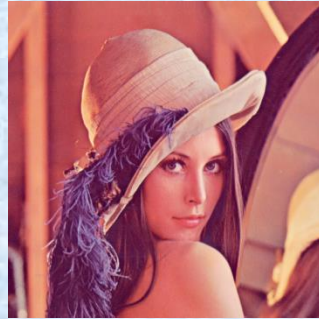


Licence 3 – Semestre 1 – Université de Bourgogne



Traitement des Images II

Jean-Luc Baril

Barjl@u-bourgogne.fr

<http://jl.baril.u-bourgogne.fr/>

Plan du cours

Chapitre 1

- **Qu'est ce que la vision?**
 - + Vision humaine
 - + Vision par ordinateur
- **Rappels d'optique géométrique (Foyer, lentilles)**
- **Instruments optiques (œil, caméra, app. Photo)**

Chapitre 2

- **Problème général de la vision du relief**
 - + Cas où on dispose d'une seule image
 - + Cas où on dispose de plusieurs images
 - + Bilan des différents domaines de vision

Plan du cours

Chapitre 3

- **Cas élémentaire de stéréovision**
 - + Le calibrage : phase 1
 - + L'appariement : phase 2
 - + La triangulation : phase 3

Chapitre 4

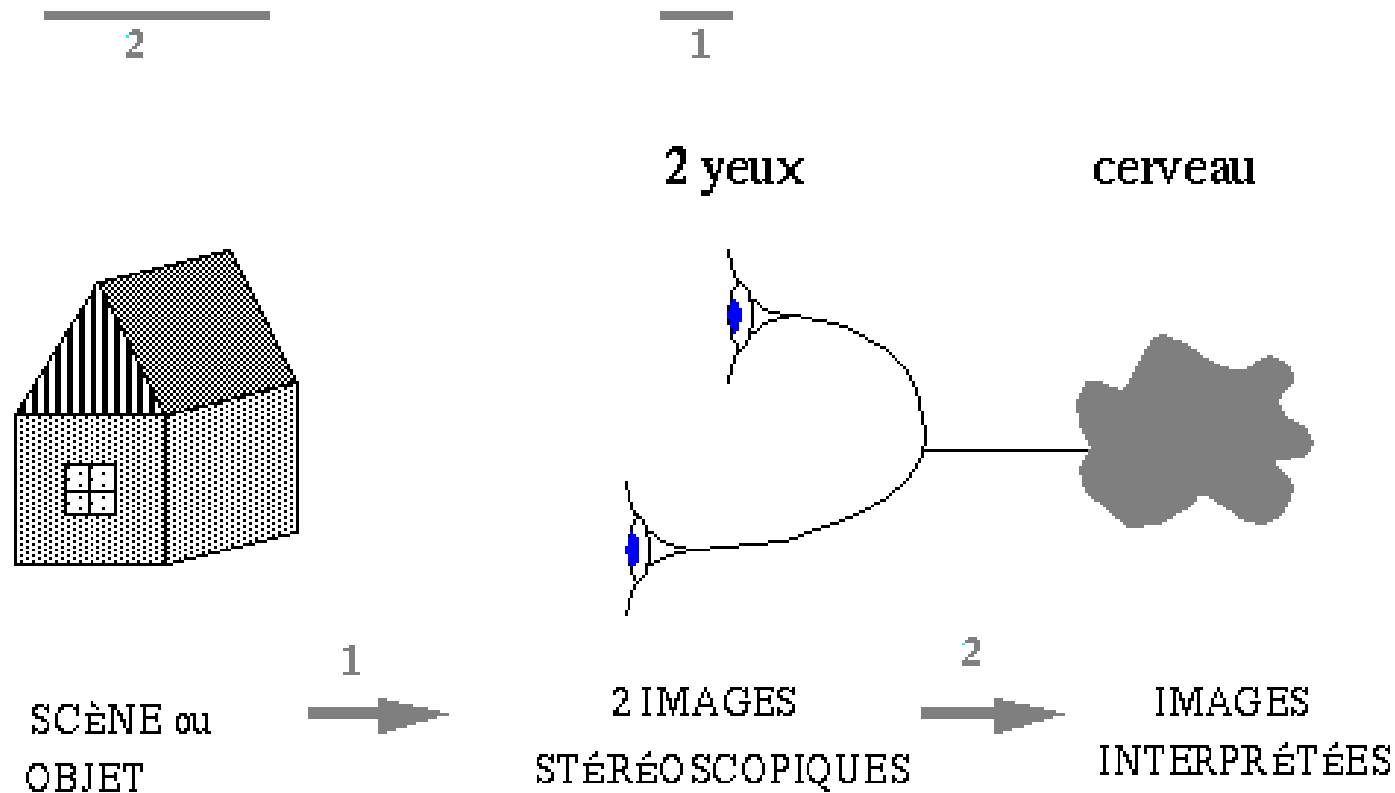
- **Cas général de la stéréovision**
 - + Le calibrage
 - + L'appariement
 - + La triangulation
 - + Deux exemples

1. Qu'est-ce que la vision ?

1.1. La vision humaine.

1.1.a. Définition.

“Perception par l’organe de la vue” (Larousse)



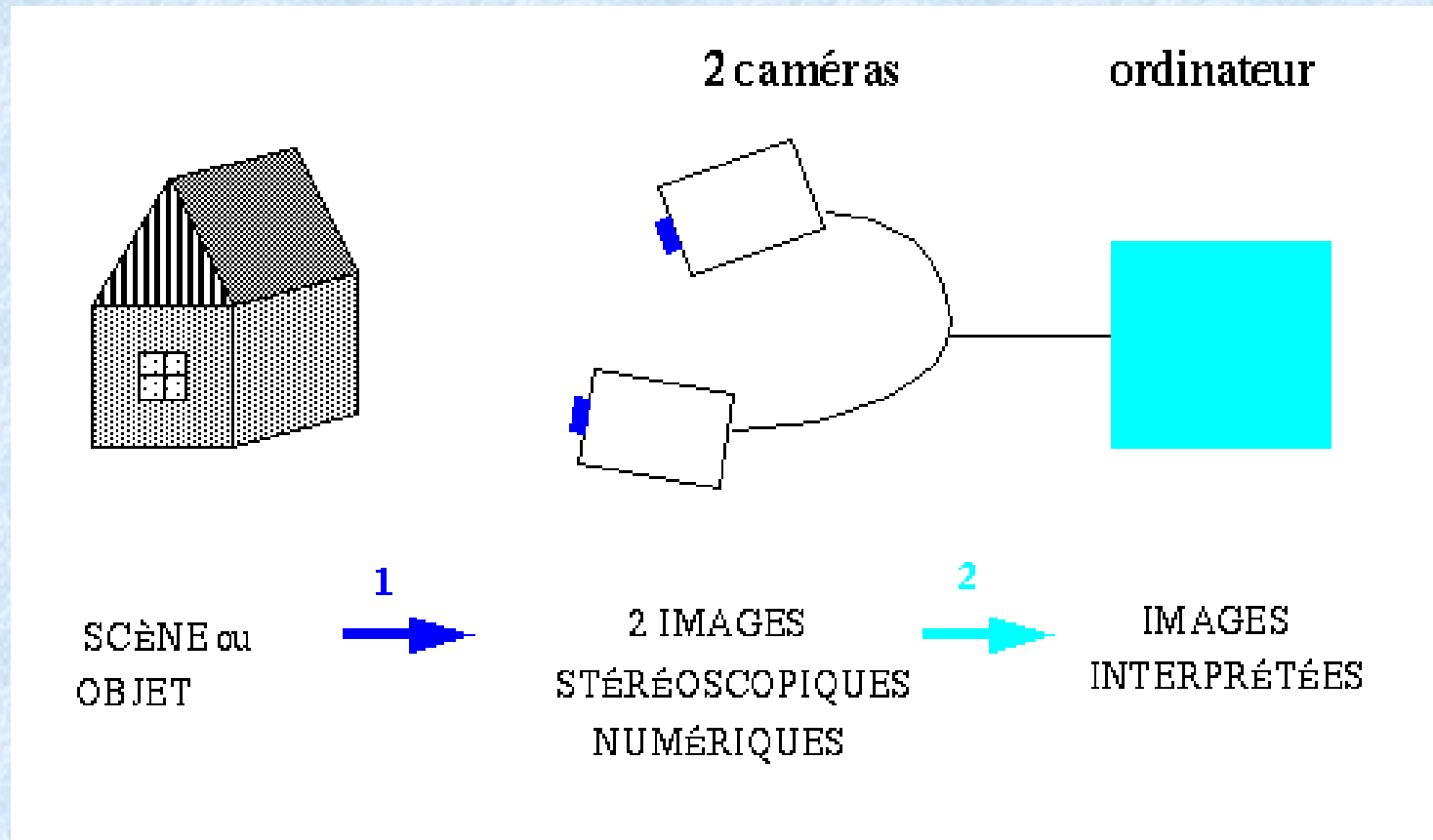
vision = **vue** + perception

1.1.b. Les quatre fonctions de la vision humaine.

- Vision des formes (1 oeil \sim 2 yeux).
- **Vision des distances (1 oeil \ll 2 yeux) -> ce qui nous intéresse dans ce cours.**
- Vision des couleurs (1 oeil \sim 2 yeux).
- Vision des mouvements (1 oeil \ll 2 yeux).

1.2. La vision par ordinateur.

1.2.a. Définition.



On essaie de reproduire la vision humaine pour donner de l'autonomie à un robot équipé de 2 caméras

1.2.b. Les différents domaines de la vision par ordinateur.

- Reconnaissance des formes (1 image ~ 2 images)
- Reconnaissance du relief (1 image \ll 2 images) **objet de ce cours.**
- Couleur (1 image ~ 2 images)
- Mouvement (1 image \ll 2 images)

1.3. Vision humaine et vision par ordinateur.

La vision humaine est :

- ce qui motive les recherches en vision par ordinateur ("si l'homme y parvient ...") ;
- le modèle de vision vers lequel la vision par ordinateur essaie de tendre, car :
 - **l'oeil est une caméra** très performante ;
 - **le cerveau est un ordinateur** très performant.

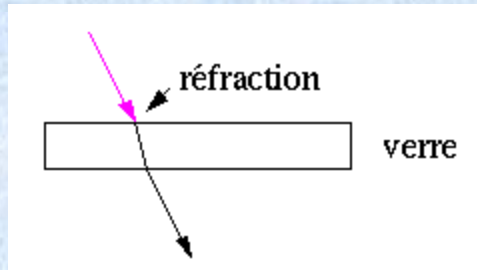
Décomposition en deux étapes de la vision

- : formation d'images numériques
- : interprétation d'images numériques

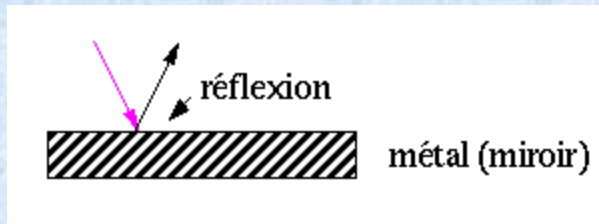
2. Rappels d'optique géométrique.

2.1. Les trois types d'objets vis-à-vis de la réémission de la lumière.

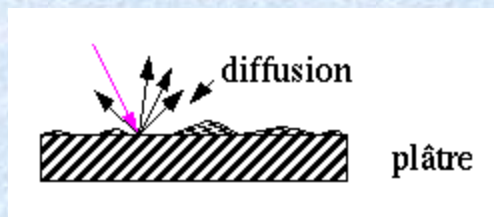
- Objet transparent ou "réfractant" :



- Objet parfaitement "réfléchissant" :

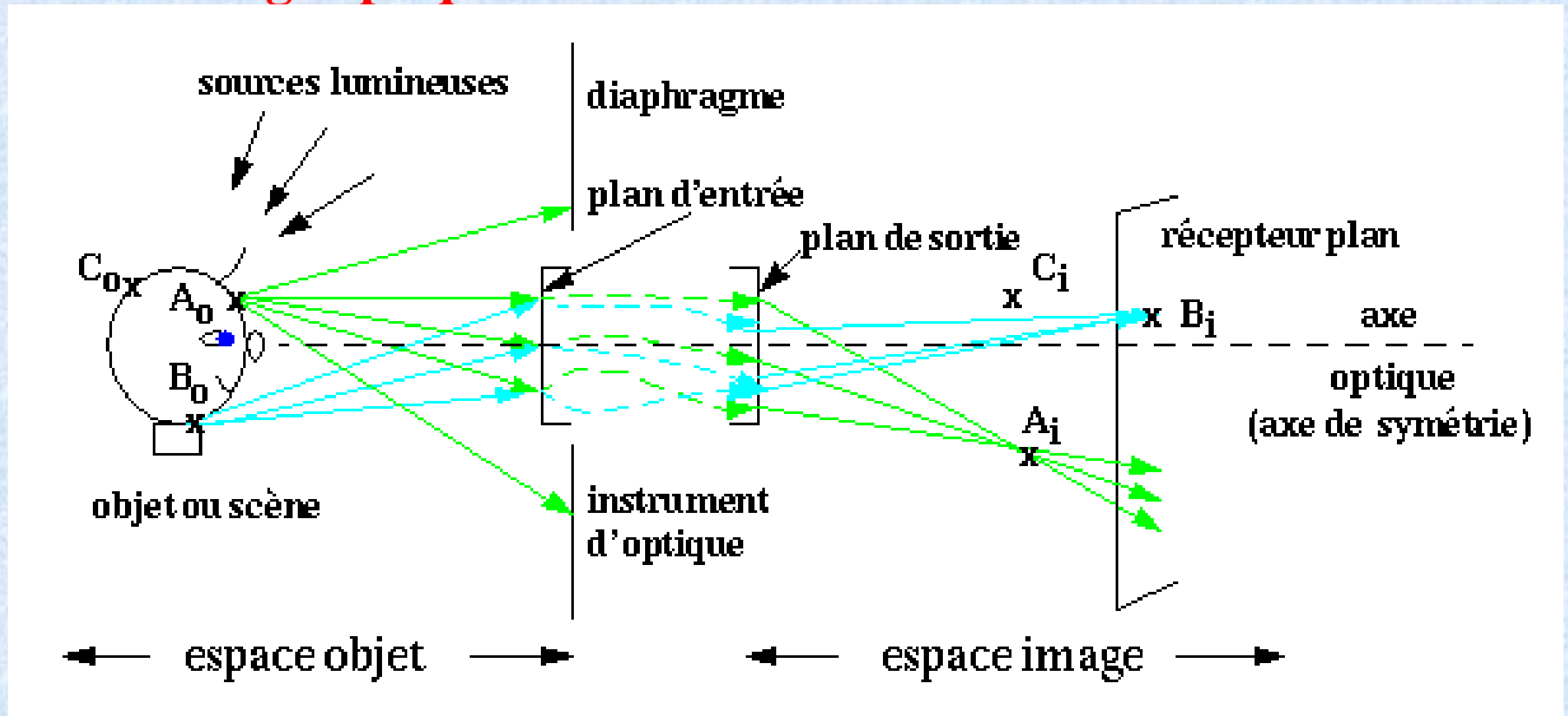


- Objet parfaitement "diffusant" :



2.2. Le stigmatisme : propriété fondamentale n° 1.

2.2.a. Montage optique usuel.



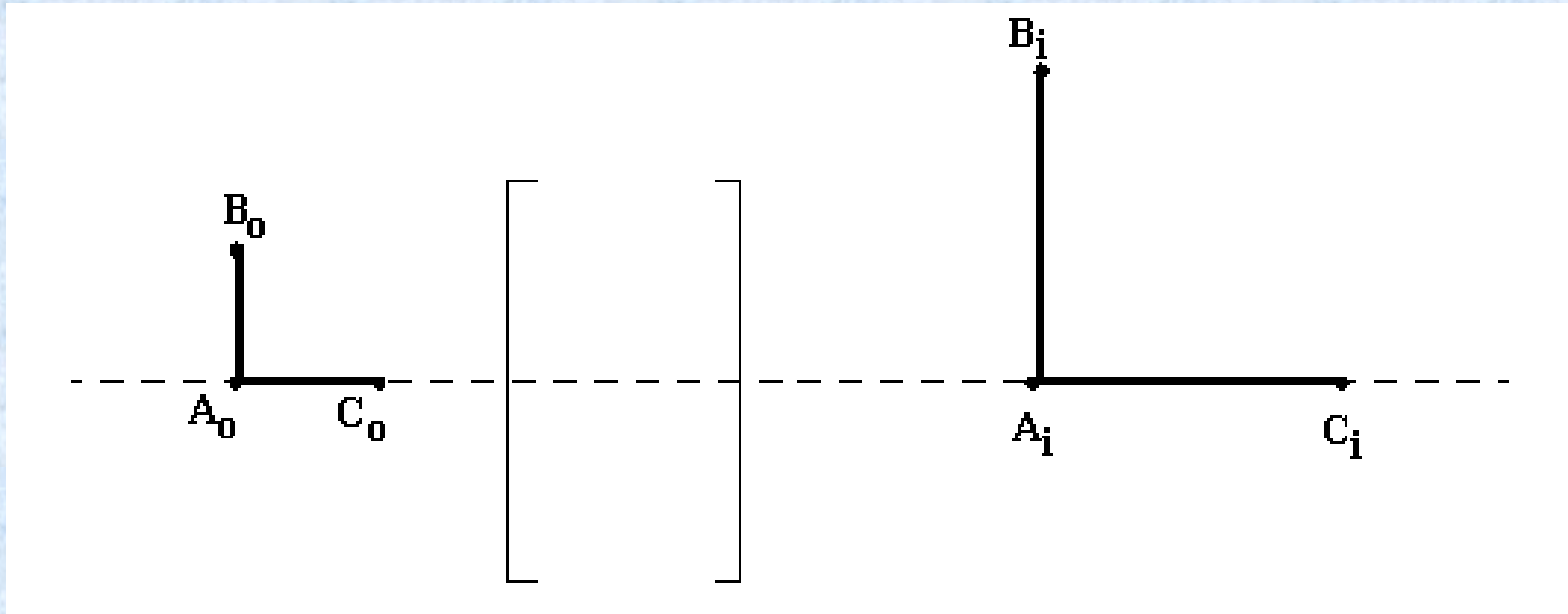
Tous les rayons réémis qui entrent dans le système optique en ressortent, après quelques réfractions, et se croisent **en un même point A_i** .

Cette propriété fondamentale s'appelle le "*stigmatisme*".

S'il y a stigmatisme, il y a une **bijection** entre les points de l'espace objet et les points de l'espace image.

2.3. La formation des images.

2.3.a. Grandissements.

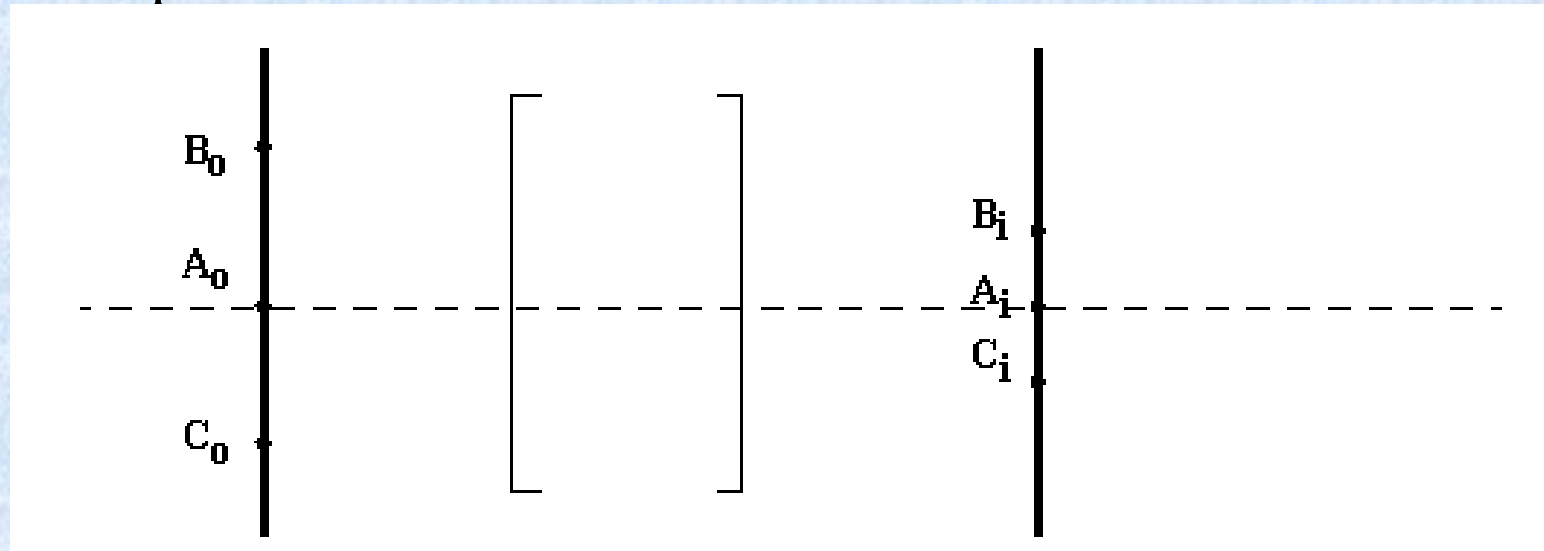


• Grandissement transversal au point A_0 : $gt = \frac{A_i B_i}{A_0 B_0}$

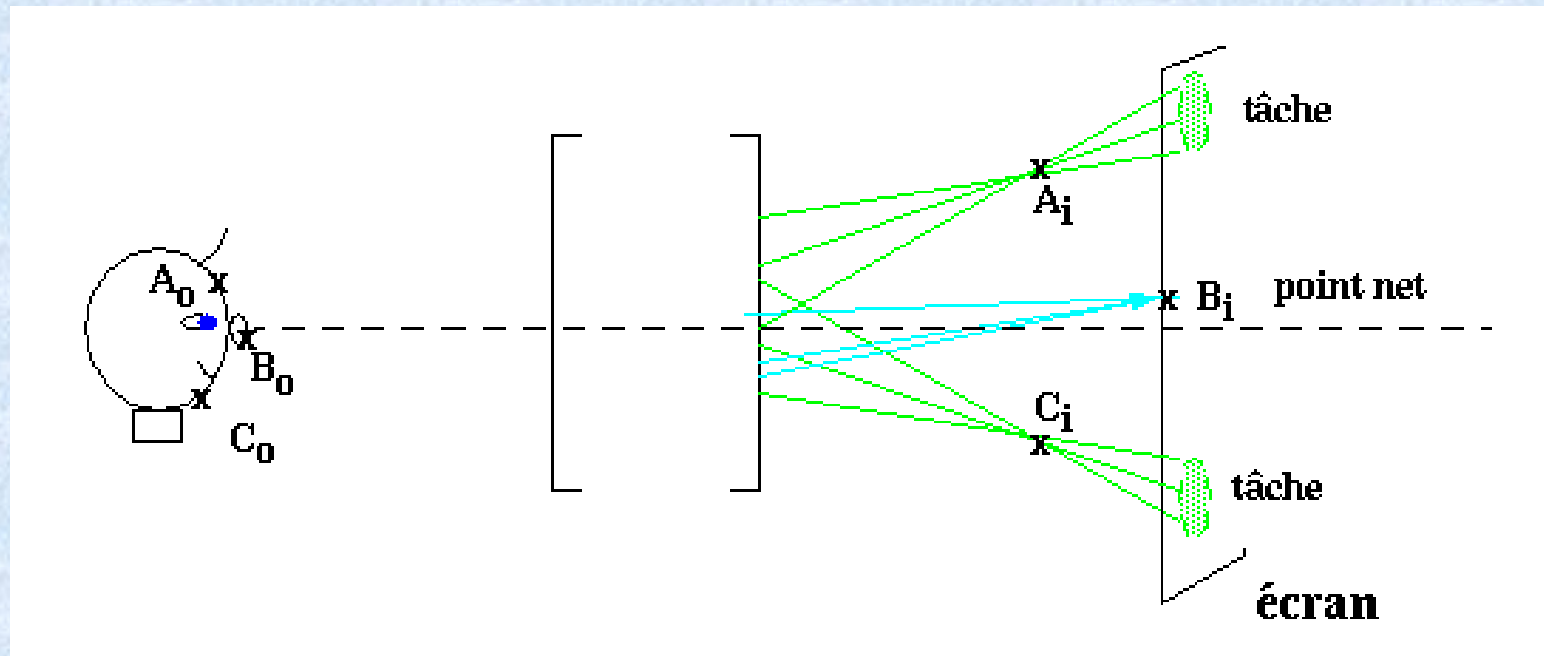
• Grandissement axial au point A_0 : $ga = \frac{A_i C_i}{A_0 C_0}$

2.3.b. Où placer le récepteur ?

