Formation et Analyse d'Images

James L. Crowley

ENSIMAG 3 Premier Bimestre 2002/2003
<u>Séance 7</u> 21 novmbre 2002

Stéreo et la Géometrie Epipolaire

<u>Plan de la Séance:</u>

La Vision Stéréoscopique

La vision stéréoscopique est l'estimation de la position des points dans la scène à partir de deux images prises de deux position différentes.

Il y a deux problèmes en stéréo :

1) Représentation des Images.
1) La recherche des Correspondances
2) Inférence de position 3D.

La recherche de points de correspondance entre deux images et une
Grand problème en Vision depuis 30 ans.
Par conséquent il existe une grande diversité d'approches
pour la mise en correspondance.

Les Techniques d'Appariement

Les techniques de mise en correspondance cherches à apparier deux "choses".
Il existe des diverses manifestations des techniques pour le stéréo :

Au Niveau Pixels :

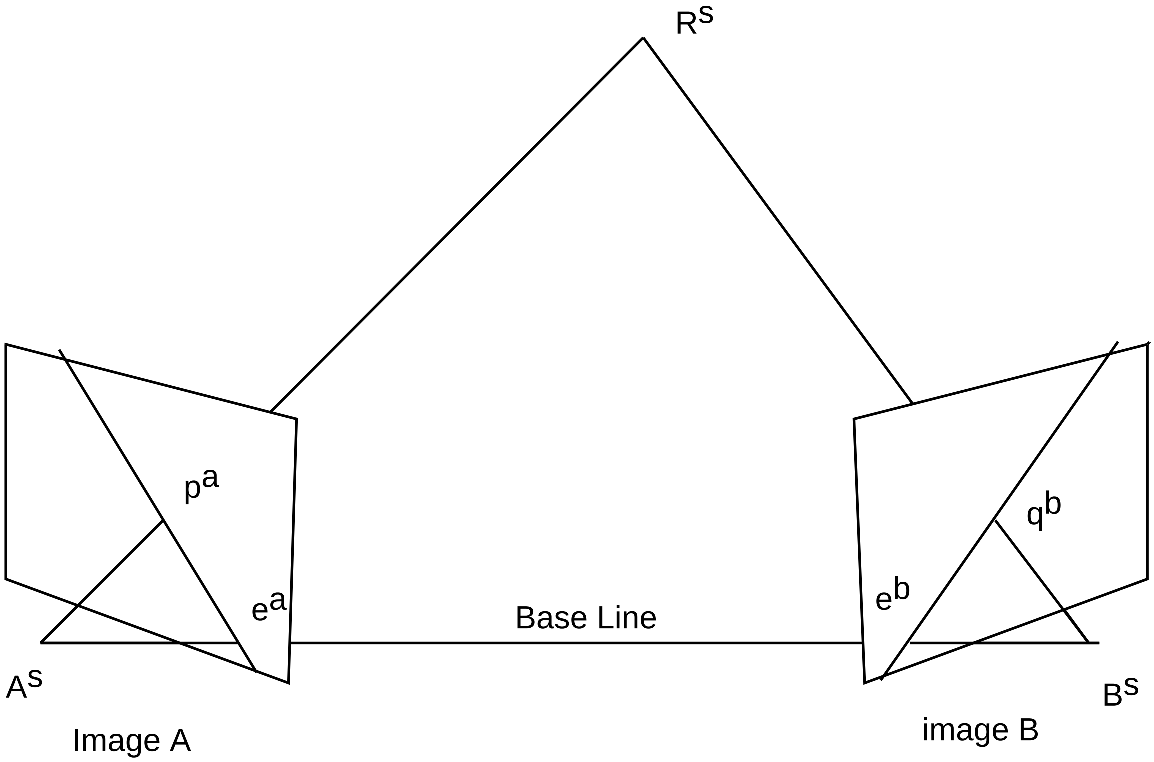
- Inter-correlation des Voisinages (NCC et SSD)
  - Dans l'image
  - Avec une pyramide multi-echelle
- Appariement après projection sur les champs réceptifs.
  - tous les points
  - Sur les maximums locaux.

Au niveau Contours :

- Par appariement de graphe
- Par appariement de group de contours

Dans tous les algorithmes, la clé est d'appliquer les <u>contraintes.</u>
La géométrie multi-image nous donne une contrainte puissante :
La Contrainte Epipolaire.

