Vision Artificielle 3D

D'après Pascal Monasse, Loic Landrieu, Aslan Artykov

7 octobre 2025

Table des matières

Résumé

`mailto:pascal.monasse@enpc.fr`, `mailto:loic.landrieu@enpc.fr`, `mailto:arslan.artykov@enpc.fr`

Introduction

Le but principal de la classe est de démontrer et d'expliquer l'équation suivante :

$$\arg \min_{\{R_i\},\{T_i\},\{X_j\}} \sum_{i,j} \varepsilon_{i,j} d\left(x_{i,j}, \Pi_i(R_i X_j + T_i)\right)^2$$

où, pour le j-ème point 3D, observé dans la i-ème image :

- $\varepsilon_{i,j} \in \{0,1\}$ est la visibilité ;
- $x_{i,j}$ est le point 2D dans l'image ;
- $X_{i,j}$ est le point 3D ;
- Π_i est la fonction de projection de l'image i ;
- R_i est la rotation de caméra de l'image i ;
- T_i est la translation de caméra de l'image i.

1 Géométrie Projective

1.1 Géométrie Projective

On s'adaptera au modèle de caméra dit *sténopé*. On suppose ainsi que la caméra est un objectif idéal modélisé par un point. Ceci ne permet pas de prendre en compte le flou ni la distortion géométrique de la lentille.

Dans ce modèle, les rayons qui viennent du point m sur l'image sont les mêmes que ceux qui viennent du point réel M : $\vec{Cm} = \lambda \vec{CM}$. Dans le repère de coordonnées de la caméra CXYZ :

$$\begin{pmatrix} x \\ y \\ f \end{pmatrix} = \lambda \begin{pmatrix} X \\ Y \\ Z \end{pmatrix}$$

où f est la distance focale de la caméra. On a $\lambda = f/Z$ et donc :

$$\begin{pmatrix} x \\ y \end{pmatrix} = f \begin{pmatrix} X/Z \\ Y/Z \end{pmatrix}$$

En coordonnées de pixels :

$$\begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} \alpha x + c_x \\ \alpha y + c_y \end{pmatrix}$$

αf est la distance focale en pixel et c_x, c_y est la position du point principal en pixel.

Définition 1.1 L'espace projectif $\mathbb{P}^2$ est défini comme le quotient de $\mathbb{R}^3 \setminus 0$ par la relation de colinéarité.

C'est donc l'ensemble des rayons passant par l'origine.

Il y a deux types de points dans $\mathbb{P}^2$:

- Ceux de la forme $\begin{pmatrix} x & y & 1 \end{pmatrix}$
- Ceux de la forme $\begin{pmatrix} x & y & 0 \end{pmatrix}$ (qui sont les points à l'infini dans la direction de pente y/x).

Le plan de $\mathbb{R}^3$ contenant 0 d'équation $aX + bY + cZ$ correspond à une ligne de $\mathbb{P}^2$ défini en coordonnées homogènes par (a, b, c) d'équation :

$$\begin{pmatrix} a & b & c \end{pmatrix} \cdot \begin{pmatrix} X & Y & Z \end{pmatrix} = 0$$

On peut alors aussi trouver des équations pour :

- la droite entre deux points x_1 et x_2 : $l = x_1 \times x_2$;
- l'intersection de deux droites l_1, l_2 : $x = l_1 \times l_2$;
- la droite à l'infini : $\ell_\infty = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$;
- l'intersection de deux droites "parallèles" :

$$\begin{pmatrix} a \\ b \\ c_1 \end{pmatrix} \times \begin{pmatrix} a \\ b \\ c_2 \end{pmatrix} = (c_2 - c_1) \begin{pmatrix} b \\ -a \\ 0 \end{pmatrix} \in \ell_\infty$$

L'équation de projection devient alors :

$$\begin{pmatrix} u \\ v \end{pmatrix} = \frac{1}{Z} \begin{pmatrix} fX + c_x Z \\ fY + c_y Z \end{pmatrix}$$

Et donc :

$$Z\begin{pmatrix}u\\v\\1\end{pmatrix} = x = \begin{pmatrix}f & 0 & c_x\\0 & f & c_y\\0 & 0 & 1\end{pmatrix}\begin{pmatrix}X\\Y\\Z\end{pmatrix}$$

Le point 3D étant exprimé dans un autre système orthonormal de coordonnées :

$$x = \begin{pmatrix}f & 0 & c_x\\0 & f & c_y\\0 & 0 & 1\end{pmatrix}\begin{pmatrix}R & T\end{pmatrix}\begin{pmatrix}X\\Y\\Z\\1\end{pmatrix}$$

On définit la matrice de calibration interne :

$$K = \begin{pmatrix}f & 0 & c_x\\0 & f & c_y\\0 & 0 & 1\end{pmatrix}$$

et la matrice de projection $P = K\,(R\;T)$. Ici, la matrice de calibration interne utilise la distance focale f exprimée en pixels.

