

Stéréovision: Cours

Annexe au travail d'automne

Sol Rosca (INF3b)

Automne 2019

Table des matières

1. Modélisation de caméra	1
1.1. Modélisation géométrique	1
1.2. Modélisation algébrique	3
1.2.1. Prise en compte de la Distance focale	3
1.2.2. Prise en compte des pixels	5
1.2.3. Prise en compte du repère monde	7
1.2.3.1. Rotation	7
1.2.3.2. Translation	9
1.2.3.3. Rotation et translation	9
1.3. Modèle complet	10
1.3.1. Paramètres extrinsèques et intrinsèques	11
2. Distorsion	12
2.1. Distorsion radiale	12
2.2. Distorsion tangentielle	12
3. Vision stéréoscopique	13
3.1. Disparité binoculaire	13
3.2. Imagerie stéréo	14
3.3. Triangulation	15
3.4. Rectification	17
3.5. Géométrie épipolaire	19
3.5.1. Définitions	20
3.5.2. Propriétés	20
3.5.3. Contrainte épipolaire	21
3.6. Matrices essentielles et fondamentales	22
3.6.1. Matrice essentielle	23
3.6.2. Matrice fondamentale	26
3.7. Calibration stéréo (p427)	27
3.8. Reprojection	27
3.8.1. Application de la méthode de Bouguet	27
4. Références	28
4.1. Papier	28
4.2. En Ligne	28
4.2.1. Articles, Cours & Tutos	28
4.2.2. Cours vidéo	28
4.2.3. Interactif	28

1. Modélisation de caméra

Pour permettre à une machine d'effectuer des calculs ou des raisonnements géométriques à partir d'images numériques, un modèle qui décrit comment l'espace tridimensionnel se projette sur une image est nécessaire. Le modèle principalement utilisé en vision par ordinateur est le **modèle sténopé** (pinhole en anglais) qui représente une **projection perspective**. Ce dernier modèle présente un bon compromis entre la simplicité des équations et leur association aux phénomènes physiques modélisés.

1.1. Modélisation géométrique

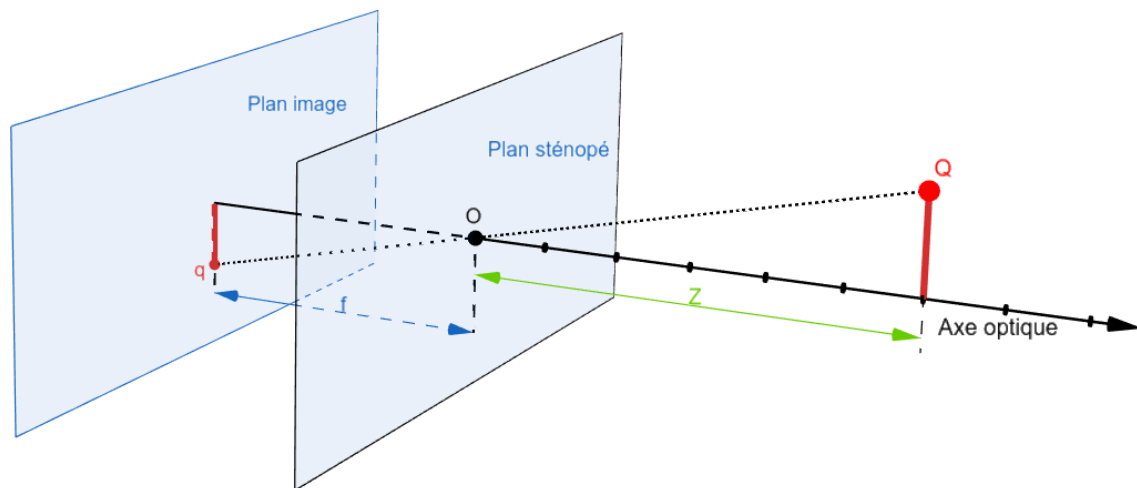


Figure 1 Modèle sténopé

Sur le plan géométrique, le modèle sténopé peut être décrit comme l'ensemble composé d'un **centre de projection** (centre optique) O et d'un **plan image** (rétine). Un point 3D est projeté le long du rayon qui le lie avec le centre de projection et son **point image** est l'intersection de ce rayon avec le plan image.

Dans ce modèle, l'image projetée sur le plan image est toujours nette et sa taille q dépend de la taille de l'objet Q , de la distance focale f et de la distance Z entre la caméra et l'objet. Ainsi, et en utilisant les propriétés des triangles semblables, il en ressort que:

$$-q = f \frac{Q}{Z}$$

Il est possible de "réarranger" le modèle sténopé pour un modèle géométrique *équivalent*¹ qui va permettre de simplifier les calculs.

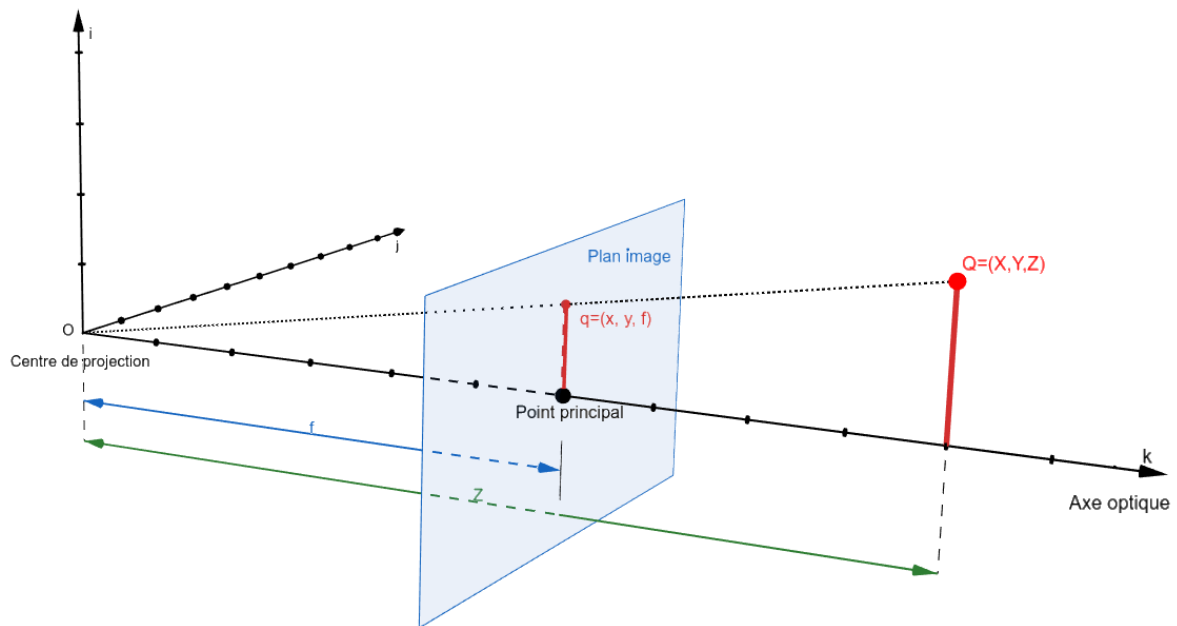


Figure 2 Un point $Q(X,Y,Z)$ projeté dans le plan image en suivant le rayon passant par le centre de projection et dont le point résultant est $q(x,y,f)$. Cette façon de voir les choses est équivalente à la précédente.

Dans ce nouveau modèle, l'intersection du rayon avec le plan sténopé est réinterprétée comme étant le **centre de projection** et le **plan image** est placé devant le plan sténopé. La droite qui passe par le centre de projection et qui est perpendiculaire au plan image est appelé *axe optique*². Le point d'intersection de l'axe optique avec le plan image est le **point principal**.

Tout comme dans la figure du modèle sténopé, l'objet distant est projeté sur le plan image qui se trouve à distance f du centre de projection et les mêmes relations apparaissent:

$$q = f \frac{Q}{Z}$$

En plus de simplifier visuellement la situation, cette façon de modéliser permet de redresser la projection et donc de changer le signe de q .

1. Modèle sténopé non-inverseur. Ce modèle n'a pas d'équivalent physique.

2. Aussi appelé rayon principal en fonction du contexte

1.2. Modélisation algébrique

Pour effectuer des calculs à partir du modèle géométrique, il est nécessaire d'établir une description algébrique analogue. Pour ce faire il est nécessaire d'identifier les **paramètres intrinsèques** d'une caméra. De plus, c'est à l'aide de coordonnées qu'il est possible de faire des calculs avec des points ou qui dit coordonnées, dit repères.

