

MASTER IPCC



Nicolas Loménie

Vision par Ordinateur 3D en Mouvement

MASTER IPCC



Vision par Ordinateur

Nicolas Loménie

STEREOVISION

STEREOVISION

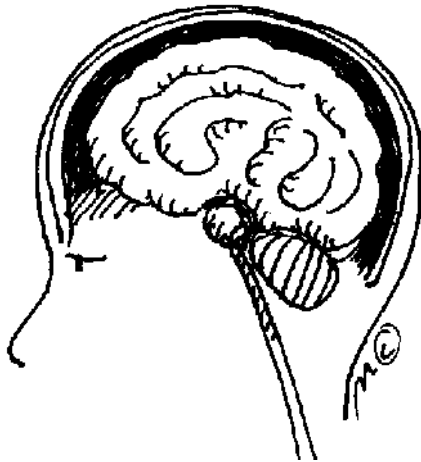
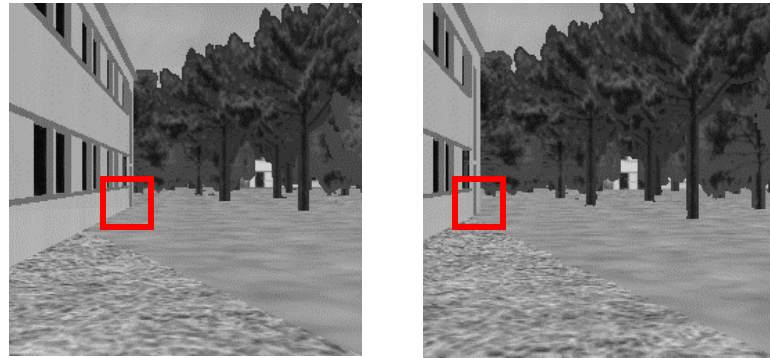
Objectifs :

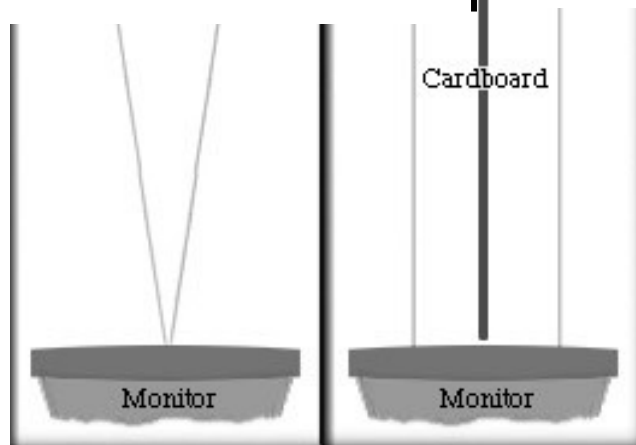
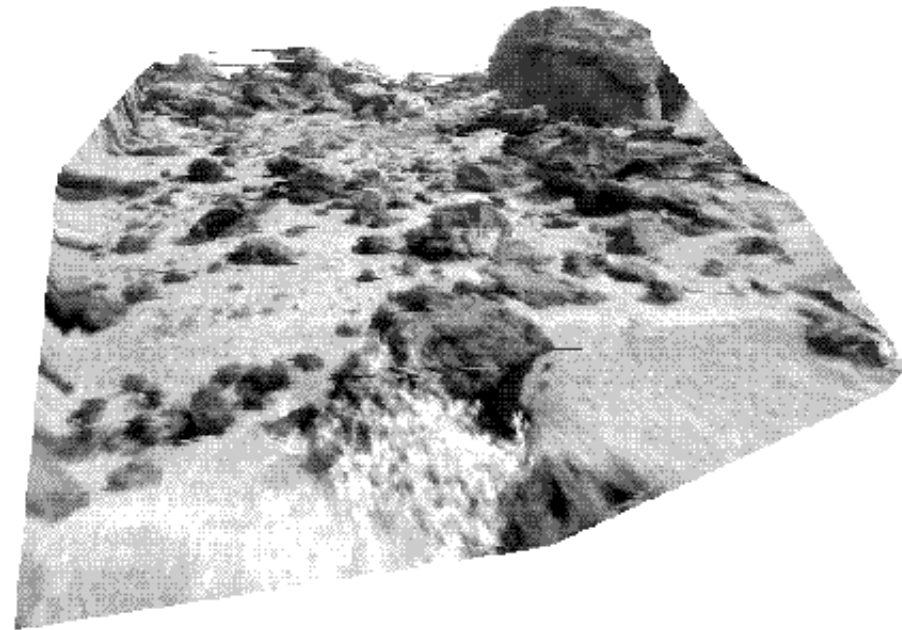
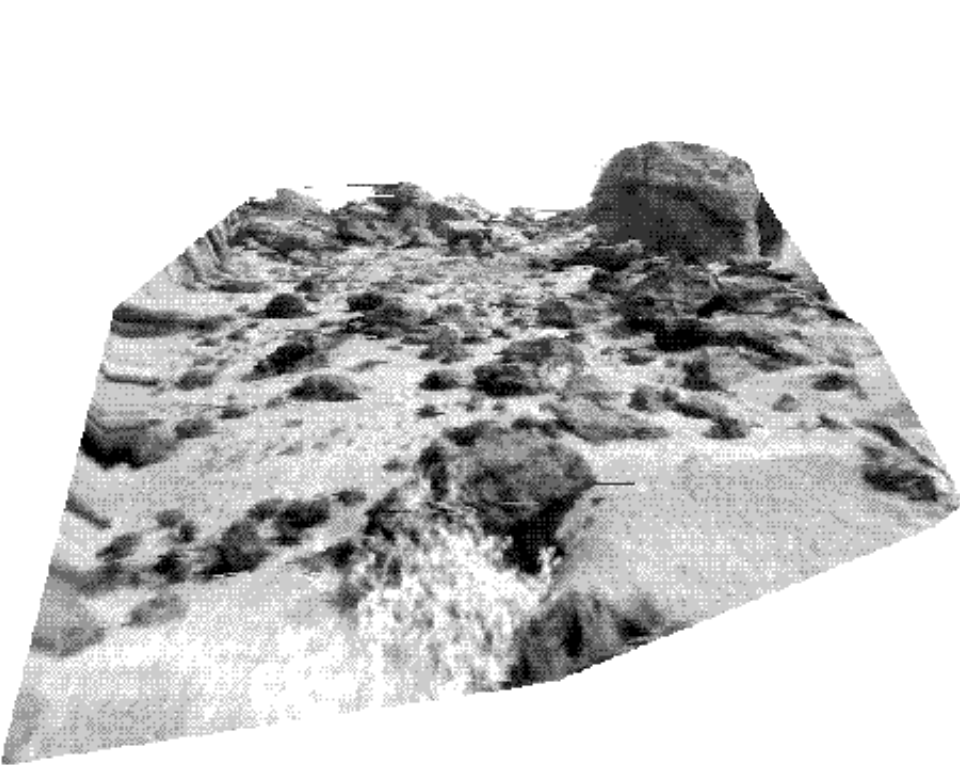
- Comprendre les deux sous-problèmes liés à la stéréovision : mise en correspondance et reconstruction;
- Être capable d'implémenter des algorithmes de base de mise en correspondance d'éléments d'image d'une paire stéréoscopique;
- Comprendre le concept de géométrie épipolaire, et être capable de travailler dans ce cadre;
- Exposer quelques méthodes de reconstruction 3D;



Problématique :

La vision stéréoscopique fait référence à la capacité d'inférer de l'information sur la structure 3D et les distances 3D d'une scène à partir d'au moins deux images optiques prises de points de vue différents

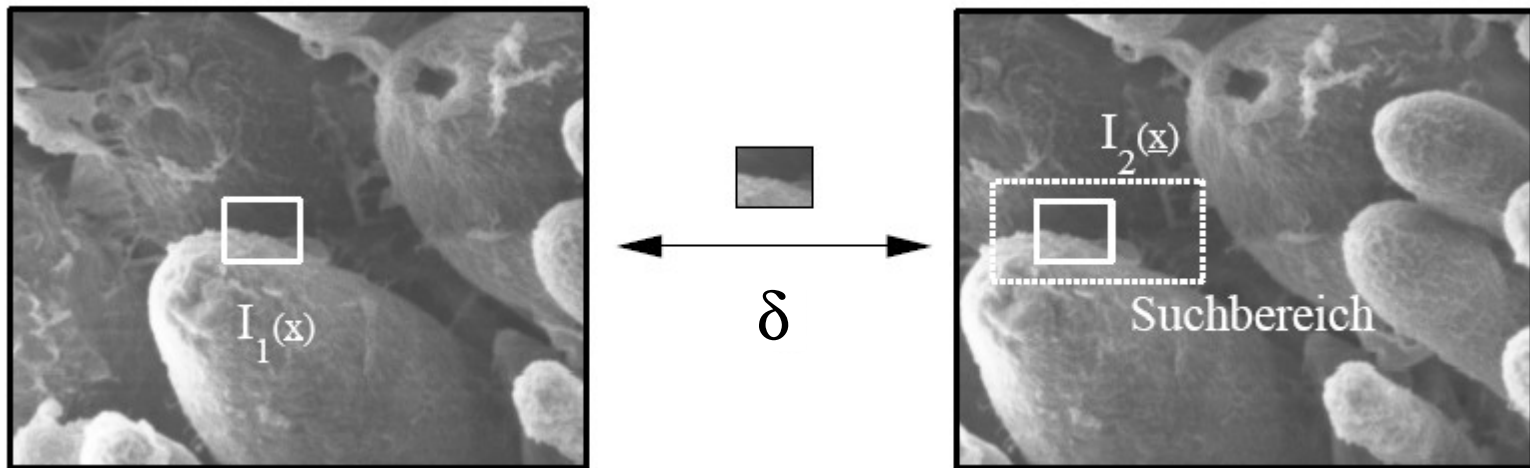




Stéréogrammes

Les 2 Problématiques Informatiques

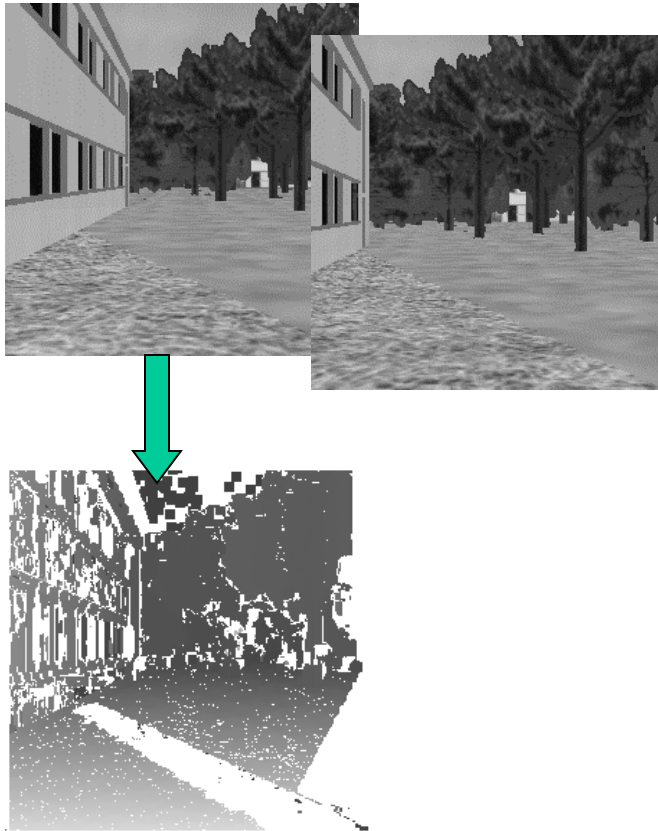
1. Le problème de la mise en correspondances (recalage) :
Quelles parties des images droite et gauche sont des projections du même élément de la scène ?



Remarque : problème des occlusions du fond par exemple par votre doigt et donc des parties qui ne doivent pas être mises en correspondances

Obtention d'une carte des disparités δ :

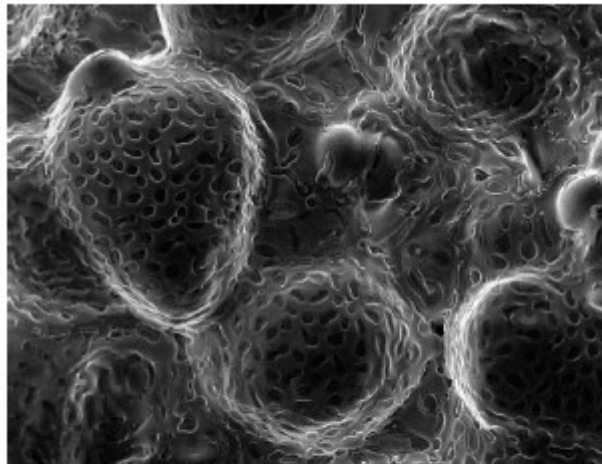
chaque pixel correspond à une information de profondeur estimée pour chaque point 3D correspondant, cad sa coordonnée manquante en Z pour reconstruire le point.



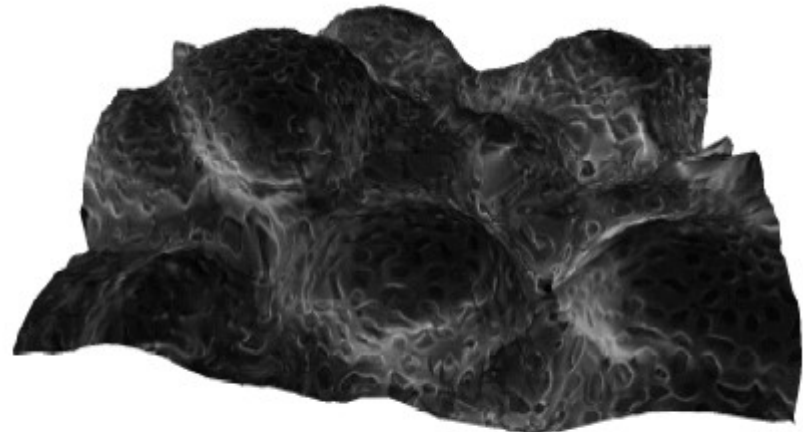
Les 2 Problématiques Informatiques

2. Le problème de la reconstruction :

Étant donné un certain nombre de parties correspondantes dans les images droite et gauche et éventuellement des informations sur la géométrie du système stéréoscopique, que pouvons-nous dire au sujet de la localisation et de la structure des objets observés ?



