméra embarquée autour de son axe optique (poignet), sans la déplacer. Peut-on trianguler la scène ? Quelle est l'homographie entre deux images ?
4. Caméra fixe verticale, $f = 800$, $h = 1\,\text{m}$: quelle taille sur la table correspond à un pixel ? Et à $0{,}8\,\text{m}$?

Réponses

1. 4 ; scène plane (caméras quelconques) ; rotation pure (scène quelconque).
2. $H = T'^{-1}\tilde{H}T$.
3. Non (pas de base, $\mathbf{t} = \mathbf{0}$) ; rotation autour de l'axe optique : $H = KR_z(\psi)K^{-1}$, indépendante de la scène.
4. $h/f = 1{,}25\,\text{mm}$; à $0{,}8\,\text{m}$: $1\,\text{mm}$.

Géométrie épipolaire

Géométrie épipolaire



- **Base** : segment $\mathbf{CC'}$ (déplacement de la caméra entre deux poses du robot, ou paire stéréo).
- **Épipoles** $\mathbf{e}, \mathbf{e'}$: image du centre de l'autre caméra.
- **Plan épipolaire** : $(\mathbf{C}, \mathbf{C'}, \mathbf{X})$.
- **Droites épipolaires** $\mathbf{l}, \mathbf{l'}$: intersections de ce plan avec les plans image.

À retenir

Le correspondant de $\mathbf{x}$ est sur la droite $\mathbf{l'}$: recherche **1D** au lieu de 2D.

Contrainte de coplanarité : matrices essentielle et fondamentale

Repère de la caméra 1 : $P = K[I \,|\, \mathbf{0}]$, $P' = K[R \,|\, \mathbf{t}]$; coordonnées normalisées $\hat{\mathbf{x}} = K^{-1}\mathbf{x}$, $\hat{\mathbf{x}}' = K^{-1}\mathbf{x}'$.

1. Un point 3D : $\mathbf{X}' = R\mathbf{X} + \mathbf{t}$, avec $\mathbf{X} = Z\hat{\mathbf{x}}$, $\mathbf{X}' = Z'\hat{\mathbf{x}}'$.
2. $\mathbf{X}'$, $\mathbf{t}$ et $R\mathbf{X}$ sont coplanaires (plan épipolaire) : $\mathbf{X}'^{\top}\big(\mathbf{t} \times R\mathbf{X}\big) = 0 \iff Z'Z\,\hat{\mathbf{x}}'^{\top}[\mathbf{t}]_{\times}R\,\hat{\mathbf{x}} = 0.$

$$\boxed{\hat{\mathbf{x}}'^{\top} E\,\hat{\mathbf{x}} = 0, \qquad E = [\mathbf{t}]_{\times}R} \quad \text{puis} \quad \boxed{\mathbf{x}'^{\top} F\mathbf{x} = 0, \qquad F = K^{-\top}EK^{-1}}$$

Matrice fondamentale F (pixels)

- droites épipolaires $\mathbf{l}' = F\mathbf{x}$, $\mathbf{l} = F^{\top}\mathbf{x}'$;
- épipoles $F\mathbf{e} = \mathbf{0}$, $F^{\top}\mathbf{e}' = \mathbf{0}$;
- $\operatorname{rang} F = 2$, 7 ddl.

Matrice essentielle E (calibrée)

- 5 ddl : 3 (rotation) + 3 (translation) − 1 (échelle) ;
- valeurs singulières $(\sigma, \sigma, 0)$ (Huang–Faugeras).

Attention

F et E n'existent que si $\mathbf{t} \neq \mathbf{0}$: en rotation pure, $[\mathbf{t}]_{\times} = 0$ (utiliser H).

Reconstruction 3D par la caméra embarquée

Pipeline (deux poses du robot $i \to j$)

1. Détection / appariement de points (ORB, SIFT) entre les deux images.
2. Mouvement de la caméra connu par le robot : ${}^{c_i}\mathbf{T}_{c_j} = X^{-1}\,{}^{b}\mathbf{T}_{e_i}^{-1}\,{}^{b}\mathbf{T}_{e_j}\,X$ ($X = {}^{e}\mathbf{T}_{c}$).
3. Donc $E = [\mathbf{t}]_\times R$ connue : appariement guidé le long des droites épipolaires.
4. Triangulation : points 3D métriques (l'échelle vient du robot).
5. Recalage du modèle CAO sur les points $\Rightarrow$ pose de la pièce.

Si la cinématique est inconnue ou douteuse

- estimer E par RANSAC (5 ou 8 points) et le décomposer : $(R, \hat{\mathbf{t}})$ avec $\|\hat{\mathbf{t}}\| = 1$;
- comparer à la cinématique : contrôle de la calibration main–œil ;
- l'échelle se retrouve avec une distance connue de la scène.

Attention

Une erreur de calibration main–œil ou du modèle du robot se propage directement dans la reconstruction.

À retenir

Configurations dégénérées : rotation pure de la caméra ($\mathbf{t} \approx \mathbf{0}$: E indéfinie, utiliser H) ; scène plane (8 points dégénéré, 5 points correct) ; base trop courte devant la distance : triangulation imprécise.

Bibliographie

- R. Hartley, A. Zisserman. *Multiple View Geometry in Computer Vision*, 2e éd., Cambridge University Press, 2004.
- Y. Ma, S. Soatto, J. Košecká, S. Sastry. *An Invitation to 3-D Vision*, Springer, 2004.
- R. Szeliski. *Computer Vision : Algorithms and Applications*, 2e éd., Springer, 2022.
- P. Corke. *Robotics, Vision and Control*, 3e éd., Springer, 2023.
- Z. Zhang. A flexible new technique for camera calibration. *IEEE TPAMI*, 22(11), 2000.
- D. Lowe. Distinctive image features from scale-invariant keypoints. *IJCV*, 60(2), 2004.
- D. Nistér. An efficient solution to the five-point relative pose problem. *IEEE TPAMI*, 26(6), 2004.
- R. Horaud, F. Dornaika. Hand-eye calibration. *Int. J. Robotics Research*, 14(3), 1995.
- E. Malis, M. Vargas. Deeper understanding of the homography decomposition for vision-based control. Rapport INRIA RR-6303, 2007.
- R. Tsai, R. Lenz. A new technique for fully autonomous and efficient 3D robotics hand/eye calibration. *IEEE Trans. Robotics and Automation*, 5(3), 1989.
- F. Park, B. Martin. Robot sensor calibration : solving $AX = XB$ on the Euclidean group. *IEEE Trans. Robotics and Automation*, 10(5), 1994.
- V. Lepetit, F. Moreno-Noguer, P. Fua. EPnP : an accurate $O(n)$ solution to the PnP problem. *IJCV*, 81(2), 2009.