1.2.1. Prise en compte de la Distance focale

La distance focale f est la distance entre le centre de projection et le plan image (donc la distance entre le centre de projection et le point principal). Les équations seront dérivées à l'aide des quatre repères présentés dans la figure suivante:

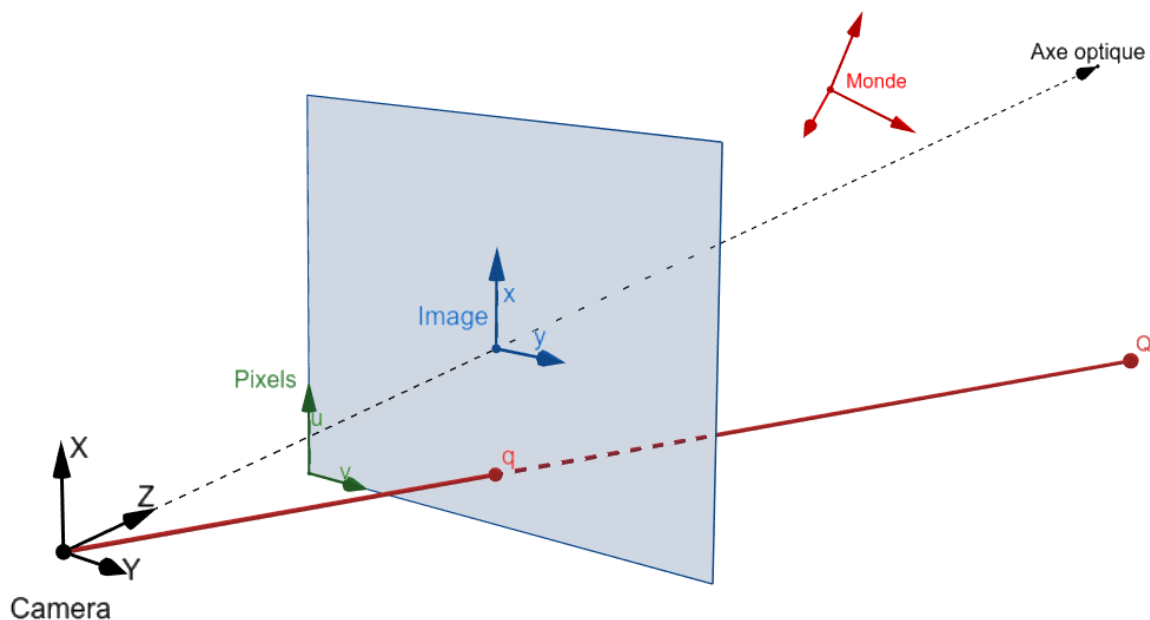


Figure 3 Repères utilisés

1. **Repère caméra:** Un repère 3D attaché à la caméra dont l'origine est le centre de projection et dont l'axe Z est l'axe optique. Les axes X et Y sont parallèles au plan image et perpendiculaires entre eux.
2. **Repère image:** Un repère 2D dont l'origine est le point principal. Ses axes x et y sont parallèles aux axes X et Y du repère caméra. Ce repère peut être vu comme la *projection orthogonale* du repère caméra.
3. **Repère pixels:** Un repère 2D donc les coordonnées sont notées u et v .
4. **Repère monde:** Un repère 3D choisi de manière arbitraire.

Coordonnées homogènes

Les *coordonnées homogènes*¹ reposent sur une notation dans laquelle un vecteur en n dimensions est représenté par un vecteur en $n + 1$ dimensions.

Soit un plan Euclidien \mathbb{R}^2 composé de points,

i

- En notation classique, un point est un vecteur: $\vec{v} = \begin{pmatrix} x \\ y \end{pmatrix}$
- En notation homogène, un point est un vecteur: $\vec{v} = \begin{pmatrix} x \\ y \\ 1 \end{pmatrix}$

Démonstration:

$$a \cdot \begin{pmatrix} x \\ y \\ 1 \end{pmatrix} = \begin{pmatrix} ax \\ ay \\ a \end{pmatrix} = \begin{pmatrix} ax/a \\ ay/a \\ a/1 \end{pmatrix} = \begin{pmatrix} x \\ y \\ 1 \end{pmatrix}$$

Soit un point Q du repère monde dont les coordonnées (en notation homogènes) sont données dans le référentiel de la caméra:

$$Q_c = \begin{pmatrix} X_c \\ Y_c \\ Z_c \\ 1 \end{pmatrix}$$

Les coordonnées x et y du point image q , peuvent être calculées à l'aide des relations entre triangles semblables:

$$x = f \frac{X_c}{Z_c} \quad (1)$$

$$y = f \frac{Y_c}{Z_c} \quad (2)$$

1. Aussi appelées coordonnées projectives

La notation homogène permet de facilement représenter ces équations par une multiplication entre une matrice et un vecteur:

$$q = \begin{pmatrix} x \\ y \\ 1 \end{pmatrix} \approx \begin{bmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{pmatrix} X_c \\ Y_c \\ Z_c \\ 1 \end{pmatrix} \quad (3)$$

Attention!



Il est important de noter l'approximation due au fait que l'égalité vectorielle ne soit définie qu'à un facteur scalaire près. C'est le prix à payer pour l'utilisation des coordonnées homogènes.

Vérification de l'équation (3):

$$\begin{bmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{pmatrix} X_c \\ Y_c \\ Z_c \\ 1 \end{pmatrix} = \begin{pmatrix} fX_c \\ fY_c \\ Z_c \end{pmatrix}$$

Pour se débarrasser de la notation homogène, il suffit de diviser les premières composantes par la dernière:

$$\begin{pmatrix} fX_c \\ fY_c \\ Z_c \end{pmatrix} \approx \begin{pmatrix} fX_c/Z_c \\ fY_c/Z_c \\ Z_c/Z_c \end{pmatrix} = \begin{pmatrix} fX_c/Z_c \\ fY_c/Z_c \\ 1 \end{pmatrix}$$

Ce résultat est bien équivalent aux équations (1) et (2).

1.2.2. Prise en compte des pixels

Comme le montre la figure 3, il est nécessaire de faire un changement de repère supplémentaire pour passer du repère image à celui des pixels (dont les coordonnées sont notées u et v). Pour effectuer ce changement, il est nécessaire d'opérer une translation suivie d'une conversion d'unités.

En effet, le repère image est un repère exprimé en unités métriques contrairement au repère pixel qui lui est exprimé dans une unité représentant une densité (px/mm). De plus dans les caméras bon marché, les pixels ne sont pas forcément carrés mais rectangulaires ce qui implique que chaque axe devra subir un changement d'unité individuel supplémentaire.

soient:

- x_0 et y_0 les coordonnées du coin inférieur gauche de l'image vis-à-vis du repère image.
- k_u la densité de pixels en direction de l'axe u .
- k_v la densité de pixels en direction de l'axe v .

Le tout en notation homogène, considérons le point q avec les coordonnées x et y dans le repère image. Ses coordonnées dans le repère pixel sont obtenues en effectuant une translation suivie d'un changements d'unités:

$$\begin{pmatrix} u \\ v \\ 1 \end{pmatrix} = \begin{bmatrix} k_u & 0 & 0 \\ 0 & k_v & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & x_0 \\ 0 & 1 & y_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{pmatrix} x \\ y \\ 1 \end{pmatrix} \quad (4)$$

Finalement, la combinaison des équations (3) et (4) correspond à la modélisation de la projection du repère caméra vers le repère pixels. Cette matrice correspond à la **matrice de projection**.

$$\begin{pmatrix} u \\ v \\ 1 \end{pmatrix} \approx \begin{bmatrix} k_u f & 0 & k_u x_0 & 0 \\ 0 & k_v f & k_v y_0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{pmatrix} X_c \\ Y_c \\ Z_c \\ 1 \end{pmatrix} \quad (5)$$



Pour donner des ordres de grandeur, la distance focale f varie typiquement de quelques mm jusqu'à une ou plusieurs centaines de mm (au delà, elle est considérée comme étant infinie). La surface photosensible d'une caméra est typiquement un rectangle avec des bords de quelques mm de longueur. Pour une image de 512×512 pixels, correspondant à une surface photosensible de $5,12 \times 5,12 mm^2$, Ce qui donne des valeurs $k_u = k_v = 100 \frac{1}{mm}$.