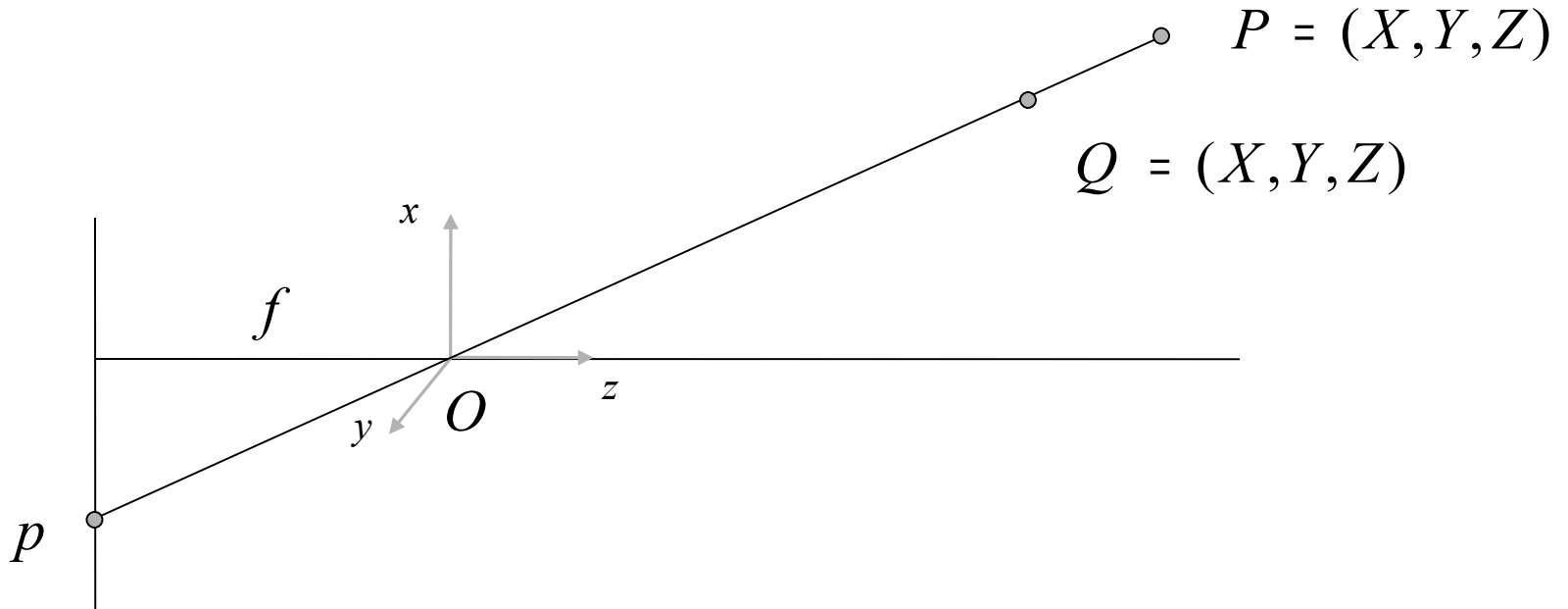
40 μm



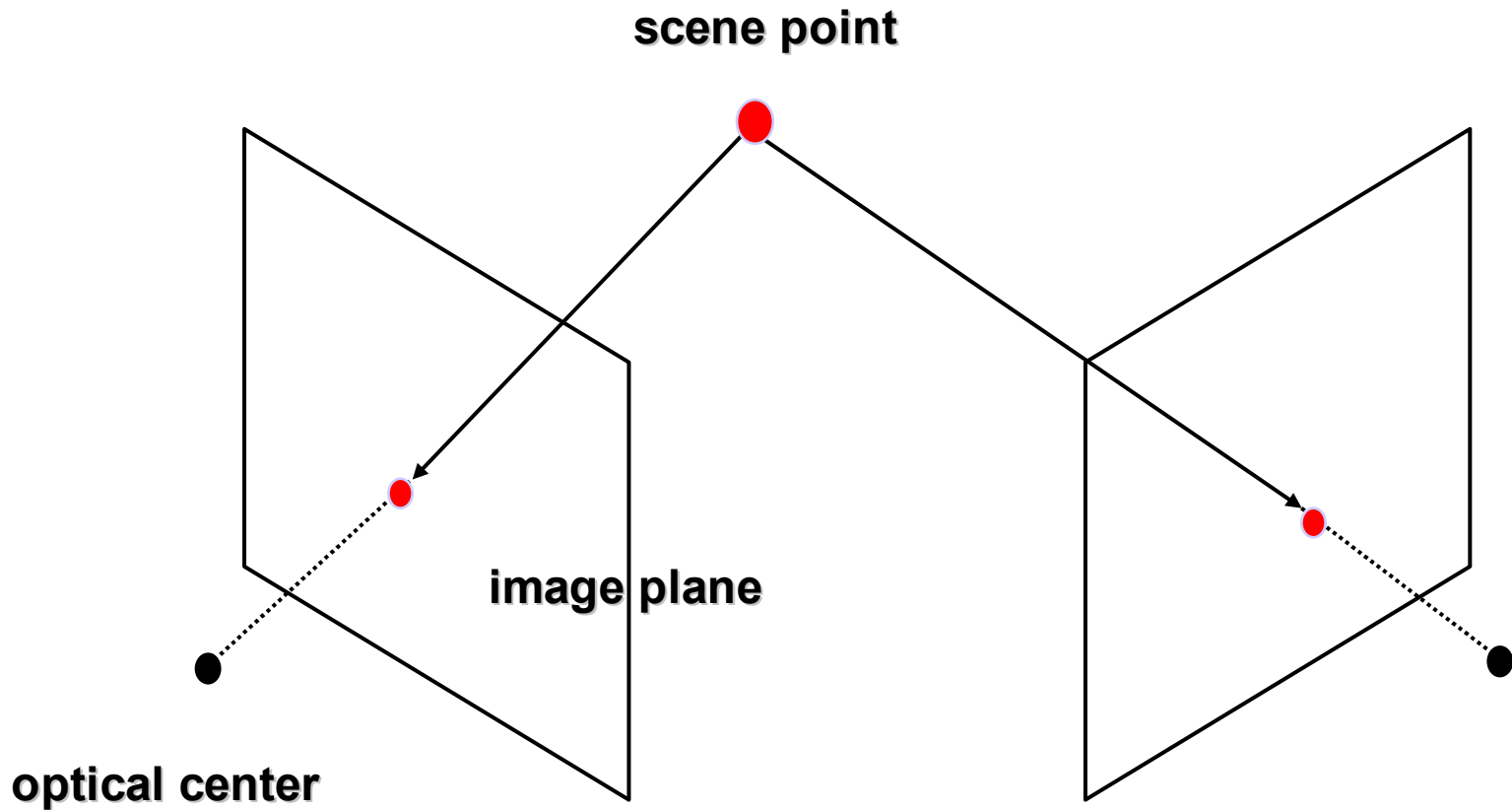
Avant la machinerie informatique, un système stéréo simple

$p \leftrightarrow P$?

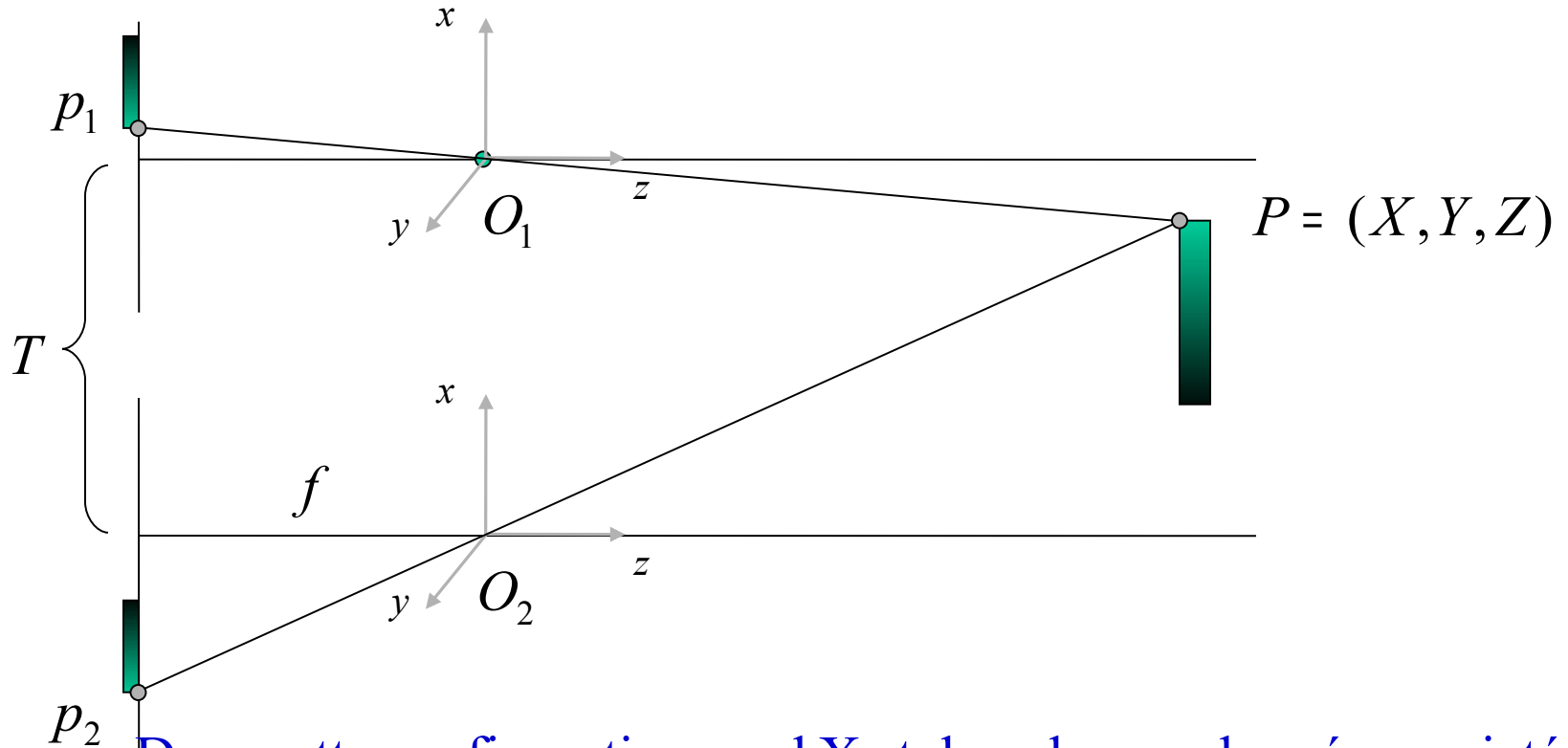
$p \leftrightarrow Q$?



Le problème est en partie réglé par l'usage de deux caméras décalés par exemple. De façon générale,



Avant la machinerie informatique, un système stéréo simple



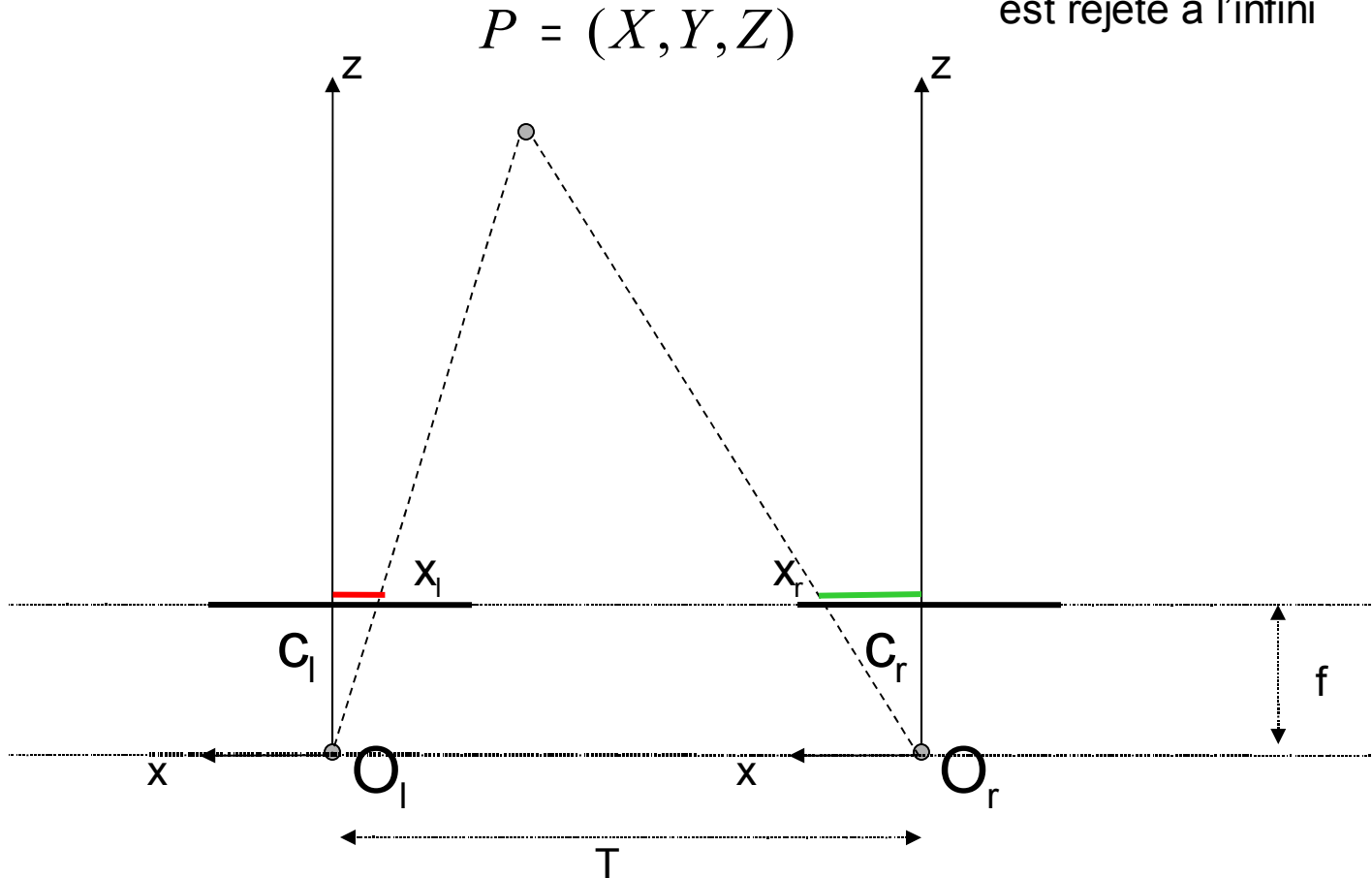
Dans cette configuration, seul X et donc la coordonnées projetée x est différente d'un repère caméra à l'autre.