La Géométrie Stéréoscopique



Notation :

Point dans le Scène R^s (Les tenseurs en 3D sont Majuscule)

Caméra A :

- Centre de Projection : A^s
- Matrice de Projection : $\mathbf{M}_s^a$
- Projection de R^s : p^a (Les tenseurs en 2D sont minuscules)
- epipole de b en a : e^a

Relations pour la caméra a :

$$p^a = \mathbf{M}_s^a R^s$$
$$e^a = \mathbf{M}_s^a B^s$$

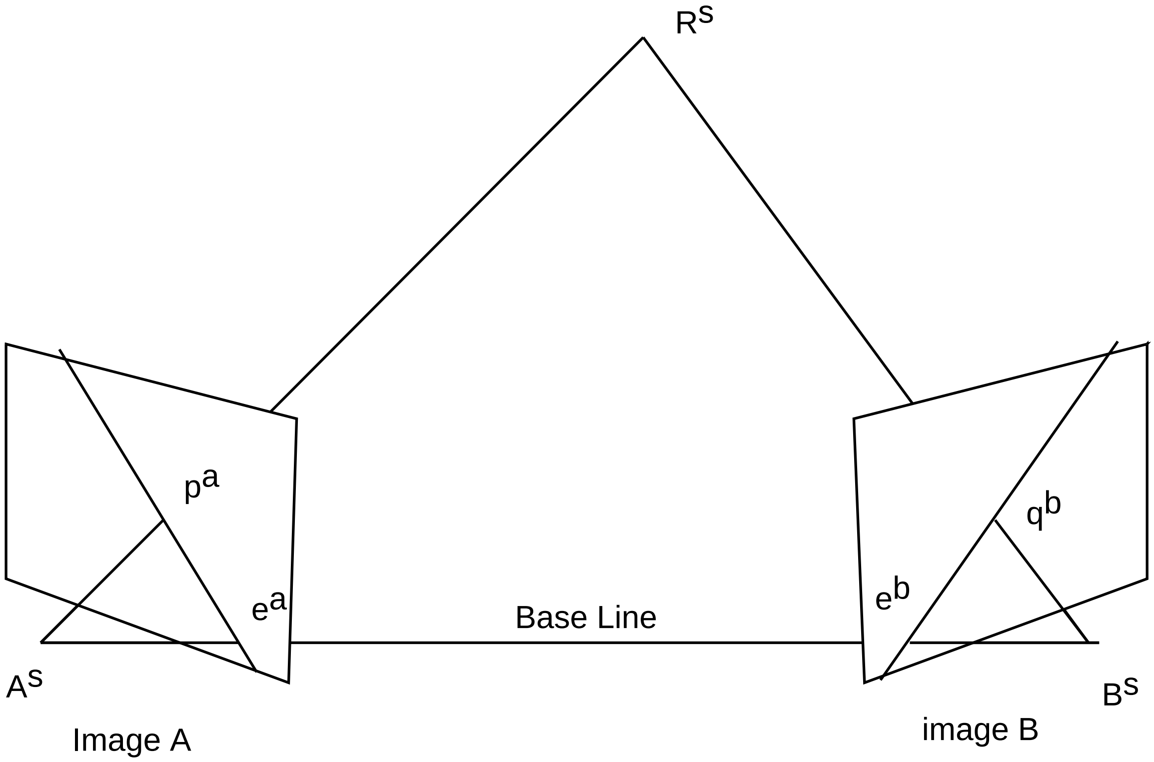
Caméra B :

- Centre de Projection : B^s
- Matrice de Projection : $\mathbf{M}_s^b$
- Projection de R^s : q^b
- epipole de a en b : e^b

Relations pour la caméra b :

$$q^b = \mathbf{M}_s^b R^s$$
$$e^b = \mathbf{M}_s^b B^s$$

Le plan Epipolair



La droite entre A^s et B^s est le "Base-line". Les épipoles sont définit par l'intersection de la base line avec les rétines.
L'épipole e^a est la projections de B^s sur l'image A. $e^a = \mathbf{M}_s^a B^s$

L'épipole e^a est la projections de B^s sur l'image A. $e^b = \mathbf{M}_s^b A^s$

Les trois points A^s, B^s et R^s définies un plan en S. : $P_s = (a, b, c, d)$.

Notation Classique : $a x + b y + c z + d = 0$
Notation Tensoriel : $P_s R^s = 0$.

Un plan est une contrainte sur quatre points tels que leurs déterminant est nulle.
Pour tout point dans la scène P^s sur le plan épipolaire :

Calcul : $$\det \begin{pmatrix} P^s \\ R^s \\ A^s \\ B^s \end{pmatrix} = \det \begin{pmatrix} x & y & z & 1 \\ R^1 & R^2 & R^3 & R^4 \\ A^1 & A^2 & A^3 & A^4 \\ B^1 & B^2 & B^3 & B^4 \end{pmatrix} = 0.$$

La Matrice Fondamentale

Dans un plan Euclidienne en R^2, en notation "classique", une droite est définie par une équation $a x + b y + x = 0$.

