UE Vision, Réalité Augmentée et Applications

Matière : Vision

Partie I

(Cours)

Sylvie Chambon
schambon@toulouse-inp.fr

29 septembre 2023

Plan

Plan

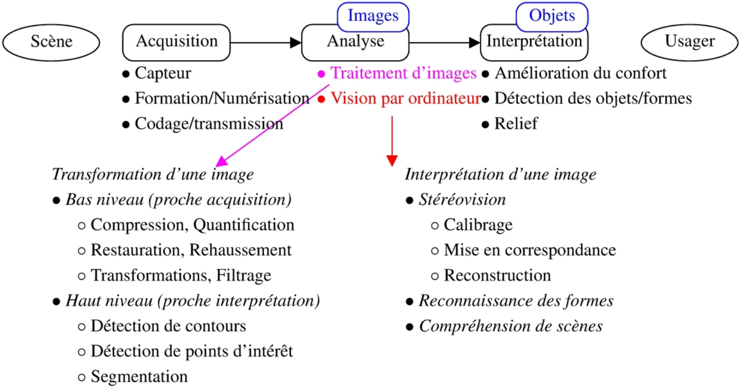
Organisation de ce cours

- Pensez à lire/relire les objectifs
  1. Apprendre les notions géométriques fondamentales pour aborder un système de vision
  2. Découvrir et apprendre les différents éléments liés à la détection de primitives d'intérêt
  3. Découvrir et apprendre différentes techniques de suivi/mise en correspondance
  4. Mettre en œuvre une chaîne de traitement permettant de répondre à un besoin spécifique/une application particulière (APP)
- Des indications sur les temps de lecture
- Aujourd'hui on doit aborder environ 30% du cours
- Pour la classe inversée :

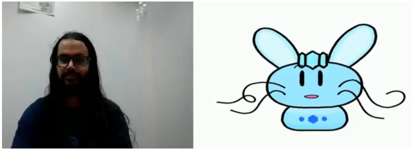
  Les questions à poser concernent la partie à lire (+ les précédentes)

Plan

Chaîne de traitement

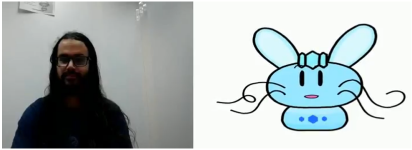


Éléments essentiels pour le suivi



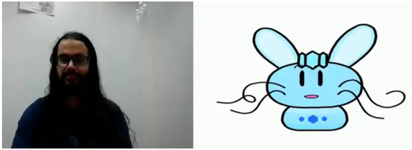
- *Éléments à suivre :* points/pixels, régions, objets (*features, regions/points of interest*)
- *Critères utilisés pour le suivi :* intensité, couleur, forme (**photométriques** ou **géométriques**)

Éléments essentiels pour le suivi



- *Hypothèses sur les* **mouvements 2D apparents** *(caméra, objets, scène)*
  - Déplacements (transformations) rigides (conservation des distances et des orientations) : translation, rotation
  - Déplacements (transformations) non-rigides (non conservation des distances et des orientations) : affine, homographie (effet de zoom, mise à l'échelle), transformations quadratique, polynomiale, à base radiale

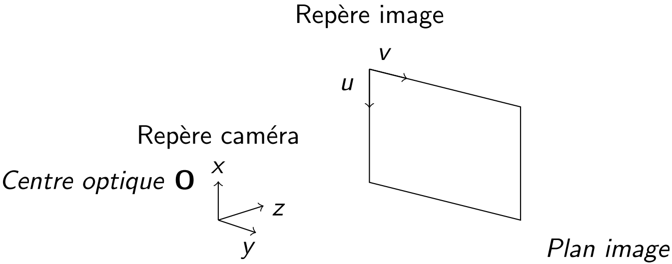
Éléments essentiels pour le suivi



- *Hypothèses sur les déformations :* (objets déformables ou non)
- *Hypothèses sur les variations photométriques :* modèle affine ou non
- *Méthode de suivi ou choix de la méthode de suivi*
    - Dense/éparse
    - Éparse suivie de propagation
      (détection de points d'intérêt, d'objets d'intérêt)