Pour des caméras avec une optique bien ajustée, le point principal se trouve en principe proche du centre de la surface photosensible, donc:

$$x_0 \approx y_0 \approx 2.56 mm$$

1.2.3. Prise en compte du repère monde

Jusqu'ici les points 3D étaient représentés dans un repère attaché à la caméra. En réalité, ces objets se trouvent dans leur propre repère, le repère *monde* (voir **figure 3**). Pour chaque image d'un objet dans le repère *monde*, il est possible de décrire cet objet en terme de **rotation** et **translation** vis-à-vis du repère *caméra*:

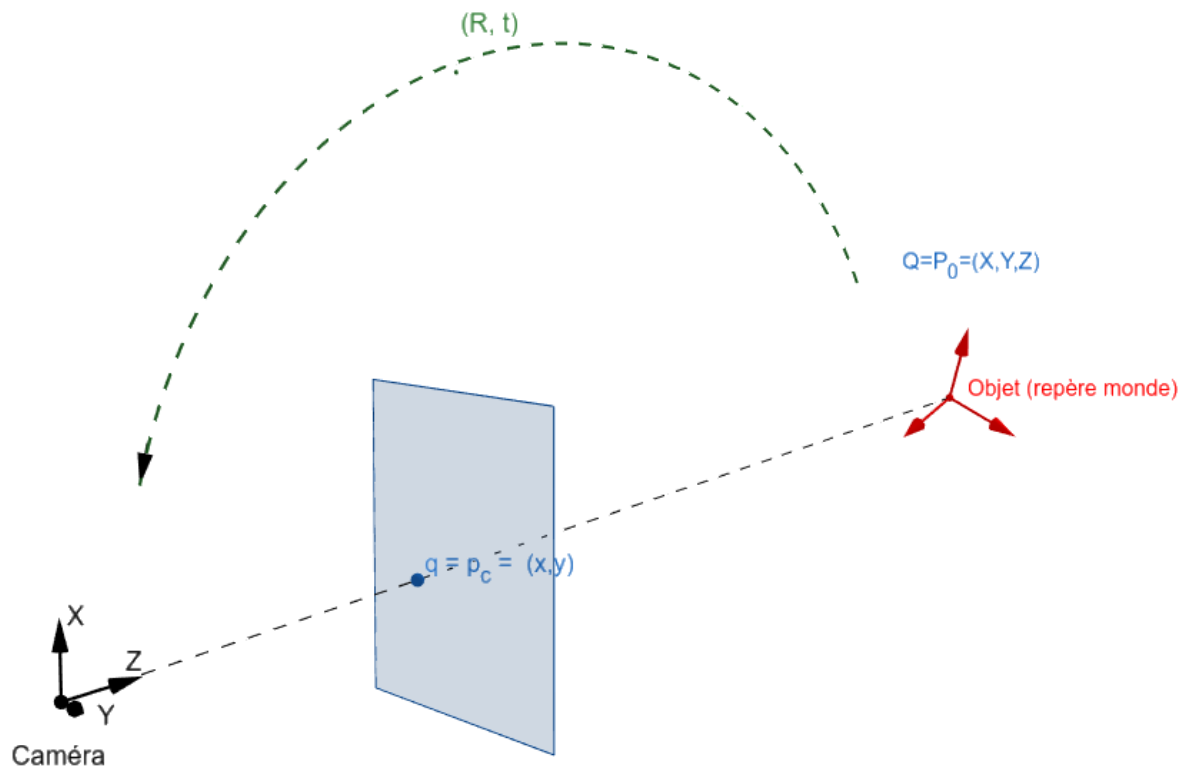


Figure 4 Conversion des coordonnées monde vers les coordonnées caméra en appliquant une matrice de rotation R et un vecteur de translation t sur le point P .

1.2.3.1. Rotation

En général, une rotation (dans n'importe quel nombre de dimensions) peut être décrite par le produit d'un vecteur par une matrice carré de taille appropriée.



Fondamentalement, une rotation est équivalente à la description de la position d'un point dans un système de coordonnées différent.

Une rotation dans l'espace, peut être décomposée en une rotation dans le plan autour de chaque axe.

Le résultat d'une rotation autour des trois axes x , y et z des angles respectifs α , β et γ de façon *séquentielle*¹ est exprimé par la matrice de rotation R donnée par le produit des trois matrices $R_x(\alpha)$, $R_y(\beta)$ et $R_z(\gamma)$ où:

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

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

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

$$R = \begin{bmatrix} \cos(\gamma) & \sin(\gamma) & 0 \\ -\sin(\gamma) & \cos(\gamma) & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos(\alpha) & \sin(\alpha) \\ 0 & -\sin(\alpha) & \cos(\alpha) \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos(\beta) & \sin(\beta) \\ 0 & -\sin(\beta) & \cos(\beta) \end{bmatrix}$$

$$R = R_z(\gamma)R_y(\beta)R_x(\alpha)$$

Matrices de rotation



Les matrices qui servent à représenter des rotations, sont les matrices dites *orthonormales*. Ce sont des matrices carrées où les colonnes (et lignes) sont des vecteurs mutuellement orthogonaux (des vecteurs dont le produit scalaire s'annule) et de norme 1. Ceci implique que le déterminant vaut ± 1 . Pour une matrice de rotation, le déterminant doit valoir nécessairement $+1$ (dans le cas où il vaudrait -1 , la matrice représenterait une réflexion). Tout cela implique que l'inverse d'une matrice de rotation est sa transposée:

$$R^T R = R R^T = I$$

Où I est la matrice identité.

1. La rotation décrite ici s'effectue d'abord autour de l'axe z , ensuite autour de la nouvelle position de l'axe y et finalement autour de la nouvelle position de l'axe x

1.2.3.2. Translation

Le vecteur de translation \vec{t} représente le décalage entre l'origine d'un premier repère et l'origine d'un second repère. Pour passer du repère *monde* (voir **figure 4**) d'un objet au repère *caméra*, donc:

$$\vec{t} = o_{\text{objet}} - o_{\text{camera}}$$

Un point P_0 du repère *monde* possèdera les coordonnées P_c dans le repère *caméra*:

$$P_c = R(P_0 - \vec{t})$$

1.2.3.3. Rotation et translation

Soient X^m, Y^m et Z^m les coordonnées d'un point Q exprimées dans le repère monde (voir **figure 4**) et X_c, Y_c et Z_c les coordonnées du centre de projection, alors, changement de repère du repère monde vers le repère caméra peut être écrit de la façon suivante:

$$\begin{pmatrix} X_c \\ Y_c \\ Z_c \end{pmatrix} = R \left(\begin{pmatrix} X^m \\ Y^m \\ Z^m \end{pmatrix} - \vec{t} \right) = R \begin{pmatrix} X^m \\ Y^m \\ Z^m \end{pmatrix} - R\vec{t}$$

Ce qui représente en coordonnées homogènes:

$$\begin{pmatrix} X_c \\ Y_c \\ Z_c \\ 1 \end{pmatrix} = \begin{bmatrix} R & -R\vec{t} \\ 0^T & 1 \end{bmatrix} \begin{pmatrix} X^m \\ Y^m \\ Z^m \\ 1 \end{pmatrix} \quad (6)$$

1.3. Modèle complet

En combinant les résultats des points précédents, c'est à dire les équations (5) et (6), il en ressort le modèle algébrique de caméra au complet:

$$\begin{pmatrix} u \\ v \\ 1 \end{pmatrix} \approx \begin{bmatrix} k_u f & 0 & k_u x_0 & 0 \\ 0 & k_v f & k_v y_0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} R & -R\vec{t} \\ 0^T & 1 \end{bmatrix} \begin{pmatrix} X^m \\ Y^m \\ Z^m \\ 1 \end{pmatrix} \quad (7)$$

En passant aux coordonnées non homogènes (en divisant par la dernière coordonnée), il est possible d'identifier la matrice M :

$$M = \begin{bmatrix} k_u f & 0 & k_u x_0 \\ 0 & k_v f & k_v y_0 \\ 0 & 0 & 1 \end{bmatrix}$$

Ce qui permet d'écrire l'équation (7) sous la forme suivante:

$$\begin{aligned} P &\approx (MR - MR\vec{t}) \\ P &\approx MR(I - \vec{t}) \end{aligned} \quad (8)$$

Où P est la matrice de projection associée à la caméra, R la matrice de rotation et \vec{t} le vecteur de translation. M correspond aux **paramètres intrinsèques** de la caméra.