Si l'objet est plan :



Sinon



2.3.c. Relation entre ces deux grandissements : propriété fondamentale n° 2.

Les deux grandissements sont liés par la relation fondamentale suivante :

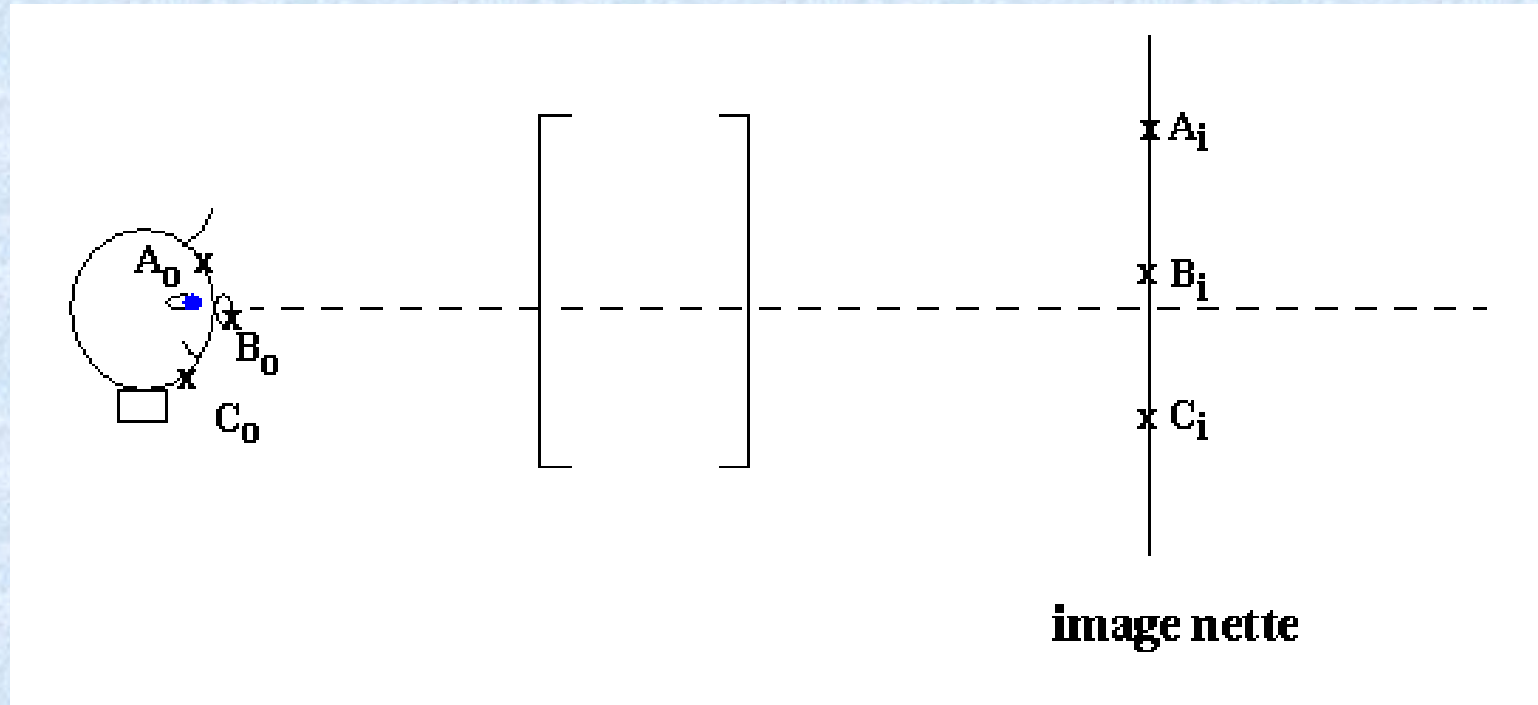
$$ga = gt^2$$

Cette relation est à la base de la formation des images car :

$$\text{Si } gt \ll 1, \text{ alors } ga = \frac{ga}{gt} \ll 1$$

Donc, un objet non plan a une image quasi-plane, et il suffit de placer le récepteur dans ce quasi-plan, et on aura une image quasiment nette de l'objet.

Il y a alors bijection entre points objets visibles et points du plan image :



On dit "qu'on perd une dimension" ou "qu'on passe du 3D au 2D".

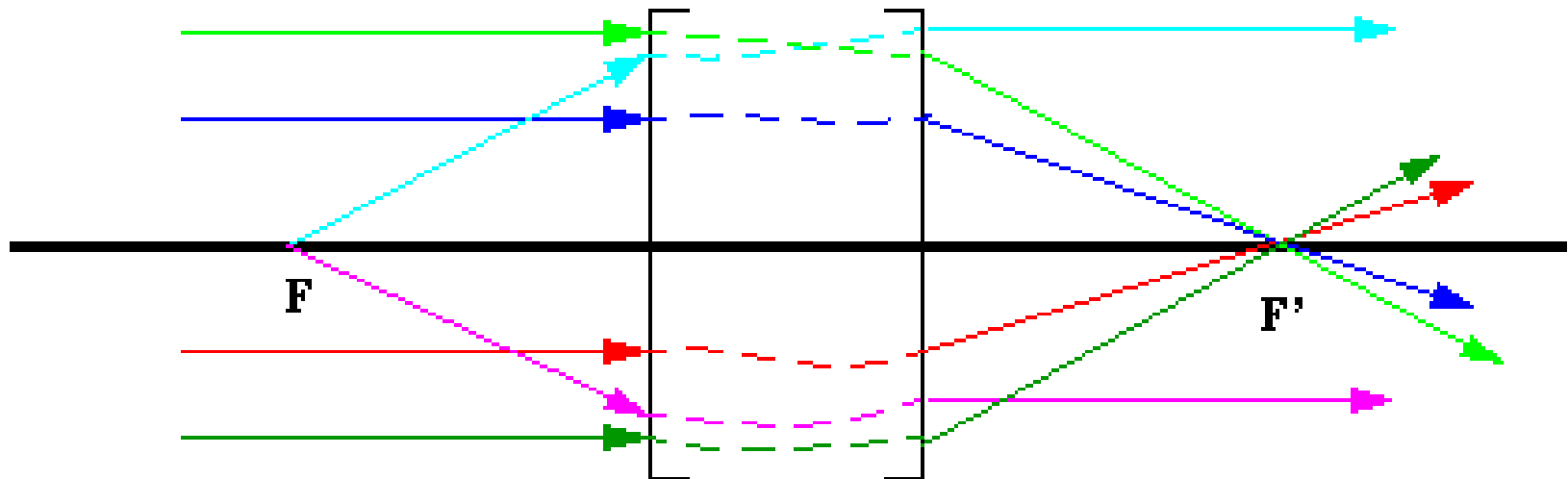
La reconnaissance du relief consiste à réaliser l'opération inverse de la formation des images : on souhaite "retrouver la 3^{ème} dimension", ou encore "passer du 2D au 3D".

2.4. Points particuliers d'un instrument optique.

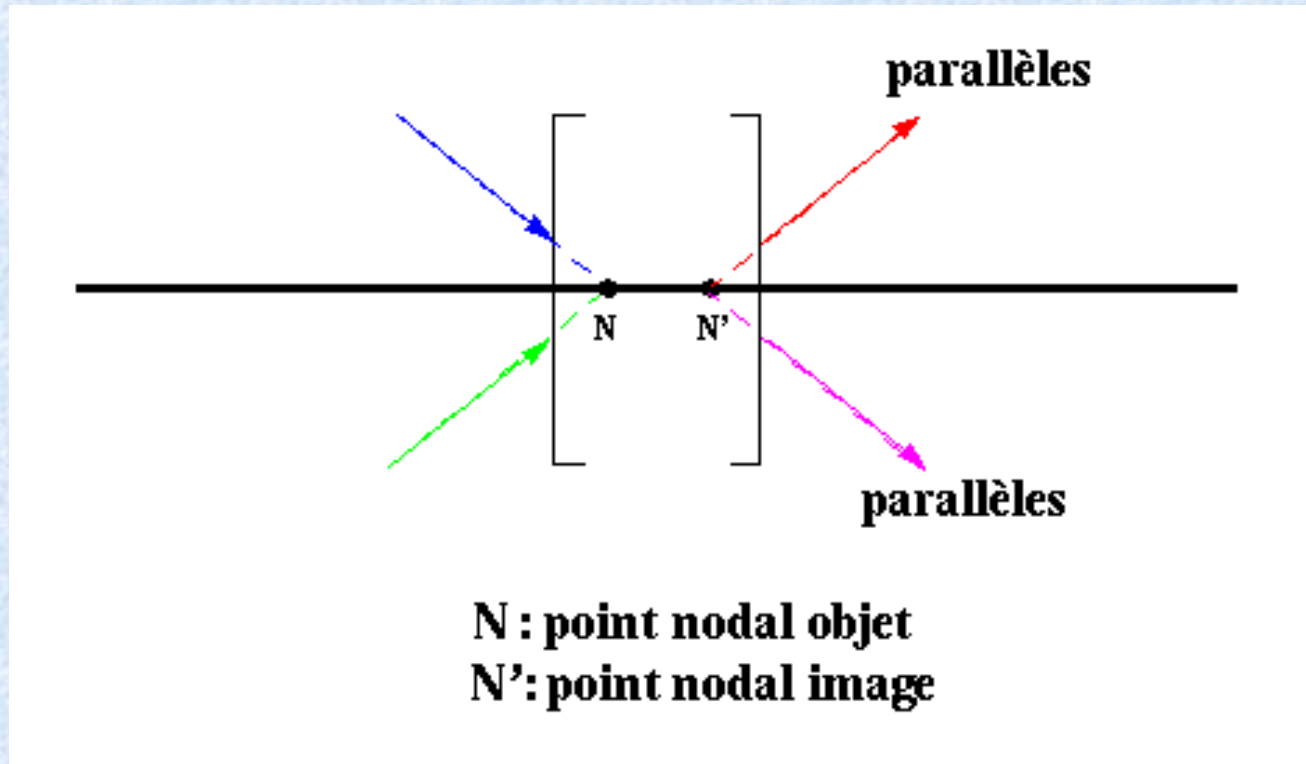
Pour tout instrument d'optique, il existe différents types de points remarquables :

- Les foyers (sur l'axe optique).
- Les points nodaux (sur l'axe optique).
- Les plans principaux.

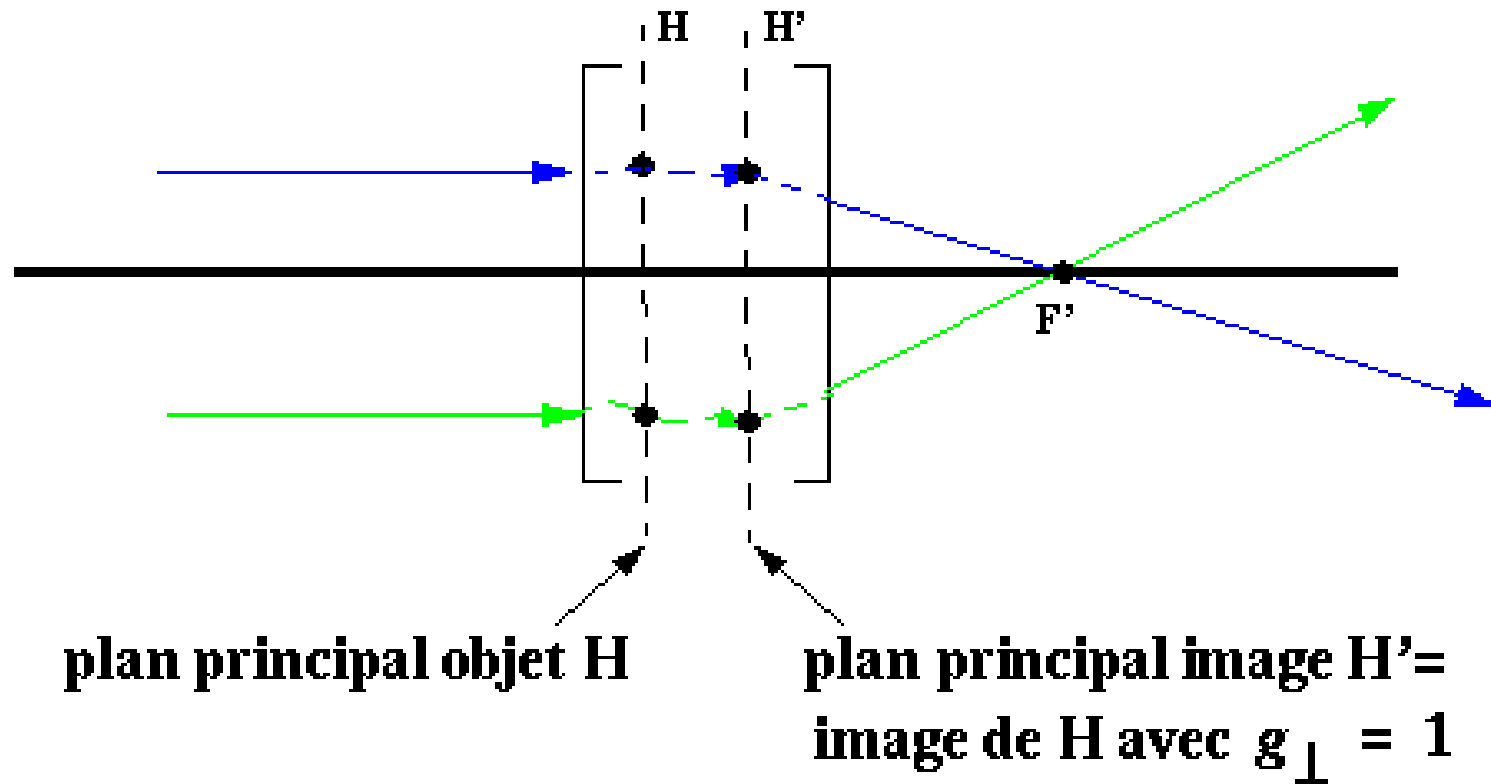
2.4.a. Foyers.



2.4.b. Points nodaux.

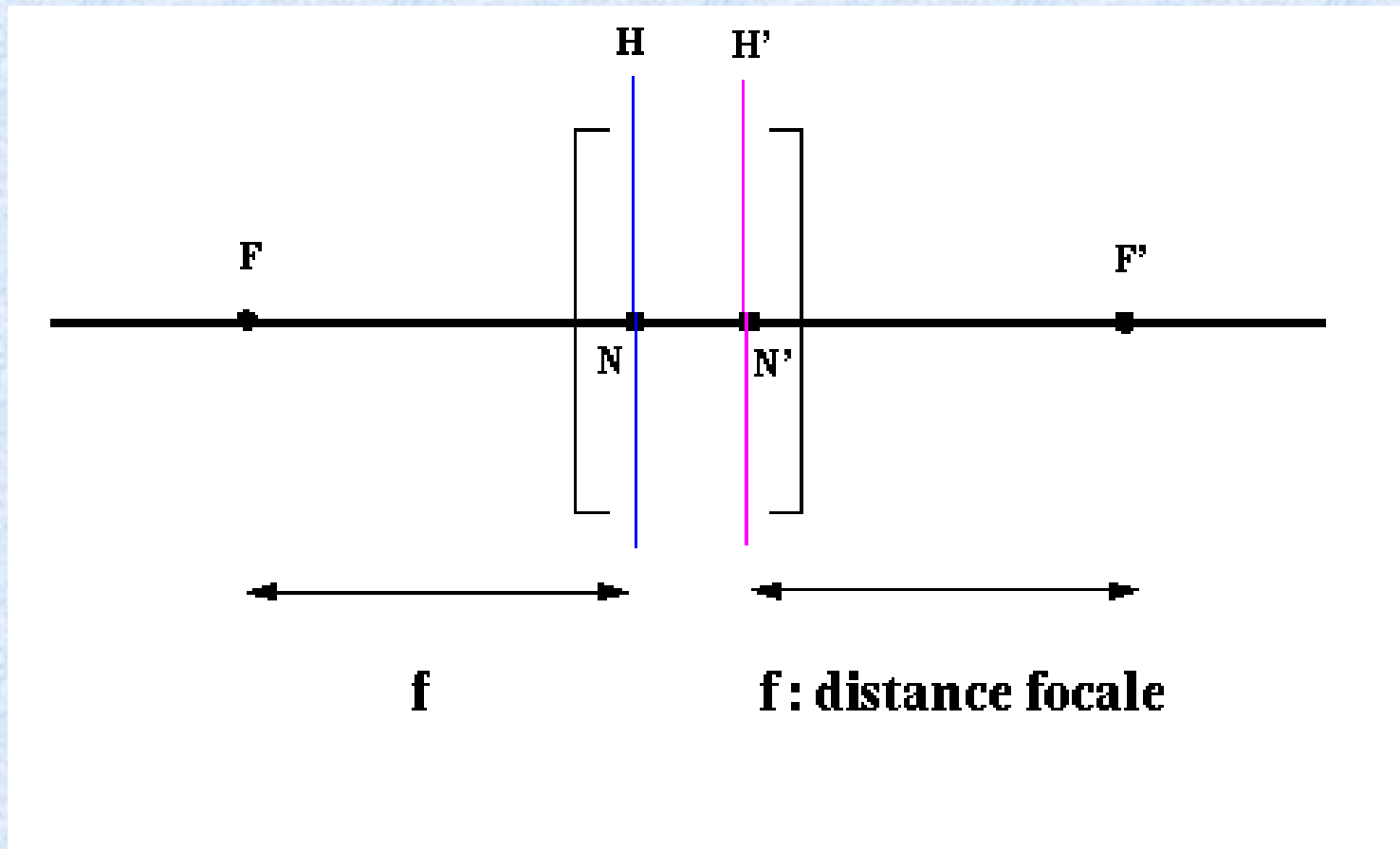


2.4.c. Plans principaux.



Remarque : $N \in H$
 $N' \in H'$

2.4.d. Distance focale.

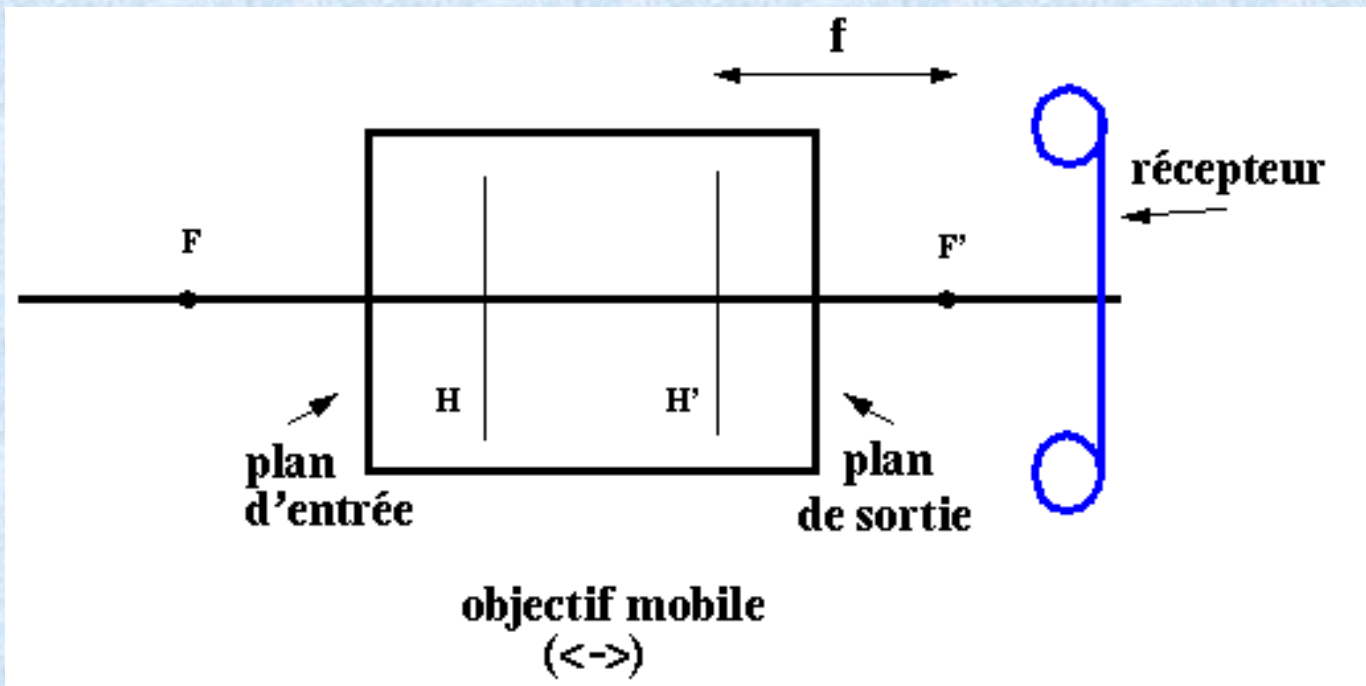


3. Quelques instruments d'optique usuels.

3.1. L'oeil.

3.2. Appareil photographique et caméra CCD (Coupled Charge Device).

3.2.a. Schéma.



3.2.b. Les objectifs.

Dans un appareil photographique, l'instrument d'optique est en fait l'objectif, composé de 7, 8 ou 9 lentilles, qui est mobile par rapport au plan du récepteur. Le réglage de cette position s'appelle la "mise au point" ("autofocus" = mise au point automatique).

Remarque :

Dans l'appareil, la mise au point s'effectue sans changer la focale, contrairement à l'oeil. La distance focale standard est de 50 mm, le grandissement transversal < 1 .

3.2.c. Cas particuliers d'objectifs.