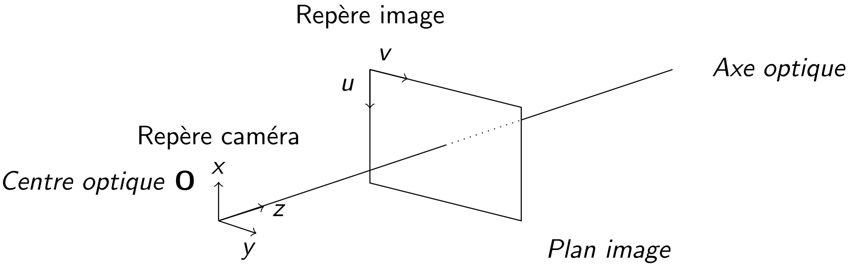
Plan

Modèle sténopé



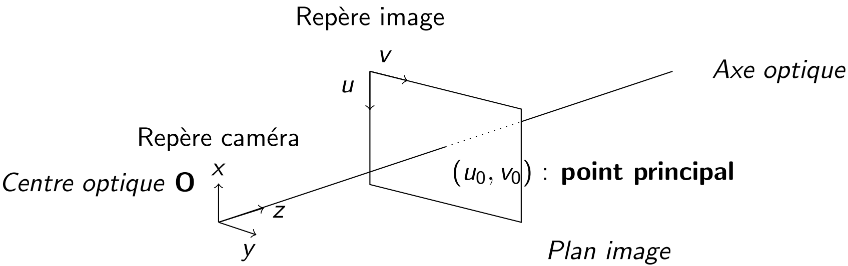
O, centre optique de la caméra, avec son repère associé et le plan image
Plan image = plan de projection

Modèle sténopé



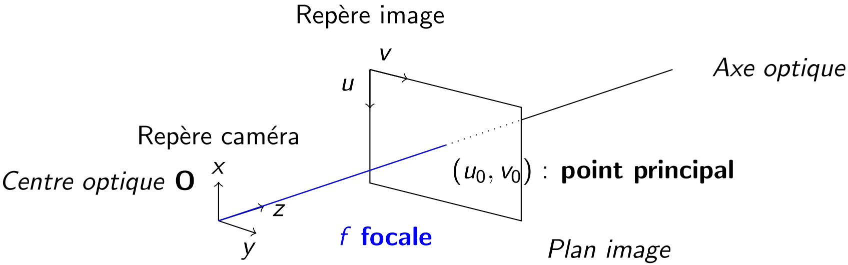
Droite passant par **O** et perpendiculaire au plan image = *axe optique*

Modèle sténopé



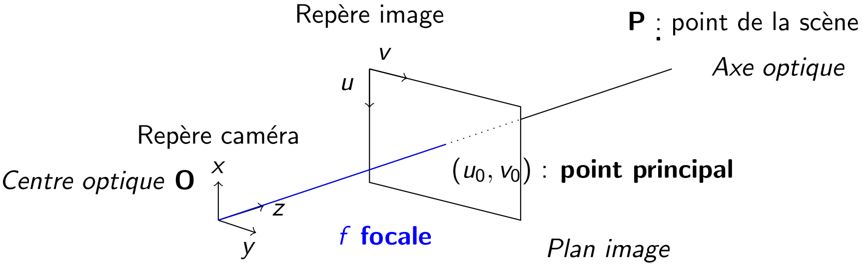
Intersection *axe optique* et plan image, notée (u_0, v_0) = **point principal**

Modèle sténopé



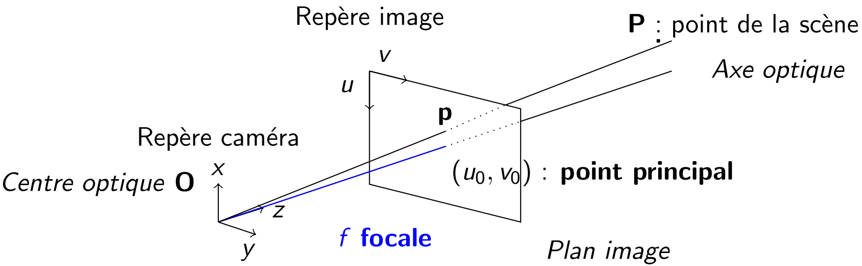
Centre optique **O**, à une distance f du plan image = **focale**