On peut exprimer cette équation comme la produit de deux vecteurs :

$$\vec{L} \cdot \vec{P} = 0 \qquad \text{ou} \quad \vec{L} = (a \;\; b \;\; c \;\;) \quad \text{et} \quad \vec{P} = \begin{pmatrix} x \\ y \\ 1 \end{pmatrix}$$

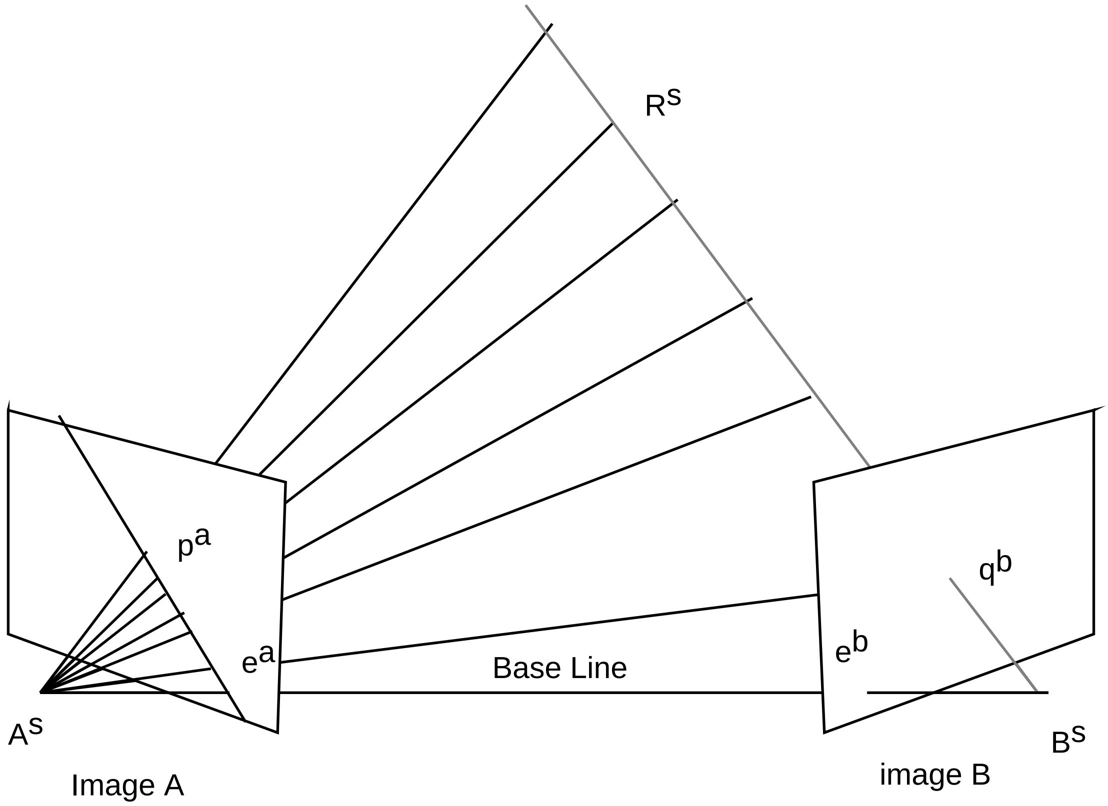
En notation tensorielle, cette équation est exprimée : $L_i P^i = 0$ avec i=1, 2, 3.

Pour chaque paire de Caméras, (a, b) il existe une matrice 3 x 3, F_{ab}, qui projet les points, p^a d'une sur une droite, l_b de l'autre

$$l_b = F_{ab}\, p^a \qquad\qquad m_a = F_{ab}\, q^b$$

L'interception de la plane épipolaire avec une image est une droite dans l'image.

Cette droite correspond à la projection de la droite $\overline{B^s q^b}$ sur la deuxième image. La matrice F_{ab} est connue comme la matrice "fondamentale".



La matrice fondamentale est de taille 3 x 3

- Taille 3 x 3 9 coefficients
- Coordonnées Homogènes $F_{33} = 1$ 8 coefficients
- 8 coefficients 4 points en correspondance

Pour calculer la matrice fondamentale : On note que $m_b\, q^b = 0$.

Parce que $m_b = F_{ab}\, p^a$ on a $(p^a F_{ab})\, q^b = p^a F_{ab}\, q^b = 0$.

Algorithme de 8 points :

a) Déterminer $K \geq 8$ paires de points correspondants (p^a_k, q^b_k)

b) Ecrire K : $p^a_k\, F_{ab}\, q^b_k = 0$.