Zoom : objectif à distance focale variable, qui change sans changer la mise au point (assez compliqué à réaliser en respectant cette contrainte) ;

Téléobjectif : pour une scène éloignée, on veut un grandissement grand. Il faut donc augmenter la distance focale .

3.2.d. Les récepteurs.

- **Appareil photo** : film photo -> signal analogique.
- **Caméra CCD** : réseau de cellules photosensibles -> signal discret
 $E_{i,j} \geq 0$ ("niveau de gris") si image noir et blanc. Si image en couleurs, 3 "canaux" : RVB.

Chapitre 2

Problème général de la vision du relief.

Dorénavant, nous ne nous intéresserons plus qu'à la reconnaissance du relief d'une scène à partir d'une ou de plusieurs images (numériques) de cette scène.

Les images seront supposées représentées en niveaux de gris (allant du noir au blanc).

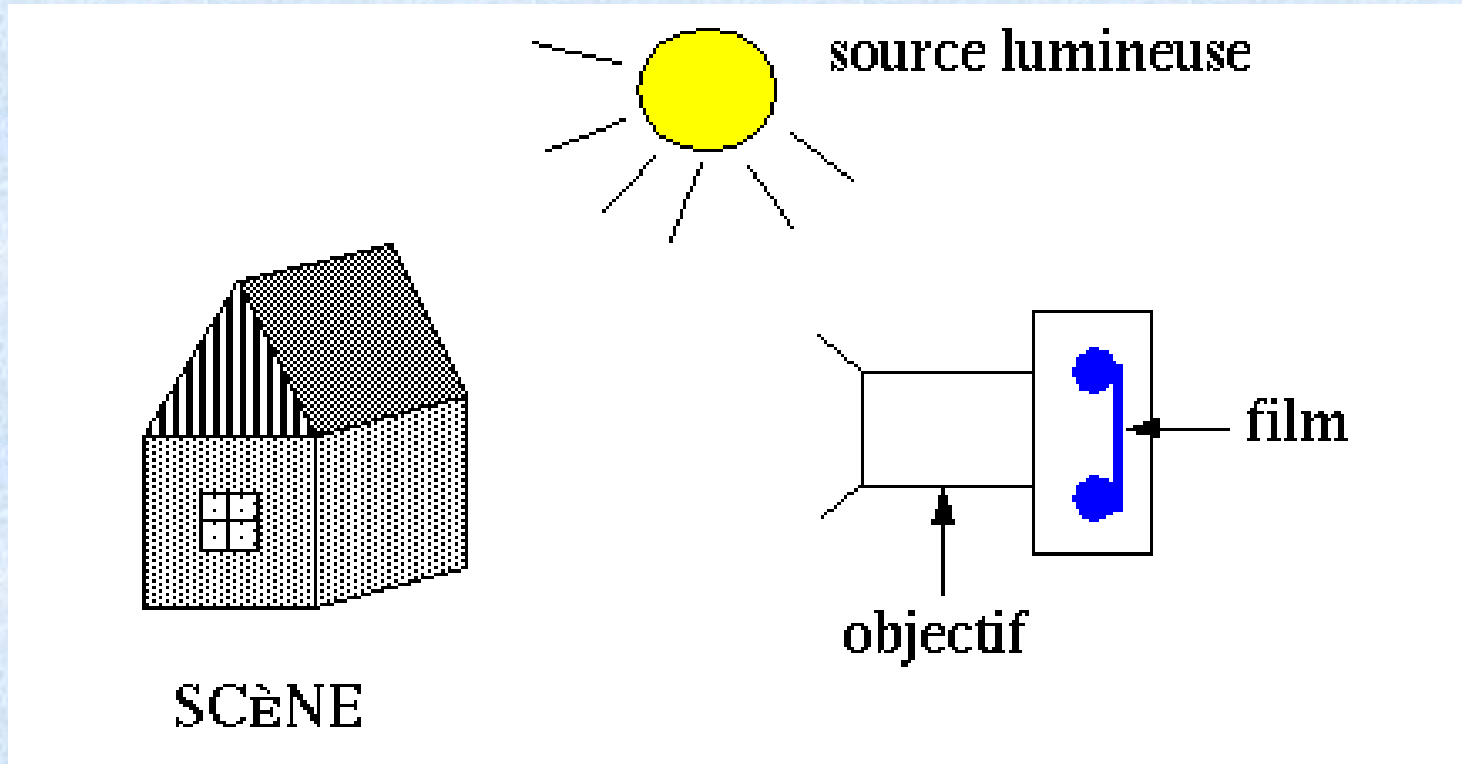
Soit une image ayant **M** lignes et **N** colonnes. Chaque pixel de l'image a un niveau de gris $E_{i,j}$ tel que :

- pour **i** appartenant à $[1,M]$ et **j** à $[1,N]$, $E_{i,j}$ est positif ou nul ;
- $E_{i,j}$ est de type **integer, float, double, ...**

1. Problème général de la vision du relief.

1.1. Les trois éléments en présence.

- Scène (ou objet) dont on veut connaître le relief.
- Sources lumineuses.
- Appareil photo (ou observateur).



1.2. But de la reconnaissance du relief.

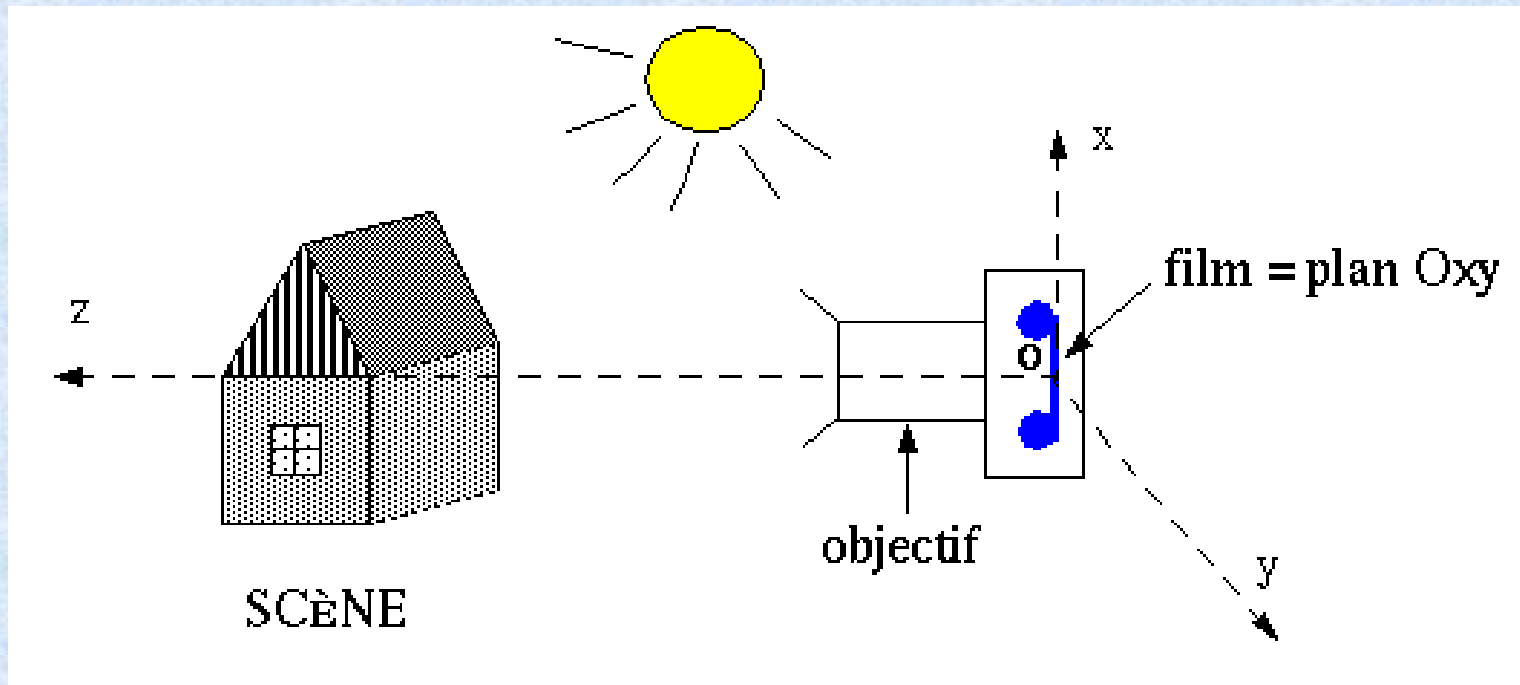
On suppose que l'image est **nette** (c'est-à-dire qu'elle ne présente pas de flou de mise au point), et on veut connaître le relief de la partie de la scène à la fois éclairée par les sources lumineuses et visible par l'appareil photo (ou par l'observateur).

En général, on ne connaît ni la position, ni les caractéristiques photométriques des sources lumineuses.

2. Cas où on ne dispose que d'une seule image.

Ce cas est celui de la "monovision".

2.1. Formulation du problème.



On cherche à calculer z pour tout point de l'image. En fait, on calcule z à une constante additive près

2.2. Les différentes méthodes.

On peut utiliser différents critères :

- flou de la mise au point : *marginal* ;
- texture : *champ d'application limité* ;
- ombres portées : *marginal* ;
- contours : *marginal* ;
- ombres propres : *cas le plus général* ("shape from shading").

Ces méthodes présentent toutes un problème majeur : beaucoup d'hypothèses leur sont nécessaires, et elles sont donc peu fiables. Elles ne sont pas utilisées industriellement (par exemple en robotique). Toutes les méthodes utilisées en robotique utilisent plusieurs images.

3. Cas où on dispose de plusieurs images.

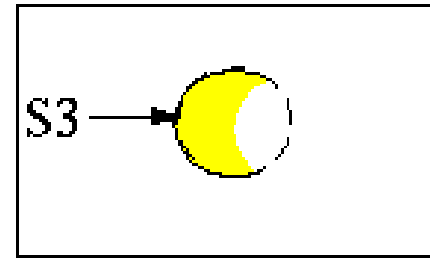
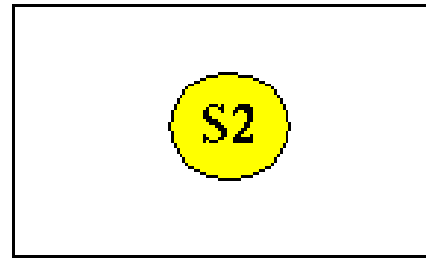
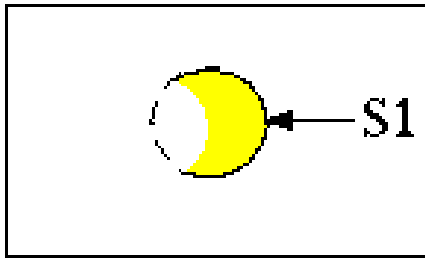
Différentes situations sont possibles :

- **Cas 1** : la scène et l'appareil sont fixes, les sources lumineuses se déplacent ("**stéréophotométrie**").
- **Cas 2** : les sources lumineuses et l'appareil sont fixes, la scène se déplace ("**mouvement**").
- **Cas 3** : la scène et les sources lumineuses sont fixes, l'observateur se déplace ("**stéréovision**").

3.1. Cas 1 : stéréophotométrie.

On prend une photo avec une source lumineuse S1, puis une photo avec une source lumineuse S2, etc... La scène est fixe.

Exemple de la Lune :

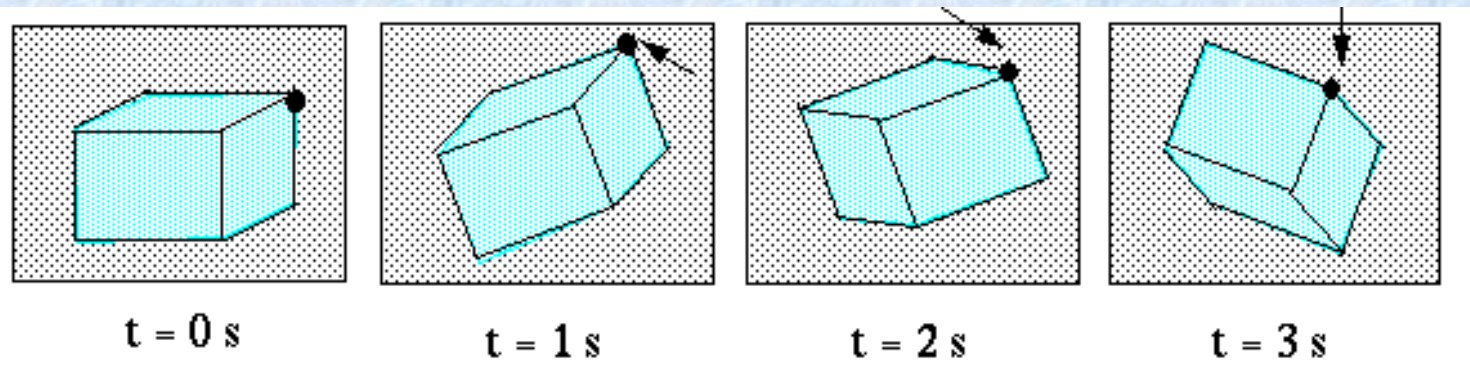


Problème :

Si la scène bouge, ce qui est souvent le cas en robotique, il faut prendre deux photos simultanément avec deux éclairages différents, ce qui n'est pas toujours facile à réaliser.

3.2. Cas 2 : mouvement.

On prend plusieurs photos successives d'une scène mobile, mais cela ne rajoute pas beaucoup d'information vis-à-vis de la reconnaissance du relief. Il présente cependant un point commun avec la vision du relief, car il faut réaliser la mise en correspondance d'un même point physique sur une séquence d'images :

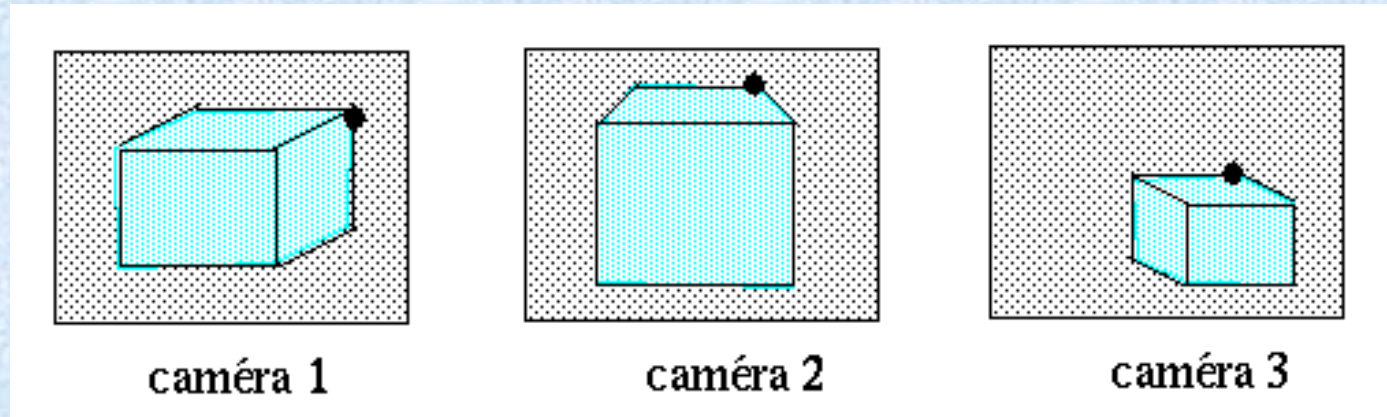


Problème :

Le but n'est pas le calcul du relief, mais celui du mouvement. Le **cas 2** correspond donc à l'analyse du mouvement en **monovision**.

3.2. Cas 3 : stéréovision.

On change la position de la caméra entre les différentes prises de vue. On peut aussi utiliser plusieurs caméras et prendre les photos simultanément, ce qu'on ne pouvait pas faire dans les cas 1 et 2. Cela est d'ailleurs nécessaire pour l'analyse du mouvement en stéréovision.



En général, on se trouve dans l'un des deux cas suivants :

- 2 caméras : stéréovision **binoculaire** ;
- 3 caméras : stéréovision **trinoculaire**.

Difficultés :

- Si on utilise plusieurs caméras, elles ne sont pas caractérisées par les mêmes paramètres (exemples : g_a , g_t ,.....).

- **Phase 1 = calibrage** : il faudra maîtriser les paramètres des caméras.

- **Phase 2 = appariement** : mise en correspondance des pixels correspondant à un même point physique (comme en vision du mouvement).

- **Phase 3 = triangulation** : calcul du relief d'une paire de pixels appariés.

La phase la plus difficile à réaliser est la phase 2. Nous y reviendrons dans le chapitre 3.

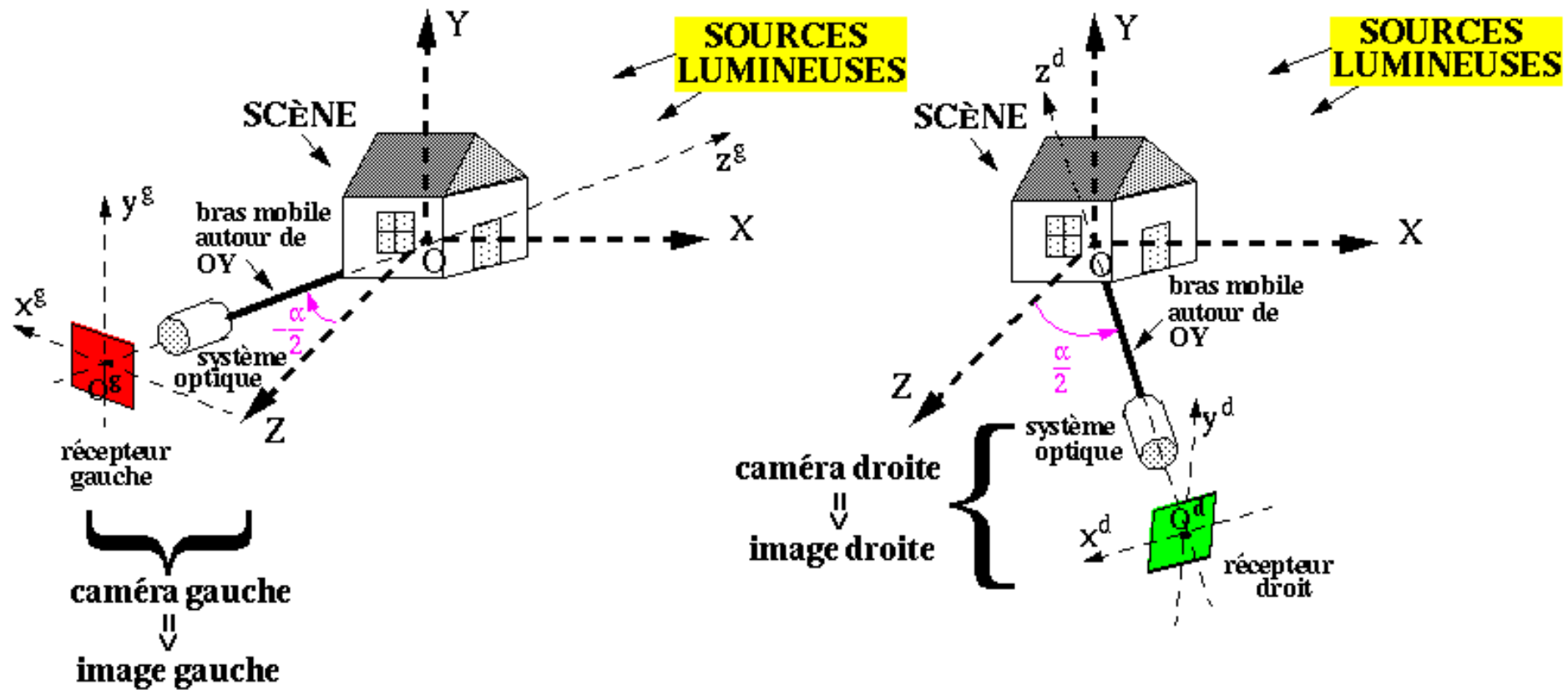
1. Cas élémentaire de stéréovision.

Ce cas survient lorsqu'on n'utilise qu'une caméra, ce qui impose à la scène d'être **statique**.

L'intérêt de n'utiliser qu'une caméra est que les paramètres de la caméra sont les mêmes entre les images gauche et droite. En d'autres termes, la phase de calibrage est nettement simplifiée.

Dans le cas général, par contre, ces paramètres diffèrent, et doivent être déterminés les uns par rapport aux autres.

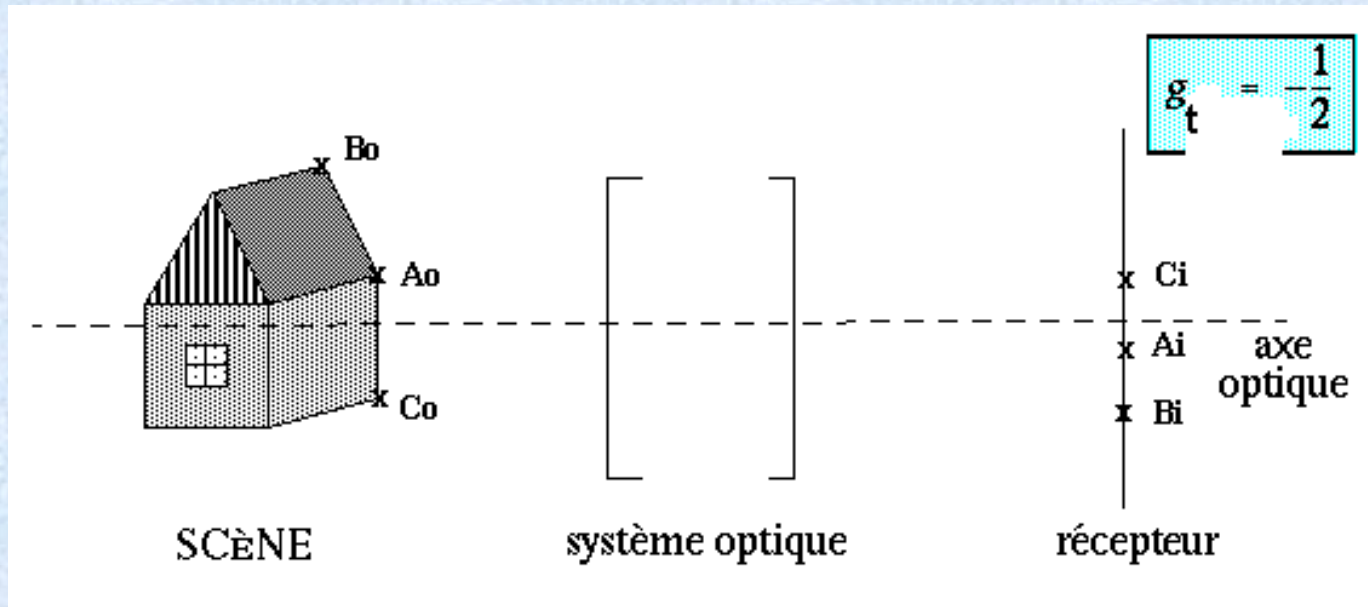
1.1. Réalisation pratique.