Exprimer la profondeur du point Z en fonction de x_1, x_2, f, T ?

Avant la machinerie informatique, un système stéréo simple

En coupe 2D,

Le point de fixation,
intersection des deux axes optiques,
est rejeté à l'infini



Exprimer la profondeur du point Z en fonction de x_l, x_r, f, T

Avant la machinerie informatique, un système stéréo simple

La stéréovision détermine les positions spatiales de P par triangulation mais attention à la mise en correspondance des points.

Si la m.e.c. est correcte, pour la reconstruction 3D, il est intéressant d'écrire les équations de la triangulation.

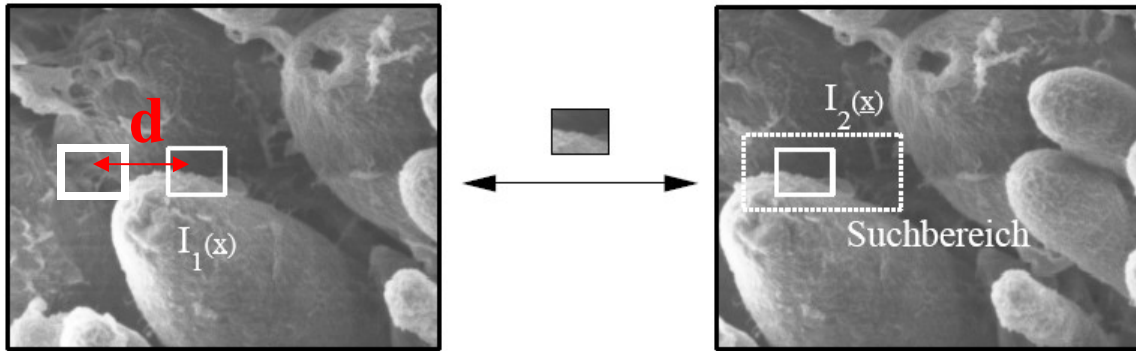
Dans les triangles similaires $p_l P p_r$ et $O_l P O_r$, on a :

$$\frac{T + x_l - x_r}{Z - f} = \frac{T}{Z}$$

$$d'où \quad Z = f \frac{T}{d} \quad \text{avec} \quad d = x_r - x_l$$

Avant la machinerie informatique, un système stéréo simple

La **disparité** $d=x_r-x_l$ mesure la différence de position rétinienne entre des points se correspondant dans deux images.



La **profondeur est inversement proportionnelle à la disparité** :
des objets distants semblent bouger moins vite que des objets proches.

Les coordonnées x_l et x_r font références aux points principaux c_l et c_r . En conséquence, les paramètres d'un système stéréo sont f , T , c_l et c_r .

Paramètres d'un système stéréo

Les paramètres intrinsèques caractérisent la transformation qui associe pour un point image les coordonnées caméra aux coordonnées pixel, dans chaque caméra.

Les paramètres extrinsèques décrivent les position et orientation relatives des deux caméras.

Déterminer ces paramètres constitue le problème de la calibration d'un système stéréoscopique.

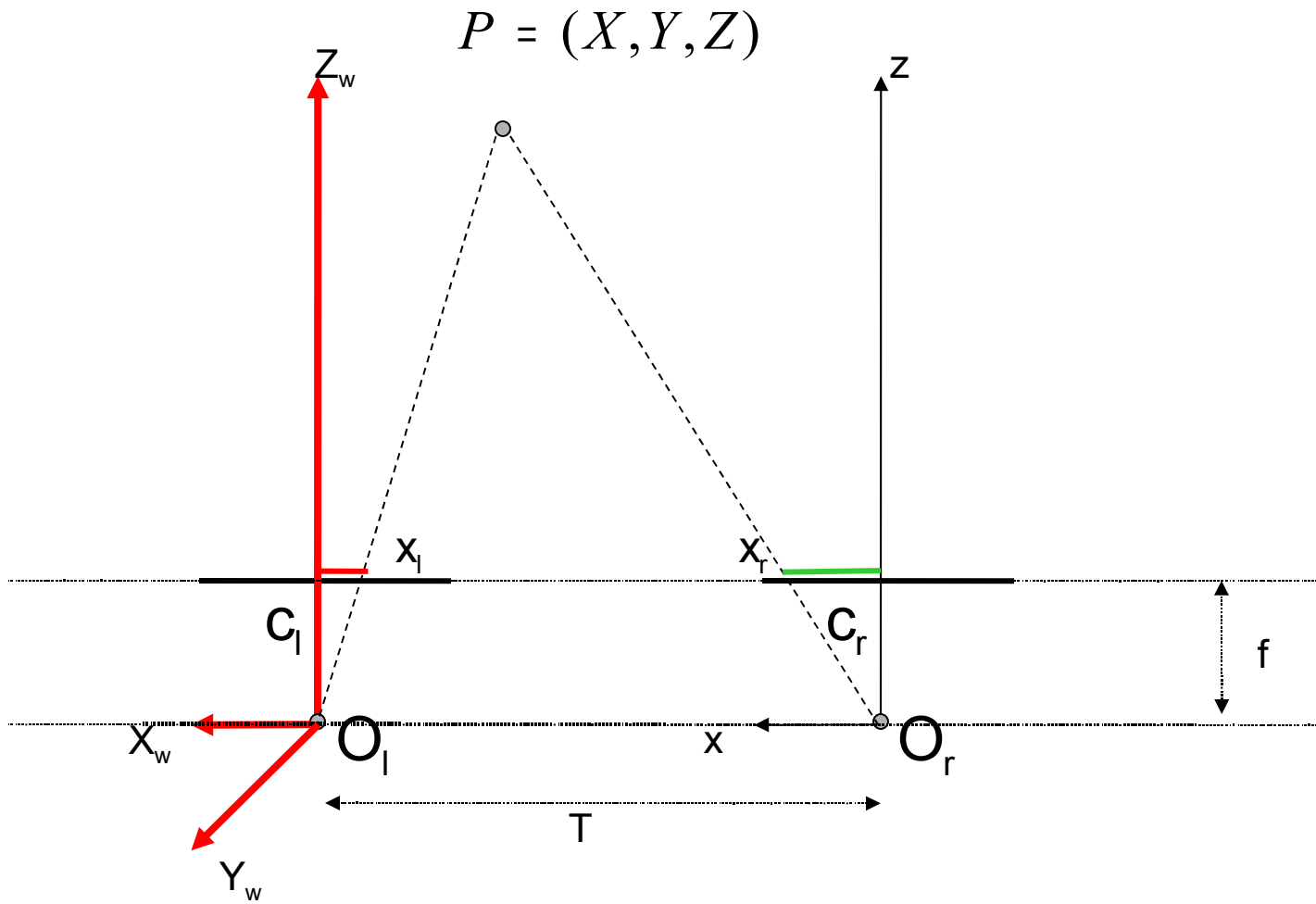
Paramètres extrinsèques d'un système stéréo

Seuls ces paramètres sont légèrement différents de ceux associés à une caméra seule.

Ici, le référentiel monde est le référentiel d'une des deux caméras.

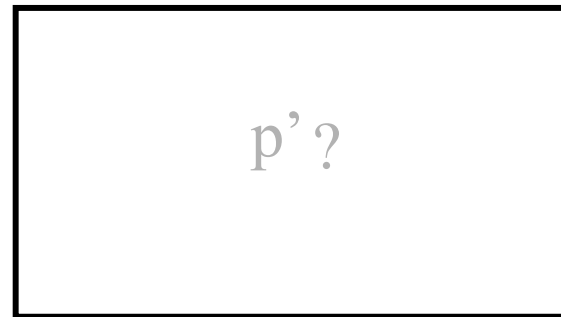
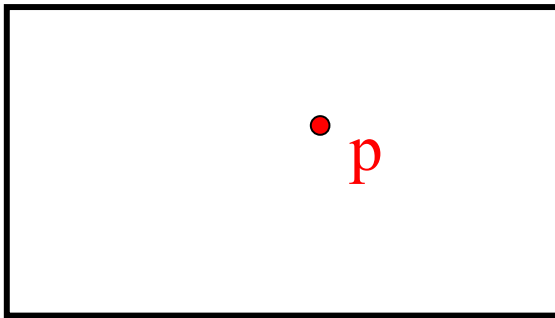
Comme souvent, la plupart de ces paramètres est inconnue. Et la reconstruction est souvent un problème de calibration.

Pourtant, on verra qu'un système stéréo peut calculer un grand nombre d'informations 3D sans connaissance *a priori* des paramètres stéréo (*uncalibrated stereo*).



Le problème de la correspondance

Supposant le problème de calibration résolu, étant donné un point p dans l'image gauche, comment trouver son correspondant dans l'image droite ?



Hypothèses

3. La plupart des points scène sont visibles des deux points de vue.
5. Les régions image correspondantes sont similaires.

Remarque : Validité si $d(\text{Point de Fixation}, \text{Caméra}) \gg T$

Etant donné un élément dans l'image gauche, on cherche l'élément correspondant dans l'image droite.

Pour cela, il faut prendre deux décisions :

3. Quel élément d'image mettre en correspondance ?
4. Quelle mesure de similarité adopter ?

Le problème de la correspondance

On classifie les algorithmes de mise en correspondance en
2 catégories :

5. Ceux **Basés sur la corrélation** (correlation based) : ces méthodes s'appliquent de façon dense à la totalité des points images :

-> obtention d'une **carte dense de disparités**;

7. Ceux **Basés sur les primitives** (feature based) : ces méthodes s'appliquent à un ensemble éparé de primitives images;

Méthodes orientées corrélation

Dans ces méthodes, les éléments à mettre en correspondance sont des fenêtres images de taille fixe, et le critère de similarité est une mesure de la corrélation entre fenêtres dans les deux images.

L'élément correspondant est donné par la fenêtre qui maximise le critère de similarité à l'intérieur d'une région de recherche.

Le problème de la correspondance

Recherche exhaustive ?

Left



Right



Algo CORR_MATCHING

INPUT :

- Paire stéréo d'images I_l et I_r .
- Soient p_l et p_r les pixels dans l'image droite et gauche. $2W+1$ la taille en pixels de la matrice de corrélation. $R(p_l)$ la région de recherche dans l'image droite associée à p_l .
- Soit $\Psi(u,v)$ une fonction de deux valeurs de pixel.

Pour chaque pixel $p_l=[i,j]^T$ de l'image gauche :

7. Pour chaque déplacement $d=[d_1, d_2]^T \in R(p_l)$ calculer :

$$c(d) = \sum_{k=-W}^W \sum_{k=-W}^W \Psi (I_l(i+k, j+1), I_r(i+k-d_1, j+1-d_2))$$

8. La disparité de p_l est le vecteur $\bar{d} = [\bar{d}_1, \bar{d}_2]^T$ qui maximise $c(d)$ sur $R(p_l)$:

$$\bar{d} = \underset{d \in R}{\text{Arg max}} \{c(d)\}$$

OUTPUT : tableau de disparités (disparity map), une pour chaque pixel de I_l .

$\Psi(u, v) = uv$: corrélation croisée entre la fenêtre dans l'image gauche et la région de recherche dans l'image droite

$\Psi(u, v) = -(u - v)^2$: somme des différences au carré ou block matching

Remarque : stocker les valeurs off-line des $\Psi(u,v)$ dans une LUT permet de se rapprocher du temps réel. Le choix de W est très variable. En ce qui concerne R , il y a plus de choses à dire (*voire la géométrie épipolaire plus loin*).

On a $\Psi(u,v) = -\frac{(u-v)^2}{2uv} = \frac{2uv - u^2 - v^2}{2uv}$: donc si l'énergie dans la fenêtre de recherche droite est constante ($\sum v^2$), les deux mesures de corrélations fournissent le même résultat.

C'est rarement le cas. On préférera la deuxième mesure qui n'est pas biaisée par la présence de régions à très petites ou au contraire très grandes variations d'intensité.

On remarquera également qu'une mesure C normalisée type variance serait préférable. Mais, le stockage dans une LUT pour le temps réel n'est plus alors possible.