Ceci donne $(p^a_k\ \ q^b_k)\, F_{ab} = 0$

On note que $p^3 = 1$ et que $q^3 = 1$ et que $F_{33} = 1$.

c) on écrit :

$$\begin{pmatrix} p^1_k q^1_k & p^1_k q^2_k & p^1_k & p^2_k q^1_k & p^2_k q^2_k & q^2_k & p^1_k & p^2_k \end{pmatrix} \begin{pmatrix} F_{11} \\ F_{12} \\ F_{13} \\ F_{21} \\ F_{22} \\ F_{23} \\ F_{13} \\ F_{23} \end{pmatrix} = 0$$

A cause des instabilités numériques et les imprécisions des pixels, il est conseillé de la calculer avec plus que 8 points par un calcul de moindre de carrée.

Méthode Moindre de Carrée.

Avec $N \geq 8$ couples de points $(p^a_k\ \ q^b_k)$
trouver F_{ab} qui minimise

$$\min\{\, \| (p^a_k\ \ q^b_k)\, F_{ab} \| \,\} = \min\{\, \| F \| \,\}$$

Le min est trouvé par SVD de A.

SVD(A) = $U \cdot D \cdot V^T$ Le dernier ligne de V est le F_{ab} qui minimise $\|AF\|$

Détermination de la Correspondance par Inter Corrélation

Méthode : Pour un point p^a de l'image A, on prend le voisinage X(m, n) de taille MxN comme un "motif". On recherche des voisinages W(m, n) de l'image B(i, j) qui "ressemble" X(n,m).

Hypothèses :

- bruit additif Gaussien.
- Pas de rotation dans l'image
- Pas de rotation dans l'espace 3D
- Pas de changement d'échelle.

Si on peut respecter ces quatre hypothèses, une norme Euclidienne est la méthode "optimale" (min probabilité d'erreur) pour détecter le voisinage d'une image qui ressemble à un motif.

Il existe des variations de la technique pour la rendre robuste au
rotations 2D, rotations 3D et changements d’échelle.

La norme Euclidienne est connue comme la SSD ("Sum of Squared Distances").
Il s'agit d'une opération efficace et précise, mais fragile.

Définition :

Soit un motif X(m, n) pour m [0, M–1], n [0, N–1].
Soit une image B(i, j) pour i [0, I–1], j [0, J–1]. (M << I, N << J)

Placer X(m, n) à chaque position possible (i, j) et
mesurer la distance Euclidienne entre X et les MN pixels de P à (i, j).

$$SSD(i, j) = \| X(m,n) - B(i+m, j+n) \|^2$$

$$= \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} (B(i+m, j+n) - X(m, n))^2$$

Si les pixels de B, à la position (i, j) ressemblent à X, la distance est nulle.
Sinon, la position ayant le maximum de vraisemblance est celle correspondant au minimum de la fonction SSD.

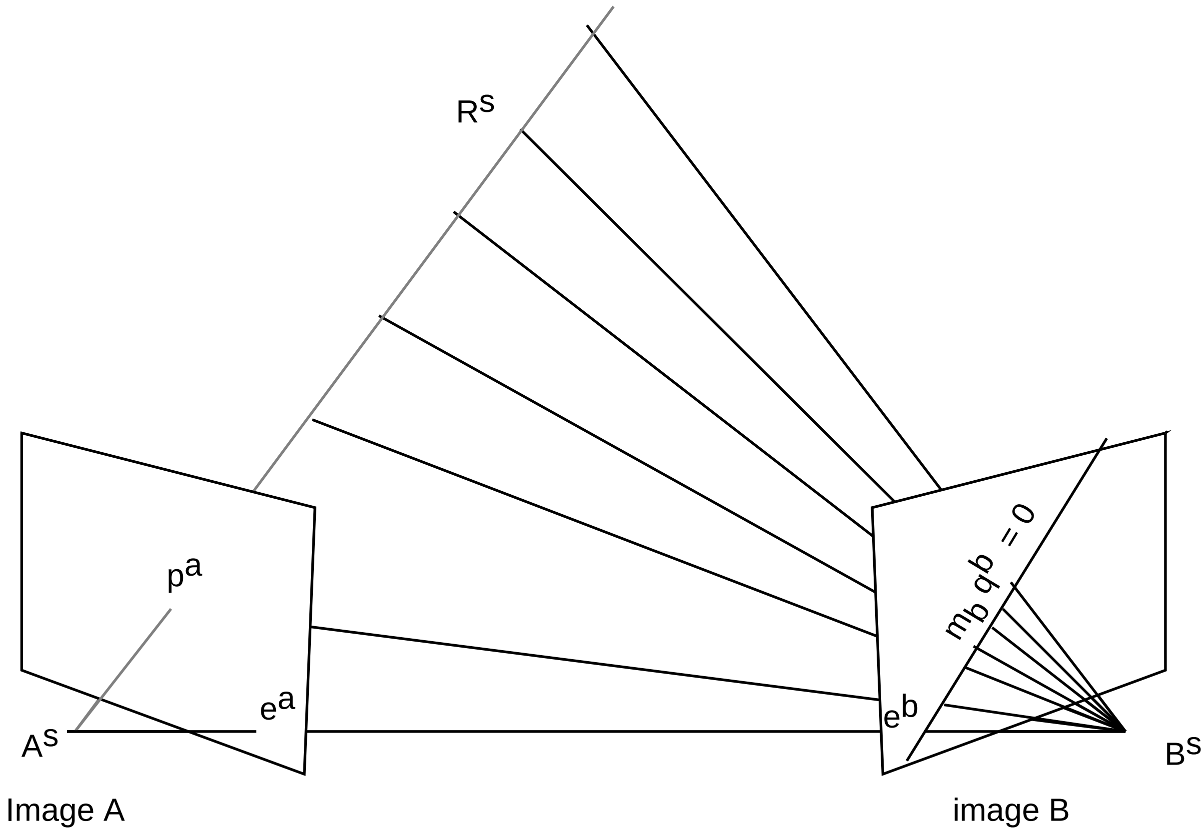
$$\underset{i\ j}{Min} \left\{ \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} (B(i+m, j+n) - X(m, n))^2 \right\}$$