1.2. Correspondance entre points objets et pixels d'une image.

Si l'image est **nette**, la correspondance entre les points objets et les **pixels** des images peut être considérée comme la succession de :

- Une **homothétie** de rapport g_t (éventuellement négatif) et de centre O ;
- Une **projection** sur le plan du récepteur ;



2. Le calibrage : phase 1.

Il va être très simple ici.

2.1. Relation mathématique entre points objets et pixels.

Soit un point objet $P = \begin{pmatrix} X_p \\ Y_p \\ Z_p \end{pmatrix}$ visible sur les deux images.

On cherche $I^g = \begin{pmatrix} x_p^g \\ y_p^g \end{pmatrix}$ et $I^d = \begin{pmatrix} x_p^d \\ y_p^d \end{pmatrix}$, car $z_p^g = z_p^d = 0$

Une remarque évidente est que :
(3.1)

$$y_p^g = y_p^d = g_{\perp} Y_p$$

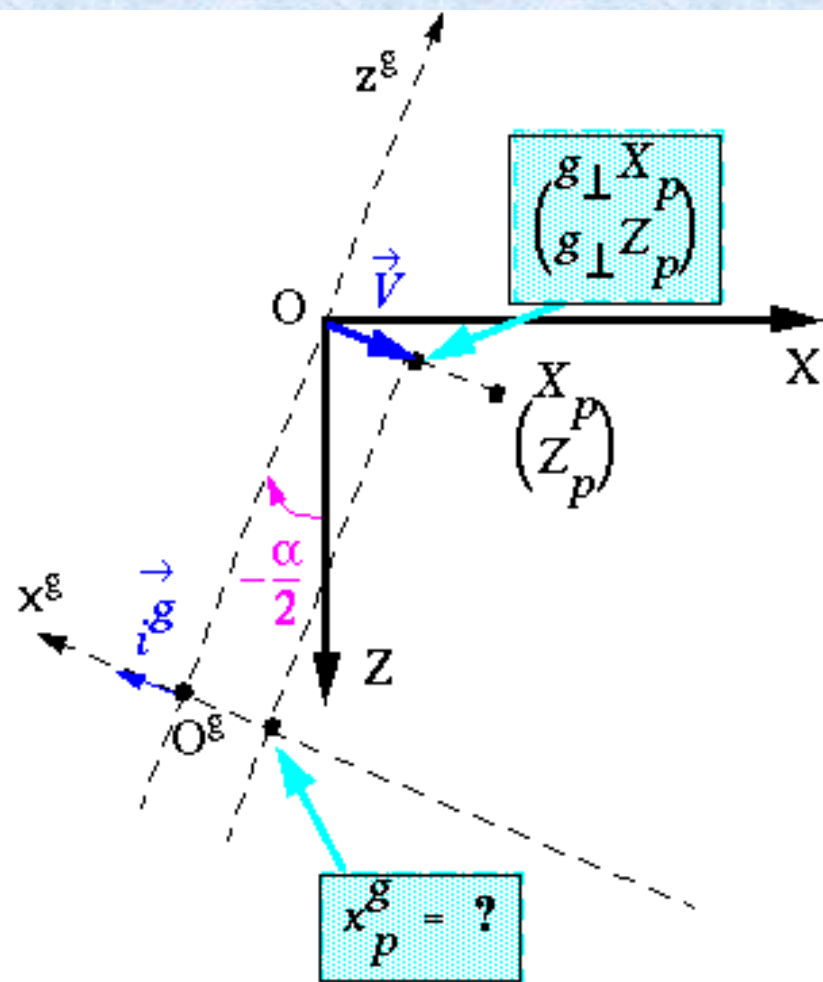


IMAGE GAUCHE

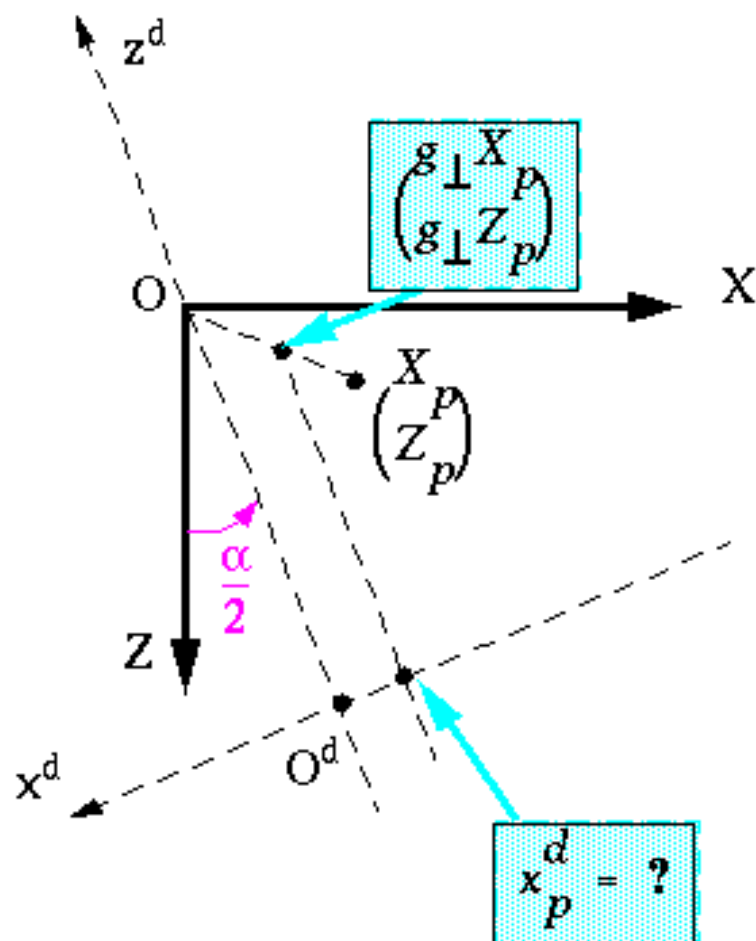


IMAGE DROITE

Si on sait calculer x_p^g , on aura x_p^d en remplaçant $-\frac{\alpha}{2}$ par $\frac{\alpha}{2}$.

• Calcul de x_p^g :

$$\vec{i}^g = \begin{pmatrix} 1 \\ 0 \end{pmatrix} (x^g, z^g) \Rightarrow \vec{i}^g = \begin{pmatrix} -\cos \frac{\alpha}{2} \\ -\sin \frac{\alpha}{2} \end{pmatrix} (X, Z)$$

$$\vec{V} = \begin{pmatrix} g_{\perp} X_p \\ g_{\perp} Z_p \end{pmatrix} (X, Z)$$

Donc :

$$\vec{i}^g \cdot \vec{V} = g_{\perp} \left(-\cos \frac{\alpha}{2} X_p - \sin \frac{\alpha}{2} Z_p \right)$$

Or ce produit scalaire est la coordonnée de \vec{V} le long de $O^g x^g$, c'est-à-dire que :

$$x_p^g = -g_{\perp} \left(\cos \frac{\alpha}{2} X_p + \sin \frac{\alpha}{2} Z_p \right) \quad (3.2)$$

•De la même manière :

$$x_p^d = g_{\perp} \left(-\cos \frac{\alpha}{2} X_p + \sin \frac{\alpha}{2} Z_p \right) \quad (2 \text{ bis})$$

•**Bilan :**

$$I_p^g = g_{\perp} \begin{pmatrix} -\cos \frac{\alpha}{2} X_p - \sin \frac{\alpha}{2} Z_p \\ Y_p \end{pmatrix} \quad (3.3)$$

$$I_p^d = g_{\perp} \begin{pmatrix} -\cos \frac{\alpha}{2} X_p + \sin \frac{\alpha}{2} Z_p \\ Y_p \end{pmatrix} \quad (3.3 \text{ bis})$$

2.2. Détermination des paramètres de calibrage.

Calibrer consiste à déterminer les coefficients permettant de calculer I^g et I^d pour tout point objet

$$P = \begin{pmatrix} X_p \\ Y_p \\ Z_p \end{pmatrix}.$$

D'après les relations (3.3) et (3.3 bis), il y a dans notre cas deux paramètres à déterminer : g_{\perp} et α .

2.2.a. Détermination de g_{\perp} .

D'après (3.3) et (3.3 bis), g_{\perp} est un facteur d'échelle,

dont on observe toujours la présence en stéréovision. C'est même le seul paramètre de calibrage qui apparaisse dans tous les cas.

2.2.b. Détermination de α .

L'angle α ne dépend pas des caractéristiques internes de la caméra. Pour cette raison, on dit que α est un paramètre "extrinsèque" de calibrage.

Soit un point objet pour lequel on connaît :

- les coordonnées X_p , Y_p et Z_p ;
- les points images correspondants I^g et I^d .

On a donc :

$$\left\{ \begin{array}{l} x_p^g = g_{\perp} \left(-\cos \frac{\alpha}{2} X_p - \sin \frac{\alpha}{2} Z_p \right) \\ y_p^g = g_{\perp} Y_p \\ x_p^d = g_{\perp} \left(-\cos \frac{\alpha}{2} X_p + \sin \frac{\alpha}{2} Z_p \right) \\ y_p^d = g_{\perp} Y_p \end{array} \right. \rightarrow \begin{array}{l} \text{+} \Rightarrow x_p^g + x_p^d = -2g_{\perp} \cos \frac{\alpha}{2} X_p \\ \text{-} \Rightarrow x_p^g - x_p^d = -2g_{\perp} \sin \frac{\alpha}{2} Z_p \end{array} \quad (3.4)$$

$$\left\{ \begin{array}{l} x_p^g = g_{\perp} \left(-\cos \frac{\alpha}{2} X_p - \sin \frac{\alpha}{2} Z_p \right) \\ y_p^g = g_{\perp} Y_p \\ x_p^d = g_{\perp} \left(-\cos \frac{\alpha}{2} X_p + \sin \frac{\alpha}{2} Z_p \right) \\ y_p^d = g_{\perp} Y_p \end{array} \right. \rightarrow \begin{array}{l} \text{+} \Rightarrow x_p^g + x_p^d = -2g_{\perp} \cos \frac{\alpha}{2} X_p \\ \text{-} \Rightarrow x_p^g - x_p^d = -2g_{\perp} \sin \frac{\alpha}{2} Z_p \end{array} \quad (3.5)$$

Le produit du membre gauche de (3.4) par le membre droit de (3.5) est égal au produit du membre droit de (3.4) par le membre gauche de (3.5), ce qui donne, en divisant à gauche et à droite par -2 :

$$X_p \cos \frac{\alpha}{2} (x_p^g - x_p^d) = Z_p \sin \frac{\alpha}{2} (x_p^g + x_p^d) \quad (3.6)$$

Cette équation se réécrit :

$$x_p^g \left(-\cos \frac{\alpha}{2} X_p + \sin \frac{\alpha}{2} Z_p \right) = x_p^d \left(-\cos \frac{\alpha}{2} X_p - \sin \frac{\alpha}{2} Z_p \right) \quad (3.6 \text{ bis})$$

Dans (3.6) ou (3.6 bis), seul α est inconnu.

II.2.c. Nombre d'angles α solutions de l'équation (3.6)

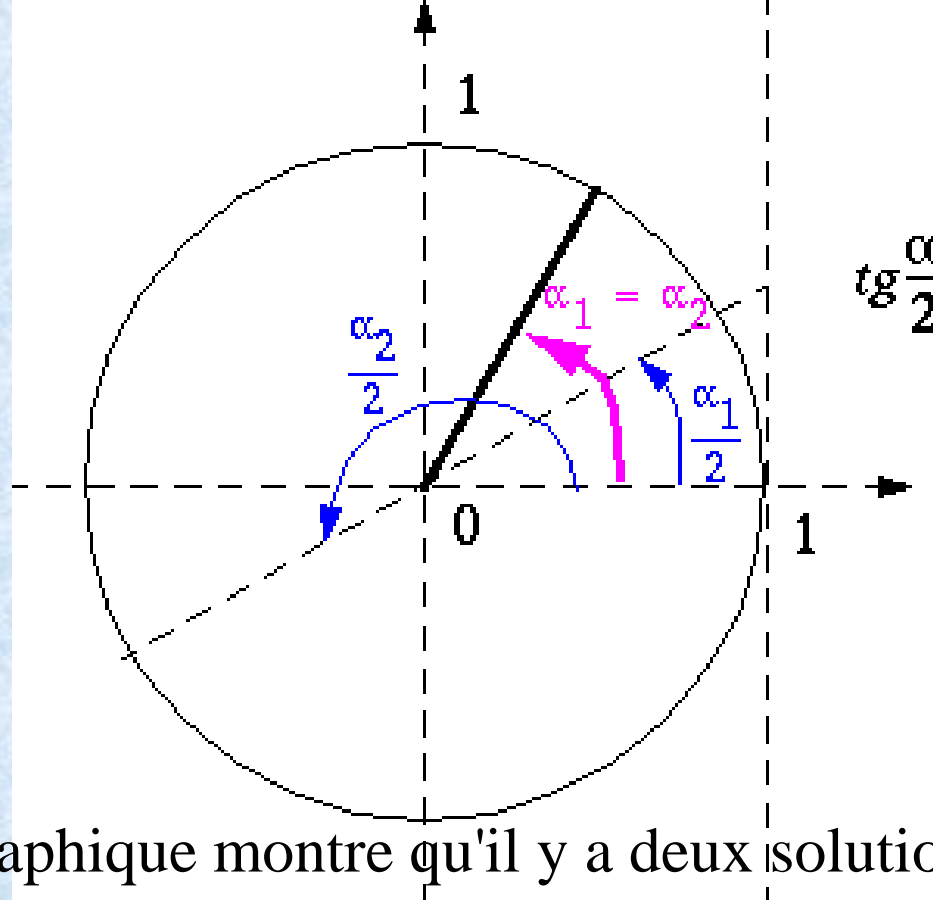
Remarque :

Comme α est forcément différent de 0 (sinon, les deux images seraient identiques !) et différent de π (sinon, aucun point ne serait visible simultanément sur les deux images !), alors $\cos \frac{\alpha}{2} \neq 0$ et $\sin \frac{\alpha}{2} \neq 0$.

Si $X_p \neq 0$ et $Z_p \neq 0$, alors, d'après la remarque précédente et les équations (3.4) et (3.5), on voit que $x_p^g - x_p^d \neq 0$ et $x_p^g + x_p^d \neq 0$. On peut donc, dans ce cas, transformer (3.6) de la manière suivante :

Résolution graphique de (3.7) :

$$\operatorname{tg} \frac{\alpha}{2} = \frac{X_p}{Z_p} \frac{x_p^g - x_p^d}{x_p^g + x_p^d}$$



Cette résolution graphique montre qu'il y a deux solutions

en $\frac{\alpha}{2}$: α_1 et α_2 . Mais, comme $\frac{\alpha_2}{2} = \frac{\alpha_1}{2} + \Pi$, il est clair que les deux angles $\frac{\alpha_1}{2}$ et $\frac{\alpha_2}{2}$ sont égaux modulo 2Π , c'est-à-dire qu'il y a une solution unique α à (3.7), donnée par exemple par :

$$\alpha = 2 \operatorname{Arctg} \left[\frac{\frac{X_p}{Z_p} \frac{x_p^g - x_p^d}{x_p^g + x_p^d}}{\frac{X_p}{Z_p} \frac{x_p^g - x_p^d}{x_p^g + x_p^d}} \right] \quad (3.8)$$

- Si $X_p = 0$ et $Z_p = 0$, les équations (3.4) et (3.5) ne donnent aucune information sur α . Un tel point P n'est donc pas utile pour le calcul de α .

- Si $X_p = 0$ et $Z_p \neq 0$, l'équation (3.5) donne :
$$\sin \frac{\alpha}{2} = \frac{x_p^g - x_p^d}{-2g_{\perp} Z_p}$$

Cette équation permet la détermination de α , mais elle fait intervenir g_{\perp} . Il est préférable d'éviter cette situation pour qu'une erreur sur le calcul de g_{\perp} n'engendre pas une erreur sur le calcul de α .

- Si $X_p \neq 0$ et $Z_p = 0$, l'équation (3.4) donne :
$$\cos \frac{\alpha}{2} = \frac{x_p^g + x_p^d}{-2g_{\perp} X_p}$$

Ce cas est similaire au précédent.

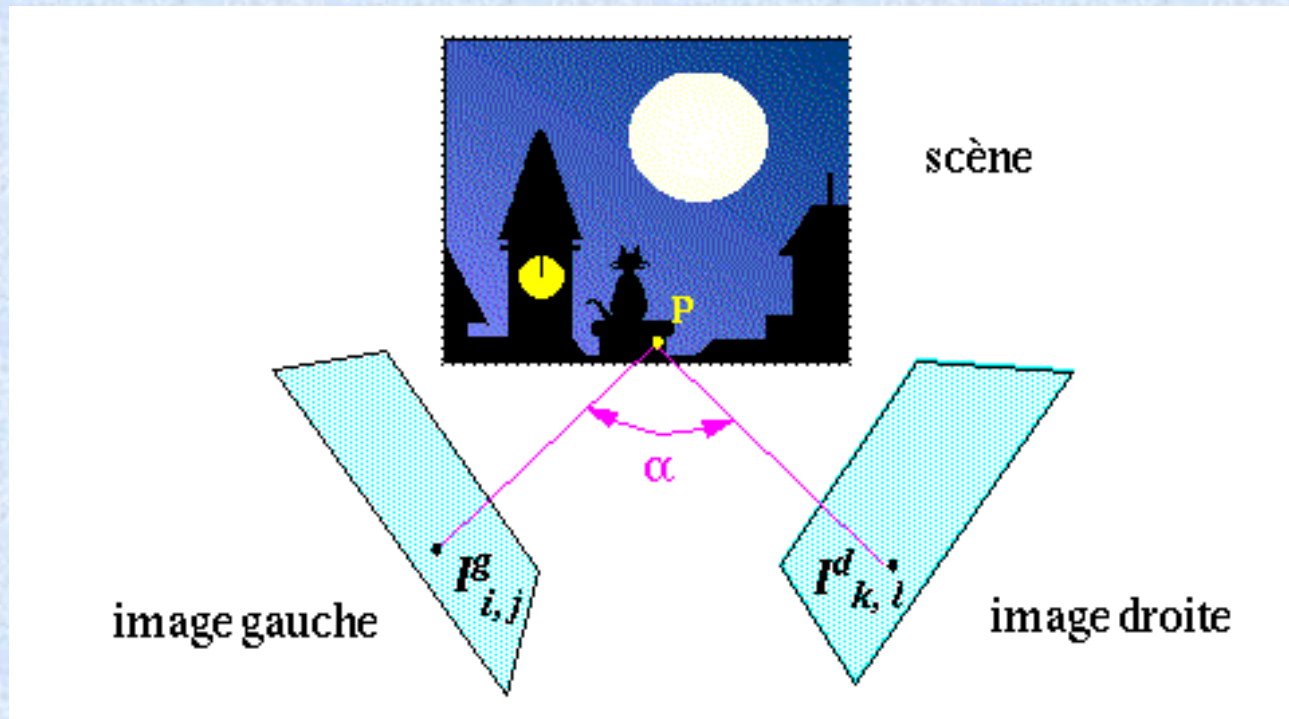
3. L'appariement : phase 2.

L'appariement ou "mise en correspondance" est la phase du traitement la plus difficile, car elle n'est pas purement géométrique. Cette phase fait appel aux niveaux de gris des deux images, contrairement aux phases 1 et 3, dans lesquelles on n'a jamais à utiliser les niveaux de gris $E^g_{i,j}$ et $E^d_{k,l}$.

Nous allons commencer par comparer les niveaux de gris de deux pixels appariés.

3.1. Comparaison des niveaux de gris de deux pixels appariés.

On dit que deux pixels $I^g_{i,j}$ et $I^d_{k,l}$ sont appariés s'ils correspondent au même point physique P de la scène. Pour trouver de telles paires à partir de la donnée de deux images, on ne peut que chercher une similarité entre les niveaux de gris $E^g_{i,j}$ et $E^d_{k,l}$ correspondant à ces deux pixels :



En fait, P ne réémet pas la même énergie lumineuse dans toutes les directions, ce qui signifie que $E^g_{i,j}$ et $E^d_{k,l}$ ne sont pas égaux. On voit aussi que si α diminue, les valeurs $E^g_{i,j}$ et $E^d_{k,l}$ se rapprochent. De toutes façons, on aura beaucoup de paires erronées si on se contente d'apparier un pixel $I^g_{i,j}$ avec le pixel $I^d_{k,l}$ qui a le niveau de gris $E^d_{k,l}$ le plus proche de $E^g_{i,j}$.

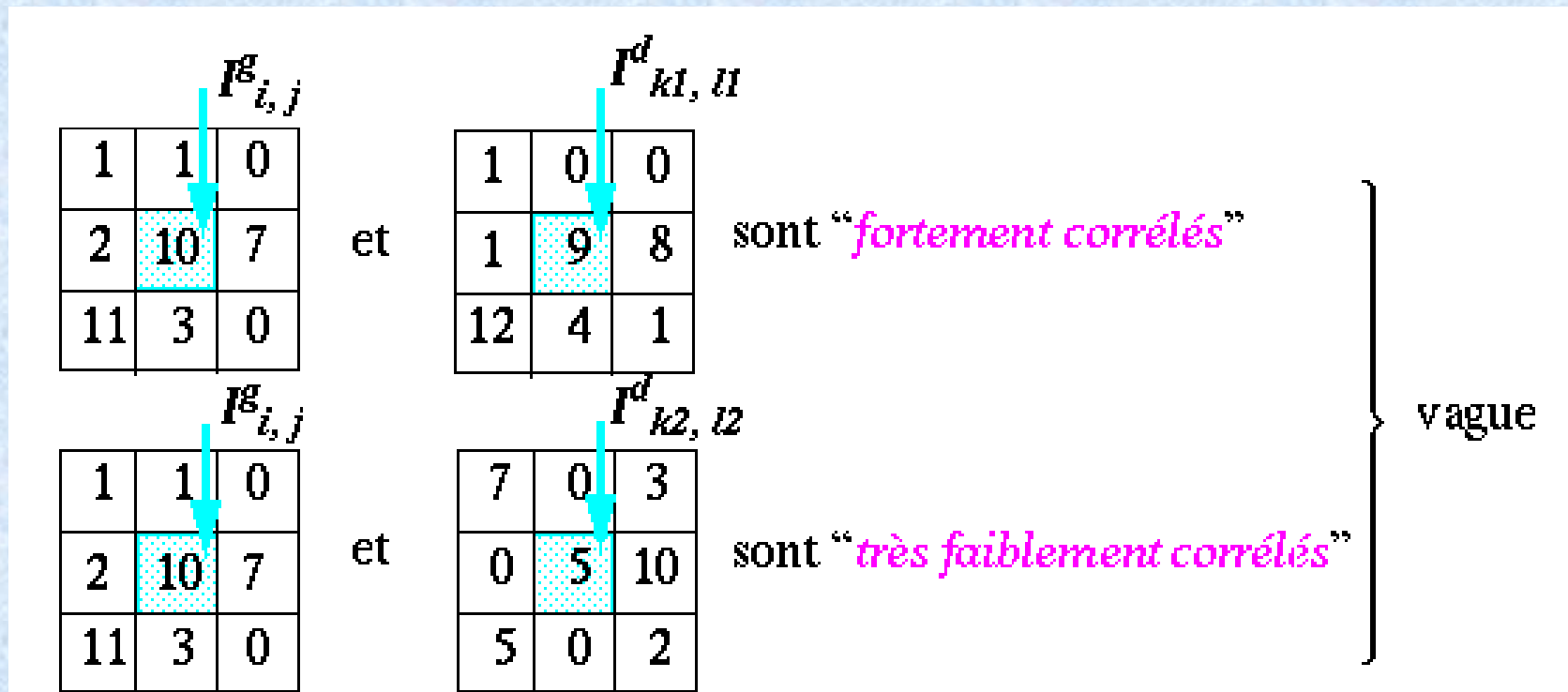
Pour cette raison, on va comparer non pas un pixel de l'image gauche avec un pixel de l'image droite, mais un pixel de l'image gauche et son voisinage, avec un pixel de l'image droite et son voisinage. C'est ce qu'on appelle la "**corrélation**".