Modèle sténopé



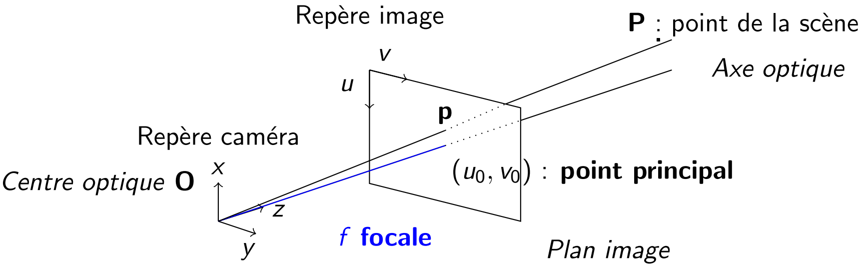
P $(X \;\; Y \;\; Z)$ point de la scène, dans le repère caméra

Modèle sténopé



Projection sur le plan image en $\mathbf{p}$ $(u\ v)$, dans le repère image = Intersection de $(\mathbf{OP})$ avec le plan image

Modèle sténopé



M = matrice de projection perspective des paramètres du modèle

$$\lambda \begin{pmatrix} u & v & 1 \end{pmatrix}^T = \mathbf{M} \begin{pmatrix} X & Y & Z & 1 \end{pmatrix}^T, \; \lambda \in \mathbb{R}^*$$

Calibrage d'une caméra = estimer la matrice **M**

Notions de coordonnées homogènes (1)

- **Notations**
  - $\mathbf{p}$ $(x\ y)$ dans $\mathbb{R}^2$
  - Si $\mathbb{R}^2$, un espace vectoriel alors $(x\ y)$: un vecteur
  - un point dans le plan représenté par $(x\ y)^T$
- **Représentation homogène d'une droite**
  - $ax + by + c = 0$: $(a\ b\ c)^T$
  - $ax + by + c = 0$ et $(ka)x + (kb)y + (kc) = 0$: la même droite, pour tout $k \neq 0$

Ainsi $(a\ b\ c)^T$ et $k(a\ b\ c)^T$: strictement la même droite donc vecteurs strictement équivalents

et classe d'équivalence de vecteurs = vecteur homogène

Conclusion : Ensemble des vecteurs de cette classe d'équivalence dans $\mathbb{R}^3 - (0\ 0\ 0)^T$ = espace projectif

Notions de coordonnées homogènes (2)

- **Représentation homogène d'un point**
  - $\mathbf{p} = (x \; y)^T$ appartient à une droite $\mathbf{l} = (a \; b \; c)^T$ si et seulement si $ax + by + c = 0$

Ainsi $(x \; y \; 1)(a \; b \; c)^T = (x \; y \; 1)\mathbf{l} = 0$

donc $\mathbf{p}$ peut être représenté par un vecteur de taille 3 dont la dernière coordonnée est 1

  - $\mathbf{p} = (x \; y \; 1)^T$ = **coordonnées homogènes** de $\mathbf{p}$

De plus $\forall k \neq 0$, $(kx \; ky \; k)\mathbf{l} = 0$ si et seulement si $(x \; y \; 1)\mathbf{l} = 0$

Ainsi $\left\{(kx \; ky \; k)^T\right\}$ = représentation de $\mathbf{p}$ dans $\mathbb{R}^2$

  - Les points sont représentés par des vecteurs homogènes

Étapes pour estimer **M**

1. Repère scène $\rightarrow$ repère caméra

$=$ Matrice des paramètres extrinsèques

2. Repère caméra $\rightarrow$ repère image

$=$ Projection perspective

$+$ Matrice des paramètres intrinsèques

Repère scène $\rightarrow$ repère caméra

Soit $(x\ y\ z)$ les coordonnées de **P**, dans le repère caméra

$$(x\ y\ z)^T = \mathbf{R}(X\ \ Y\ \ Z)^T + \mathbf{t} \tag{1}$$

Nous notons :

$$\mathbf{R} = \begin{pmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{pmatrix} \quad \text{et} \quad \mathbf{t} = \begin{pmatrix} t_x \\ t_y \\ t_z \end{pmatrix}$$

Matrice des paramètres extrinsèques

$$(x\ y\ z\ 1)^T = \mathbf{A}(X\ \ Y\ \ Z\ 1)^T \tag{2}$$

avec :

$$\mathbf{A} = \begin{pmatrix} r_{11} & r_{12} & r_{13} & t_x \\ r_{21} & r_{22} & r_{23} & t_y \\ r_{31} & r_{32} & r_{33} & t_z \\ 0 & 0 & 0 & 1 \end{pmatrix} \tag{3}$$

- **$\mathbf{A}$ = matrice des paramètres extrinsèques**
- 12 paramètres à estimer
- $\mathbf{R}$ = 9 éléments mais sachant que $\mathbf{R}\ \mathbf{R}^T = 1 \rightarrow 3$ éléments indépendants : 3 angles polaires
- $\mathbf{t}$ = 3 éléments

Donc 6 paramètres indépendants

= 6 correspondances de points

- En pratique : une centaine de points, voire plusieurs centaines !