1.3.1. Paramètres extrinsèques et intrinsèques

Il est possible de regrouper les paramètres définissant la projection effectuée par la caméra en deux ensembles:

- **Les paramètres extrinsèques:** La matrice de rotation R et le vecteur de translation \vec{t} . Qui ensemble représentent la position et l'orientation de la caméra vis-à-vis du "monde extérieur".
- **Les paramètres intrinsèques:** La matrice M qui représente les caractéristiques internes de la caméra (f , k_u , k_v , x_0 et y_0). Ces paramètres sont indépendants de la position de la caméra.

Pour simplifier M , il est possible de représenter ses 5 paramètres par 4 paramètres intermédiaires qui auront l'avantage d'être sans unités:

- $\alpha_u = k_u f$
- $\alpha_v = k_v f$
- $u_0 = k_u x_0$
- $v_0 = k_v y_0$

Dans la notation traditionnelle, c'est ces 4 paramètres qui sont considérés comme les paramètres intrinsèques de la caméra et leur interprétation est la suivante:

- α_u et α_v expriment respectivement la **distance focale** en nombre de pixels horizontalement et verticalement.
- u_0 et v_0 sont les coordonnées du point principal exprimé dans le repère pixels.

La matrice M devient alors:

$$M = \begin{bmatrix} \alpha_u & 0 & u_0 \\ 0 & \alpha_v & v_0 \\ 0 & 0 & 1 \end{bmatrix}$$

Sous cette forme, la matrice M correspond à la **matrice de calibrage** de la caméra.

2. Distorsion

Avec le modèle sténopé il serait nécessaire d'attendre un temps considérable pour que la lumière s'accumule sur le capteur photosensible. Pour s'affranchir de ce problème, les caméras utilisent des lentilles. Une lentille peut concentrer une grande quantité de lumière sur le point de projection et ainsi accélérer le processus de captation mais au prix d'une distorsion de la réalité.

Il existe deux types principaux de distorsion:

- **Radiale:** induite par la forme de la lentille.
- **Tangentielle:** induite par l'alignement de la lentille et du capteur photosensible.

Il existe de nombreux autres types de distorsion dans les systèmes de capture d'images mais leur effet est négligeable comparativement à celui des deux précédents.

2.1. Distorsion radiale

Dans le contexte de la distorsion radiale, la distorsion vaut 0 au niveau du centre optique du capteur photosensible et accroît en direction de la périphérie. En pratique, cette distorsion est faible et peut être caractérisée par les premiers termes d'une série de MacLaurin ($r = 0$). Pour donner une idée, pour une webcam de bureau sont généralement utilisés les deux premiers termes (k_1, k_2). Pour des lentilles à grosse distorsion, comme des lentilles fish-eye, un troisième terme (k_3) sera utilisé. Ainsi, la mise à l'échelle de la position radiale d'un point sur un capteur photosensible sera exprimée par les équations suivantes:

$$x_{correction} = x(1 + k_1 r^2 + k_2 r^4) \quad (6)$$

$$y_{correction} = y(1 + k_1 r^2 + k_2 r^4) \quad (7)$$

2.2. Distorsion tangentielle

La distorsion tangentielle se caractérise par deux nouveaux paramètres p_1 et p_2 tel que:

$$x_{correction} = x + [2p_1 y + p_2(r^2 + 2x^2)] \quad (7)$$

$$y_{correction} = y + [p_1(r^2 + 2y^2) + 2p_2 x] \quad (8)$$

3. Vision stéréoscopique

La perception humaine du relief (et donc de la profondeur) est due à ses **deux** yeux. En effet, lorsqu'un objet est perçu, chaque oeil en reçoit une image différente. L'image reçue par l'oeil gauche est légèrement décalée par rapport à celle reçue par l'oeil droit, et c'est la combinaison des deux images faite par le cerveau qui produit l'effet de relief.



La vision stéréoscopique peut être définie comme la possibilité de percevoir les distances relatives des objets sur base de l'information visuelle provenant de la **disparité binoculaire**.

3.1. Disparité binoculaire

Les yeux sont espacés en moyenne de 60mm ¹. Chaque oeil perçoit une image légèrement différente de l'autre. Cette différence est appelée **disparité binoculaire** et c'est elle qui nous permet de percevoir la profondeur d'un point 3D.

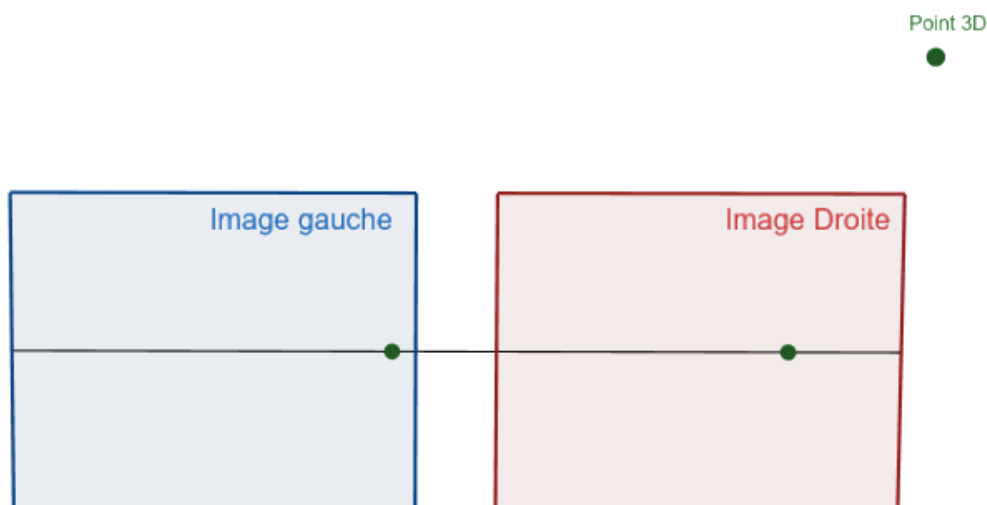


Figure 5 La disparité (d) est la distance entre les deux points sur l'axe

$$d = x_{\text{gauche}} - x_{\text{droite}}$$

À quel point est-il possible d'émuler cette capacité dans le contexte d'un système informatique ?

1. De pupille à pupille, d'après <https://choisir-ses-lunettes.com>

3.2. Imagerie stéréo

Une machine peut accomplir cette tâche en cherchant une correspondance entre un point perçu par un premier capteur ainsi que son point homologue sur un second capteur. Avec cette correspondance et la distance entre les deux capteurs, il est possible de déduire les coordonnées 3D de ce point. La recherche de correspondances est une opération coûteuse en ressources et il est impératif d'élaguer le problème en utilisant au mieux les informations sur le système géométrique du problème.

Dans la pratique, l'imagerie stéréo implique 4 étapes principales:

1. **Suppression de la distorsion** radiale et tangentielle des lentilles.
2. **Rectification** des angles et des distances entre les caméras pour faire en sorte que les deux images soient coplanaires et que les points homologues de chaque image soient alignés en y .
3. **Recherche de correspondance** entre les points homologues des deux images. Cette étape produit une **carte de disparité**.
4. **La reprojection** utilise la connaissance de l'arrangement géométrique des caméras pour déduire par triangulation les distances de chaque point de la carte de disparité et produire une **carte de profondeur**.

Ces étapes seront détaillées dans la partie pratique du présent document après avoir fait le point sur plusieurs notions théoriques nécessaires à leurs compréhension.