3.2. Corrélation entre deux pixels et leurs voisinages.

Un pixel $I^g_{i,j}$ de l'image gauche et un pixel $I^d_{k,l}$ de l'image droite sont **corrélés** si les niveaux de gris ont des valeurs similaires au voisinage de ces deux pixels.

Attention :

La corrélation n'a rien à voir avec la convolution ni avec la covariance.

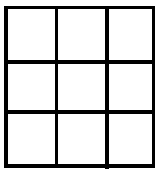


3.2.a. Définition du voisinage : fenêtre de corrélation.

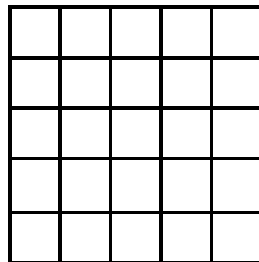
On peut chercher la corrélation entre deux pixels et leurs voisinages avec une taille de voisinage plus ou moins importante. On appelle ce voisinage la "fenêtre de corrélation".

- prendre la même fenêtre de corrélation sur l'image gauche et l'image droite ;
- prendre une fenêtre de corrélation connexe (en un seul morceau) ;
- la fenêtre de corrélation contient le pixel qu'on cherche à apparier.

Exemples :



fenêtre 3x3



fenêtre 5x5 ; 7x7 ; 9x9 ; 11x11 ; ...

3.2.b. Mesure de corrélation euclidienne.

Il existe différentes méthodes pour mesurer numériquement la corrélation. La plus élémentaire est la mesure de corrélation euclidienne. Avec une fenêtre de corrélation 3 x 3, elle s'écrit :

$$\|I_{i,j}^g - I_{k,l}^d\|^2 = \sum_{p=-1}^1 \sum_{q=-1}^1 \left[E_{i+p, j+q}^g - E_{k+p, l+q}^d \right]^2$$

Avec l'exemple précédent :

$$\|I_{i,j}^g - I_{k1,l1}^d\|^2 = 1 + 1 + 1 + 1 + 1 + 1 + 1 = 7$$

=> pixels "*fortement corrélés*".

$$\|I_{i,j}^g - I_{k2,l2}^d\|^2 = 36 + 1 + 9 + 4 + 25 + 9 + 36 + 9 + 4 = 133$$

=> pixels "*très faiblement corrélés*".

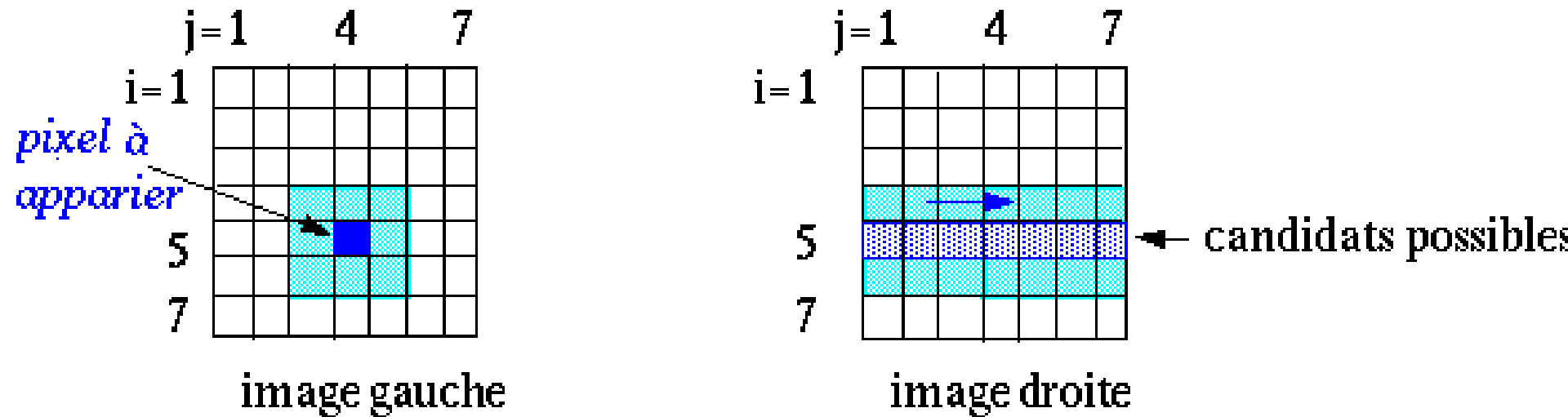
Plus la mesure de corrélation euclidienne entre deux pixels et leurs voisinages est faible, plus ces pixels sont corrélés. Cela nous procure une méthode d'appariement.

3.3. Méthode d'appariement.

Rappel :

Dans le cas où nous nous sommes placés, nous avons vu que $y^g_p = y^d_p$. Ceci signifie que deux pixels appariés se trouvent forcément sur des lignes de même numéro : cette propriété est la traduction de la "contrainte épipolaire".

Exemple :



En fait, on se limite aux pixels pour lesquels la fenêtre de corrélation est entièrement contenue dans l'image. Cela élimine les candidats situés sur les bords.

3.3.a. Appariement d'un pixel de l'image gauche.

Soit $I^g_{i,j}$ un pixel de l'image gauche tel que la fenêtre de corrélation soit contenue dans l'image gauche. On cherche à apparier $I^g_{i,j}$. Les candidats sont les pixels $I^d_{i,l}$ pour lesquels la fenêtre de corrélation est contenue dans l'image droite. On apparie $I^g_{i,j}$ avec le pixel $I^d_{i,l}$ pour lequel la mesure de corrélation euclidienne est minimale.

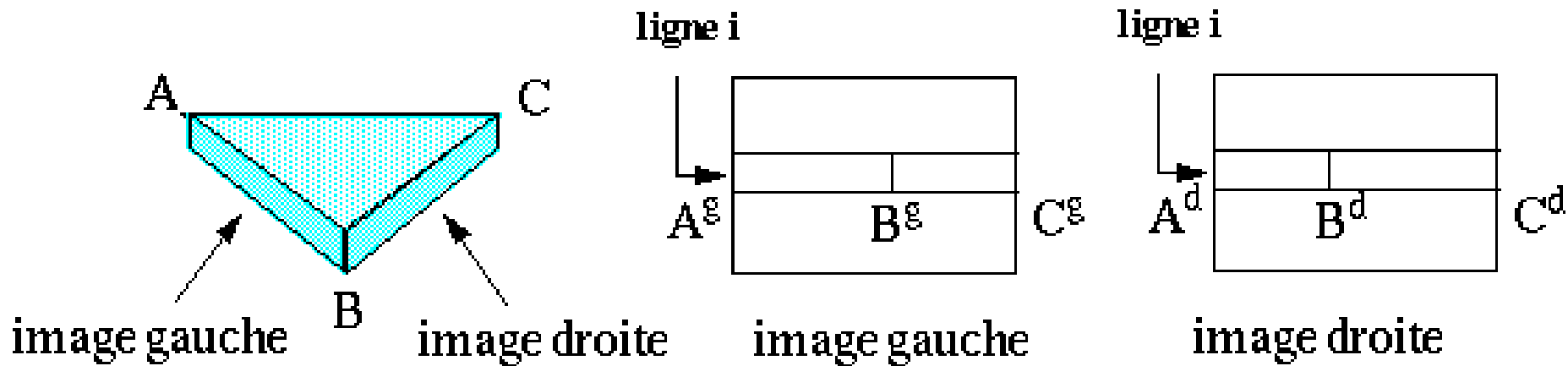
Remarques :

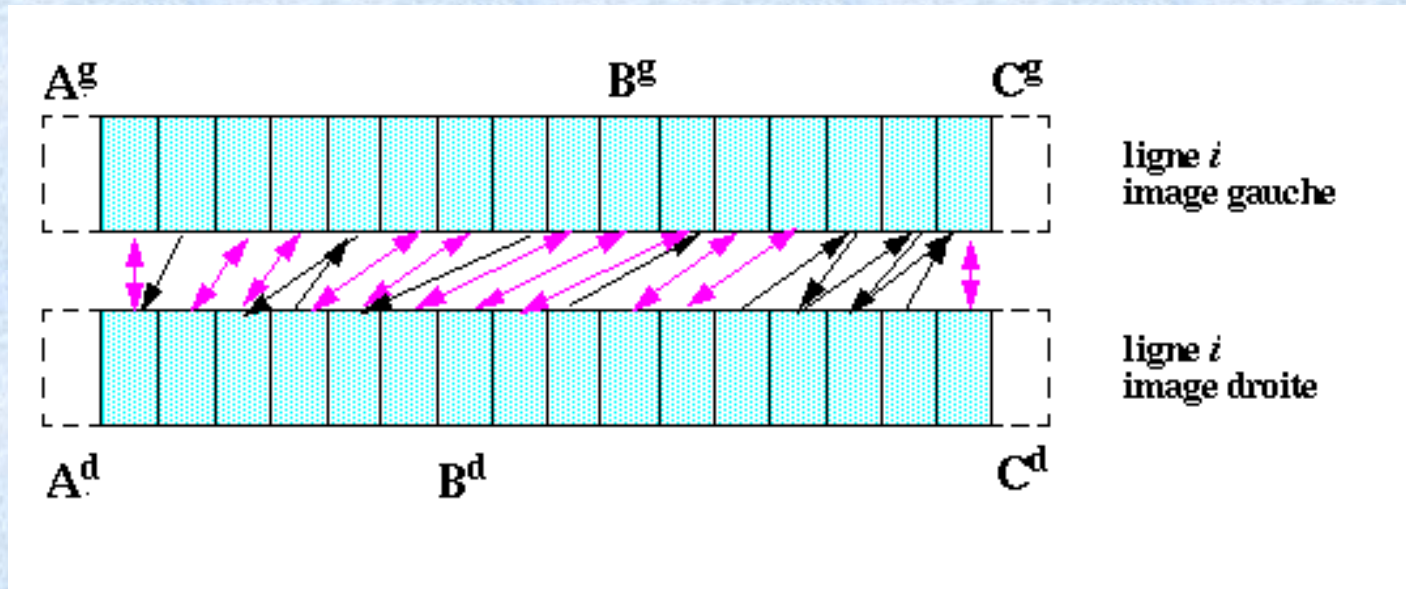
- Il peut y avoir des erreurs d'appariement.
- Dans le cas exceptionnel où plusieurs candidats donnent la même mesure de corrélation minimale, on ne se détermine pas, puisqu'on court le risque d'effectuer un appariement erroné.

3.3.b. Appariement d'un pixel de l'image droite.

On a privilégié les pixels de l'image gauche sans raison. On peut donc réaliser la même opération pour les pixels de l'image droite. Si $I^g_{i,j}$ a été apparié avec le pixel $I^d_{i,l}$, il se peut très bien que lors de l'opération inverse, $I^d_{i,l}$ ne soit pas apparié avec $I^g_{i,j}$.

Exemple :





Chaque simple flèche noire désigne un appariement unilatéral, c'est-à-dire n'ayant pas débouché sur la formation d'une paire. Les doubles flèches roses indiquent les paires formées d'un pixel de l'image gauche et d'un pixel de l'image droite, c'est-à-dire les paires apparaissant dans les deux appariements : ces paires vérifient ce qu'on appelle la "**contrainte d'unicité**". Dans notre exemple, après application de la contrainte d'unicité, il ne reste plus que 11 paires (sur les 16 possibles).

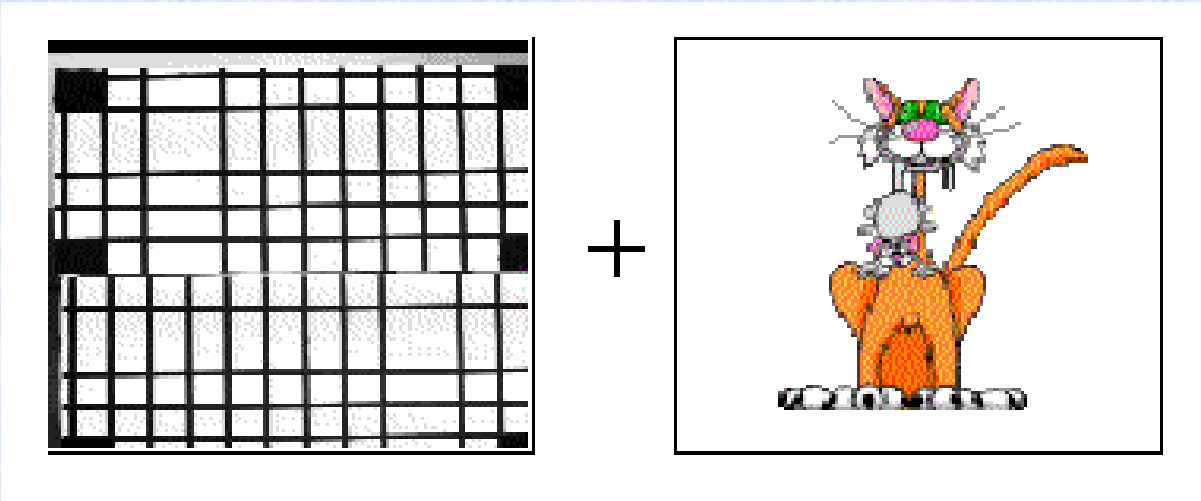
3.3.c. Conclusion.

Cette méthode d'appariement consiste à mettre en bijection nb pixels de l'image gauche ($nb < N \times M$) avec nb pixels de l'image droite.

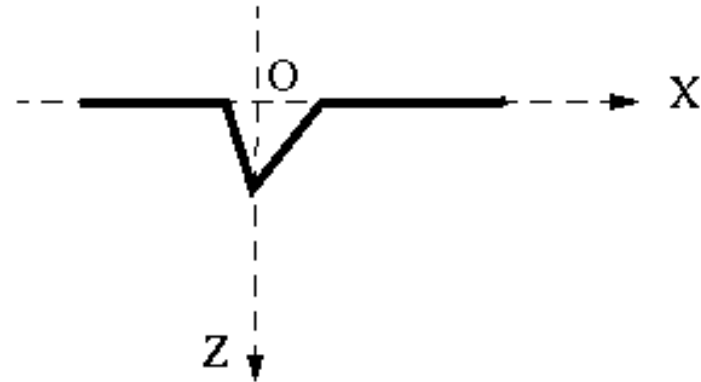
Un certain nombre de pixels restent non appariés, en particulier sur les bords. Cela n'est pas un problème, et nous verrons plus loin comment traiter ces pixels non appariés. Un problème plus important est celui des appariements erronés, que nous allons évoquer dans le paragraphe suivant.

3.4. Appariements erronés.

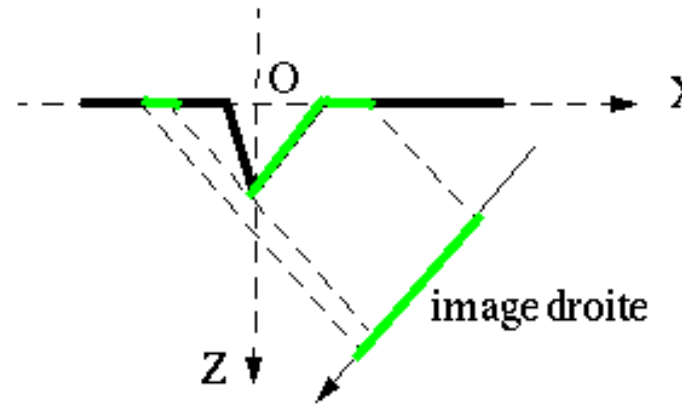
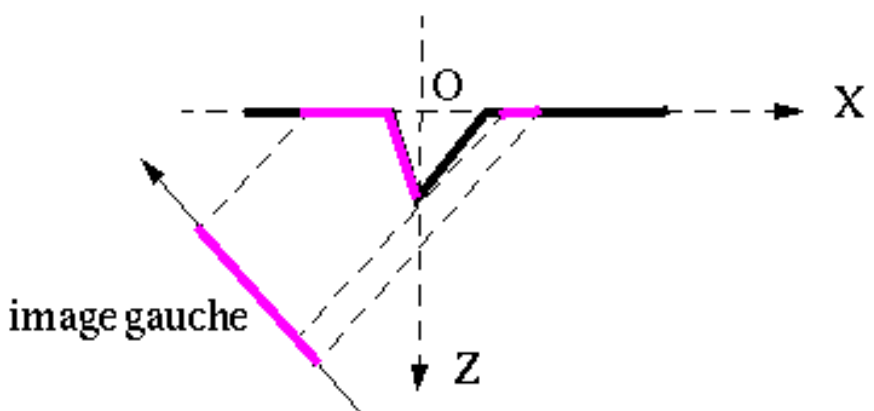
Avec la méthode que nous venons de voir, il est possible de traiter un couple de deux images quelconques :



=> couples de pixels appariés ?????!!!!

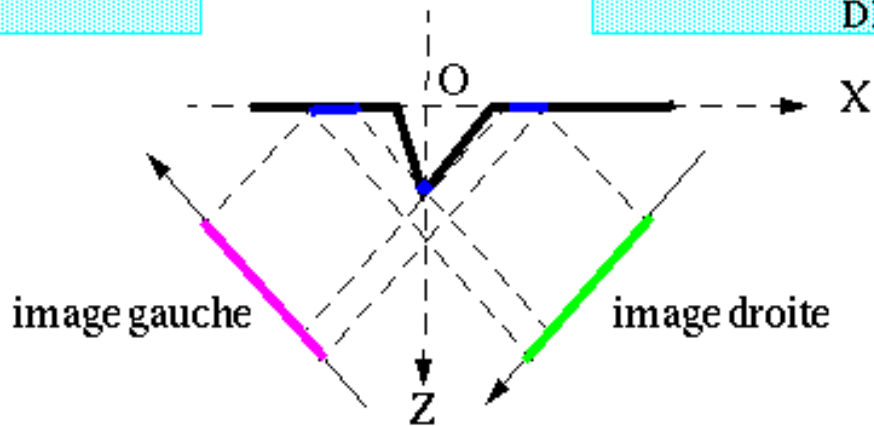


POINTS OBJETS



POINTS OBJETS VISIBLES SUR L'IMAGE GAUCHE

POINTS OBJETS VISIBLES SUR L'IMAGE DROITE



POINTS OBJETS VISIBLES SUR LES 2 IMAGES

Il se peut très bien qu'un pixel correspondant à une des zones non visibles sur une des images ait été apparié. Il faut éliminer si possible un tel appariement, puisqu'il est forcément erroné.

Pour ce faire, nous introduisons un seuil S , et considérons que deux niveaux de gris $E_{i,j}^g$ et $E_{i,l}^d$ sont "proches" si et seulement si :

$$\Delta(E_{i,j}^g, E_{i,l}^d) \leq S \quad (3.9)$$

avec :

$$\Delta(E_{i,j}^g, E_{i,l}^d) \quad \text{"écart relatif"} = \frac{|E_{i,j}^g - E_{i,l}^d|}{E_{i,j}^g + E_{i,l}^d} \quad \text{si} \begin{cases} E_{i,j}^g \neq 0 \\ \text{ou} \\ E_{i,l}^d \neq 0 \end{cases}$$
$$= 0 \quad \text{si} \quad E_{i,j}^g = E_{i,l}^d = 0$$

Exemple : $E^g_{ij} = 12$

$E^d_{i,l}$	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
écart relatif	1	0.85	0.72	0.6	0.5	0.41	0.33	0.26	0.2	0.14	0.09	0.04	0	0.04	0.08	0.11	0.14	0.17	0.2

Si $S = 0,1$, alors parmi ces valeurs entières de $E^d_{i,l}$, seules 5 sont considérées comme "proches" de E^g_{ij} : 10, 11, 12, 13 et 14.

Comme la similarité entre deux pixels est jugée sur les voisinages de ces deux pixels, on va considérer que deux pixels $(I^g_{ij}, I^d_{i,l})$ forment une paire acceptable si (avec une fenêtre 3 x 3) :

$$\sum_{p=-1}^1 \sum_{q=-1}^1 \Delta(E^g_{i+p, j+q}, E^d_{i+p, l+q})^2 \leq \sum_{p=-1}^1 \sum_{q=-1}^1 S^2 = 9S^2$$

Il est possible de montrer qu'une condition suffisante pour que cette inégalité soit vérifiée s'écrit :

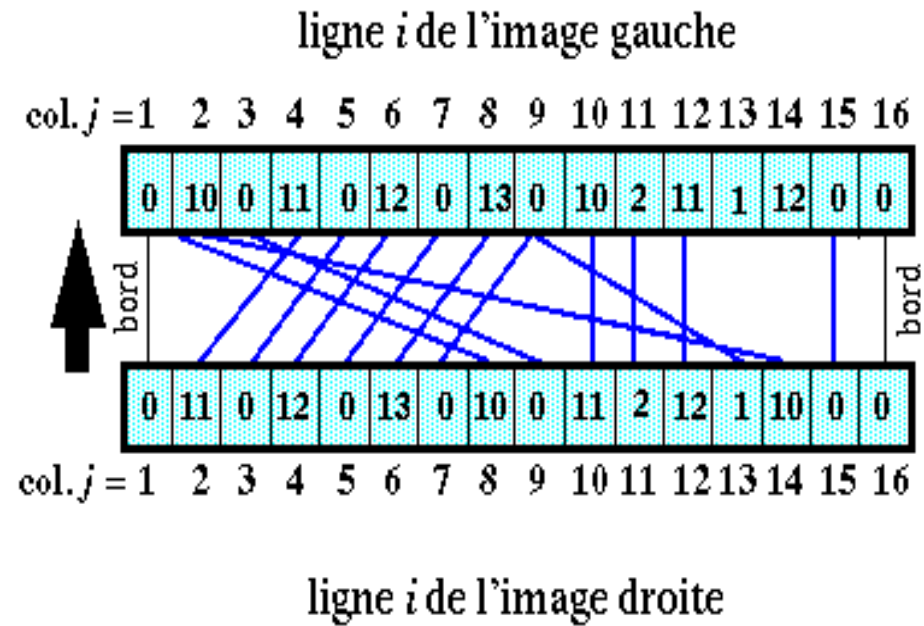
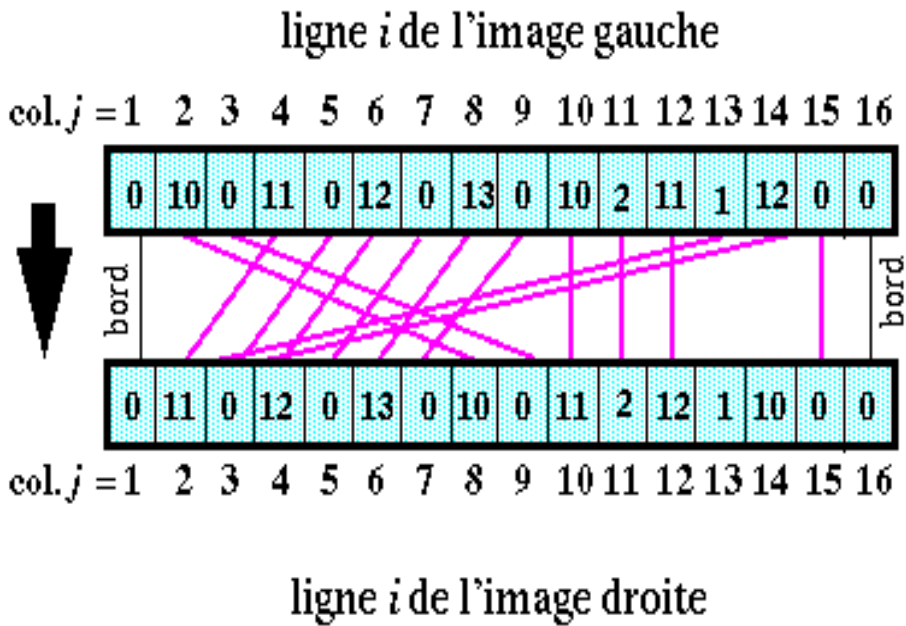
$$\|I_{i,j}^g - I_{i,l}^d\|^2 \leq 9S^2 \times \sum_{p=-1}^1 \sum_{q=-1}^1 \left(E_{i+p, j+q}^g + E_{i+p, l+q}^d \right)^2$$

(3.10)

Les paires de pixels ne vérifiant pas cette égalité sont rejetées car trop suspectes. Ceci revient à imposer une **contrainte de seuil**.