Dans le cas où les pixels sont des trapézoïdes, on peut généraliser K par :

$$\begin{pmatrix}f_x & s & c_x\\0 & f_y & c_y\\0 & 0 & 1\end{pmatrix}$$

où $s = -f_x \text{cotan}\theta$ et θ est l'angle du parallélograme.

Théorème 1.1 Si $P \in \mathcal{M}_{3,4}(\mathbb{R})$ dont la sous-matrice 3×3 gauche est inversible, elle se factorise de manière unique en $P = K(R\;T)$ où K est inversible, R est une matrice de rotation et T une matrice 3×4 de translation.

Démonstration. On décompose la sous-matrice de gauche de P via le procédé de Gram-Schmidt. L'unicité décolue de l'unicité de la décomposition QR. ■

Des lignes parallèles dans l'espace se projètent dans un faisceau (ensemble de lignes parallèles ou concourrentes). Si on se donne un vecteur d de direction :

$$\begin{aligned}K(X + \lambda d) &= KX + \lambda K d\\ l_x &= (KX) \times (Kd)\\ \forall X, {}^{\mathsf{t}}\ell_X v &= 0, \text{ for } v = Kd\end{aligned}$$

On dit que v est le point de fuite des droites de direction d. Si $v_1 = Kd_1$ et $v_2 = Kd_2$ sont des points de fuites de droites "horizontales", un autre ensemble de droites horizontales a direction $\alpha d_1 + \beta d_2$ et donc son point de fuite est $\alpha v_1 + \beta v_2$ qui appartient à la droite d'horizon $v_1 \times v_2$.

1.2 Homographies

Si on étudie deux images différentes, la matrice pour passer de la première image à la seconde est une homographie 3×3, i.e. une matrice inversible.

Proposition 1.1 Une homographie préserve l'alignement.

Démonstration. En effet, on a :

$$\begin{vmatrix} Hx_1 & Hx_2 & Hx_3 \end{vmatrix} = |H| \begin{vmatrix} x_1 x_2 x_3 \end{vmatrix}$$

■

Dans le cadre de transformation simples, on a des formes plus simples pour les homographies :

Type	Matrice	Condition	Invariants
Rigide	$\begin{pmatrix} c & -s & t_x \\ s & c & t_y \\ 0 & 0 & 1 \end{pmatrix}$	$c^2 + s^2 = 1$	Distances
Similaritude	$\begin{pmatrix} c & -s & t_x \\ s & c & t_y \\ 0 & 0 & 1 \end{pmatrix}$	$c^2 + s^2 \neq 0$	|Angles|, ratio des distances.
Homographie	H inversible	$\|H\| \neq 0$	Birapport de 4 points

Théorème 1.2 Soient $e_1, \ldots, e_{d+1}, f_1, \ldots, f_{d+1} \in \mathbb{R}^d$ de sorte que tous d vecteurs e_i et f_i soient linéairement indépendants. Alors, il existe (à facteur près), un unique isomorphisme H et un unique ensemble de scalaires $\lambda_i \neq 0$ tels que pour tout i, $He_i = \lambda_i f_i$.

Démonstration. On écrit $e_{d+1} = \sum_{i \leq d} \mu_i e_i$ et $f_{d+1} = \sum_{i \leq d} \nu_i f_i$, o vérifie que : $\lambda_i = \frac{\nu_i}{\mu_i} \lambda_{d+1}$. Réciproquement, en fixant $\lambda_{d+1} = 1$, on trouve bien l'existence d'un unique isomorphisme H. ■

En particulier, ce théorème montre que pour $n+2$ paires dans $\mathbb{P}^n$, on a une unique homographie les transportant. Ceci va nous permettre de regrouper des images en corrigeant des homographies. On utilise l'estimation $\lambda x' = Hx \Rightarrow x' \times (Hx) = 0$ ce qui nous donne deux équations linéairement indépendantes par paire de points correspondants. 4 correspondances suffisent pour estimer H, mais si l'on en a plus on minimise une erreur :

- L'erreur algébrique :

$$\varepsilon = \sum_i \|x_i' \times (Hx_i)\|^2$$

- L'erreur de transfert (ou sa symétrisée) :

$$d^2 = d(x', Hx)^2, d^2 + d'^2 = d(x, H^{-1}x')^2 + d(x', Hx)^2$$

En considérant x, x' comme des observateurs bruités de $\hat{x}$ et $\hat{x}' = H\hat{x}$, on considère plutôt $\varepsilon(H, \hat{x}) = d(x, \hat{x})^2 + d(x', H\hat{x}')$. Puisque cette méthode a beaucoup de paramètres, on considère plutôt l'erreur de Sampson :

$$\varepsilon(H, \hat{x}) = \varepsilon(H, x) + J(\hat{x} - x), J = \frac{\partial \varepsilon}{\partial x}(H, x)$$

On adapte ici en résolvant pour $\hat{x}$ minimisant $\|x - \hat{x}\|^2$ sous la contrainte $J(x - \hat{x}) = \varepsilon$. On obtient un bon estimateur de l'erreur algébrique, mais avec un produit scalaire adapté.