grace à la contrainte épipolaire, on peut limiter la recherche
à la droite $m_b = F_{ab}\, p^a$.

$$\underset{i\ j \quad M}{\mathrm{Min}} \left\{ \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} (P(i+m,\ j+n) - X(m,\ n))^2 \right\}$$

M est la droite $m_1 i_b + m_2 j_b + m_3 = 0$

Chaque point de la scène R^s sur $\overline{A^s p}$ correspond à une point sur $m_b q^b = 0$



Pixels d'un voisinage comme vecteur charactéristique

Les pixels X(m, n) peuvent être vus comme un vecteur $\vec{X} = X_k$ ou $k = nM+m$.
Les pixels de chaque voisinage W(i, j) de B peuvent aussi être vus comme un vecteur.

$$\vec{W} = W_k \quad \text{ou } k = (j+n)I+m+i$$

L'opération SSD est la norme de la différence de ces deux vecteurs :

$$SSD(i, j) = \| \vec{X} - \vec{W} \|^2 = \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} (W(i+m, j+n) - X(m, n))^2 \quad \}$$

Une autre méthode de comparaison est le produit scalaire
(CC pour "Cross Corrélation") :

$$CC(i, j) = < \vec{X}, \vec{W} > = \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} W(i+m,j+n)X(m, n)$$

Si les vecteurs $\vec{X}$ et $\vec{W}$ ont une longueur unitaire, le produit scalaire est un cosinus de l'angle entre les vecteurs.

$$X_u(m, n) = \frac{\vec{X}}{\|\vec{X}\|} = \frac{X(m, n)}{\sum_{m=0}^{M-1} \sum_{n=0}^{N-1} X(m, n)^2}$$

$$W_u(m, n) = \frac{\vec{W}}{\|\vec{W}\|} = \frac{W(i+m, j+n)}{\sum_{m=0}^{M-1} \sum_{n=0}^{N-1} W(i+m, j+n)^2}$$

On obtient un inter corrélation "normalisée" par l'énergie (NCC) :

$$NCC(i, j) = < \vec{X}_u, \vec{W}_u > = \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} \frac{W(i+m,j+n)}{\|\vec{W}\|} \frac{X(m, n)}{\|\vec{X}\|}$$

Le NCC est le cosinus entre $\vec{X}_u$ $\vec{W}_u$. Sa valeur est entre –1 et 1.

Estimation du point R dans la scène

Le cas des rétines coplanaires

Les axes "optiques" sont parallèles
Les images sont dans le même plan.

Les deux caméras sont séparées par une Transformation T_a^b dans la direction X_S.

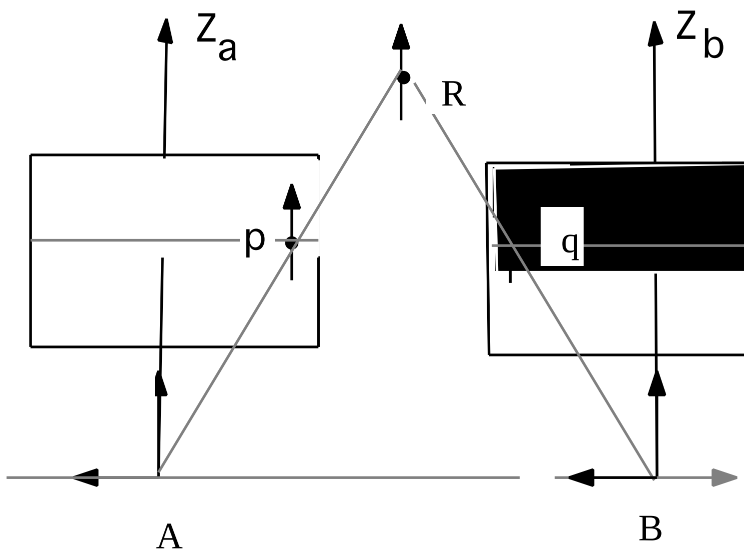
C.-à-d.