3.3. Triangulation

Soit le cas idéal d'une paire de caméras parfaitement alignées, ne souffrant d'aucune distorsion et dont toutes les mesures sont connues:

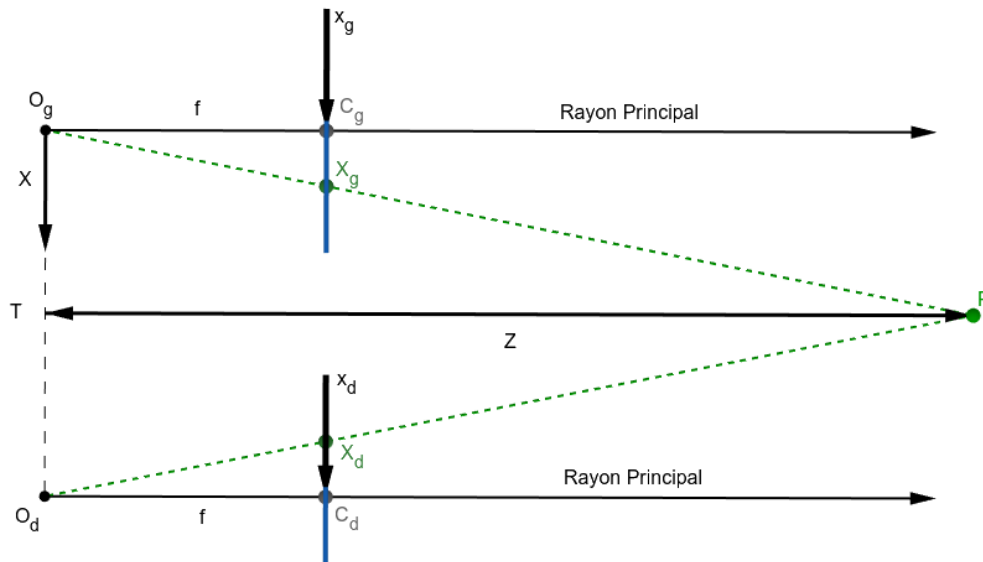


Figure 6 Modèle idéal avec Y perpendiculaire à la page

Donc deux caméras qui partagent les propriétés suivantes:

- Les distances focales f sont identiques. $f_g = f_d = f$
- Les plans image sont parfaitement coplanaires.
- L'écart 1T les séparant est connu.
- Les rayons principaux (partant des centres de projection O_g et O_d et passant par les points principaux C_g et C_d) sont parallèles.
- Les points principaux C_g et C_d sont calibrés pour faire en sorte d'avoir les mêmes coordonnées pixel dans leurs images respectives.
- Les images sont rectifiées et chaque rangée de pixels d'une image correspond à celle de l'autre.

1. Aussi appelé Baseline

Un tel arrangement se nomme *frontal parallèle*¹. Dans cette arrangement, avec X_g et X_d étant la position horizontale de la projection du point P dans leurs plans images respectifs, en utilisant les triangles semblables, il est possible de déduire la valeur de Z :

$$\frac{T - (X_g - X_d)}{Z - f} = \frac{T}{Z} \Rightarrow Z = \frac{Tf}{X_g - X_d}$$

Avec:

$$d = X_g - X_d$$

Cela implique que:

1. Quand la disparité tend vers 0:
 - Les petites différences de disparité traduisent de grandes différences de profondeur.
2. Quand la disparité est grande:
 - Les petites différences de disparité traduisent de faibles différences de profondeur.

Autrement dit, les systèmes d'imagerie stéréo ont une **haute précision** de la profondeur uniquement pour les objets relativement **proches** des caméras.



Il est important de retenir que la **profondeur** z est inversement proportionnelle à la disparité. Voir page n pour une démonstration du cas pratique

1. Bradski et Kaehler (2008), Learning OpenCV, "O'Reilly Media"

3.4. Rectification

Le modèle idéal précédent permet de comprendre la stratégie employée pour retrouver la profondeur. Il est maintenant nécessaire de trouver un moyen de configurer un système de caméras du monde physique dans une géométrie qui lui ressemble le plus possible. En effet, dans le monde physique, deux caméras ne jouiront jamais des propriétés idéales de l'arrangement *frontal parallèle* de la figure précédente.

À défaut, il est possible de retrouver mathématiquement les images projetées ainsi que les images sans distorsion et rectifiées pour simuler un arrangement (*frontal parallèle*) qui correspond à ce qui est décrit dans la figure suivante:

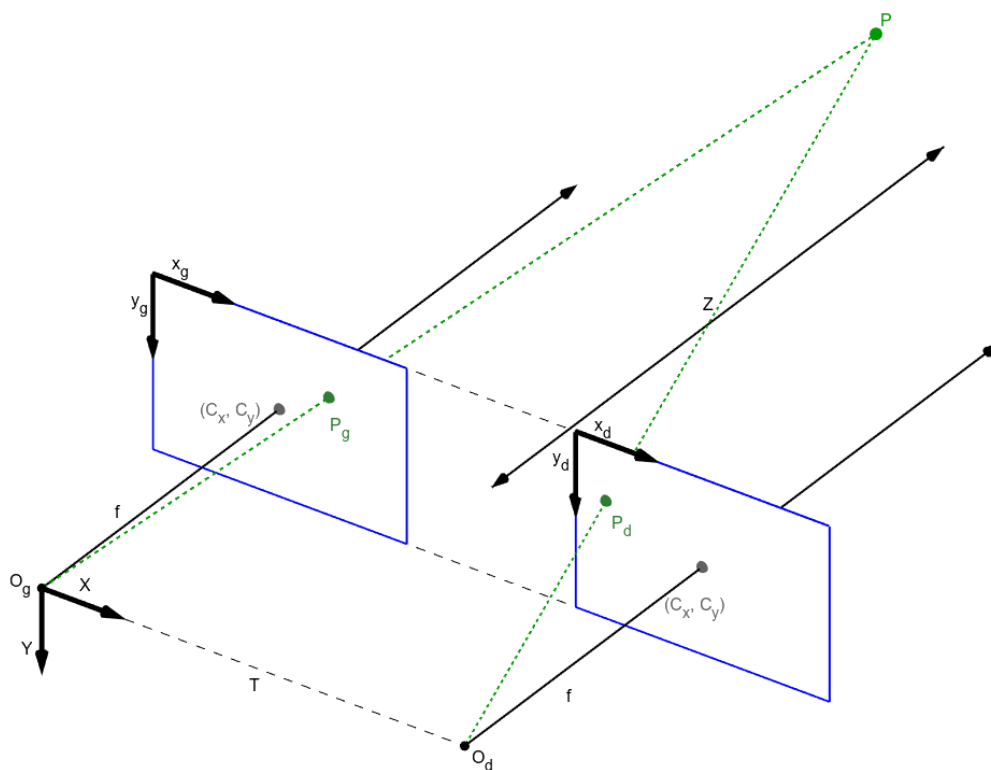


Figure 7 Un système réel (imparfait) après rectification.



Dans la pratique, il est important d'aligner au mieux possible les deux caméras physiques. Moins elles le sont, plus l'alignement simulé mathématiquement produira de distorsion ce qui aura pour effet de réduire voir d'éliminer la zone de superposition des deux images et donc de réduire la précision.

La figure suivante illustre la situation réelle (bleu) entre deux caméras et l'alignement mathématique obtenu après rectification (rouge).

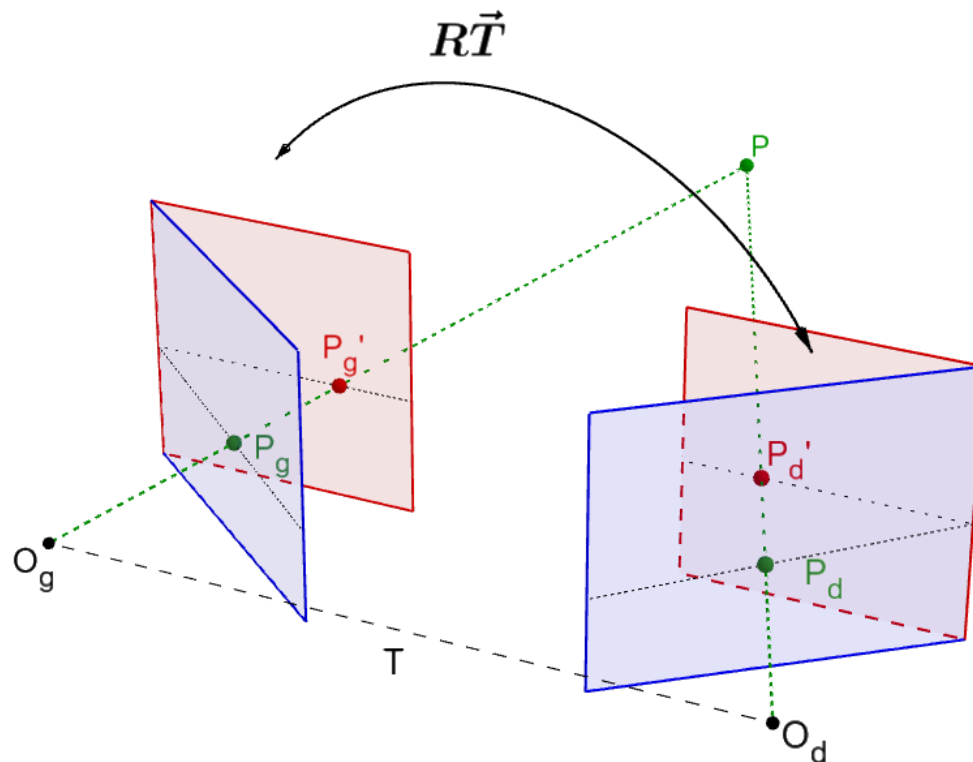


Figure 8 Alignement virtuel des deux caméras en un plan image où les rangées de pixels de chaque image sont alignées. Bleu: Situation initiale, rouge: Situation finale

Pour comprendre comment arriver à ce résultat il est nécessaire d'introduire la géométrie qui décrit deux caméras observant une même scène.