On peut ensuite appliquer l'homographie à l'image en poussant les pixels sur l'image transformée et arrondissant au pixel le plus proche, ou en tirant les pixels de l'image originelle par interpolation.

1.3 Algorithmes et Calibration

On a deux méthodes principales pour calibrer les paramètres internes de la caméra K :

1. Par resection, on calcule des points dont on connaît les coordonnées en 3D : on a un système linéaire qu'on peut résoudre avec 6 points, s'il ne sont pas dans un plan.

2. Par planarité : on utilise une surface dont on connaît la position réelle et on calcule l'homographie entre la position et la vision par la caméra. Une fois ceci fait, on peut, avec 3 vues, avoir 6 équations pour 5 paramètres.

Cependant, à petite distance focale, on ne peut pas ignorer la distortion géométrique liée à la courbure de la lentille. On peut la modéliser par un modèle radial polynomial, ou en minimisant par des méthodes itératives.

2 Matrices Fondamentales et Essentielles

2.1 Définition

On va chercher, étant données deux images rectifiées (avec un plan d'image commun et des mouvements de caméra parallèles), des correspondances de point et le calcule de leur décalage apparant, à calculer la profondeur relative de la salle.

Les contraintes épipolaires, décrivent le fait qu'il est suffisant de chercher x sur sa ligne épipolaire : Les vecteurs $\vec{Cx}, \vec{C'x'}$ et T sont complanaires : $|x\ T\ Rx'| = 0$, ce qu'on peut réécrire ${}^{\mathsf{t}}x(T \times Rx') = 0$. On note $[T]_\times x = T \times x$ et on obtient : ${}^{\mathsf{t}}xEx' = 0$ où $E = [T]_\times R$.

Définition 2.1 La matrice E est appelée matrice essentielle des contraintes.

La conversion en coordonnées de pixel se fait en mulitpliant par l'inverse de la matrice de calibration K. On peut réécrire la contrainte épipolaire en :

$$ {}^{\mathsf{t}}xFx' = 0 \text{ où } F = {}^{\mathsf{t}}K^{-1}EK'^{-1} = {}^{\mathsf{t}}K^{-1}[T]_\times RK'^{-1} $$

Définition 2.2 La matrice F est appelée matrice fondamentale des contraintes épipolaires.

Définition 2.3
- $e = KT$ vérifie ${}^{\mathsf{t}}eF = 0$, c'est l'épipole gauche.
- $e' = K'R^{-1}T$ vérifie $Fe' = 0$, c'est l'épipole droit.
- Fx' est la ligne épipolaire associée à x' dans l'image de gauche.
- ${}^{\mathsf{t}}Fx$ est la ligne épipolaire associée à x' dans l'image de gauche.

2.2 Calcul

2.2.1 Matrice Fondamentale

La méthode des 8 points est la plus simple car elle est linéaire. On impose $\|f\| = 1$ dans :

$$ \min_f \|Af\|^2 \text{ tel que } \|f\| = 1 $$

avec $A = ({}^{\mathsf{t}}A_i)$ telle que ${}^{\mathsf{t}}x_iFx'_i = 0$ si et seulement si ${}^{\mathsf{t}}A_if = 0$. Autrement dit, f est une valeur propre de ${}^{\mathsf{t}}AA$. Pour forcer que F soit de rang 2, on calcul sa SVD et on fixe sa troisième valeur propre à 0.

Toutefois, forcer $\det F = 0$ après optimisation n'est pas optimal. La méthode des 7 points force ceci. On a un système linéiare $Af = 0$ de taille 7×9. On se donne f_1, f_2 deux vecteurs linéairement indépendants de $\ker A$. On cherche une solution $f_1 + xf_2$ avec $\det F = 0$. L'intérêt principal n'est pas de calculer f avec moins de points (puisqu'on perd en précision) mais de ne pas sélectionner de fausses correspondances. Pour s'assurer qu'on n'incorpore pas de mauvaises correspondances, puisque les coefficients de F sont de taille si différentes, on normalize les points pour que leurs coordonnées soient d'ordre 1.

2.2.2 Matrice Essentielle

Une matrice est essentielle si et seulement si ses valeurs singulières sont 0 et deux valeurs positives égales :

$$2E^{\mathsf{t}}EE - \mathrm{Tr}(E^{\mathsf{t}}E)E = 0 \wedge \det E = 0$$

L'algorithme des 5 points (Nistér) cherche a résoudre $Ae = 0$ où A est de taille 5×9 avec une solution de la forme $xX + yY + zZ + W$ où X, Y, Z, W est une base du noyau de A. Les contraintes donnent 10 équations polynomiales de degré 3 en x, y, z. On applique le pivot de Gauss pour éliminer les termes de degré ≥ 2 en x, y puis on obtient $B(z)(x\ y\ 1)^\top = 0$, c'est à dire $\det B(z) = 0$ une équation polynomiale de degré 10.