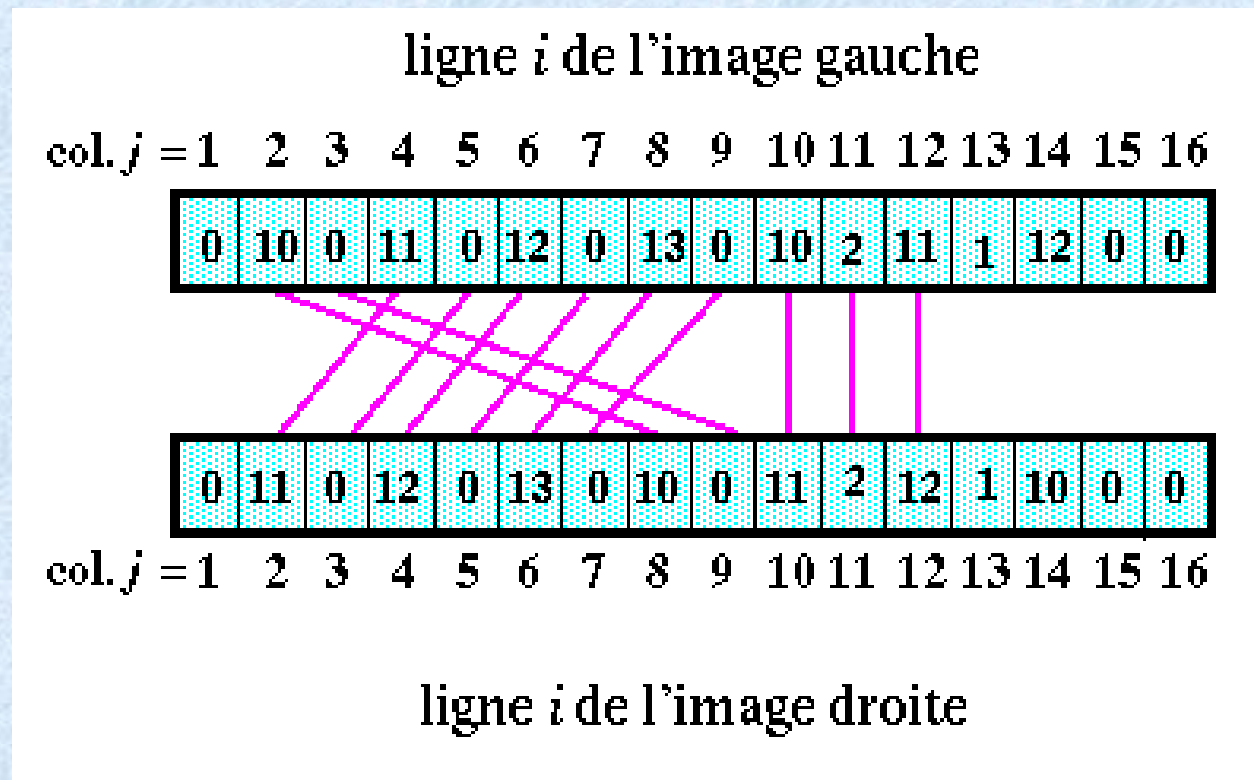
Les trois contraintes que nous avons citées (épipolaire, d'unicité et de seuil) fournissent un ensemble de paires de pixels $(I^g_{i,j}, I^d_{i,l})$ assez bon en général. Il subsistera des erreurs dans le cas d'un motif périodique ou presque périodique.

Exemple 1 :

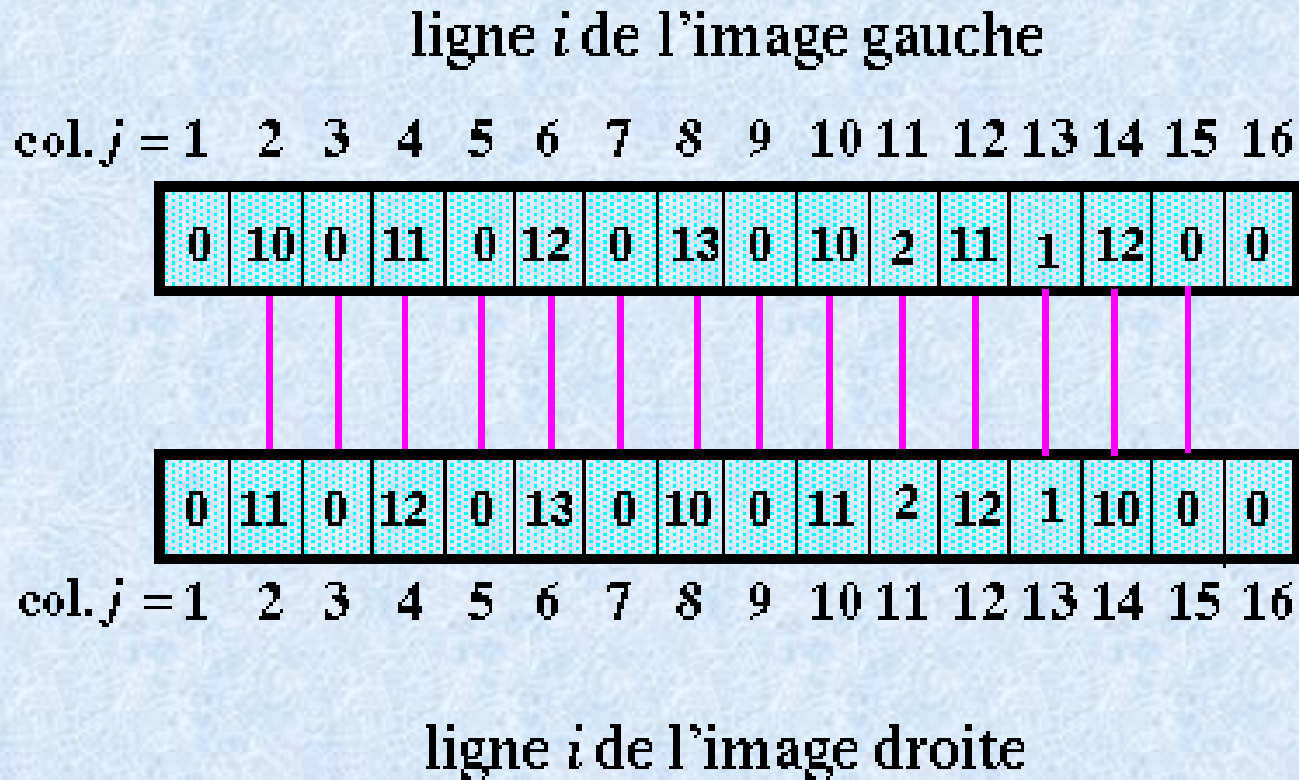


Sur le schéma précédent, le tableau de gauche correspond à la phase d'appariement des pixels de l'image gauche et le tableau de droite correspond à la phase d'appariement des pixels de l'image droite.

En appliquant les contraintes d'unicité et de seuil (avec $S = 1$), il ne reste que les 11 paires suivantes :



Or, on aurait plutôt tendance à penser que deux pixels formant une paire ont des numéros de colonnes identiques :



Il existe deux améliorations possibles à la méthode d'appariement décrite dans les paragraphes précédents :

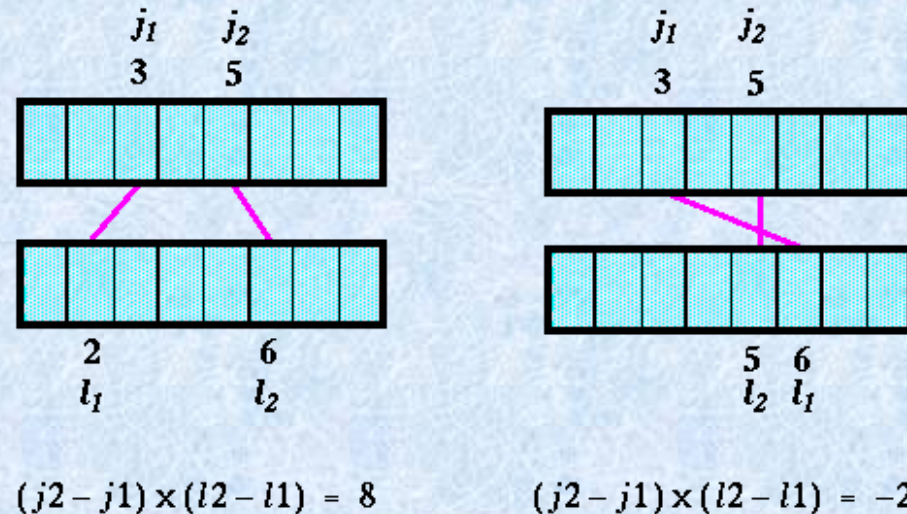
- la **contrainte d'ordre** ;
- l'**étude de la courbe de corrélation**.

3.5.a. Contrainte d'ordre.

Soit $(I^g_{i,j1}, I^d_{i,l1})$ et $(I^g_{i,j2}, I^d_{i,l2})$ deux paires de pixels, vérifiant les trois contraintes citées précédemment (épipolaire, d'unicité et de seuil).

On impose en plus au produit $(j2 - j1) \times (l2 - l1)$ d'être supérieur ou égal à 0. Considérons l'exemple suivant :

Exemple 2 :



accepté

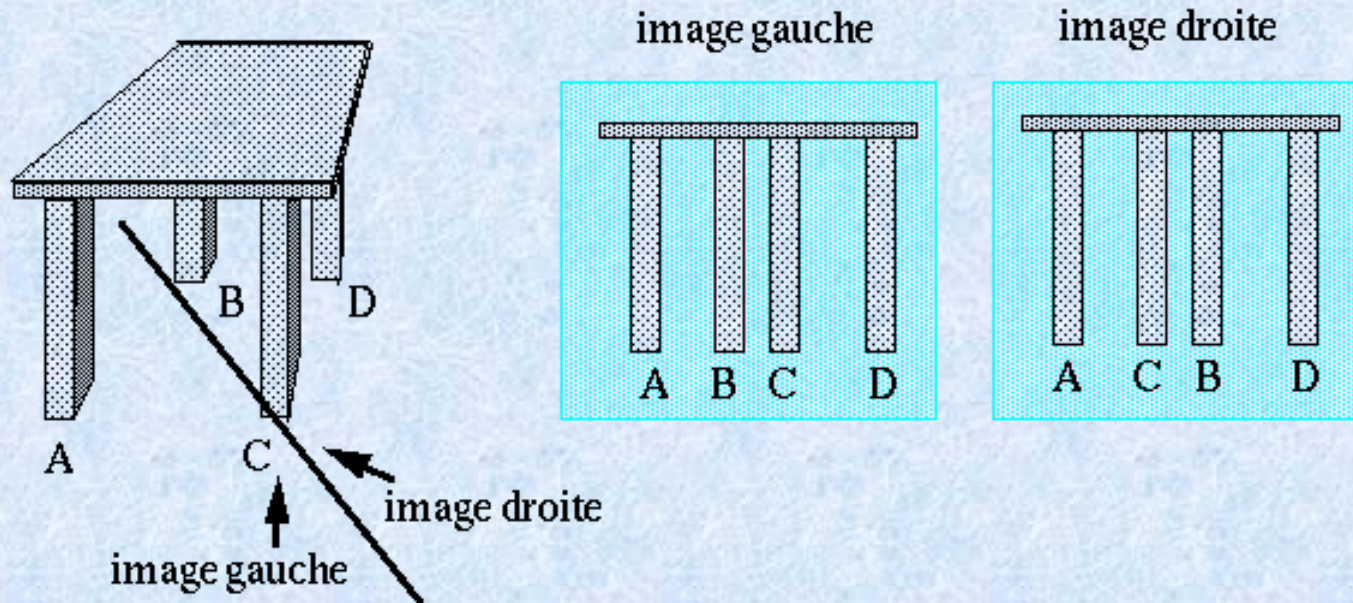
refusé

Si un pixel $I^g_{i,j1}$ est à gauche d'un pixel $I^g_{i,j2}$ sur l'image gauche, alors $I^d_{i,l1}$ doit être à gauche de $I^d_{i,l2}$ sur l'image droite : c'est la **contrainte d'ordre**.

Dans l'exemple 2, la figure de droite présente deux paires qui ne vérifient pas cette contrainte.

Remarque :

La contrainte d'ordre est parfois fausse. Par exemple :

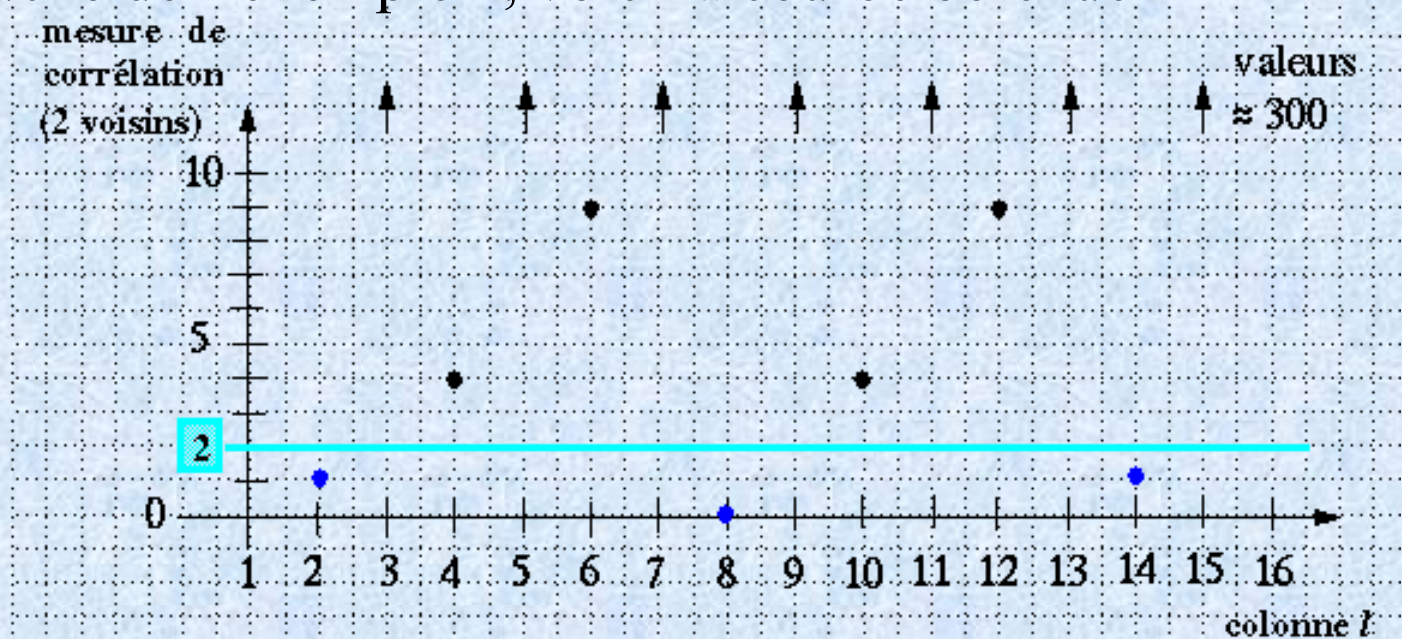


Dans la pratique, on cherchera à respecter cette contrainte en éliminant un minimum de paires. Dans l'[exemple 1](#), il suffit d'enlever 2 paires pour respecter la contrainte d'ordre : $(I^g_{i,2}, I^d_{i,8})$ et $(I^g_{i,3}, I^d_{i,9})$. Mais cela ne change pas les appariements erronés. Que faire ?

3.5.b. Étude de la courbe de corrélation.

Pour chaque pixel à apparier, on trace une courbe avec, en abscisse, les indices de colonnes des candidats et, en ordonnée, les mesures de corrélation. Pour le pixel $I^g_{i,2}$ situé sur la ligne i et la colonne 2, sur l'image gauche de l'exemple 1, voici la courbe obtenue

Exemple :



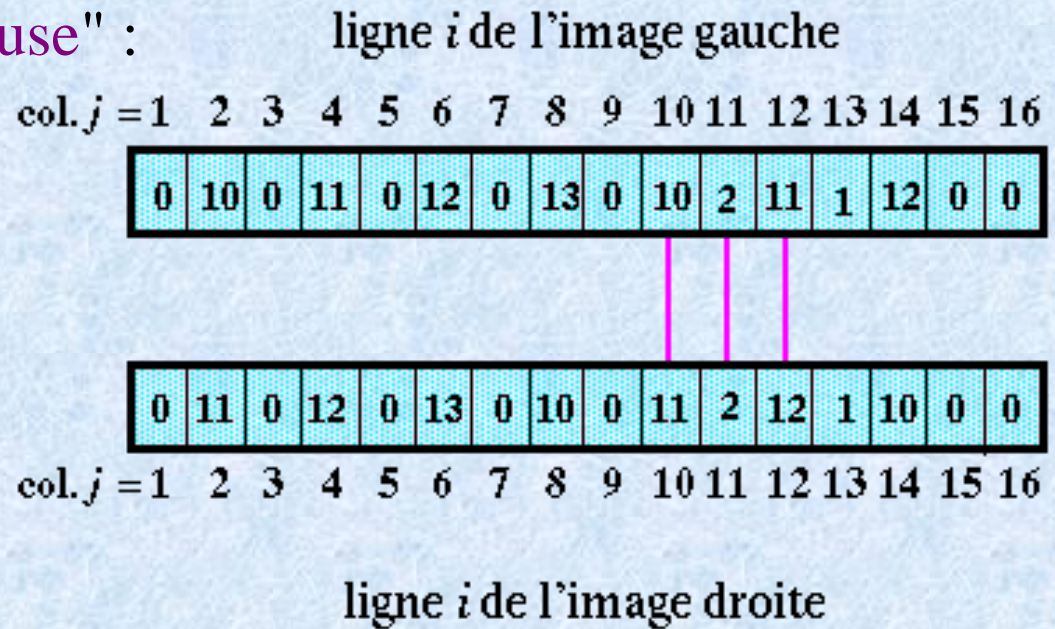
La courbe présente bien un minimum absolu unique égal à 0, mais plusieurs minima relatifs. On a été trop affirmatif en choisissant $l = 8$. On peut prendre en compte les candidats pour lesquels la corrélation est inférieure ou égale à 2 (seuil $S' = 2$). Dans ce cas, il reste 3 pixels candidats situés sur les colonnes suivantes : $l = 2, 8$ et 14 .

Deux attitudes sont possibles :

- méthode "peureuse": s'il existe plusieurs candidats, on n'en choisit aucun => pixel non apparié (problème déjà évoqué) ;
- méthode "hardie" : on cherche à vérifier à la fois la contrainte d'ordre et à réaliser un maximum d'appariements, en choisissant un candidat qui ne correspond pas forcément au minimum absolu (algorithme relativement complexe).

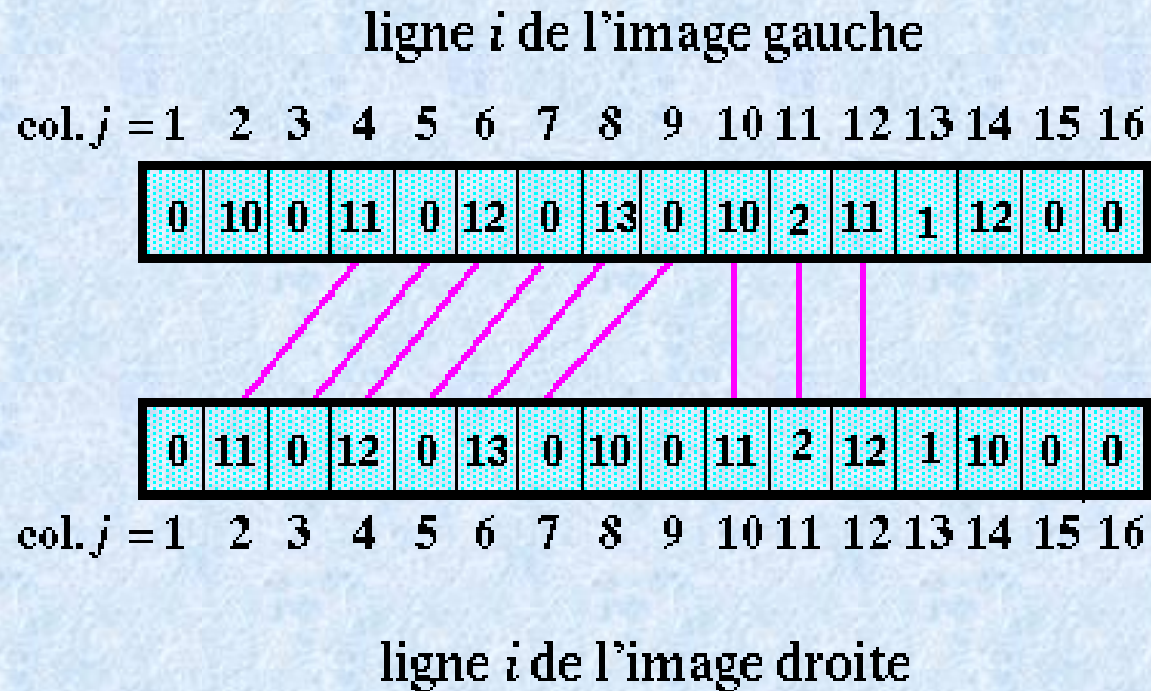
On obtiendra, avec l'exemple 1 ($S' = 2$) :

- Par la méthode "peureuse" :



Il ne reste que 3 couples, ce qui n'est pas très satisfaisant.