3.5. Géométrie épipolaire

La géométrie épipolaire est un modèle qui décrit les relations géométriques entre deux vues d'une même scène prises de points de vue différents. Ce modèle dépend des paramètres intrinsèques et extrinsèques des caméras et ne s'occupe pas de la structure ou du contenu de la scène.

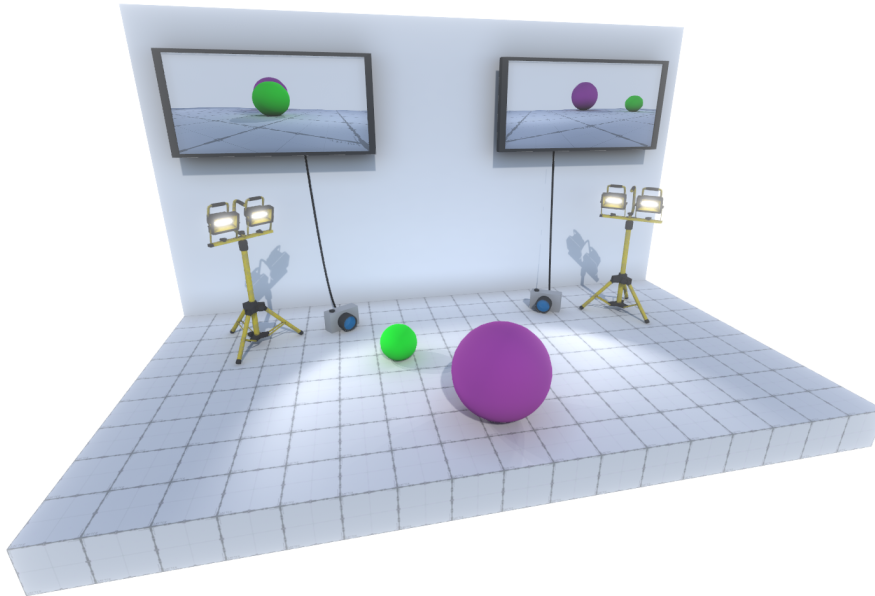


Figure 9 Deux appareils photo observent la même scène de deux positions différentes. La géométrie épipolaire décrit les relations entre les deux images.

Cette géométrie combine deux modèles sténopé ainsi que des points d'intérêt nommés **épipôles**. Le principal intérêt de cette géométrie est le fait de permettre de considérablement réduire la zone de recherche d'un point de correspondance.

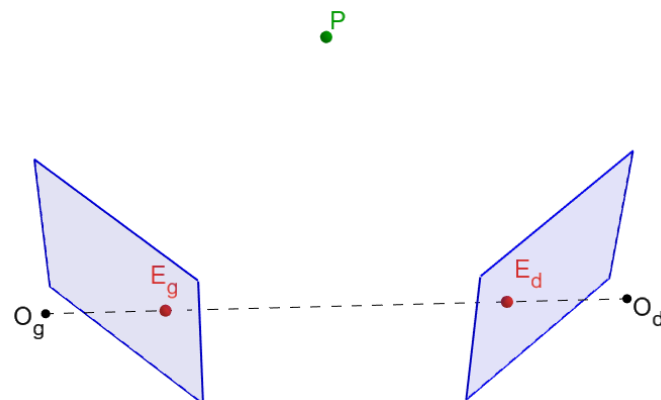


Figure 10 Épipôles

3.5.1. Définitions

- Un **épipôle** est le centre de projection d'une caméra vue dans le plan image de l'autre caméra. La droite qui relie les deux épipôles est la **baseline**.
- Le **plan épipolaire** Π_e est défini par le point considéré P ainsi que le centre de projection des deux caméras ou les deux points épipolaires.
- Les **droites épipolaires**: e_g et e_d sont les droites résultant de l'intersection du plan épipolaire avec les plans image des deux caméras.

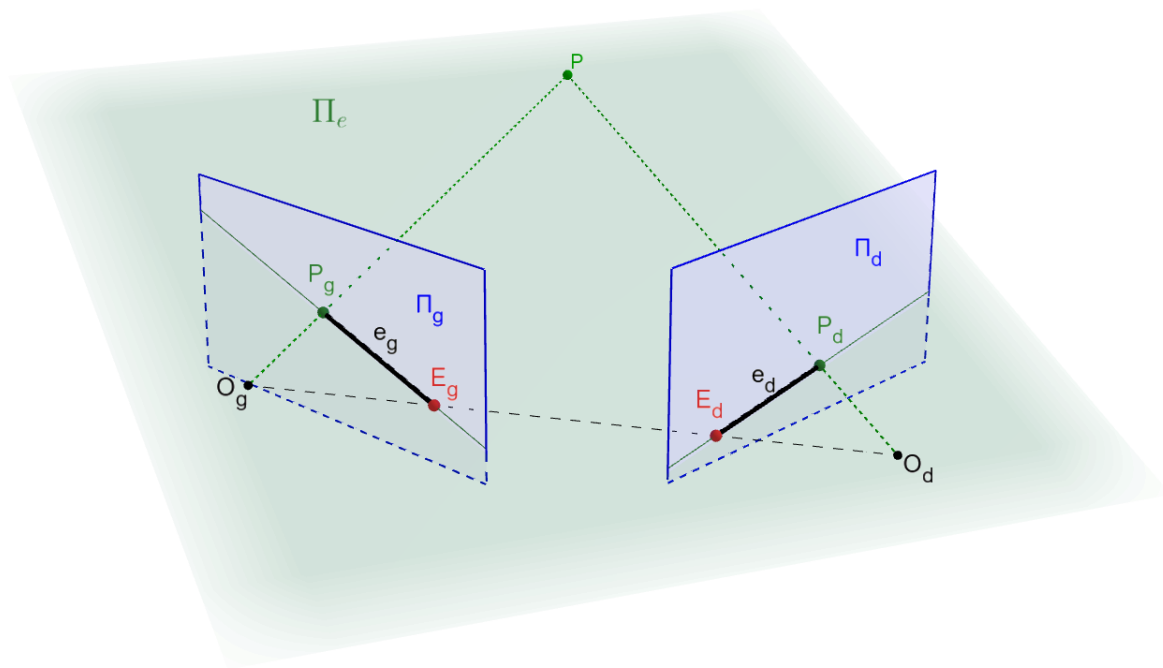


Figure 11 Géométrie épipolaire

3.5.2. Propriétés

- Il existe un plan épipolaire par point objet.
- Dans chaque plan image, il n'existe qu'une droite épipolaire par point objet.
- Toutes les droites épipolaires dans un plan image passent par l'épipôle.
- Tous les plans épipolaires passent par la baseline qui en est le centre de rotation
FAIRE UNE FIGURE.
- Quand les rayons principaux sont parallèles, les épipôles sont reportés à l'infini.
- Un point d'une image possède un point homologue sur la droite épipolaire de l'autre image. On parle de **contrainte épipolaire**.

3.5.3. Contrainte épipolaire

Pour comprendre l'utilité des épipôles, il est nécessaire de comprendre que lorsqu'un point P est projeté sur un des plans image disons Π_g , du point de vue de l'autre caméra, ce point pourrait être situé n'importe où sur une droite formée par le point et sa projection P_g .