Repère caméra → repère image

- Passage de (x, y, z) à (u, v)
- Deux transformations
  1. Projection perspective
  2. Transformation avec changement d'unités

  = **matrice des paramètres intrinsèques**

Projection perspective

- $(x_p \ y_p \ z_p)$ les coordonnées du point **p**, du plan image, dans le repère caméra
- Avec f la focale + le théorème de Thalès

$$\begin{cases} x_p & = f\dfrac{x}{z} \\ y_p & = f\dfrac{y}{z} \\ z_p & = f \end{cases} \tag{4}$$

Matrice des paramètres extrinsèques

- Pour obtenir (u, v) **en pixels** (avec (u_0, v_0) point principal) :

$$\begin{cases} u &= -x_p + u_0 \\ v &= y_p + v_0 \end{cases} \tag{5}$$

- D'où mise à l'échelle avec k_u, facteur vertical, et k_v, facteur horizontal ($pixels/mm$) :

$$\begin{cases} u &= -k_u x_p + u_0 \\ v &= k_v y_p + v_0 \end{cases} \tag{6}$$

- En reprenant la projection perspective :

$$\begin{cases} u &= -k_u f \frac{x}{z} + u_0 \\ v &= k_v f \frac{y}{z} + v_0 \end{cases} \tag{7}$$

Matrice des paramètres extrinsèques

- Sous forme matricielle, en coordonnées homogènes :

$$\begin{pmatrix}\lambda u\\ \lambda v\\ \lambda\end{pmatrix} = \begin{pmatrix}-k_u f & 0 & u_0\\ 0 & k_v f & v_0\\ 0 & 0 & 1\end{pmatrix}\begin{pmatrix}x\\ y\\ z\end{pmatrix} \tag{8}$$

- Coordonnées cartésiennes (u, v) obtenues en divisant le résultat obtenu par λ (la troisième coordonnée) :

$$\lambda(u\ v\ 1)^T = \mathbf{K}\ (x\ y\ z)^T \quad \text{avec}$$
$$\mathbf{K} = \begin{pmatrix}-k_u f & 0 & u_0\\ 0 & k_v f & v_0\\ 0 & 0 & 1\end{pmatrix} \tag{9}$$

- **$\mathbf{K}$ = matrice des paramètres intrinsèques**
- 4 paramètres : $\alpha_u = -k_u f, \alpha_v = -k_v f, u_0$ et v_0

Expression de la matrice de projection perspective **M**

$$\begin{cases} u & = -k_u f \dfrac{r_{11}X + r_{12}Y + r_{13}Z + t_x}{r_{31}X + r_{32}Y + r_{33}Z + t_z} + u_0 \\ v & = k_v f \dfrac{r_{21}X + r_{22}Y + r_{23}Z + t_y}{r_{31}X + r_{32}Y + r_{33}Z + t_z} + v_0 \end{cases} \tag{10}$$

Expression de la matrice de projection perspective **M**

- Réduction au même dénominateur commun :

$$\begin{cases} u &= \dfrac{m_{11}X + m_{12}Y + m_{13}Z + m_{14}}{m_{31}X + m_{32}Y + m_{33}Z + m_{34}} \\ v &= \dfrac{m_{21}X + m_{22}Y + m_{23}Z + m_{24}}{m_{31}X + m_{32}Y + m_{33}Z + m_{34}} \end{cases} \quad (11)$$

- On retrouve bien :

$$\lambda \begin{pmatrix} u & v & 1 \end{pmatrix}^T = \mathbf{M} \begin{pmatrix} X & Y & Z & 1 \end{pmatrix}^T, \ \lambda \in \mathbb{R}^*. \quad (12)$$