Méthodes orientées primitives

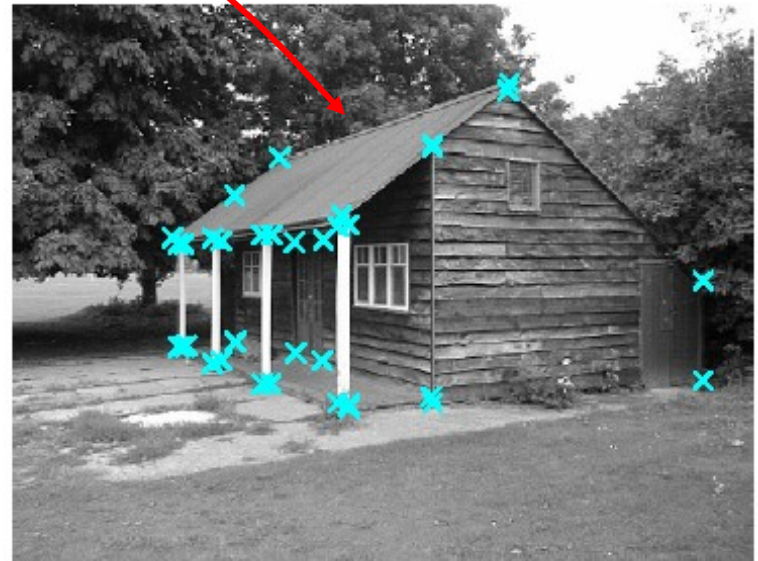
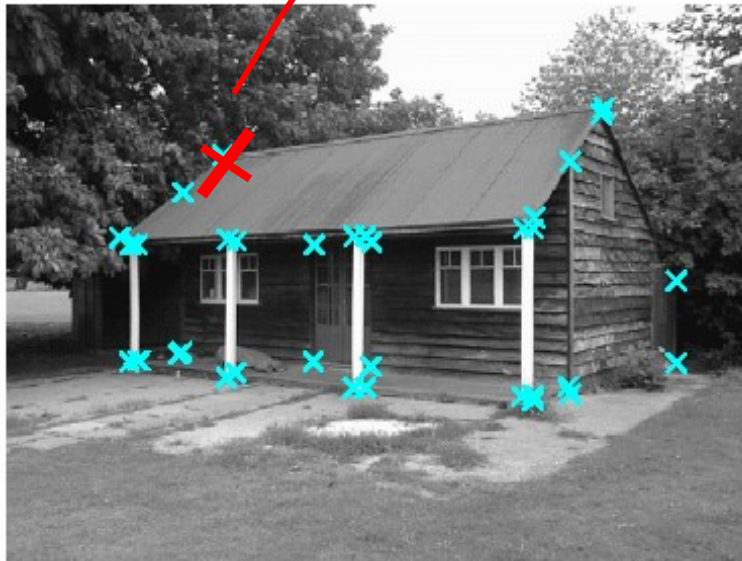
Ces méthodes restreignent la recherche des correspondances à un ensemble épars de primitives. A la place de fenêtres images, elles utilisent des propriétés symboliques et numériques des primitives, au moyen des descripteurs de primitives.

A la place des mesures de corrélation, elles utilisent des mesures de distance entre descripteurs de primitives.

Les éléments correspondants sont donnés par les paires de primitives les plus similaires.

Le problème de la correspondance

Recherche exhaustive ?



Méthodes orientées primitives

La plupart des méthodes réduit le nombre de correspondants possibles pour chaque primitive par renforcement de contraintes sur des m.e.c possibles. Ces contraintes peuvent être :

- Géométriques : comme les contraintes épipolaires
- Analytiques : comme la contrainte d'unicité ou celle de continuité (variation continue de la disparité presque partout dans l'image)

On va s'intéresser aux méthodes non contraintes. Il n'existe pas de primitive unique fonctionnant dans toutes les situations. Donc et malheureusement, pour toutes primitives possibles, il y a une longue liste d'algorithmes de m.e.c.

Méthodes orientées primitives

Considérons le cas d'une m.e.c. de droites contours avec le descripteur suivant :

- longueur l ;
- orientation θ ;
- coordonnée du milieu m du segment;
- contraste moyen c le long de la droite contour.

Un exemple de critère de similarité est :

$$S = \frac{1}{1 + w_0(l_r - l_l)^2 + w_1(\theta_r - \theta_l)^2 + w_2(m_r - m_l)^2 + w_3(c_r - c_l)^2}$$

Algo FEATURE_MATCHING

INPUT :

- Paire stéréo d'images I_l et I_r .
- Deux ensembles correspondants de descripteurs de primitives.
- Soit $R(f_l)$ la région de recherche dans l'image droite associée avec le descripteur f_l , et $d(f_l, f_r)$ la disparité entre deux primitives correspondantes f_l et f_r .

Pour chaque primitive f_l dans l'image gauche :

7. Calculer la similarité entre f_l et chaque primitive image dans $R(f_l)$.
8. Sélectionner la primitive de l'image droite qui maximise la mesure de similarité.
9. Enregistrer la correspondance et la disparité associée à f_l , cad le déplacement entre les points définissant la position de la primitive $[c_r, d_2]^T$.

OUTPUT : liste de primitives se correspondant et une carte de disparités.

Bilan Correlation vs. Features

Corrélation :

- Facile à implémenter;
- Fournit une Carte de disparité dense;
 - Nécessite des images texturés;
- Nécessite des paires pas trop éloignées;
- Assez coûteuse à cause de la précision sub-pixellique souvent nécessaire.

Features :

- Intéressant en environnement intérieur ou structuré;
 - Moins coûteuse a priori;
- Moins sensible aux changements d'illumination.

Bilan Correlation vs. Features

L'enjeu de ces méthodes est d'être robuste aux occlusions et aux m.e.c inopportunes. Des contraintes appropriées réduisent ces effets :

- Seules les correspondances trouvées de g. à d. et de d. à g. sont acceptées (*left-right consistency constraint*)
 - **Contrainte épipolaire**

GEOMETRIE EPIPOLAIRE

Utile pour définir quelles informations sont nécessaires pour réaliser la recherche d'éléments correspondants le long de droites images

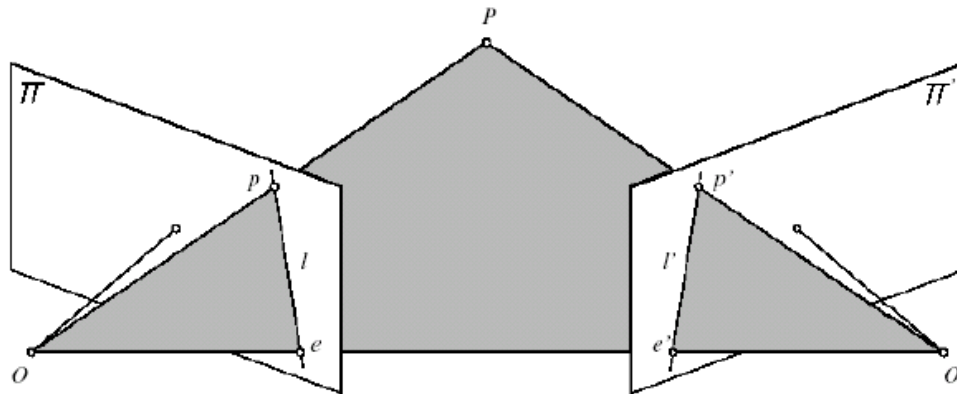
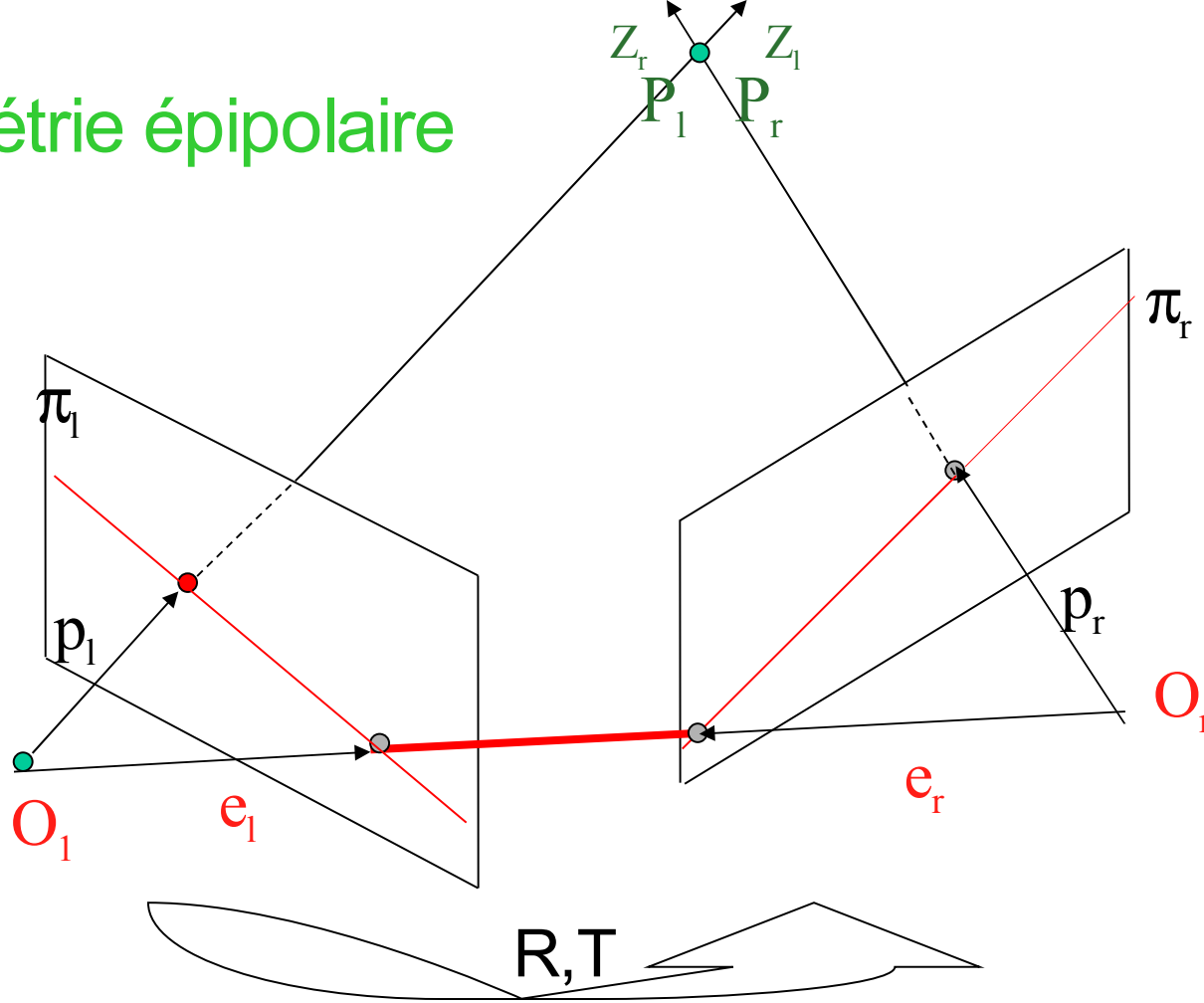


FIGURE 11.1: Epipolar geometry: the point P , the optical centers O and O' of the two cameras, and the two images p and p' of P all lie in the same plane.

All epipolar lines contain epipole, the image of other camera center.

Géométrie épipolaire



$$p_l = \frac{f_l}{Z_l} P_l$$

et $p_r = \frac{f_r}{Z_r} P_r$

$$P = P_l = P_r$$

Les vecteurs $P_l = [X_l, Y_l, Z_l]^T$ et $P_r = [X_r, Y_r, Z_r]^T$ se réfèrent au même point 3D P , vu comme un vecteur **dans les 2 repères caméras gauche et droit**.

Les vecteurs $p_l = [x_l, y_l, z_l]^T$ et $p_r = [x_r, y_r, z_r]^T$ se réfèrent aux projections 2D³⁵ de P , sur les plans images q . et d . et sont exprimés dans les mêmes

Géométrie épipolaire

Pour tous les points images, on a $z_l=f_l$ et $z_r=f_r$. Comme chaque plan image peut être vu comme un sous-ensemble de l'espace projectif P^2 , les points images peuvent également être vus comme des points de l'espace projectif P^2 .

Les repères caméras g_l et g_r sont reliés via les paramètres extrinsèques, qui définissent une transformation rigide de l'espace 3D, définie par une translation T et une matrice rotation R .

$$T = \overrightarrow{O_l O_r} = O_r - O_l$$

La relation entre P_l et P_r est ainsi : $P_r = R(P_l - T)$

La géométrie épipolaire doit son nom au fait que les points d'intersection de la droite reliant les centres de projection ($O_l O_r$) intersecte les plans images en des points appelés **épipôles**.

Remarque : si ($O_l O_r$) est parallèle à un des plans images, l'épipôle correspondant est le point à l'infini de cette droite.

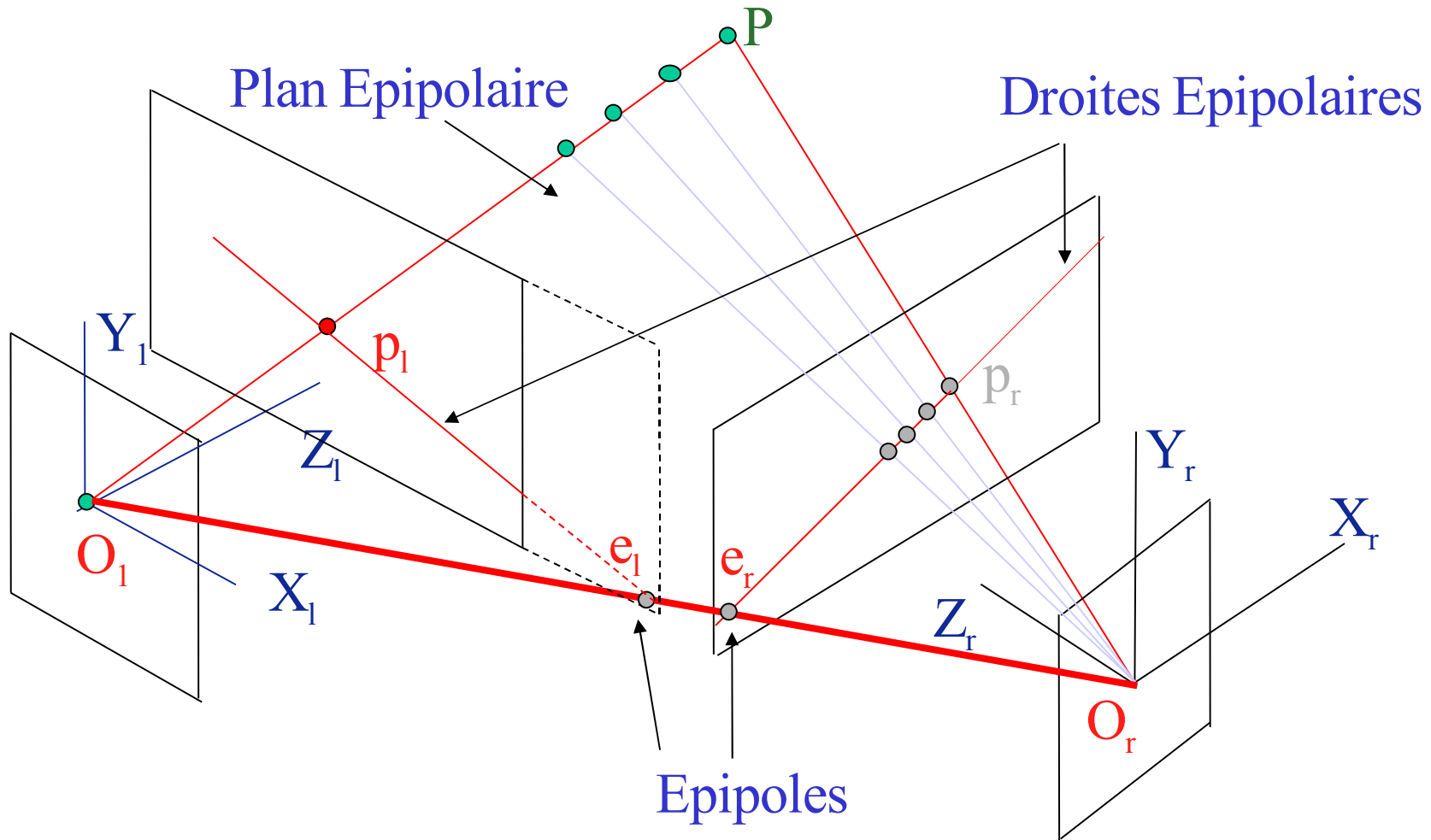
Géométrie épipolaire

Le plan (PO_1O_r) appelé **plan épipolaire** intersecte chaque plan image selon une droite appelée **droite épipolaire**.