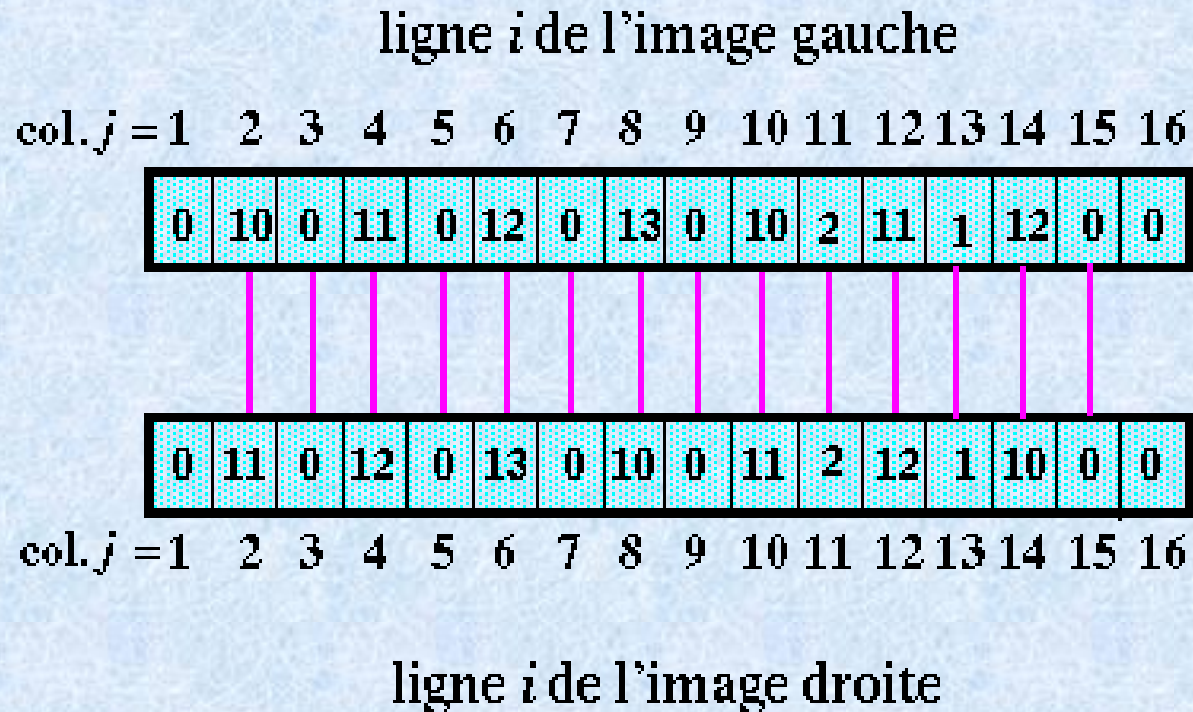
- Par la méthode "hardie" :



Avec cette méthode, il reste 9 appariements, dont 6 faux, ce qui n'est pas satisfaisant non plus !

Changeons le seuil : les candidats retenus sont maintenant ceux pour lesquels la corrélation est inférieure ou égale à $S' = 10$.

- Par la méthode "peureuse" : il ne reste aucun appariement ;
- Par la méthode "hardie" :



- Au maximum, on peut réaliser 14 appariements (on ne pourrait pas en réaliser plus), qui sont tous exacts qui plus est !

3.5.c. Conclusions sur ces améliorations.

- Les méthodes d'appariement ressemblent souvent à de la "cuisine".
Suivant la valeur du seuil S' , on obtient par la méthode "hardie" :
 - si $S' = 2$: 9 appariements , dont 6 faux ;
 - si $S' = 10$: 14 appariements exacts.
 - Par ailleurs, ces méthodes sont beaucoup plus complexes à mettre en oeuvre que la méthode décrite précédemment (avec les trois contraintes initiales).
- Passons maintenant à la phase 3 de **triangulation**, la plus facile.

4. La triangulation : phase 3.

On suppose que :

- Les paramètres α et g_{\perp} sont connus (phase 1) ; *
- Un certain nombre de paires de pixels (I_{ij}^g, I_{il}^d) ont été déterminées (phase 2).

4.1. Expression mathématique.

Soit une paire de pixels appariés (I_{ij}^g, I_{il}^d). Le point $P = (X_p, Y_p, Z_p)$ correspondant à cette paire vérifie les trois équations suivantes :

$$\begin{cases} -\cos \frac{\alpha}{2} X_p - \sin \frac{\alpha}{2} Z_p = \frac{x_p^g}{g_{\perp}} \\ -\cos \frac{\alpha}{2} X_p + \sin \frac{\alpha}{2} Z_p = \frac{x_p^d}{g_{\perp}} \end{cases}$$

$$Y_p = \frac{y_p^g}{g_{\perp}} = \frac{y_p^d}{g_{\perp}} \quad (3.11)$$

Les deux dernières équations constituent un système linéaire de deux équations à deux inconnues : X_p et Z_p . Calculons son déterminant

$$Dét = \begin{vmatrix} -\cos \frac{\alpha}{2} & -\sin \frac{\alpha}{2} \\ -\cos \frac{\alpha}{2} & \sin \frac{\alpha}{2} \end{vmatrix} = -2 \sin \frac{\alpha}{2} \cos \frac{\alpha}{2} = -\sin \alpha$$

Comme $\alpha \neq 0$ et $\alpha \neq \Pi$, il s'ensuit que $Dét \neq 0$, donc que le système est un système de Cramer :

$$\left\{ \begin{array}{l} X_p = \frac{1}{Dét} \begin{vmatrix} \frac{x_p^g}{g_{\perp}} & -\sin \frac{\alpha}{2} \\ \frac{x_p^d}{g_{\perp}} & \sin \frac{\alpha}{2} \end{vmatrix} \\ Z_p = \frac{1}{Dét} \begin{vmatrix} -\cos \frac{\alpha}{2} & \frac{x_p^g}{g_{\perp}} \\ -\cos \frac{\alpha}{2} & \frac{x_p^d}{g_{\perp}} \end{vmatrix} \end{array} \right.$$

soit encore :

$$X_P = -\frac{1}{2 \cos \frac{\alpha}{2} g_{\perp}} \left(x_P^g + x_P^d \right) \quad (3.12)$$

$$Z_P = -\frac{1}{2 \sin \frac{\alpha}{2} g_{\perp}} \left(x_P^g - x_P^d \right) \quad (3.13)$$

4.2. La disparité.

Dans (3.13) apparaît la différence :
appelée **disparité** d'un couple de pixels.

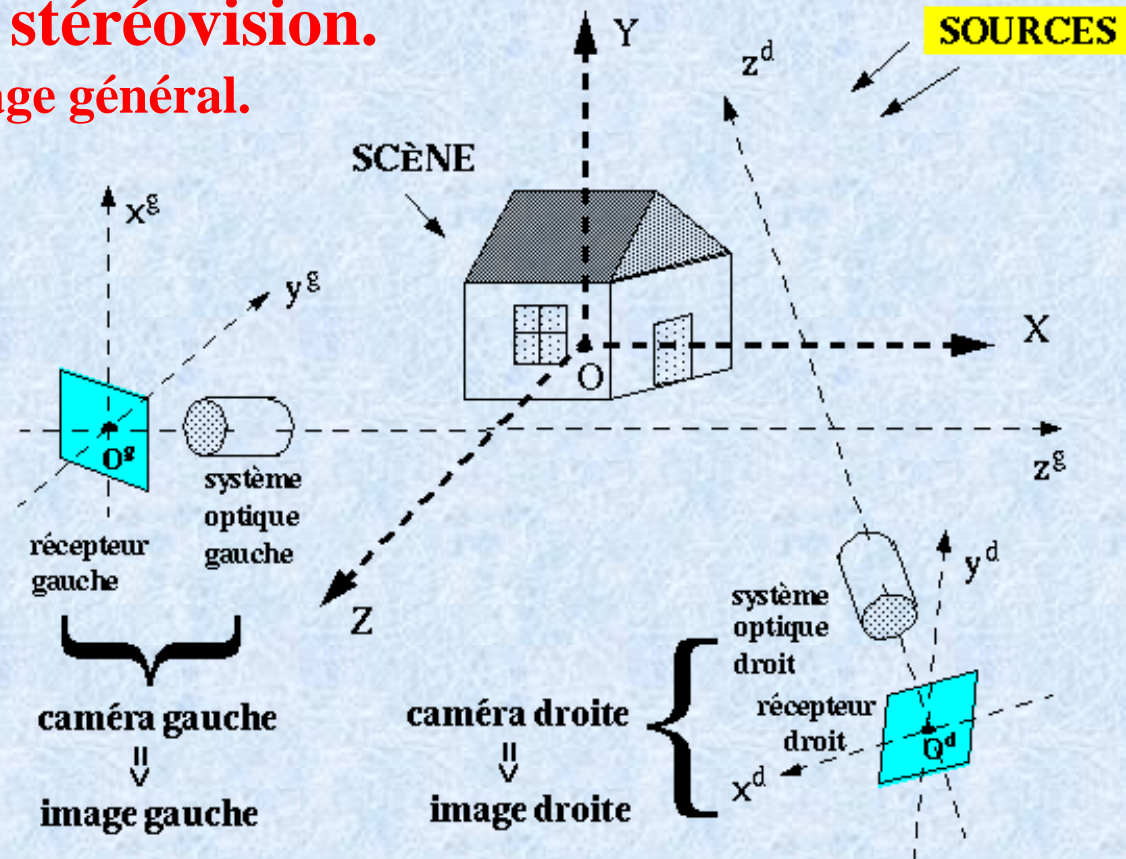
On voit que l'altitude d'un point P est proportionnelle à la disparité du couple de pixels qui lui est associé.

$$x_P^g - x_P^d = D_P$$

Dans le chapitre 3, nous nous sommes placés dans un cas de stéréovision élémentaire. Nous allons reprendre différents points du chapitre 3, et donner leur généralisation. Nous conclurons en donnant quelques exemples de stéréovision utilisés dans le milieu industriel.

1. Cas général de la stéréovision.

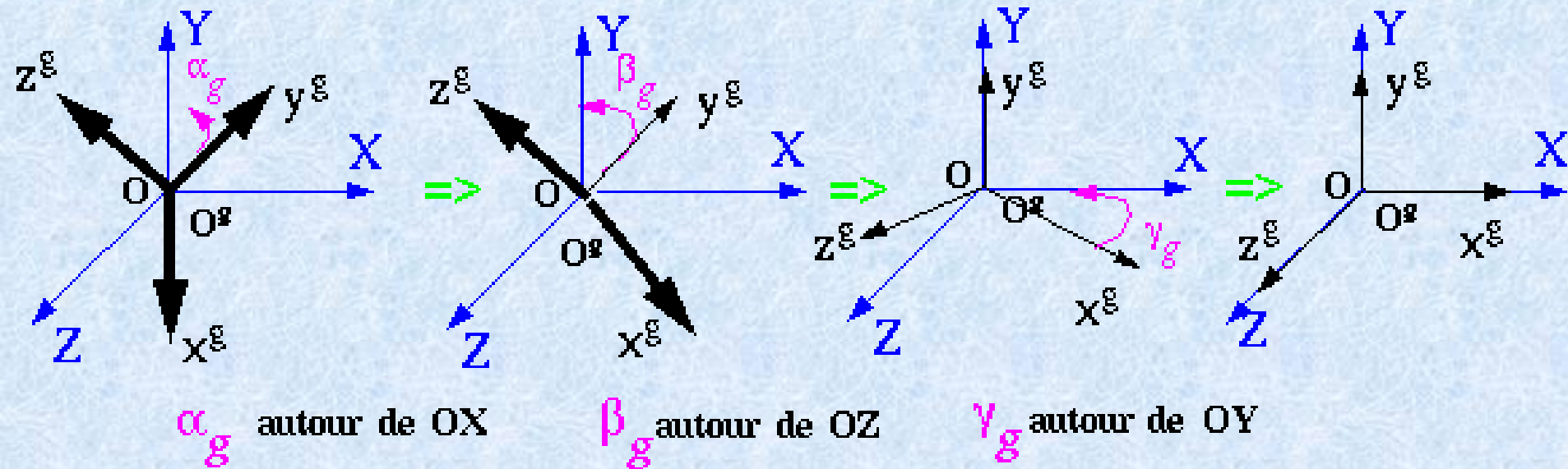
1.1. Description du montage général.



On souhaite connaître Z pour chaque point de la scène, éclairé et visible sur ces deux images numériques stéréoscopiques.

Il y a en général deux caméras distinctes. Les repères des caméras ne sont pas liés au repère OXYZ. Pour passer de OXYZ à $O^g x^g y^g z^g$, ou à $O^d x^d y^d z^d$, il faut donc effectuer :

- Une translation $\overrightarrow{OO^g}$ ou $\overrightarrow{OO^d}$
- Une rotation dans \mathbf{R}^3 (définie par 3 angles) :



1.2. Correspondance entre points objets et pixels d'une image.

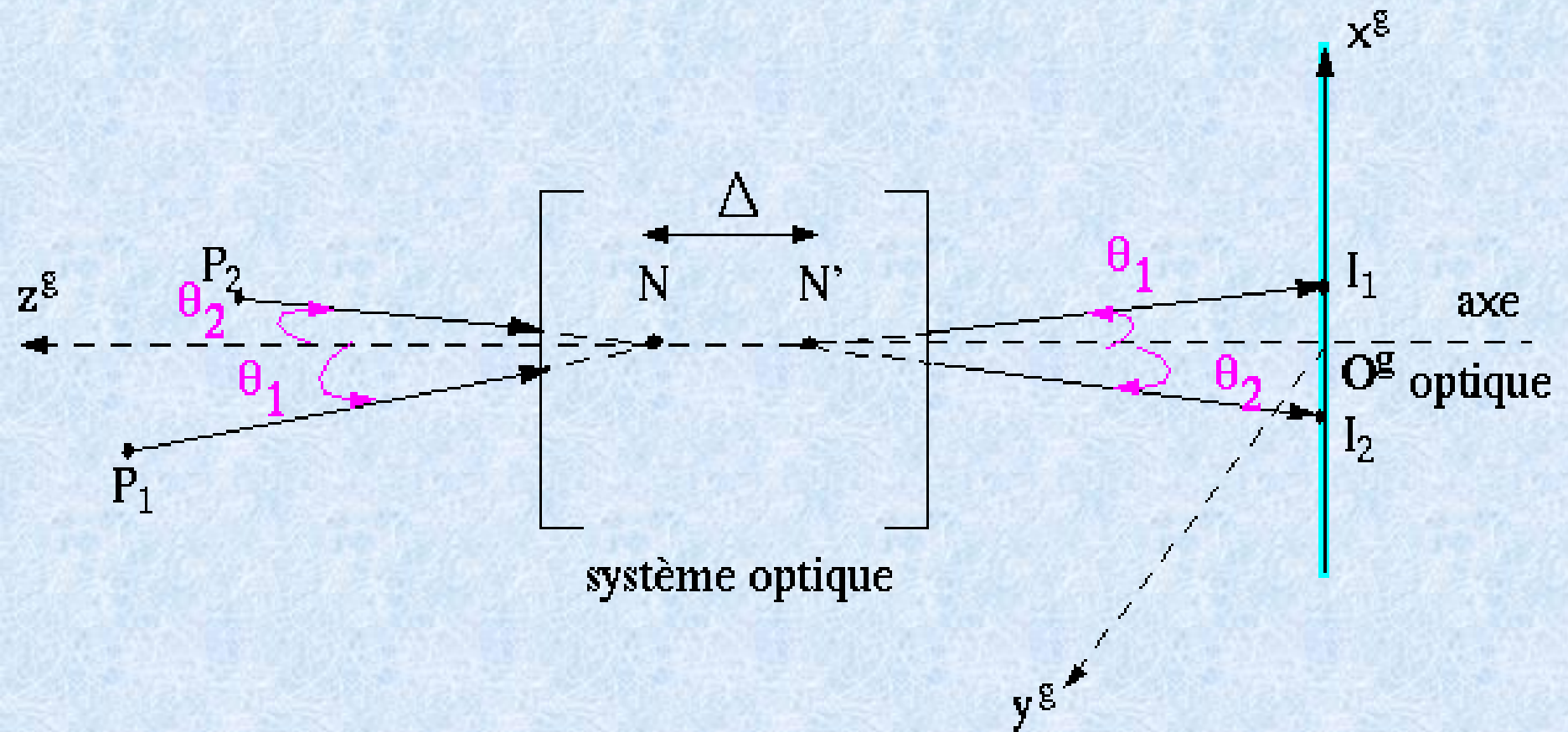
On suppose que les images sont nettes.

À ce niveau, on a fait dans le chapitre 3 l'hypothèse simplificatrice suivante :

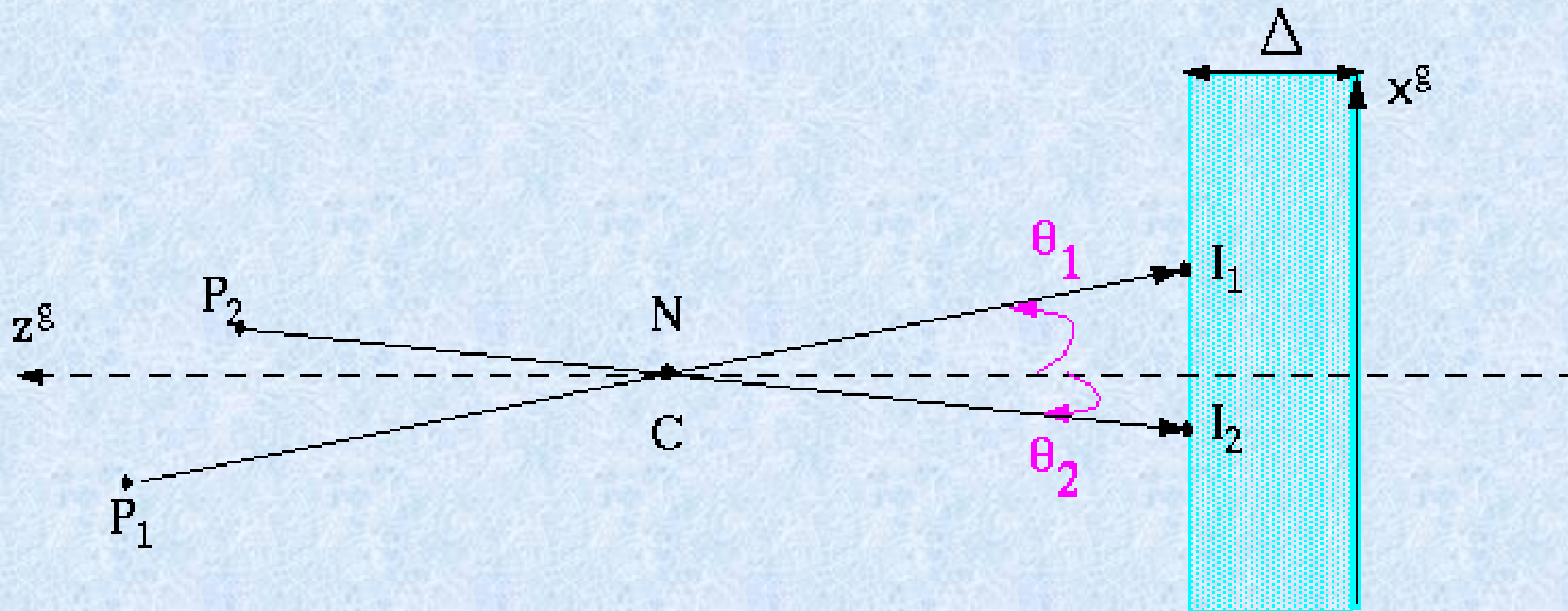
- Homothétie de rapport $g \neq 1$ et de centre O ;
- Projection sur le plan du récepteur.

Cette hypothèse est fautive. En fait, il s'agit d'une **projection perspective** et non, comme on l'avait supposé, d'une **projection orthogonale**.

1.2.a. Centre optique d'une caméra.



C'est équivalent à :

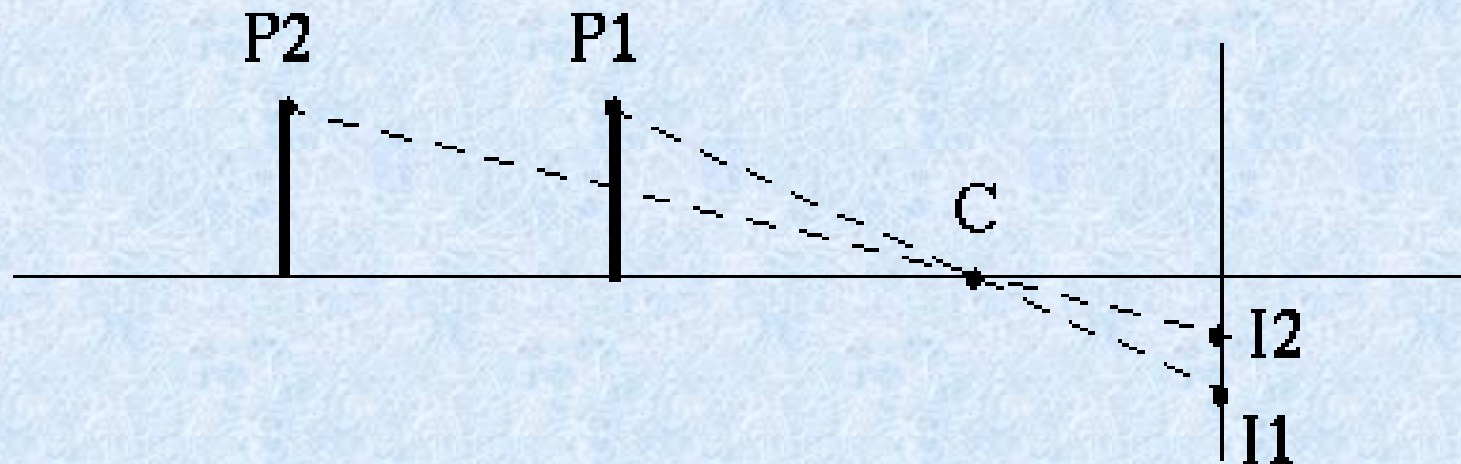


1.2.b. Projection perspective.

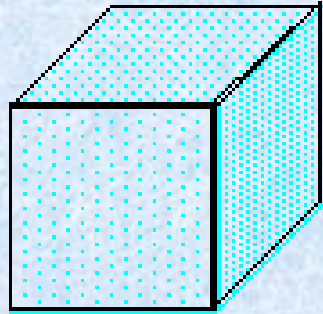
Pour trouver le pixel I correspondant à un point objet P , on prend l'intersection de la droite PN avec un plan obtenu par translation de D le long de l'axe optique du plan de la rétine.