$$\mathbf{N}_S^b = \mathbf{M}_S^a \; T_a^b$$

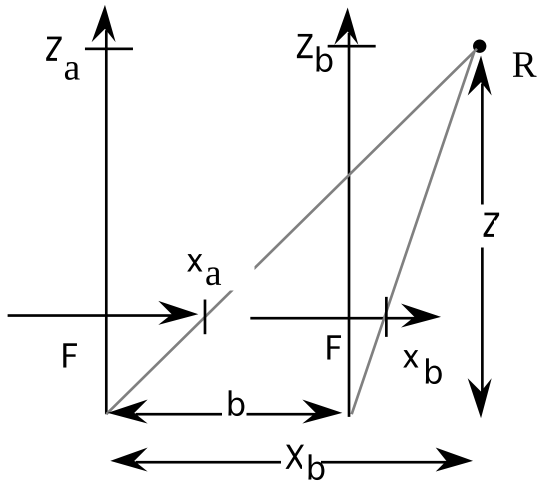
tel que

$$T_a^b = \begin{pmatrix} 1 & 0 & 0 & B \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

Dans ce cas, les droites "épipolaires" correspondent aux lignes des images.



La formule pour la profondeur peut être déduite par triangle semblable.



Écrire :

$$\frac{X_a}{Z_a} = \frac{x_{ra}}{F} \qquad \frac{X_b}{Z_b} = \frac{x_{rb}}{F}$$

Prendre la différence des équations :

$$\frac{X_a}{Z_a} - \frac{X_b}{Z_b} = \frac{x_{ra}}{F} - \frac{x_{rb}}{F}$$

Note que $Z_a = Z_b = Z$ et que $(X_a - X_b) = B$

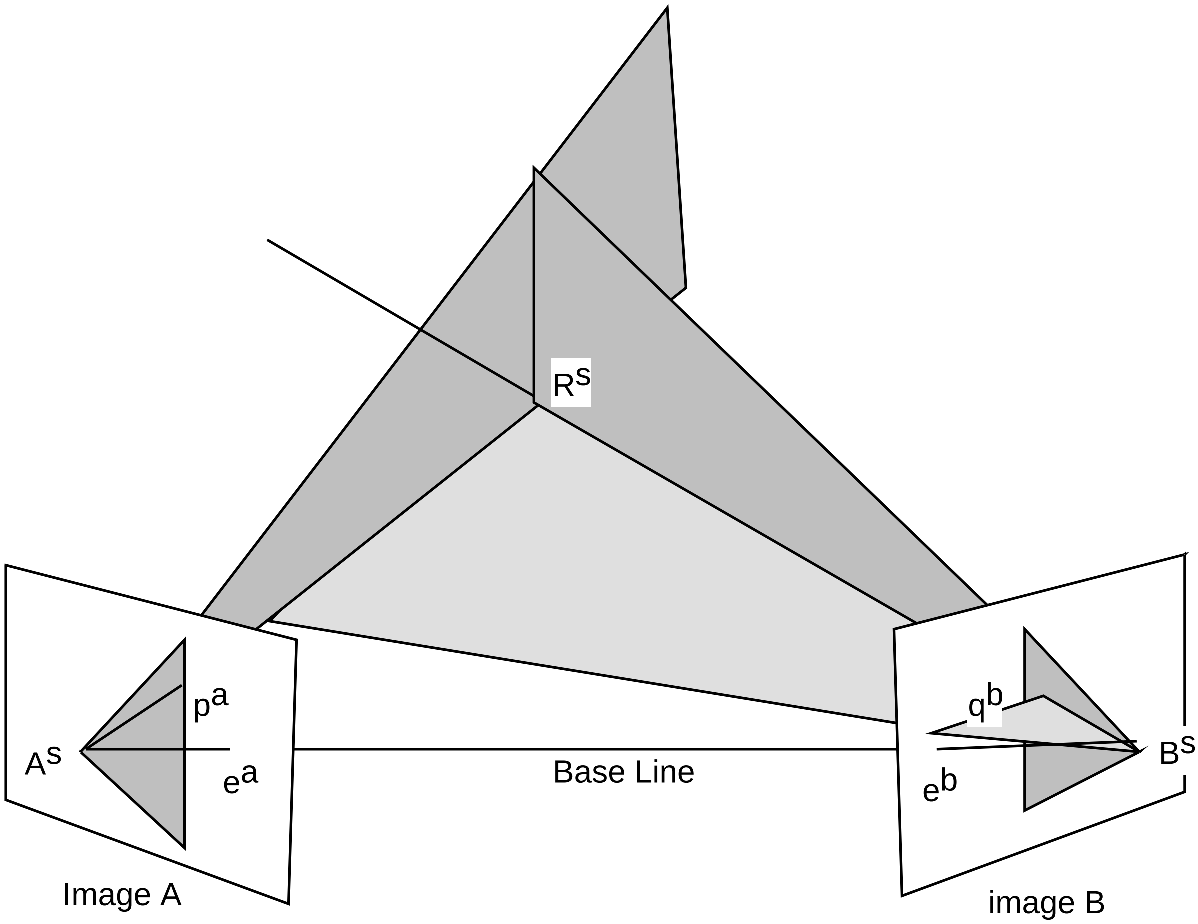
$$\Longrightarrow \quad \frac{B}{Z} = \frac{(x_{ra}-x_{rb})}{F}$$