Ainsi étant donné p_l , P peut se trouver n'importe où sur la droite O_1p_l , mais comme l'image de cette droite sur l'image droite est la droite épipolaire indiquée, $(e_r p_r)$, la m.e.c. correcte de p_l doit se trouver sur cette droite épipolaire.

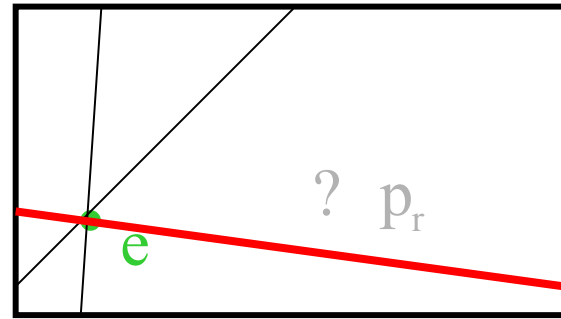
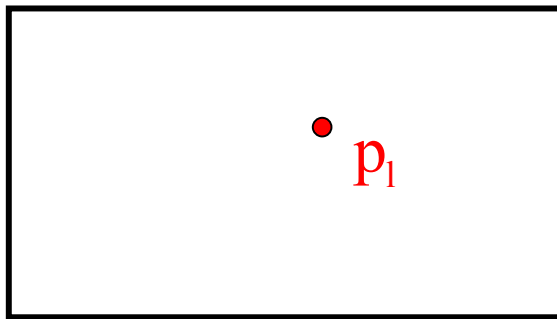
C'est la **contrainte épipolaire**, qui établit une correspondance entre les points de l'image gauche et les droites de l'image droite et *vice versa*.

Géométrie épipolaire



Géométrie épipolaire

Remarque : toutes les droites épipolaires passent par l'épipole, et le problème de la m.e.c peut se réduire à un problème de recherche 1D le long de la droite épipolaire associée à un point.



Question : comment estimer cette géométrie épipolaire ?
Trouver (e, p_r) connaissant p_l ?

Géométrie épipolaire

Etant donné une paire stéréo de caméras, tout point 3D de l'espace, P , définit un plan π_P passant par P et les centres de projection des deux caméras. Le plan π_P est appelé plan épipolaire, et les droites d'intersection avec les plans images sont appelées droites épipolaires conjuguées. L'image dans une caméra du centre de projection de l'autre caméra est appelé épipole.

Propriétés des épipoles

A l'exception de l'épipole, à travers tout point de l'image passe une seule droite épipolaire.

Toutes les droites épipolaires d'une caméra passent à travers l'épipole de la caméra.

Contrainte épipolaire

Les points conjugués doivent se trouver sur des droites épipolaires conjuguées.

La matrice Essentielle E

Dans le repère de la caméra gauche, l'existence d'un plan épipolaire indique que $(O_l P_l)$, $(O_l O_r)$, $(O_r P_l)$ forment trois droites coplanaires, d'où :

$$(P_l - T)^T T \times P_l = 0 \quad \text{avec} \quad P_l - T = \overrightarrow{O_r P_l} \quad \text{à gauche}$$

$$\text{soit} \quad (R^T P_r)^T T \times P_l = 0$$

$$\text{or} \quad T \times P_l = S P_l$$

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

La matrice Essentielle E

Finalement, on a :

$$P_r^T EP_l = p_r^T Ep_l = 0$$

$$\text{avec } E = RS$$

E est la matrice Essentielle et établit un lien naturel entre la contrainte épipolaire et les paramètres extrinsèques du système stéréoscopique.

La matrice Essentielle E

Si on revient à l'interprétation en terme de géométrie projective, on peut voir $E p_l$ comme l'équation de la droite projective dans le plan projectif droit, u_r , passant par p_r et l'épipole e_r :

$$u_r = E p_l$$

Ainsi, la matrice essentielle réalise la transformation recherchée entre les points et les droites épipolaires, en termes de coordonnées image dans le plan projectif \mathcal{P}^2 .

La matrice Fondamentale F

On va montrer à présent que cette transformation point-droite épipolaire peut s'obtenir en coordonnées pixels simplement à partir de points mis en correspondance.

Soient M_l et M_r les matrices 3x3 des paramètres intrinsèques des deux caméras gauche et droite. Ces matrices sont inversibles, la relation entre les coordonnées pixels et les coordonnées image des points projetés s'écrit :

$$\begin{cases} P_l = M_l^{-1} \bar{P}_l \\ P_r = M_r^{-1} \bar{P}_r \end{cases}$$

La matrice Fondamentale F

Finalement, on a :

$$\bar{p}_r^T F \bar{p}_l = 0$$

$$\text{avec } F = M_r^{-T} E M_l^{-1}$$

F est la matrice Fondamentale et établit un lien naturel entre la contrainte épipolaire et les paramètres extrinsèques du système stéréoscopique **en mode pixel**.

La matrice Fondamentale F

Si on revient à l'interprétation en terme de géométrie projective, on peut voir \bar{u}_r comme l'équation de la droite épipolaire projective droite \bar{u}_r , correspondant au point \bar{p}_l :

$$\bar{u}_r = F \bar{p}_l$$

Ainsi, la matrice Fondamentale réalise la transformation recherchée entre les points et les droites épipolaires, en termes de coordonnées pixels.

Matrice essentielle

Pour chaque paire de points p_l et p_r en coordonnées **caméras**, la matrice essentielle satisfait l'équation :

$$p_r^T E p_l = 0$$

Cette matrice :

5. code l'information sur les paramètres extrinsèques seulement;
6. est de rang 2, car S est de rang 2 et R est de rang complet;
7. possède deux valeurs singulières non nulles égales.

Matrice fondamentale

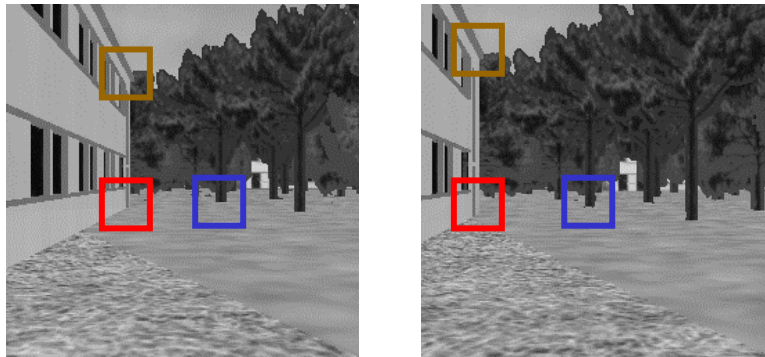
Pour chaque paire de points p_l et p_r en coordonnées **pixels**, la matrice fondamentale satisfait l'équation :

$$p_r^T F p_l = 0$$

Cette matrice :

12. code l'information sur les paramètres extrinsèques et intrinsèques;
13. est de rang 2, car E est de rang 2 et M_l et M_r sont de rang complet;
14. Se déduit de E par la formule $F = M_l^{-T} E M_l^{-1}$

Si on peut estimer la matrice fondamentale F à partir d'un certain nombre de points mis en correspondance en coordonnées pixels, on peut reconstruire la géométrie épipolaire sans aucune information *a priori* sur les paramètres intrinsèques et extrinsèques du système stéréoscopique.



L'algorithme des 8 points pour calculer E ou F

Raisonnons sur la matrice fondamentale 3x3 F.

Les coefficients de F peuvent s'écrire dans un vecteur

$$f = [f_{11}, f_{12}, \dots, f_{33}]^T$$

Chaque équation $p_r^T F p_l = 0$ pour une paire de points (p_r, p_l) se correspondant fournit une équation linéaire homogène en les 9 inconnus correspondant au vecteur f. F étant connu à un facteur multiplicatif près, on a en fait 8 inconnus indépendantes.

Par conséquent, écrire un système de 8 équations linéaires homogènes permettra de résoudre le problème et de déterminer le vecteur f à un facteur d'échelle près.

Algo HUIT_POINTS

INPUT : n couple de points (p_l, p_r) mis en correspondance dans une paire stéréo ($n \geq 8$)

3. Construire le système $Af=0$ avec A la matrice $n \times 9$ des coefficients du système linéaire homogène construit à partir de l'équation $p_r^T F p_l = 0$. Soit $A=UDV^T$ la SVD de A .
4. La solution f définie à un facteur d'échelle près correspond au vecteur colonne de V correspondant à la plus petite valeur singulière de A . (*)
5. Pour forcer la contrainte de singularité sur F , calculer la SVD de $F=UDV^T$
6. Affecter la valeur nulle à la plus petite valeur singulière sur la diagonale de D . Soit D' la matrice corrigée.
7. L'estimée corrigée F' de F est finalement donnée par :

$$F'=UD'V^T$$

OUTPUT : Estimée de la matrice fondamentale F'

(*) En raison du bruit, des erreurs numériques et de correspondances incorrectes, A risque d'être de rang complet

Très important :

Pour éviter des instabilités numériques lors du calcul de F , il faut prendre certaines précautions.

La plus importante consiste à normaliser les coordonnées des points mis en correspondance pour que les entrées de A soit de taille comparable.

En effet, les deux premières coordonnées (en pixels) d'un point image sont données par rapport au coin gauche supérieure de l'image si bien que elles peuvent varier de quelques pixels à plusieurs centaines.

Cette différence peut facilement rendre A très mal conditionnée. De plus, la troisième coordonnée homogène est habituellement fixée à 1.

Il faut donc penser à normaliser les coordonnées des points (p_l, p_r) dans un cube de taille 1 par exemple et adapter l'algorithme en conséquence...

Localiser les épipoles e_l et e_r à partir de F

Puisque l'épipole gauche e_l appartient à toutes les droites épipolaires de l'image gauche,

$$p_r^T F e_l = 0 \text{ pour tout point } p_r$$

Comme F n'est pas nulle, ceci est possible ssi :

$$F e_l = 0$$

Comme F est de rang 2,

l'épipole e_l correspond au noyau de F , symétriquement, e_r correspond au noyau de F^T .

Une localisation précise des épipoles est très utile pour :

- Raffiner la localisation des droites épipolaires correspondantes;
- Simplifier la géométrie stéréo;
- Recouvrir la structure 3D dans le cas de caméras non calibrés;

Algo LOCALISATION_EPIPOLES

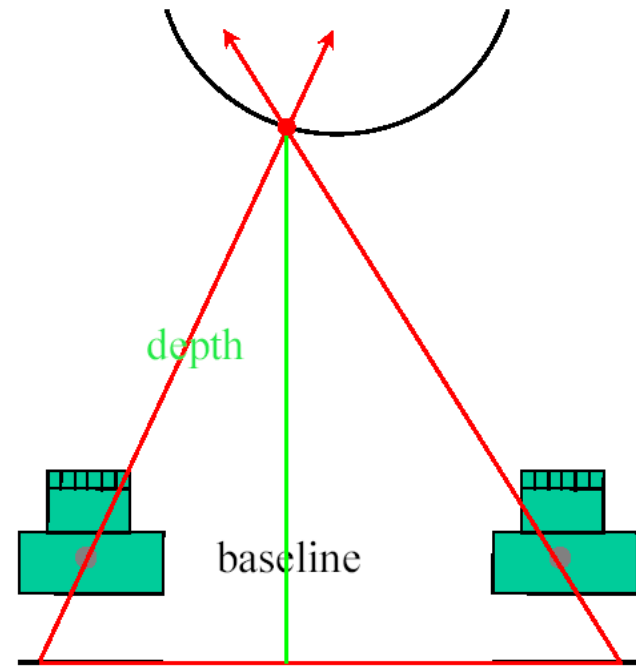
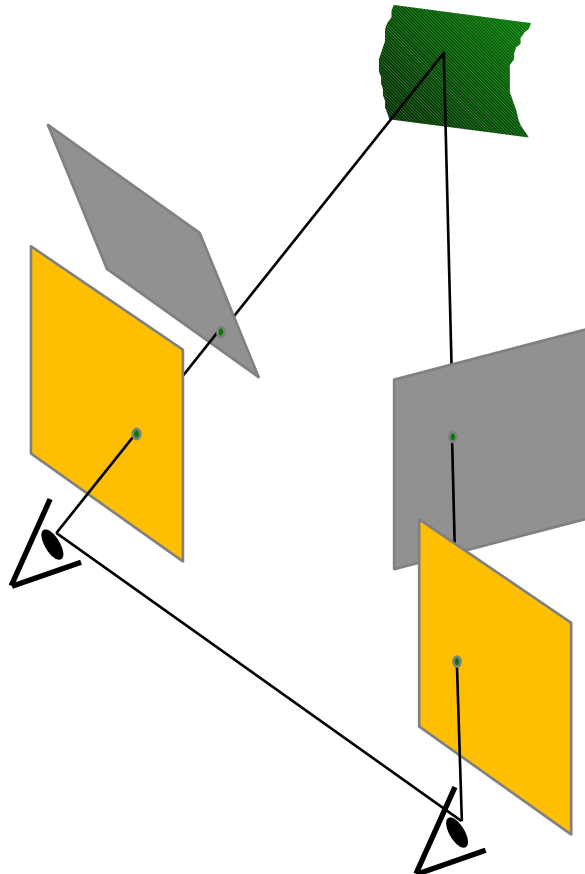
INPUT : la matrice fondamentale F

3. Déterminer la SVD de $F=UDV^T$
4. L'épipole e_l est la colonne de V correspondant à la valeur singulière nulle;
5. L'épipole e_r est la colonne de U correspondant à la valeur singulière nulle;

OUTPUT : Les épipoles e_l et e_r

La rectification

Etant donnée une paire d'images stéréo, la rectification détermine une transformation (ou warping) de chaque image si bien que les paires de droites épipolaires conjuguées deviennent colinéaires à l'axe image horizontal



L'importance de la rectification tient au fait que le problème de la mise en correspondance devient un problème de recherche 1D le long d'une droite trivialement identifiée.