On appelle pour cette raison N (point nodal objet) le **centre optique** de la caméra, noté C .

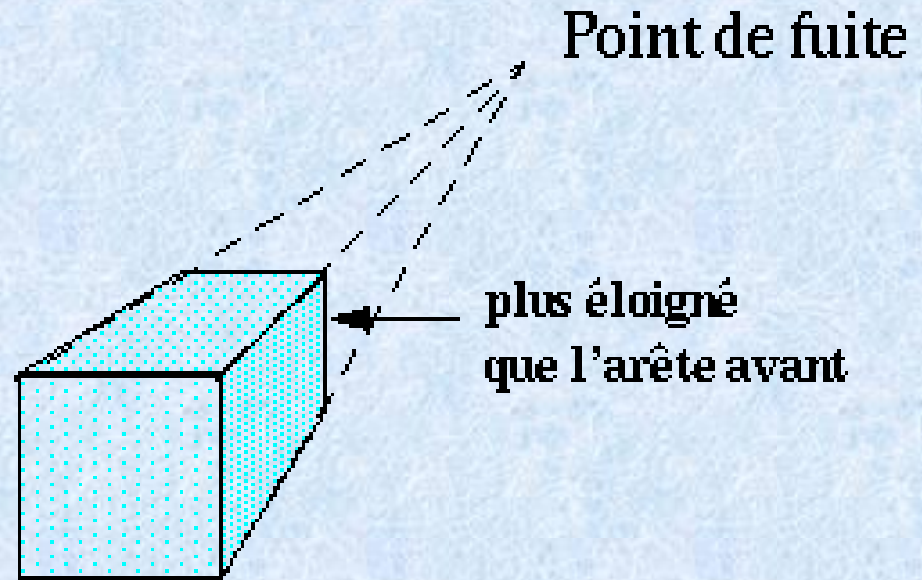
Cette transformation est appelée **projection perspective**. Une propriété élémentaire de la projection perspective est qu'entre deux segments identiques, le segment le plus proche de la caméra a une image plus grande :



Exemple d'un cube :



Projection orthogonale



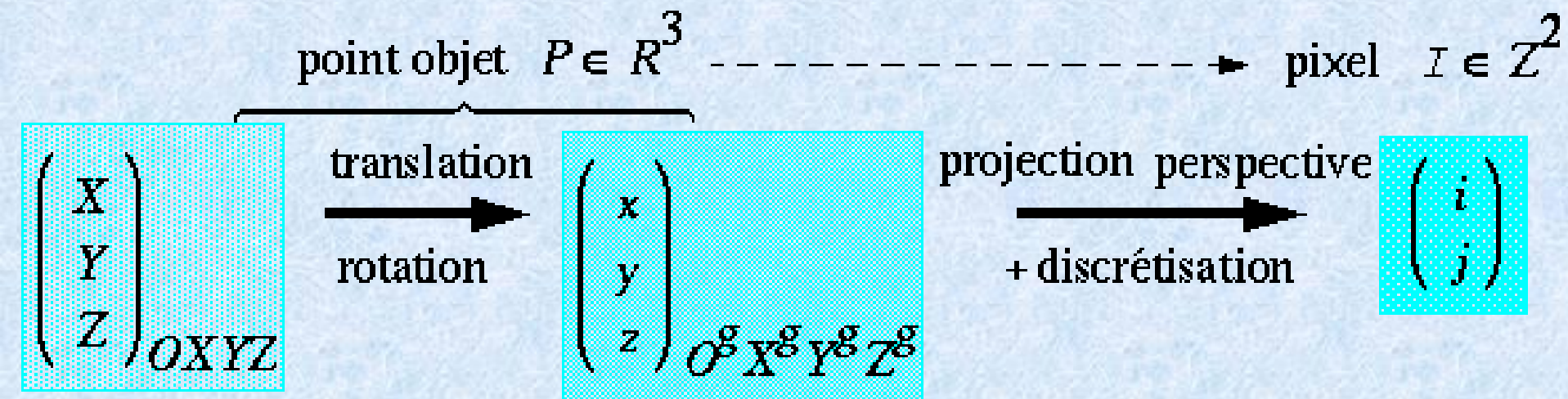
Projection perspective

Les propriétés de la projection perspective (exemple : point de fuite) sont nombreuses et compliquées. Nous n'en parlerons pas dans ce cours (voir un cours de DEA sur la "géométrie projective").

2. Le calibrage.

2.1. Relation mathématique entre points objets et pixels.

On s'intéresse par exemple à l'image gauche.



2.1.a. Translation + rotation.

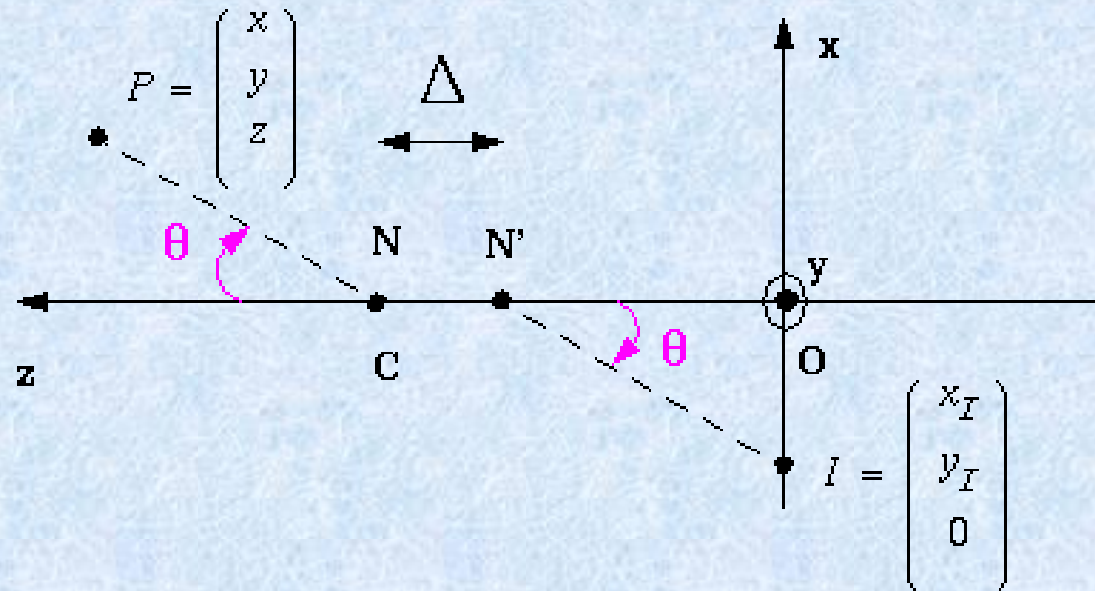
$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} \\ \\ \end{pmatrix} \begin{pmatrix} X \\ Y \\ Z \end{pmatrix} + \begin{pmatrix} \\ \\ \end{pmatrix} \quad (4.1)$$

matrice 3x3 M : vecteur \vec{V} :

rotation translation

Il faut connaître 6 paramètres (3 angles pour la matrice M, 3 coordonnées pour le vecteur \vec{V}), appelés "paramètres extrinsèques" de la caméra.

2.1.b. Projection perspective.



Les expressions de x_I et de y_I en fonction de x , y et z sont les suivantes :

(4.2)

$$\operatorname{tg}\theta = \frac{x_I}{z_{N'}} = \frac{x}{z - z_N} \Rightarrow x_I = z_{N'} \frac{x}{z - z_N}$$

(4.3)

$$= \frac{y_I}{z_{N'}} = \frac{y}{z - z_N} \Rightarrow y_I = z_{N'} \frac{y}{z - z_N}$$

2.1.c. Discrétisation.

Si O correspond à (i_0, j_0) , alors I correspond à (i, j) tel que :

$$(4.4) \quad x_I = (i - i_0)\delta_x \quad \Leftrightarrow \quad \begin{cases} i = \frac{x_I}{\delta_x} + i_0 = \frac{z_{N'}}{\delta_x} \frac{x}{z - z_N} + i_0 \\ j = \frac{y_I}{\delta_y} + j_0 = \frac{z_{N'}}{\delta_y} \frac{y}{z - z_N} + j_0 \end{cases}$$
$$(4.5) \quad y_I = (j - j_0)\delta_y$$

Problème :

Les formules (4.4) et (4.5) ne sont pas linéaires en x , y et z , contrairement à (4.1). La projection perspective n'est donc pas linéaire, contrairement à la projection orthogonale.

Astuce :

On utilise de nouvelles coordonnées, appelées "**coordonnées projectives**", définies par :

$$(4.6) \quad \left\{ \begin{array}{l} u = \frac{z_{N'}}{\delta_x} x + i_0(z - z_N) \end{array} \right.$$

$$(4.7) \quad \left\{ \begin{array}{l} v = \frac{z_{N'}}{\delta_y} y + j_0(z - z_N) \end{array} \right.$$

$$(4.8) \quad \left\{ \begin{array}{l} s = z - z_N \end{array} \right.$$

Remarques :

- Ces définitions sont linéaires en x, y, z .
- **6 paramètres intrinsèques**

2.1.d. Conclusion : matrice perspective.

Grâce aux calculs précédents :

$$\begin{pmatrix} u \\ v \\ s \end{pmatrix} = M' \begin{pmatrix} x \\ y \\ z \end{pmatrix} + \vec{V}' = M' M \begin{pmatrix} X \\ Y \\ Z \end{pmatrix} + M' \vec{V} + \vec{V}' \\ = N \begin{pmatrix} X \\ Y \\ Z \end{pmatrix} + \vec{W}$$

où N est une matrice 3 x 3.

On l'écrit plutôt :

$$(4.9) \quad \begin{pmatrix} u \\ v \\ s \end{pmatrix} = T \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix}$$

où T est une matrice à 3 lignes et 4 colonnes, appelée "**matrice perspective**" de la caméra. Ses 12 coefficients dépendent des 12 paramètres de la caméra :

- 6 paramètres **extrinsèques** ;
- 6 paramètres **intrinsèques**.

Par ailleurs :

$$(4.10) \quad i = \frac{u}{s}$$

$$(4.11) \quad j = \frac{v}{s}$$

Remarque :

$$\text{si } T' = \lambda T, \text{ alors } T' \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix} = \begin{pmatrix} \lambda u \\ \lambda v \\ \lambda s \end{pmatrix} \quad \text{et} \quad \begin{cases} \frac{\lambda u}{\lambda s} = i \\ \frac{\lambda v}{\lambda s} = j \end{cases}$$

donc on obtient le même couple (i,j) .

Cela signifie que T est déterminable à un facteur près. On peut fixer un coefficient à 1 :

$$T = \begin{pmatrix} \mathbf{x} & \mathbf{x} & \mathbf{x} & \mathbf{x} \\ \mathbf{x} & \mathbf{x} & \mathbf{x} & \mathbf{x} \\ \mathbf{x} & \mathbf{x} & \mathbf{x} & \mathbf{1} \end{pmatrix}$$

Il ne reste que 11 coefficients inconnus, dépendant des 12 paramètres de la caméra.

2.2. Détermination des paramètres de calibrage.

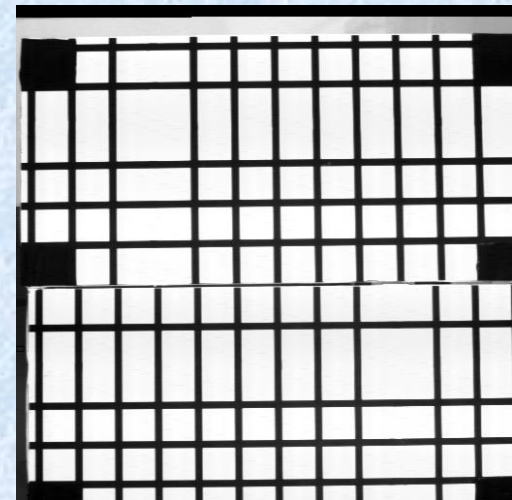
On doit déterminer les 12 paramètres de la caméra.

En général, on se contente de déterminer la matrice T (c'est-à-dire 11 coefficients) pour chaque caméra.

Il suffit de connaître 6 points P_k de la scène, leurs coordonnées $\begin{pmatrix} X_k \\ Y_k \\ Z_k \end{pmatrix}$,
et les pixels $\begin{pmatrix} i \\ j \end{pmatrix}_k^g$ et $\begin{pmatrix} i' \\ j' \end{pmatrix}_k^d$ qui leur sont associés.

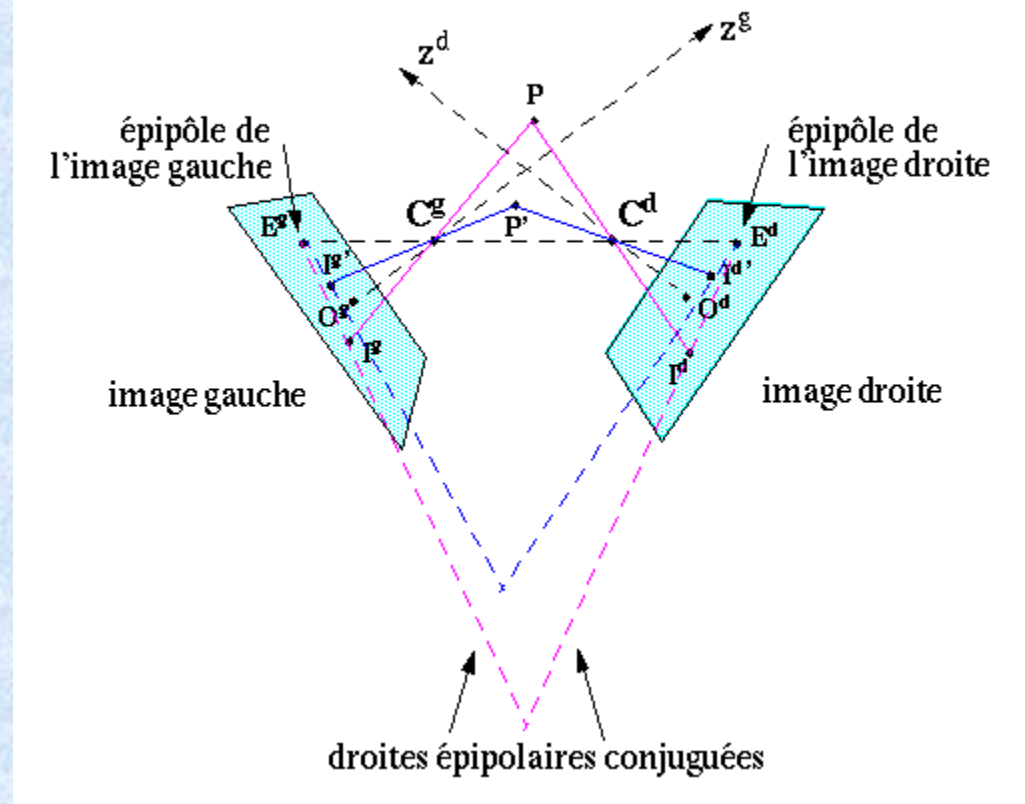
On peut montrer que ces points ne doivent pas tous être coplanaires.

À l'IRIT, on utilise une grille de calibrage, sur laquelle il est facile de prendre des repères :



3. L'appariement.

3.1. La contrainte épipolaire : cas général.



Si on cherche le pixel I^d correspondant à un pixel I^g de l'image gauche, on peut dire que I^d appartient nécessairement à une droite de l'image droite, appelée "**droite épipolaire**" associée à I^g , et qui est l'intersection du plan $I^g C^g C^d$ et de l'image droite.

Comme il y a symétrie entre les deux images, on construit ainsi des paires de droites épipolaires, qui passent toutes par un des épipôles E^g et E^d .

Dans la pratique :

- On commence par calculer la position des épipôles E^g et E^d .
- Ensuite, pour un pixel I^g par exemple, on calcule un pixel I^d situé sur la droite épipolaire associée à I^g .
- On cherche ensuite à apparier I^g avec un pixel de cette droite épipolaire, en utilisant :
 - la **corrélation** ;
 - la contrainte d'**unicité** ;
 - la contrainte de **seuil** ;
 - et parfois la contrainte d'**ordre**.

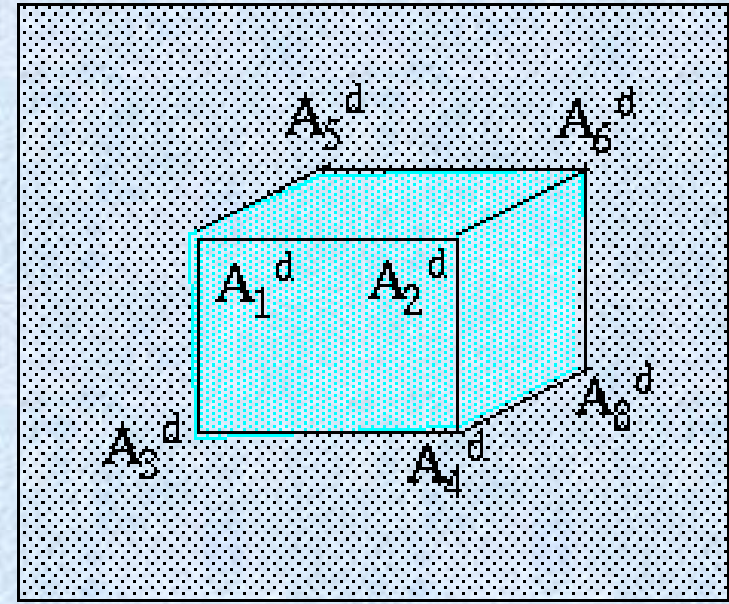
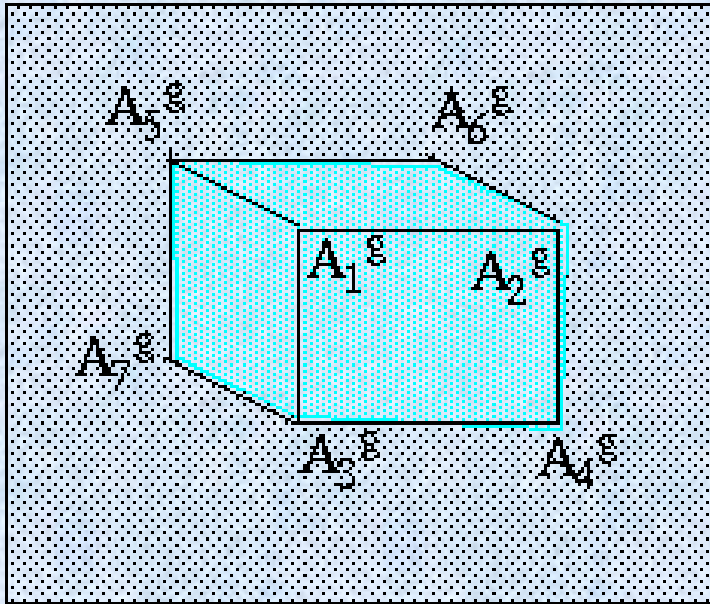
3.2. Appariement dense et appariement non dense.

Lors des séances de TP :

$$\frac{\text{nombre de paires}}{\text{nombre de pixels d'une image}} \approx 1$$

Grâce à la texture, il est possible d'effectuer un appariement dense. En revanche, si l'image n'est pas texturée, ce quotient peut être très faible.

Exemple :



Entre ces deux images, on peut mettre en correspondance :

- Les angles A_1 , A_2 , A_3 , A_4 , A_5 , A_6 .
- Les arêtes A_1A_2 , A_1A_3 , A_2A_4 , A_3A_4 , A_1A_5 , A_2A_6 , A_5A_6 .

L'appariement réalisé est un appariement non dense.

En général, on apparie les contours, les arêtes, les angles, lorsqu'on fait de l'appariement non dense.

La tendance actuelle est d'effectuer plutôt de l'appariement dense, car les performances des calculateurs ne cessent d'augmenter.

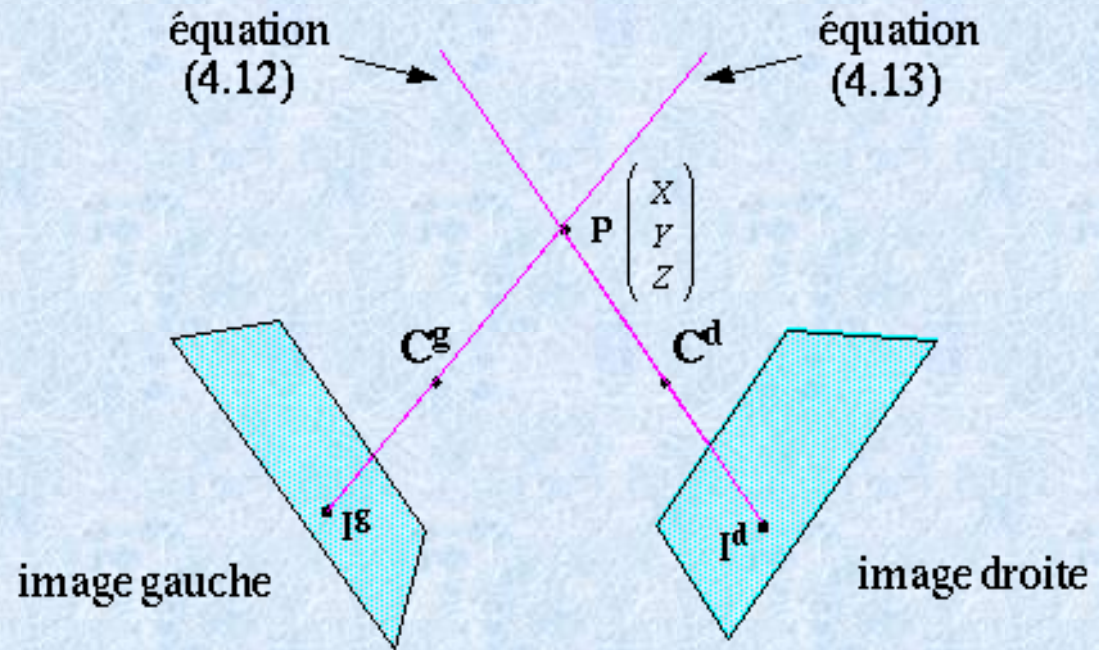
4. La triangulation.

On a les relations suivantes :

$$T^g = \begin{pmatrix} u^g \\ v^g \\ s^g \end{pmatrix} = T^g \begin{pmatrix} X \\ Y \\ Z \end{pmatrix} \quad (4.12)$$

$$T^d = \begin{pmatrix} u^d \\ v^d \\ s^d \end{pmatrix} = T^d \begin{pmatrix} X \\ Y \\ Z \end{pmatrix} \quad (4.13)$$

On sait qu'un point $P = \begin{pmatrix} X \\ Y \\ Z \end{pmatrix}$ correspond à cette paire de pixels, en coordonnées projectives. On veut connaître X , Y et Z .



Le système composé des équations (4.12) et (4.13) a une solution et une seule : P , le point cherché. L'expression mathématique des coordonnées de P est un peu complexe, et nous ne la donnerons pas ici.

Dans le prochain paragraphe, nous allons citer deux exemples de stéréovision provenant du milieu industriel.