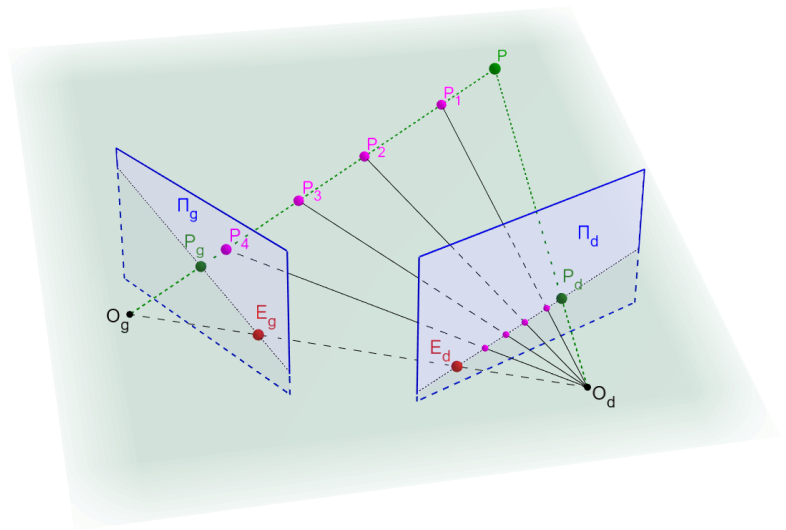


Figure 12 Contrainte épipolaire

À ce stade, il est intéressant de se demander à quoi ressemble la projection de la droite P, P_g sur Π_d . La réponse à cette question, comme illustré par la figure précédente, est que cette projection est la droite formée par P_d et E_d . C'est ce qui s'appelle la **contrainte épipolaire**. Comme l'illustre la figure suivante, cette contrainte est symétrique.

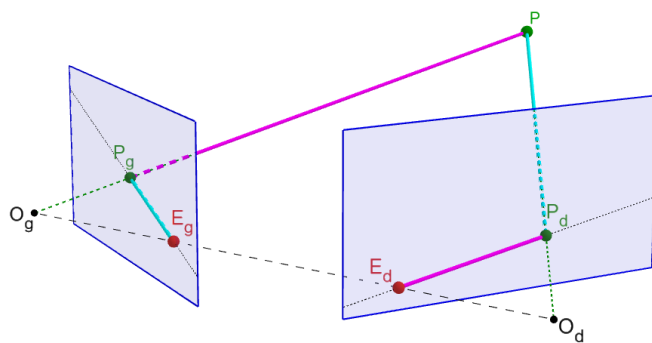


Figure 13 Symétrie de la contrainte épipolaire

La contrainte épipolaire veut dire que le problème de recherche de correspondance dans une image (2D) devient un problème de recherche le long d'une droite (1D).

Cela se traduit non seulement par de vastes économies en ressources et en temps de calcul mais aussi en une réduction considérable de faux positifs en terme de correspondance qui se seraient traduits par des artefacts lors de la génération de la carte de disparité.

3.6. Matrices essentielles et fondamentales

Ces deux matrices expriment la même chose mais dans des repères différents:

- La **matrice essentielle E (Caméra)** contient des informations sur la translation et la rotation qui lie les deux caméras dans le référentiel des caméras.
- La **matrice fondamentale F (Pixels)** contient les mêmes informations que E en y ajoutant les paramètres intrinsèques des deux caméras pour pouvoir exprimer la translation et la rotation qui lie les caméras mais cette fois dans le référentiel lié au plan image.

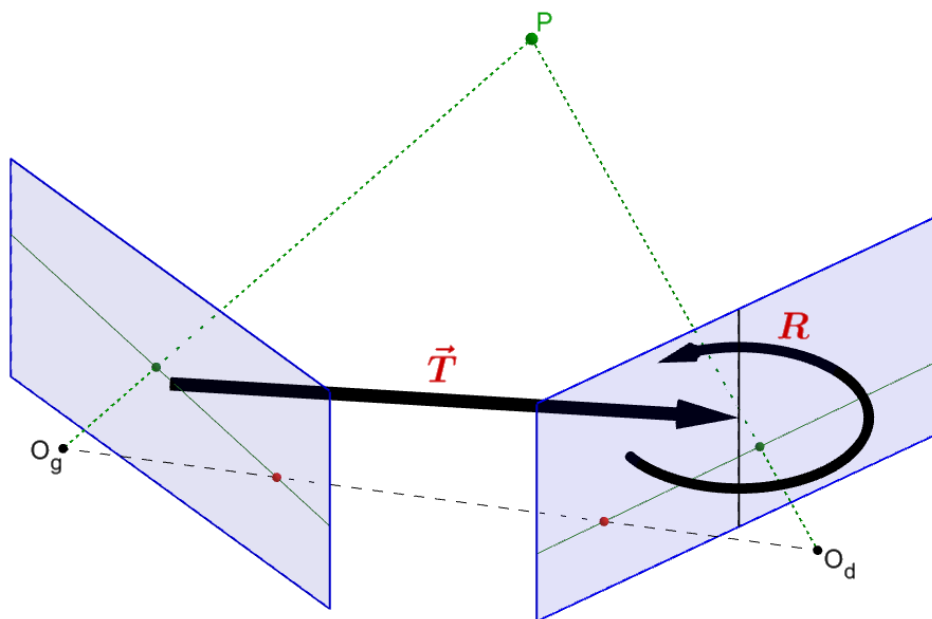


Figure 14 Translation et rotation pour mapper mathématiquement un plan image sur l'autre.

La matrice essentielle est donc purement géométrique, elle lie la position d'un point P de la scène du point de vue d'une caméra à la position du même point mais cette fois du point de vue de l'autre caméra. La matrice fondamentale quant à elle lie la position de la projection de P sur un plan image à la projection du même point mais sur l'autre plan image.

Nuance importante



Il est nécessaire de bien comprendre que dans le cas de la matrice essentielle on ne parle pas de projection de P sur un plan image. En effet la matrice essentielle n'a aucune information concernant ce repère !

3.6.1. Matrice essentielle

Les positions Q_g et Q_d , c'est à dire la position du point P dans le repère respectif de chaque caméra (**et non pas la projection de P sur les plans image**). La relation qui les lie peut être déduite grâce à la géométrie épipolaire.

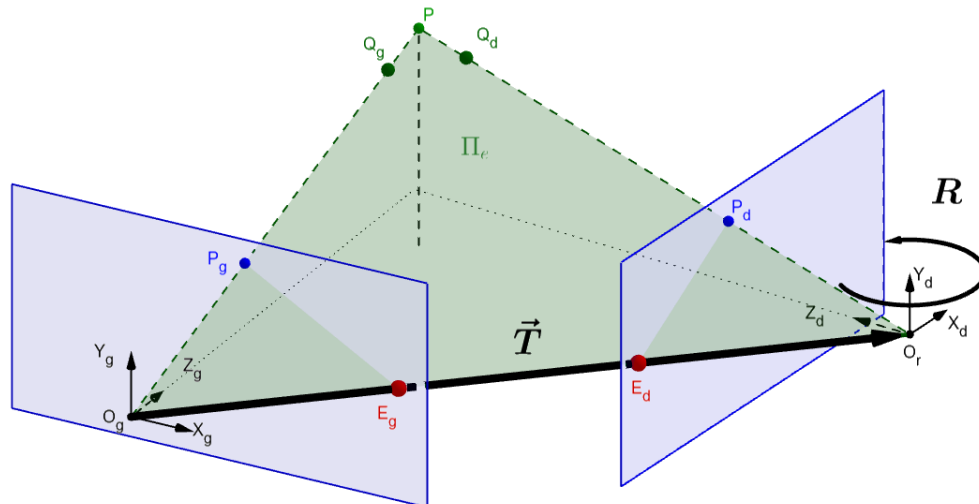


Figure 15 Opérations de la matrice fondamentale

Considérons le repère (caméra) centré en O_g . Dans ce système de coordonnées, la position du point P est Q_g et l'origine de l'autre caméra se trouve en T . Dans ce référentiel, le point P vu par la caméra de droite est Q_d avec:

$$\begin{aligned} Q_d &= R(Q_g - T) \\ R^{-1}Q_d &= (Q_g - T) \end{aligned} \quad (8)$$



Rappel

- Si $\vec{u} \perp \vec{v}$ alors $\vec{u} \circ \vec{v} = 0$
- $\vec{u} \circ \vec{v} = \vec{u}^T \cdot \vec{v}$
- Soit R une matrice de rotation, alors: $R^T = R^{-1}$