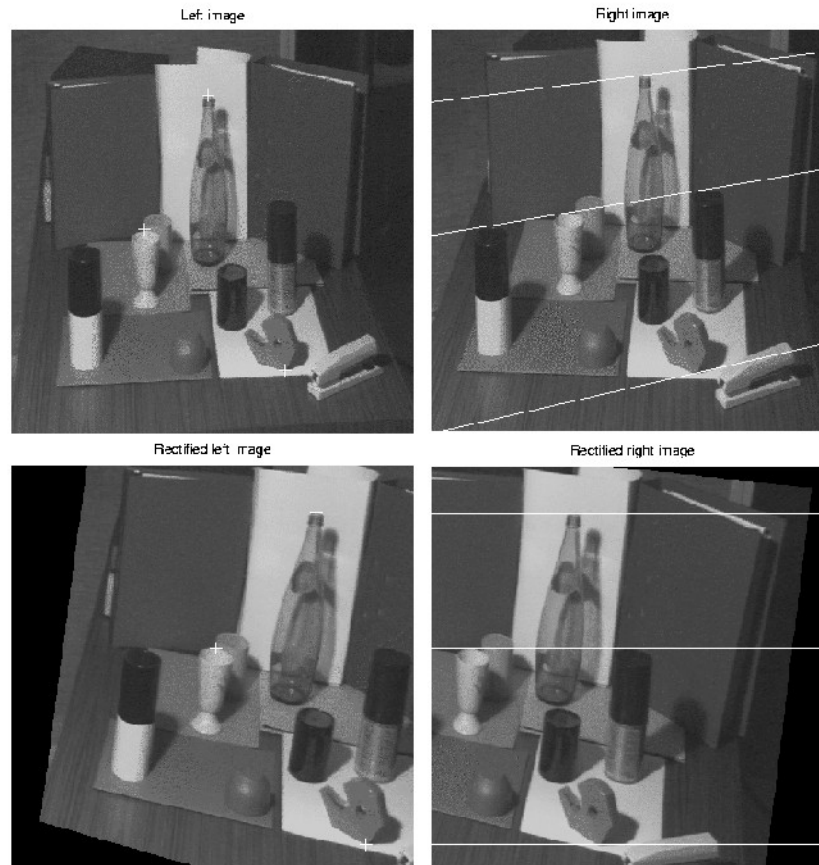
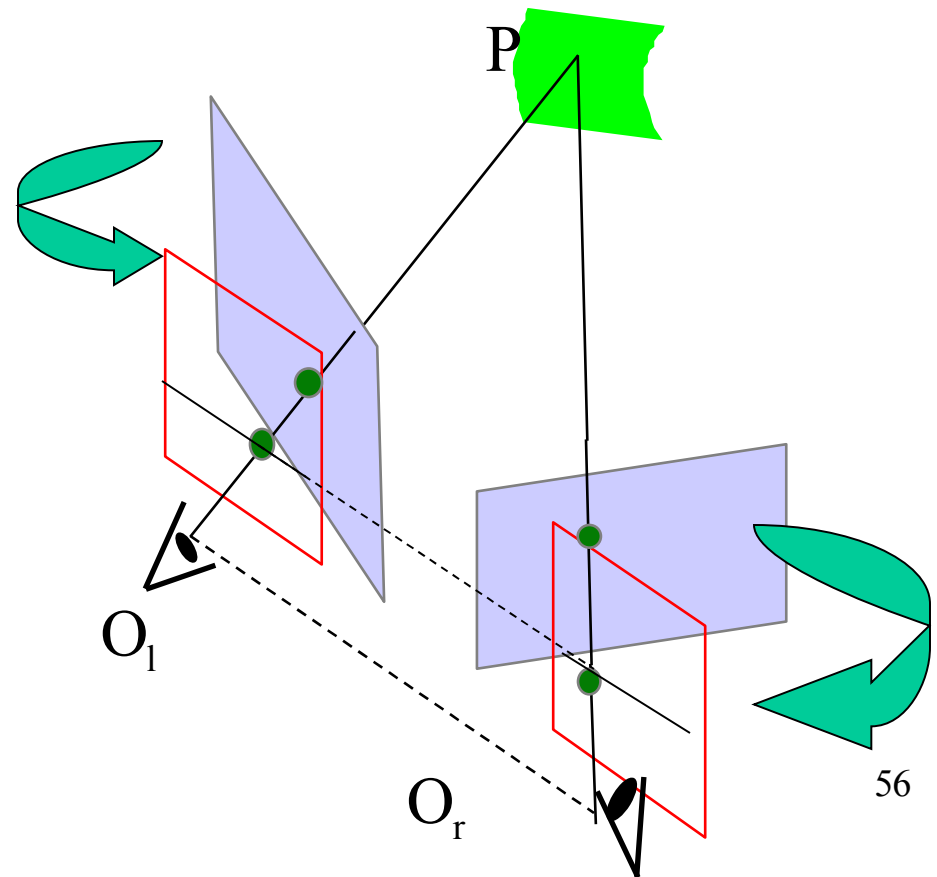


Figure 2: Original “Color” stereo pair (top) and rectified pair (bottom). The left pictures plot the epipolar lines corresponding to the point marked in the right pictures.

Position du problème et hypothèses

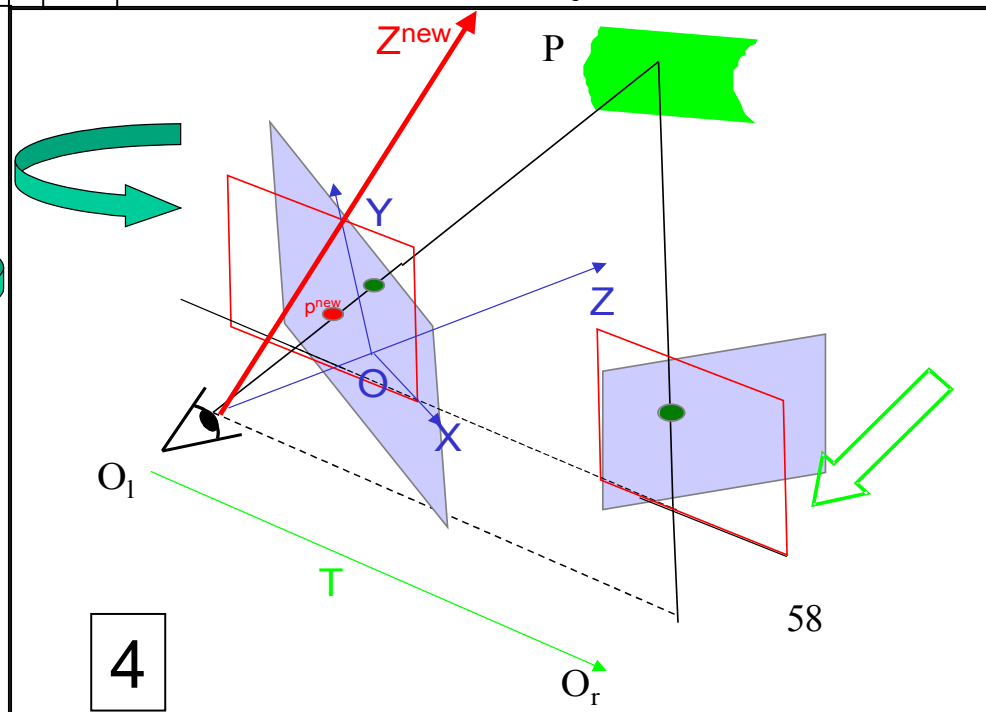
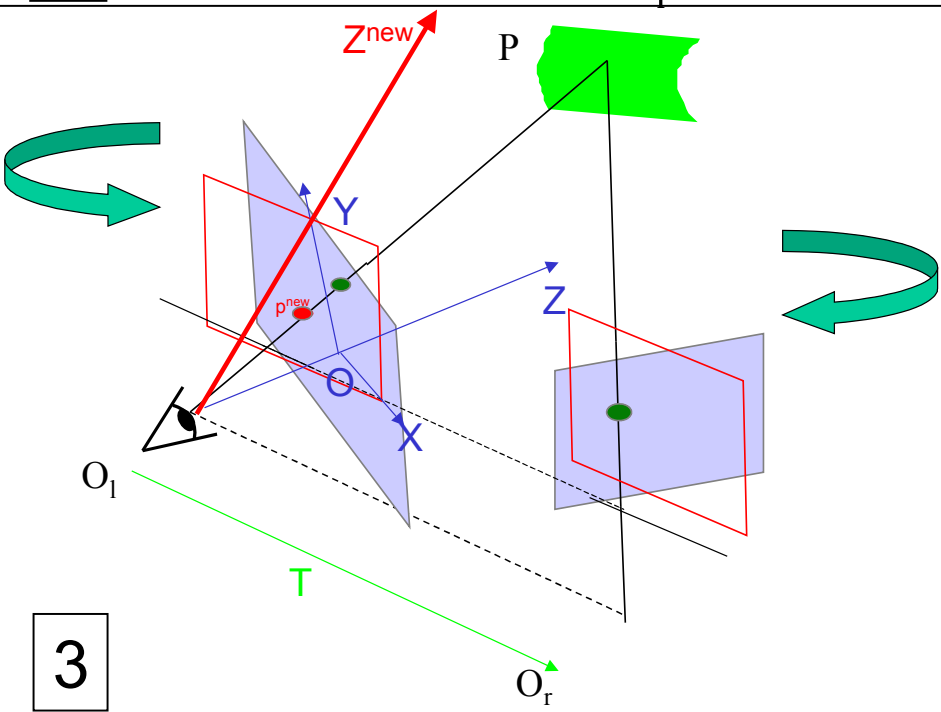
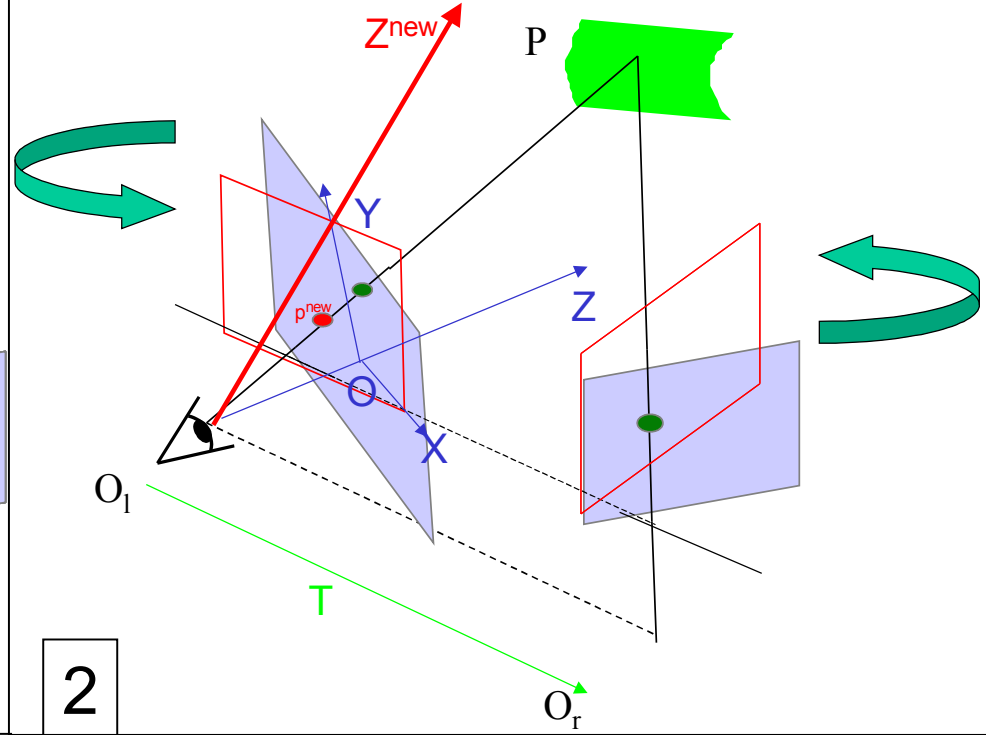
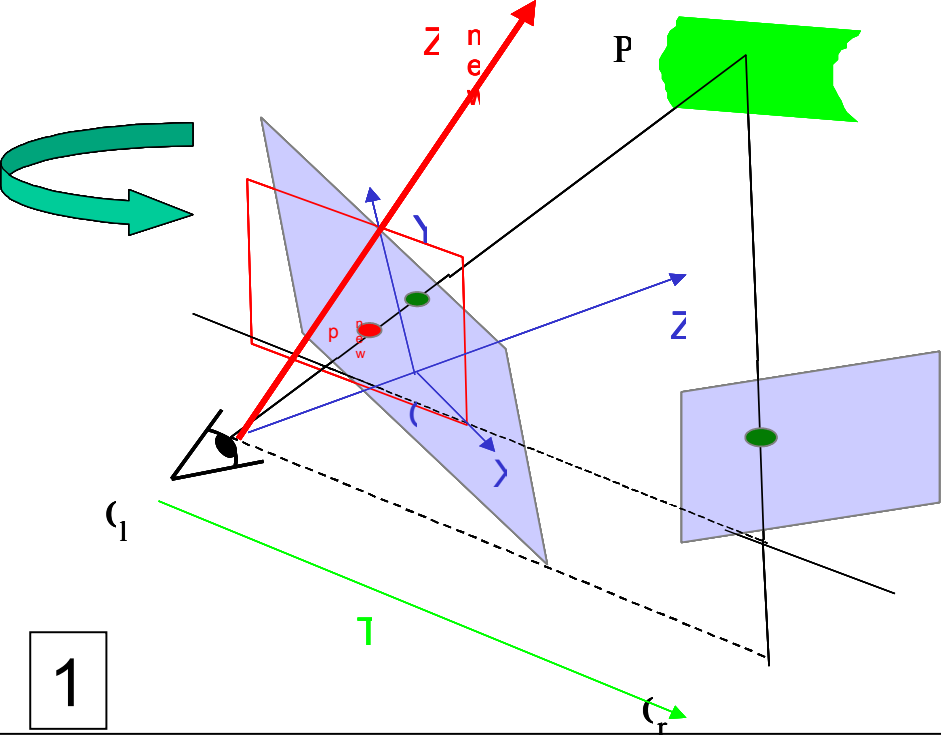
Etant donné une paire stéréo d'images, les paramètres intrinsèques de chacune des caméras, et les paramètres extrinsèques du système R et T, calculer la transformation image qui rende les droites épipolaires conjuguées parallèles à l'axe image horizontal.