$\Delta x = x_{ra}-x_{rb}$ est la Disparité

ou bien : $Z = \dfrac{BF}{\Delta x}$

NOTE : La formule de profondeur est INDEPENDANTE des valeurs de x_{rg} et x_{rd}. Ça dépend de Δx!

Le cas generales



La point dans la scène est l'intersection de trois plans.

Soit $\mathbf{M}_s^a$ la matrice de la caméra A. Pour un point inconnu R^s nous observons p^a

En tensorielle :

$$p^a = \mathbf{M}_s^a R^s \qquad \mathbf{M}_s^1 R^s - p^1 \mathbf{M}_s^3 R^s = 0$$
$$\text{et} \qquad \mathbf{M}_s^2 R^s - p^2 \mathbf{M}_s^3 R^s = 0$$

en notation classique $\quad \mathbf{M}_s^1 R^s - i_a \mathbf{M}_s^3 R^s = 0$

et $\quad \mathbf{M}_s^2 R^s - j_a \mathbf{M}_s^3 R^s = 0$

Soit $\mathbf{N}_s^b$ la matrice de la caméra B. Pour le point inconnu R^s nous observons q^b

$$q^b = \mathbf{N}_s^b R^s \qquad \mathbf{N}_s^1 R^s - q^1 \mathbf{N}_s^3 R^s = 0$$
$$\text{et} \qquad \mathbf{N}_s^2 R^s - q^2 \mathbf{N}_s^3 R^s = 0$$

en notation classique $\quad \mathbf{N}_s^1 R^s - i_b \mathbf{N}_s^3 R^s = 0$

et $\quad \mathbf{N}_s^2 R^s - j_b \mathbf{N}_s^3 R^s = 0$

Nous avons quatre équations et trois inconnu $R^s = (x_s, y_S, z_s, 1)$.

On peut trouver le point ${}^{s}P_L$ avec deux équations du A et une du B.

$$\begin{pmatrix} \mathbf{M}_s^1 - i_a\ \mathbf{M}_s^3 \\ \mathbf{M}_s^2 - j_a\ \mathbf{M}_s^3 \\ \mathbf{N}_s^1 - i_b\ \mathbf{N}_s^3 \end{pmatrix} R^s = 0$$