- Définition à un facteur multiplicatif près
- Si **M** connue = possibilité d'estimer les paramètres intrinsèques **K** et extrinsèques **A**
- Inversement :

$$\mathbf{M} = (\mathbf{K}\ \mathbf{0}_{3\times1})\,\mathbf{A}. \quad (13)$$

Expression des paramètres à partir de $\mathbf{M}$

$$\mathbf{A} = \begin{pmatrix} \mathbf{r}_1 & t_x \\ \mathbf{r}_2 & t_y \\ \mathbf{r}_3 & t_z \\ \mathbf{0} & 1 \end{pmatrix} \text{ avec } \mathbf{r}_i = \left(r_{i1} \;\; r_{i2} \;\; r_{i3}\right). \tag{14}$$

$$\mathbf{M} = \begin{pmatrix} \mathbf{m}_1 & m_{14} \\ \mathbf{m}_2 & m_{24} \\ \mathbf{m}_3 & m_{34} \end{pmatrix} \text{ avec } \mathbf{m}_i = \left(m_{i1} \;\; m_{i2} \;\; m_{i3}\right). \tag{15}$$

Expression des paramètres à partir de **M**

- En utilisant $\mathbf{M} = (\mathbf{K}\ \mathbf{0}_{3\times 1})\,\mathbf{A}$
- les propriétés d'othogonalité de la matrice de rotation
- $\alpha_u < 0$

alors

$$
\begin{aligned}
\mathbf{r}_3 &= \mathbf{m}_3 \\
u_0 &= \mathbf{m}_1 \cdot \mathbf{m}_3 \\
v_0 &= \mathbf{m}_2 \cdot \mathbf{m}_3 \\
\alpha_u &= -\|\mathbf{m}_1 \wedge \mathbf{m}_3\| \\
\alpha_v &= \|\mathbf{m}_2 \wedge \mathbf{m}_3\| \\
\mathbf{r}_1 &= \frac{\mathbf{m}_1 - u_0\mathbf{m}_3}{\alpha_u} \\
\mathbf{r}_2 &= \frac{\mathbf{m}_2 - v_0\mathbf{m}_3}{\alpha_v} \\
t_x &= \frac{m_{14} - u_0 m_{34}}{\alpha_u} \\
t_y &= \frac{m_{14} - v_0 m_{34}}{\alpha_v} \\
t_z &= m_{34}
\end{aligned}
\tag{16}
$$

Plan

Calibrage d'une caméra

- Deux éléments :
  1. Estimer la matrice de projection $\mathbf{M}$ = un problème de resection ou *resection problem*
  2. Estimer, éventuellement, si nécessaire, les paramètres extrinsèques et intrinsèques de la caméra : $\mathbf{A}$ et $\mathbf{K}$
- En général, méthodes de calibrage =
  1. Estimer la matrice de projection
  2. À partir de cette matrice estimer les paramètres intrinsèques et extrinsèques
- Nombreuses solutions : sans ou **avec points connus dans la scène**

Calibrage d'une caméra : le problème

- **Hypothèse** : N correspondances de points connues $\mathbf{P}_i = (X_i \ Y_i \ Z_i)$ de la scène vers $\mathbf{p}_i = (u_i \ v_i \ 1)$ de l'image

$$\lambda_i \mathbf{p}_i = \mathbf{M} \ \mathbf{P}_i. \tag{17}$$

- **Inconnues** : 11 paramètres de $\mathbf{M}$ et les $N \lambda_i$
- Il y a $3N$ équations
- Il faut $3N \geq 11 + N$, soit $N \geq 6$

Calibrage d'une caméra : la solution simple

- Transformation linéaire directe ou *Direct Linear Transformation* (DLT)

$$\mathbf{M} = \begin{pmatrix} \mathbf{m_1}^T \\ \mathbf{m_2}^T \\ \mathbf{m_3}^T \end{pmatrix}$$

- Avec $\mathbf{0}_{1\times 4} = \mathbf{0}$

$$\begin{pmatrix} \mathbf{P}_i & \mathbf{0} & \mathbf{0} & -u_i \\ \mathbf{0} & \mathbf{P}_i & \mathbf{0} & -v_i \\ \mathbf{0} & \mathbf{0} & \mathbf{P}_i & -1 \end{pmatrix} \begin{pmatrix} \mathbf{m_1} \\ \mathbf{m_2} \\ \mathbf{m_3} \\ \lambda_i \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}. \qquad (18)$$