Les images rectifiées peuvent être vues comme acquises par un nouveau système stéréo, obtenu en faisant pivoter les caméras originales autour de leur centre optique.



L'algorithme consiste en 4 étapes :

1. Faire pivoter la caméra gauche pour que l'épipole aille à l'infini le long de l'axe horizontal;
3. Appliquer la même rotation à la caméra droite pour se retrouver dans la configuration géométrique initiale;
5. Faire pivote la caméra droite de la rotation R , correspondant au paramètre extrinsèque du système stéréo de départ;
7. Ajuster l'échelle dans les deux repères caméras.



Il faut construire un triplet de vecteurs unitaires mutuellement orthogonaux e_1 , e_2 et e_3 .

$$e_1 = \frac{T}{\|T\|}$$

$$e_2 = \frac{1}{\sqrt{T_x^2 + T_y^2}} [-T_y, T_x, 0]^T = \left(\overrightarrow{O_l O} \times T \right)_{normalisé}$$

$$e_3 = e_1 \times e_2$$

Alors la matrice R_{rect} suivante fait pivoter la caméra gauche autour du centre de projection de telle manière que les droites épipolaires deviennent parallèles à l'axe horizontal

$$R_{rect} = \begin{pmatrix} e_1^T \\ e_2^T \\ e_3^T \end{pmatrix}$$

Algo RECTIFICATION

INPUT : les paramètres intrinsèques et extrinsèques du système stéréo. La focale est la même dans les deux caméras. L'origine du repère image est le point principal.

3. Construire la matrice R_{rect} ;
4. Fixer $R_l = R_{\text{rect}}$ et $R_r = RR_{\text{rect}}$;
5. Pour chaque point de la caméra gauche, $p_l = [x, y, f]^T$ calculer
 $R_l p_l = [x', y', z']$
et les coordonnées du point rectifié correspondant, p'_l
 $p'_l = f/z' [x', y', z']$
6. Répéter l'étape précédente pour la caméra droite en utilisant R_r et p_r

OUTPUT : la paire de transformations à appliquer aux deux caméras pour rectifier les deux images, plus les deux images rectifiées.

Remarque : cet algorithme peut s'appliquer sur un sous-ensemble de pixels de chaque image. De plus, il faudra gérer le fait que les coordonnées rectifiées ne sont plus a priori entières

Remarque : les centres de projection des deux caméras n'ont jamais été déplacés, ce qui permettra de reconstruire aisément la scène en 3D à partir des images rectifiées.

Exemples de spécification logicielle

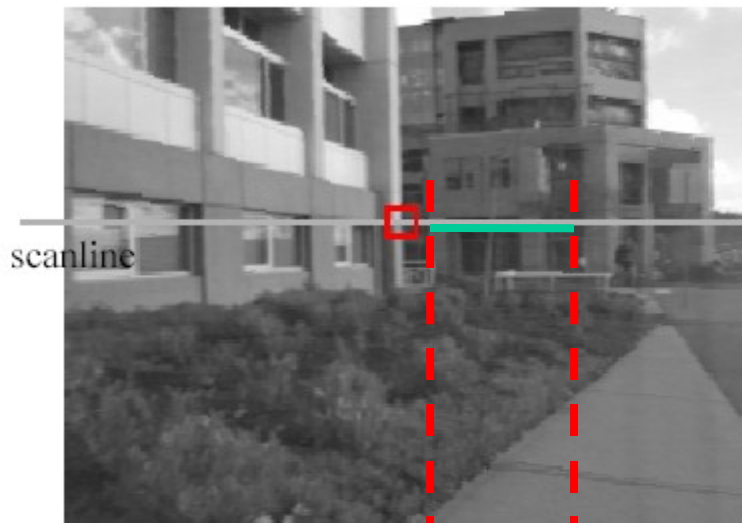
On peut définir une contrainte supplémentaire sur la plage de disparité de recherche à partir des profondeurs min et max que l'on veut reconstruire :

$$d_{\min} = \frac{b f}{Z_{\max}}$$

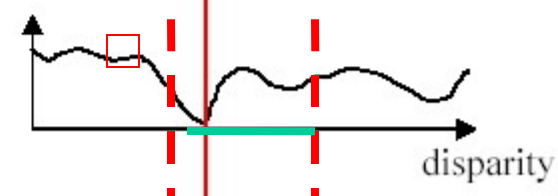
$$d_{\max} = \frac{b f}{Z_{\min}}$$

Left

Right



Criterion function:



On va utiliser une mesure de corrélation normalisée :

$$C(x, y, d) = \frac{\sum_{i,j} \left[\left(I_1(x+i, y+j) - \overline{I_1(x, y)} \right) - \left(I_2(x+i+d, y+j) - \overline{I_2(x+d, y)} \right) \right]^2}{\sqrt{\sum_{i,j} \left(I_1(x+i, y+j) - \overline{I_1(x, y)} \right)^2} * \sqrt{\sum_{i,j} \left(I_2(x+i+d, y+j) - \overline{I_2(x+d, y)} \right)^2}}$$

La taille de la fenêtre peut varier parmi les valeurs 3x3, 5x5 ou 7x7

Pour respecter la contrainte d'unicité et détecter les faux matchs, après avoir l'appariement d'un point P sur l'image 1 avec un point P' dans l'image 2, on réitère une corrélation en inversant les rôles des images 1 et 2. Si la disparité trouvée est égale en valeur absolue à la première, on retient l'appariement.

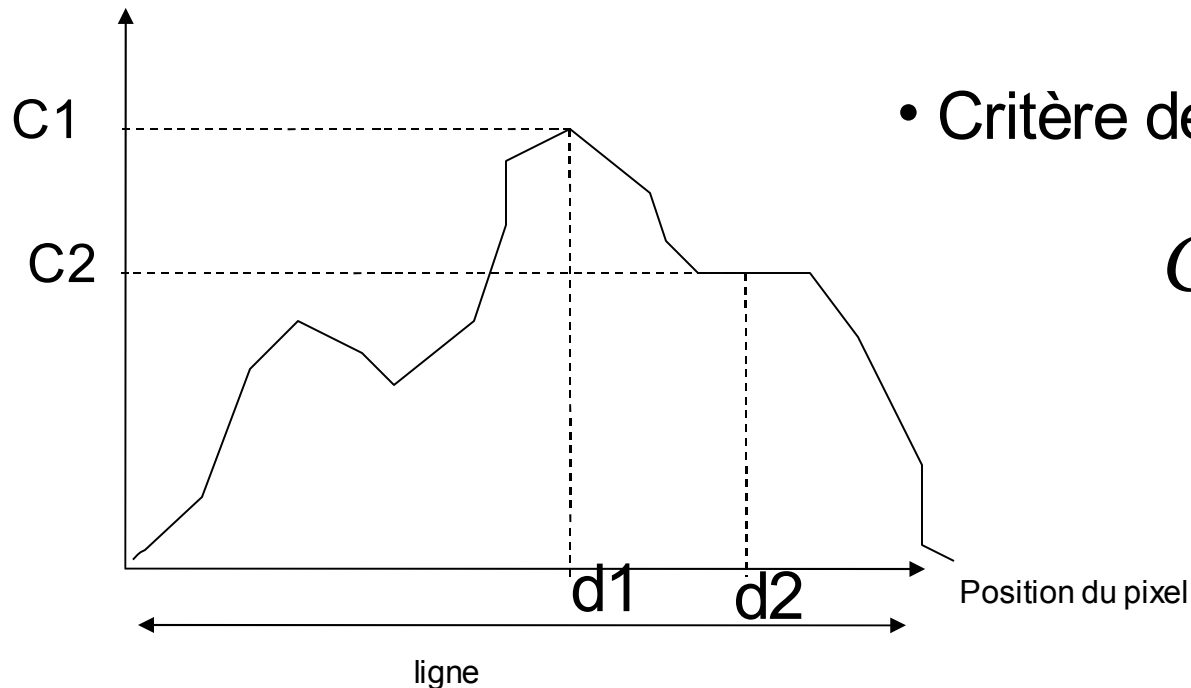
On double le temps de calcul, mais on améliore la robustesse du procédé. On fonction du critère utilisé, on peut toujours imaginer des algorithmes d'optimisation au niveau du calcul de C.

Pour augmenter la robustesse, on peut assortir un match d'une mesure de confiance C supplémentaire de l'association.
 Pour cela, on récupère les deux plus grand maxima locaux de la fonction de score de corrélation C1 et C2.

- Critère de validité :

$$|d2 - d1| \geq \text{Pas de corrélation} + \text{Taille de la fenêtre de corrélation}$$

Valeur du critère de corrélation C



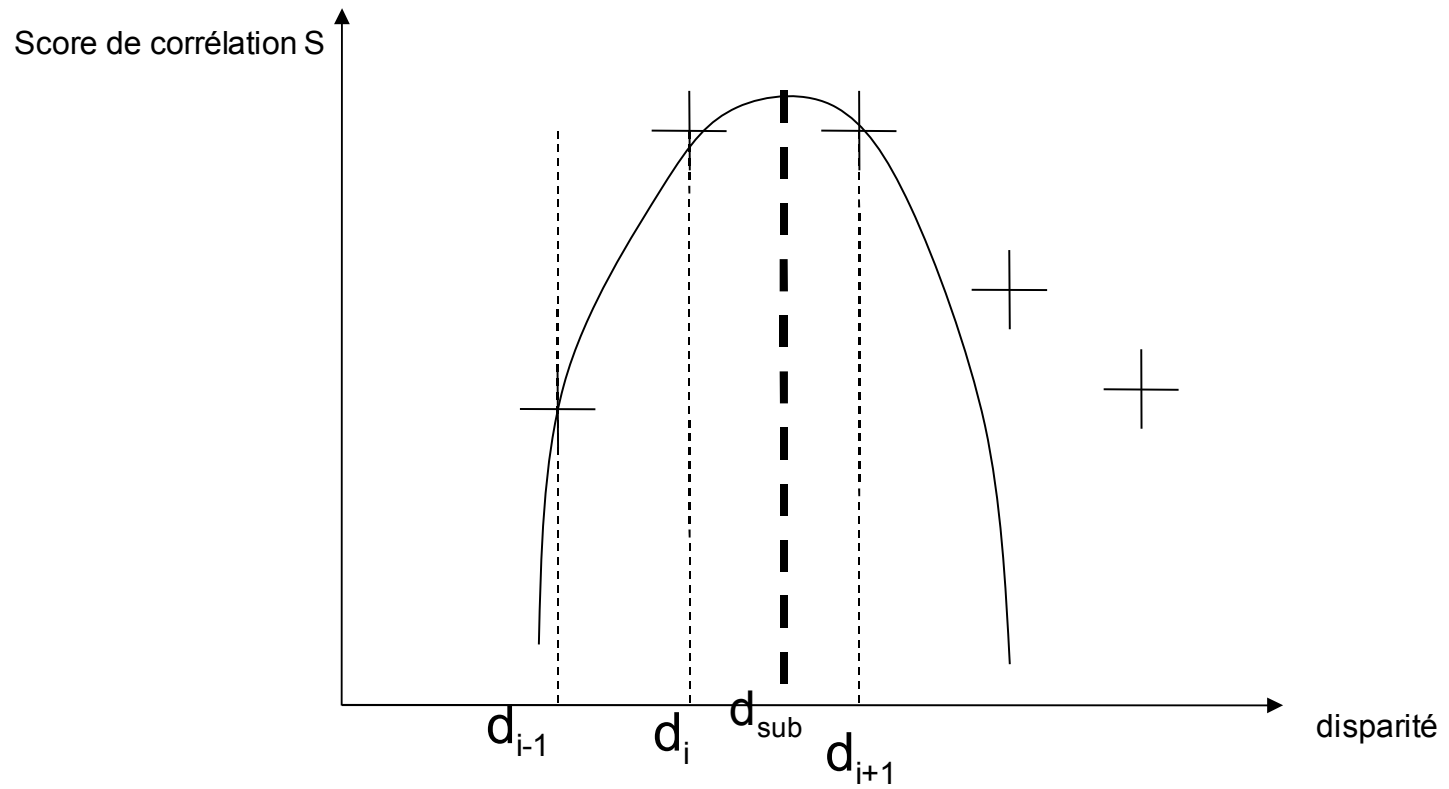
- Critère de confiance :

$$C = \frac{C1 - C2}{C1}$$

Ainsi, le critère de corrélation inverse ne se fera que sur les points validés à l'étapes précédentes selon un critère de confiance C_0 . Soient $d_{1 \rightarrow 2}$ et $d_{2 \rightarrow 1}$ les disparités obtenues par corrélation de l'image 1 vers 2 puis de l'image 2 vers 1. Une nouvelle condition de validation sur $d_{1 \rightarrow 2} + d_{2 \rightarrow 1}$ est prise en compte.