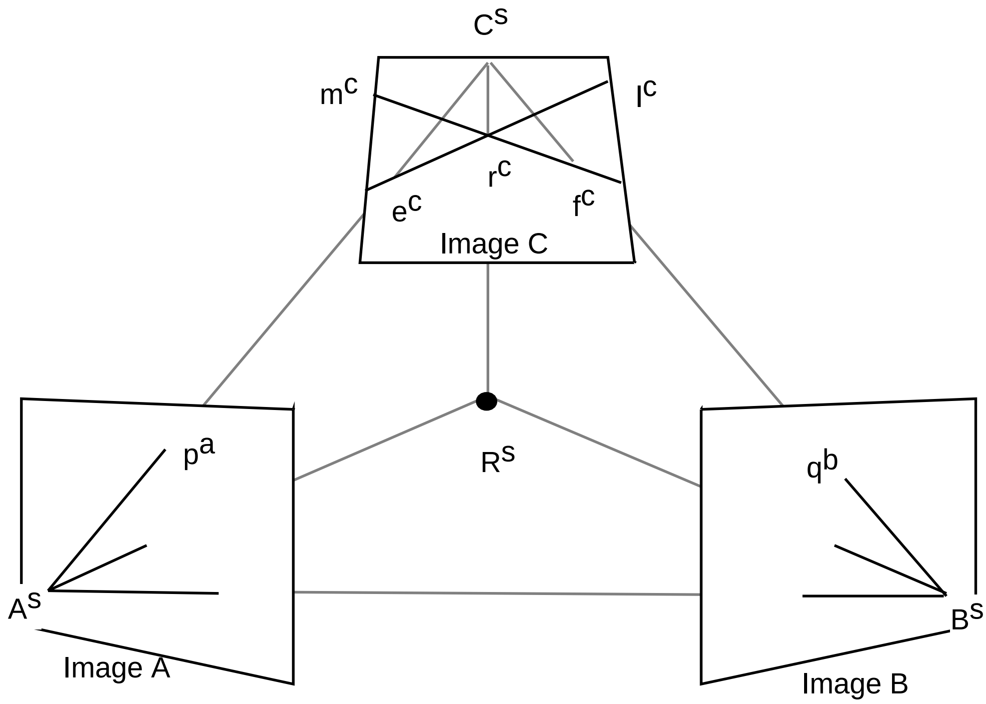
ou bien

$$\begin{pmatrix} \mathbf{M}_1^1 - i_a\ \mathbf{M}_1^3 & \mathbf{M}_2^1 - i_a\ \mathbf{M}_2^3 & \mathbf{M}_3^1 - i_a\ \mathbf{M}_3^3 \\ \mathbf{M}_1^2 - j_a\ \mathbf{M}_1^3 & \mathbf{M}_2^2 - j_a\ \mathbf{M}_2^3 & \mathbf{M}_3^2 - j_a\ \mathbf{M}_3^3 \\ \mathbf{N}_1^1 - i_b\ \mathbf{N}_1^3 & \mathbf{M}_2^1 - i_b\ \mathbf{N}_2^3 & \mathbf{N}_3^1 - i_b\ \mathbf{N}_3^3 \end{pmatrix} \begin{pmatrix} X_s \\ Y_s \\ Z_s \end{pmatrix} = \begin{pmatrix} \mathbf{M}_4^1 - i_a\ \mathbf{M}_4^3 \\ \mathbf{M}_4^2 - j_a\ \mathbf{M}_4^3 \\ \mathbf{N}_4^1 - i_b\ \mathbf{N}_4^3 \end{pmatrix}$$

et donc

$$\begin{pmatrix} X_s \\ Y_s \\ Z_s \end{pmatrix} = \begin{pmatrix} \mathbf{M}_1^1 - i_a\ \mathbf{M}_1^3 & \mathbf{M}_2^1 - i_a\ \mathbf{M}_2^3 & \mathbf{M}_3^1 - i_a\ \mathbf{M}_3^3 \\ \mathbf{M}_1^2 - j_a\ \mathbf{M}_1^3 & \mathbf{M}_2^2 - j_a\ \mathbf{M}_2^3 & \mathbf{M}_3^2 - j_a\ \mathbf{M}_3^3 \\ \mathbf{N}_1^1 - i_b\ \mathbf{N}_1^3 & \mathbf{M}_2^1 - i_b\ \mathbf{N}_2^3 & \mathbf{N}_3^1 - i_b\ \mathbf{N}_3^3 \end{pmatrix}^{-1} \begin{pmatrix} \mathbf{M}_4^1 - i_a\ \mathbf{M}_4^3 \\ \mathbf{M}_4^2 - j_a\ \mathbf{M}_4^3 \\ \mathbf{N}_4^1 - i_b\ \mathbf{N}_4^3 \end{pmatrix}$$

Le Transport Epipolair



Soit un Troisième Image C.

L'Estimation de la fondamental F_{ac} donne : $l_c = \mathbf{F}_{ac}\, p^a$

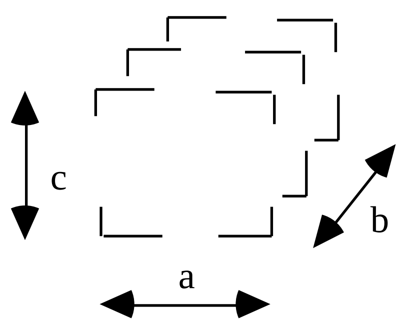
L'Estimation de la fondamental F_{ac} donne : $m_c = \mathbf{F}_{bc}\, q^b$

mais $r^c = l_c \times m_c = (\mathbf{F}_{ac}\, p^a) \times (\mathbf{F}_{bc}\, q^b)$

La notation tensoriel permet de faire :

$$r^c = (F_{ac} \times F_{bc})\, p^a q^b = \mathbf{T}^c_{ab}\, p^a q^b$$

Le produit croisé des deux tenseur d'ordre 2 est une tenseur d'ordre 3.



$$\mathbf{T}^c_{ab} = (F_{ac} \times F_{bc})$$

Le tenseur $\mathbf{T}^c_{ab}$ donne $r^c = \mathbf{T}^c_{ab}\, p^a q^b$