Calibrage d'une caméra : la solution simple

En développant

$$\underbrace{\begin{pmatrix} \mathbf{P}_1 & \mathbf{0} & \mathbf{0} & -u_1 & \dots & \dots & \dots \\ \mathbf{0} & \mathbf{P}_1 & \mathbf{0} & -u_1 & \dots & \dots & \dots \\ \mathbf{0} & \mathbf{0} & \mathbf{P}_1 & -1 & \dots & \dots & \dots \\ \mathbf{P}_2 & \mathbf{0} & \mathbf{0} & 0 & -u_2 & \dots & \dots \\ \mathbf{0} & \mathbf{P}_2 & \mathbf{0} & 0 & -u_2 & \dots & \dots \\ \mathbf{0} & \mathbf{0} & \mathbf{P}_2 & 0 & -1 & \dots & \dots \\ \mathbf{P}_3 & \mathbf{0} & \mathbf{0} & 0 & 0 & -u_3 & \dots \\ \mathbf{0} & \mathbf{P}_3 & \mathbf{0} & 0 & 0 & -u_3 & \dots \\ \mathbf{0} & \mathbf{0} & \mathbf{P}_3 & 0 & 0 & -1 & \dots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}}_{=\mathbf{D}} \underbrace{\begin{pmatrix} \mathbf{m_1} \\ \mathbf{m_2} \\ \mathbf{m_3} \\ \lambda_1 \\ \lambda_2 \\ \lambda_3 \\ \vdots \end{pmatrix}}_{=\mathbf{l}} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ \vdots \end{pmatrix}. \tag{19}$$

Calibrage d'une caméra : la solution simple

- Estimer un vecteur non nul du noyau de la matrice $\mathbf{D}$
- Comme échelle abitraire, ajout de cette contrainte : $||\mathbf{l}||^2 = 1$
- On cherche à résoudre

$$\min_{||\mathbf{l}||^2=0} ||\mathbf{Dl}||^2. \tag{20}$$

- Estimation au sens des moindres carrés
- Deux solutions possibles puisque $||\mathbf{Dl}|| = ||\mathbf{D}(-\mathbf{l})||$ et $||\mathbf{l}|| = ||-\mathbf{l}||$
- une des deux place la caméra derrière la scène $\rightarrow$ profondeurs négatives

Donc Choisir la solution où tous les λ_i sont positifs

- **Solution =** prendre le vecteur propre ayant la plus petite valeur propre de la matrice $\mathbf{D}^T\mathbf{D}$
- On peut simplement utiliser l'algorithme de décomposition en valeurs singulières, *Singular Value Decomposition*

Calibrage d'une caméra : en pratique

- **Nombreuses approches :** avec points connus de la scène ou sans
- Utilisation de photos acquises en présence d'une mire : en TP, en réalité augmenté, un damier (plein d'autres possibilités)
- Détection de points d'intérêt de cette mire : par exemple, les coins du damier bien contrastés
- À partir des correspondances de points, estimation de la matrice **M**
- À partir de **M**, calcul de **A** et **K**

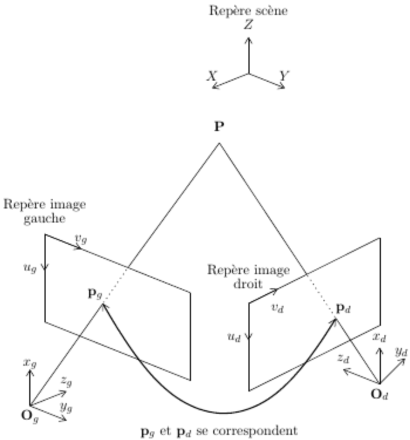
Plan

Modèle stéréoscopique

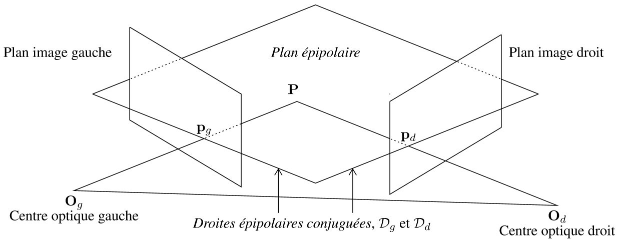
- Deux images acquises de deux points de vue différents
  - Deux caméras

  ou un déplacement
- Position relative de telle sorte qu'une grande partie de la scène soit visible dans les deux points de vue
- Pas de rotation autour du centre optique de la caméra sinon reconstruction impossible
- Deux matrices de projection perspective

Modèle stéréoscopique



Géométrie épipolaire



Contrainte et rectification épipolaire

- Contrainte épipolaire = relation entre les points de l'image 1 et les points de l'image 2 qui ne dépend que des paramètres des caméras et des coordonnées image et qui est indépendante des coordonnées du point de la scène
- Matrice fondamentale, cf. cours de vision par ordinateur avec Jean-Denis Durou
- Rectification épipolaire = transformation géométrique à appliquer aux deux images
- Difficulté = estimation des paramètres de cette transformation