Sachant que tout point x appartenant à un plan ayant un vecteur normal \vec{n} passant par un point a satisfait la relation $(x - a) \circ \vec{n} = 0$, en substituant \vec{n} avec le vecteur issu de $(Q_g \times T)$, cette relation devient:

$$\begin{aligned} (Q_g - T) \circ (T \times Q_g) &= 0 \\ (Q_g - T)^T (T \times Q_g) &= 0 \end{aligned}$$

En substituant dans (8), le résultat suivant apparaît:

$$(R^T Q_d)^T (T \times Q_g) = 0 \quad (9)$$

Soit SQ_g , le résultat du produit $T \times Q_g$ avec:

$$S = \begin{bmatrix} 0 & -T_z & T_y \\ T_z & 0 & -T_x \\ -T_y & T_x & 0 \end{bmatrix}$$

Rappel

i

- $(A^T)^T = A$
- $(A + B)^T = B^T + A^T$
- $(AB)^T = B^T A^T$
- Pour tout scalaire r , $(rA)^T = rA^T$

Finalement, en substituant ce résultat dans (9), une relation liant Q_d et Q_g est obtenue:

$$(R^T Q_d)^T SQ_g = Q_g^T RSQ_g = 0 \quad (10)$$

Mais ce qui est recherché est une relation entre P_d et P_g , des **points homologues dans les deux images**. Pour ce faire, les équations (1) et (2) du chapitre sur la modélisation de caméra sont ressuscitées et adaptées:

$$P_g = Q_g \frac{f_g}{Z_g} \Rightarrow Q_g = P_g \frac{Z_g}{f_g} \quad P_d = Q_d \frac{f_d}{Z_d} \Rightarrow Q_d = P_d \frac{Z_d}{f_d}$$

En substituant dans (10):

$$P_d^T \frac{Z_g Z_d}{f_g f_d} R S P_g = P_d^T E P_g = 0 \quad (11)$$

Où E est la **matrice essentielle**, une matrice 3x3 où les points sont **exprimés dans le repère caméra**.

$$E = \frac{Z_g Z_d}{f_g f_d} R S$$

Il est important de passer un moment à comprendre ce que signifie le résultat de (11). En effet à première vue on pourrait penser que cette équation permet de décrire un des termes P si l'autre est donné mais à cause des caractéristiques de E , (11) est l'équation d'une droite. En effet, si on développe (11) sous forme homogène, on obtient:

$$(P_d^x, P_d^y, 1) \begin{bmatrix} e_{11} & e_{12} & e_{13} \\ e_{21} & e_{22} & e_{23} \\ e_{31} & e_{32} & e_{33} \end{bmatrix} \begin{pmatrix} P_g^x \\ P_g^y \\ 1 \end{pmatrix} = 0$$

$$(P_d^x, P_d^y, 1) \begin{bmatrix} e_{11}P_g^x & e_{12}P_g^y & e_{13} \\ e_{21}P_g^x & e_{22}P_g^y & e_{23} \\ e_{31}P_g^x & e_{32}P_g^y & e_{33} \end{bmatrix} = 0$$

Et donc avec:

$$\begin{aligned} a &= (e_{11}P_g^x + e_{12}P_g^y + e_{13}) \\ b &= (e_{21}P_g^x + e_{22}P_g^y + e_{23}) \\ c &= (e_{31}P_g^x + e_{32}P_g^y + e_{33}) \end{aligned}$$

On obtient finalement l'équation de la **droite épipolaire** dans l'image de droite exprimée dans le repère caméra:

$$aP_d^x + bP_d^y + c = 0$$

Il est maintenant possible de conclure que la matrice essentielle permet à l'aide d'un point exprimé dans les coordonnées du repère d'une des caméras de trouver l'équation de la droite (épipolaire) où se trouve son point homologue dans l'autre image.

3.6.2. Matrice fondamentale

Le précédent chapitre a fait le gros du travail explicatif, **fondamentalement**, Il ne reste qu'à trouver q_d et q_g qui est l'expression dans le repère image des points P_d et P_g (qui eux sont exprimés dans le repère caméra). Pour ce faire il est nécessaire d'utiliser la matrice M encapsulant les paramètres intrinsèques (chapitre 3.3.1). Soit:

$$\begin{aligned} q &= MP \\ p &= M^{-1}q \end{aligned}$$

avec :

$$M = \begin{bmatrix} f_x & 0 & C_x \\ 0 & -f_y & C_y \\ 0 & 0 & 1 \end{bmatrix}$$

Donc:

$$\begin{aligned} (M_d^{-1} q_d)^T \cdot EM_g^{-1} q_g &= 0 \\ q_d^T (M_d^{-1})^T \cdot EM_g^{-1} q_g &= 0 \\ q_d^T F q_g &= 0 \end{aligned} \quad (12)$$

Où F est la **matrice fondamentale**, une matrice 3x3 où les points sont exprimés dans le repère image:

$$F = (M_d^{-1})^T \cdot EM_g^{-1}$$

Cette relation exprime le fait qu'à l'aide de la matrice fondamentale, il est possible de trouver la droite épipolaire où se trouve le point homologue d'un point exprimé dans le repère image de la première image.

En conclusion et pour résumer, il faut garder en tête que la matrice fondamentale fait la même chose que la matrice essentielle sauf que le fait dans le repère image alors que la matrice essentielle opère dans le repère caméra. Les deux matrices sont des matrices 3x3 de rang 2. Aussi il est important de spécifier que la matrice fondamentale possède 7 paramètres, deux pour chaque épipôle et trois pour l'homographie qui lie les deux plans image.

3.7. Calibration stéréo (p427)

La calibration stéréo est le procédé qui permet de calculer la relation géométrique entre deux caméras dans un repère 3D.

Ce procédé précède la rectification stéréo qui elle fait en sorte de corriger des images individuelles pour faire en sorte qu'elles soient comme capturées côte à côte.

Une telle configuration correspond à celle de la figure 7.

3.8. Reprojection

Après avoir compris de façon isolée comment trouver la profondeur d'un point par triangulation (chapitre 4.3) en ayant au préalable réduit la zone de recherche dans la seconde image grâce à la géométrie épipolaire (chapitre 4.5) et des matrices E et F il est maintenant possible de comprendre comment arriver au résultat décrit dans le (chapitre 4.4).



Rappel

La reprojection utilise la connaissance de l'arrangement géométrique des caméras pour déduire par triangulation les distances de chaque point de la carte de disparité et produire une carte de profondeur.

3.8.1. Application de la méthode de Bouguet

La rectification de Bouguet se déroule en 4 étapes:

1. Faire pivoter la caméra (disons gauche) pour que l'épipôle aille à l'

La méthode de Bouguet se déroule en 4 étapes:

1. Faire pivoter la caméra (disons gauche) pour que l'épipôle aille à l'infini le long de l'axe horizontal.
2. Appliquer la même rotation à l'autre caméra (droite) pour se retrouver dans la configuration géométrique initiale.
3. Faire pivoter la caméra de droite d'une rotation R correspondant au paramètre extrinsèque du système stéréo de départ.
4. Ajuster l'échelle dans les deux repères caméras.

... To be continued

4. Références

4.1. Papier

- G. Bradski et A. Kaehler, Learning OpenCV, O'Reilly, 2008

4.2. En Ligne

4.2.1. Articles, Cours & Tutos

- Coordonnées Homogènes
- Vision par ordinateur (inria.fr)
- Modélisation à partir d'images (inria.fr)
- Vision stéréoscopique (Marie-odile Berger)
- Stereo vision using one camera
- Tutorial made by the dev of Py stereovision module

4.2.2. Cours vidéo

- Disparité binoculaire
- Triangulation
- Tutoriel Vision stéréoscopique (matlab)
- William Hoff Lecture 21-1 Stereo Vision 1
- William Hoff Lecture 21-2 Stereo Vision 2
- William Hoff Lecture 21-3 Stereo Vision 3
- William Hoff Lecture 23-1 Epipolar and Essential
- William Hoff Lecture 23-2 Epipolar and Essential 2

4.2.3. Interactif

- Pinhole (stenopé) (Geogebra)
- Disparité (Geogebra)
- Disparité II (Geogebra)
- Géométrie épipolaire (Geogebra)