Enfin, on peut affiner le calcul de la disparité avec une approximation subpixellique de celle-ci.

Cela consiste à faire passer une parabole par les points d_{i-1} , d_i et d_{i+1} sur la ligne des scores de disparité où d_i est la corrélation trouvée par la corrélation et d_{i-1} et d_{i+1} les points suivants sur la ligne.



La parabole a pour équation :

$$S = ad^2 + bd + c$$

avec

$$a = \frac{(S(d_{i+1}) - S(d_{i-1})) * (d_{i-1} - d_i) + (S(d_i) - S(d_{i-1})) * (d_{i+1} - d_{i-1}))}{(d_{i+1}^2 - d_{i-1}^2)(d_{i-1} - d_i) + (d_{i+1} - d_{i-1})(d_i^2 - d_{i-1}^2)}$$

$$b = \frac{S(d_{i-1}) - S(d_i) + a * ((d_i^2 - d_{i-1}^2))}{d_{i-1} - d_i}$$

$$c = S(d_{i-1}) - a * d_{i-1}^2 - b * d_{i-1}$$

D'où la disparité subpixellique optimale

$$d_{sub} = - \frac{b}{2a}$$

On peut définir une précision subpixellique (en fonction de la forme de la parabole) :

$$\sigma_{sub} = \sqrt{\left(1 - \frac{1}{e}\right) * \frac{b^2 - 4ac}{4a^2}}$$

Récapitulatif des variables d'entrées :

- Choix de la taille de la fenêtre de corrélation
- Choix du seuil de critère de confiance
- Choix du seuil de condition de validité de corrélation inverse
- Choix du pas de corrélation (pixel)

Récapitulatifs des variables à sortir :

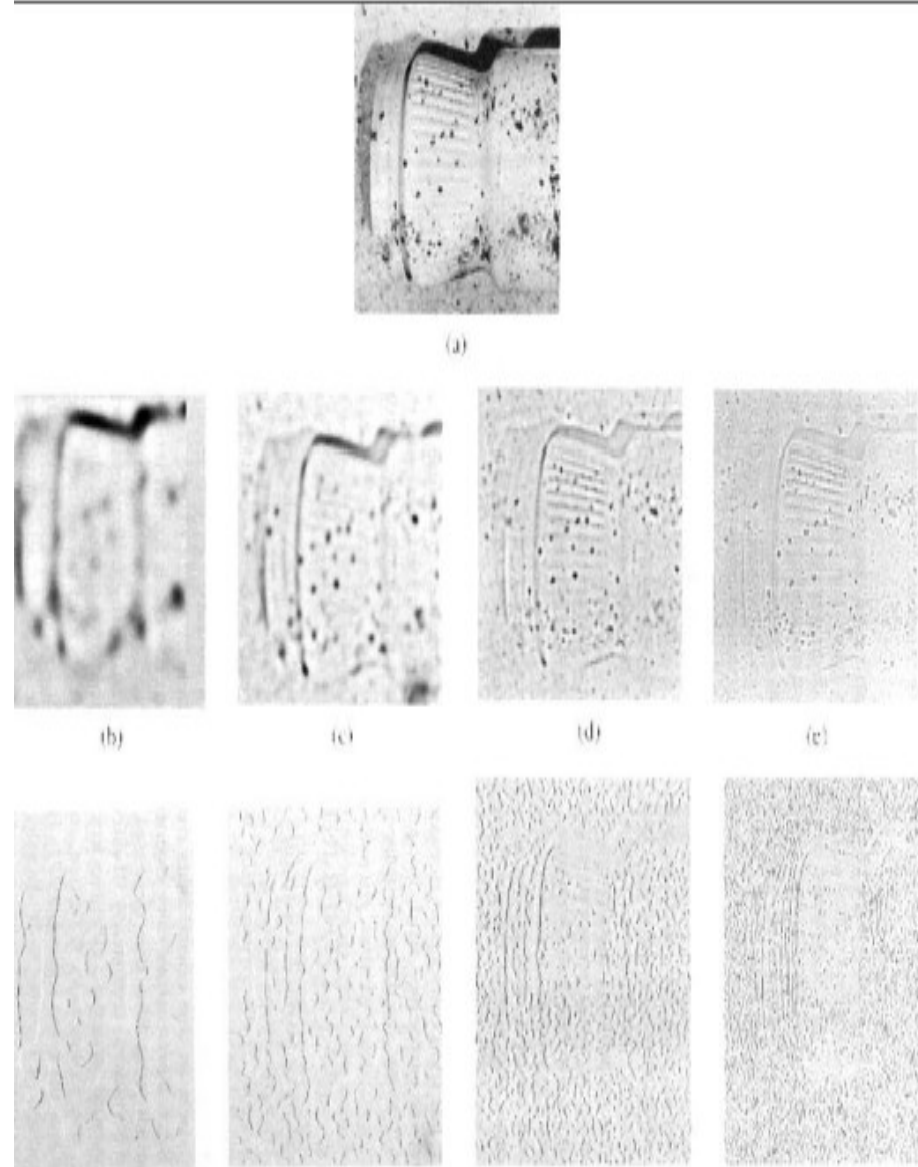
- Images de disparité
- Points x, y et disparités d (pixellique) et d_{sub} (subpixellique) correspondantes
- Critère de confiance pour chaque point
- Précision sur la disparité subpixellique

Contrainte temps réel :

- 1 seconde par couple d'images

L'apport du multiéchelle

- superposition grossière de I_l et I_r
- Lissage très fort avec une gaussienne sigma et détection des points d'intérêt
- décalage horizontal sur les points appariés
- dans une fenêtre centrée sur le pixel à gauche et de taille $4 \cdot \sqrt{2} \cdot \sigma$ (considérations statistiques), on cherche les appariements possibles pour un effet de lissage $\sigma' < \sigma$.
- ainsi de suite, on affine progressivement la vision sur des fenêtres de plus en plus étroites.



Cas d'ondelettes

La reconstruction 3D

Trois cas d'obtention de reconstruction 3D dépendant de la quantité de connaissance *a priori* disponible sur les paramètres du système stéréo :

1. Les paramètres intrinsèques et extrinsèques sont connus
-> reconstruction 3D sans ambiguïté;
2. Seuls les paramètres intrinsèques sont connus
-> reconstruction 3D à un facteur d'échelle près +
estimation des paramètres extrinsèques;
3. L'ensemble des paramètres du système stéréo est inconnu et
seules les correspondances pixel à pixel sont connus
-> reconstruction 3D à une transformation projective
globale près.

Reconstruction absolue par triangulation

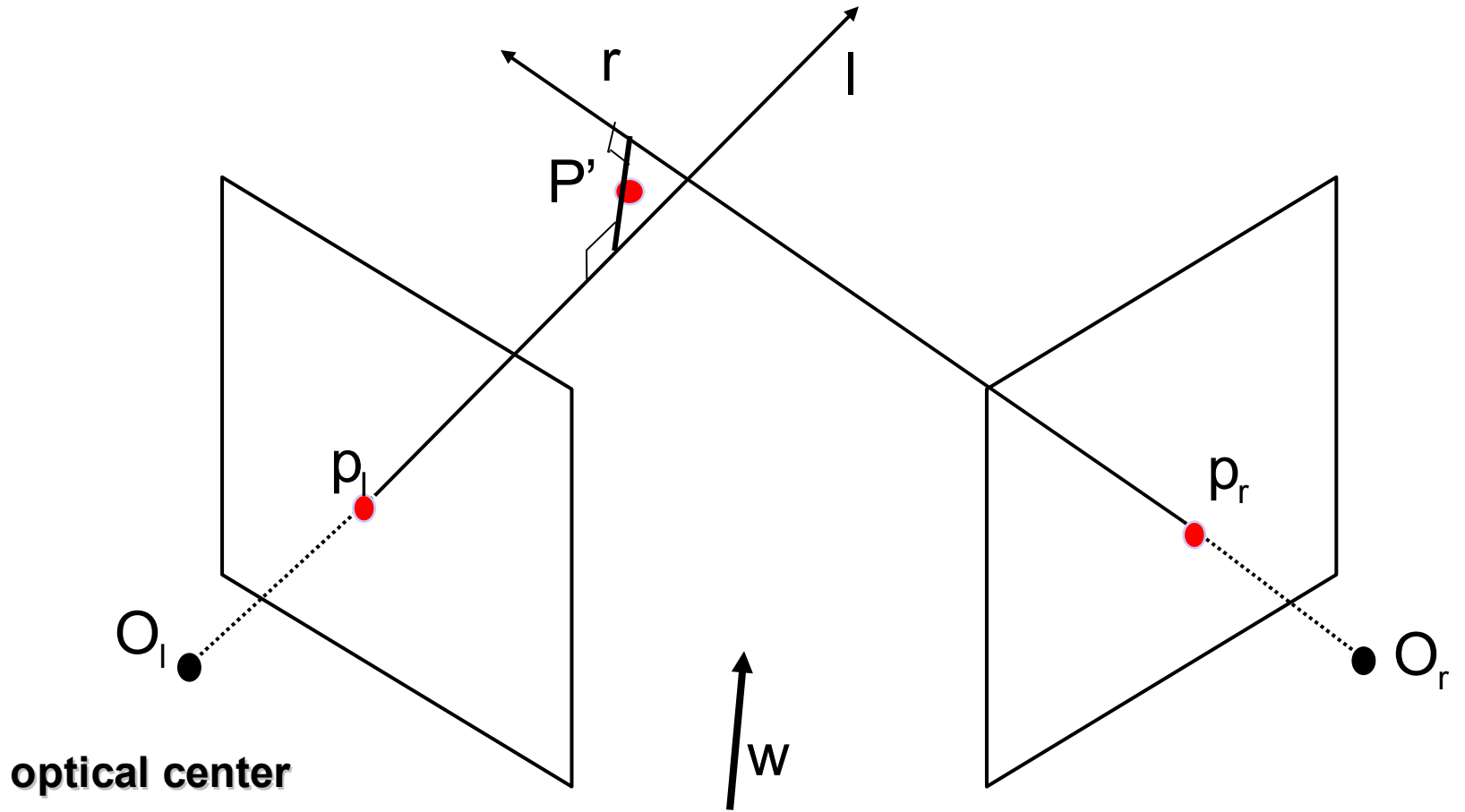
Position du problème et hypothèses

Sous l'hypothèse d'avoir totalement calibré le système stéréo, calculer la localisation des points 3D à partir de leurs projections p_l et p_r .

Il suffit d'intersecter $O_r p_r$ et $O_l p_l$.

Mais, puisque les paramètres et les localisations images sont approximatives, les deux rayons ne s'intersectent généralement exactement.

Leur intersection peut simplement être estimée comme le point de distance minimum aux deux rayons.



Soit ap_l ($a \in \mathfrak{R}$) le rayon, l , passant par O_l et p_l .

Soit $T+bR^T p_r$ ($b \in \mathfrak{R}$) le rayon, r , passant par O_r et p_r exprimé dans le référentiel gauche.

Soit w un vecteur orthogonal à la fois à l et r .

Le problème revient à déterminer le milieu P' du segment parallèle à w qui joint l et r .

Les extrémités du segment en question, soit $a_0 p_l$ et $T+b_0 R^T p_r$, peuvent être calculés en résolvant le système d'équations linéaires :

$$ap_l - bR^T p_r + c(p_l \times R^T p_r) = T$$

Algo TRIANGULATION

INPUT : tous les vecteurs et coordonnées sont exprimés par rapport au référentiel caméra gauche. Dans ce cadre, on possède en entrée un ensemble de points mis en correspondance; soit p_l et p_r une paire générique.

Soit ap_l , $a \in \mathcal{R}$, le rayon passant par O_l ($a=0$) et p_l ($a=1$). Soit $T+bR^T p_r$, $b \in \mathcal{R}$, le rayon r passant par O_r ($b=0$) et p_r ($b=1$). Soit $w = p_l \times R^T p_r$ le vecteur orthogonal à la fois à l et r , et $ap_l + cw$, c , la droite (w) passant par ap_l (pour un a fixé) et parallèle à w .

4. Déterminer les extrémités du segment s sur la droite parallèle à w et joignant l et r : soit $a_0 p_l$ et $T + b_0 R^T p_r$ en résolvant

$$ap_l - bR^T p_r + c(p_l \times R^T p_r) = T$$

8. Le point triangulé, P' , est le milieu du segment s .

OUTPUT : l'ensemble de points 3D reconstruits

Cas d'utilisation de l'algorithme :

Si la géométrie du système d'acquisition ne varie pas dans le temps, les paramètres int. et ext. de chaque caméra peuvent être estimés à travers les procédures de calibration classique.

Si T_l et R_l ainsi que T_r et R_r correspondent aux paramètres extrinsèques ainsi extraits dans le repère monde, il est aisé de montrer que les paramètres ext. du système stéréo, T et R , sont donnés par :

$$R = R_r R_l^T$$

$$T = T_l - R^T T_r$$

Une autre façon d'exprimer les choses :

$$P_{3 \times 4} = [M_{3 \times 3} \ p_4] = [p_1 \ p_2 \ p_3 \ p_4] = [P^{2T}]$$

$$P^{1T}$$

$$P^{3T}$$

Si les images sont rectifiées et tout est connu pour les deux caméras P et P', pour tout point A (X,Y,Z,T) d'image a(u,v,1) et a' , on a :

$$\begin{bmatrix} P^{1T} \\ P^{2T} \\ P^{1T} \ -P'^{1T} \\ P^{3T} \end{bmatrix} A = R_{3D}^{-1} A = \begin{bmatrix} u \\ v \\ d(u, v) \\ 1 \end{bmatrix} = m_d$$

Où d(u,v) est la disparité associée et la surface (u,v, d(u,v)) est appelée reconstruction projective de la scène 3D

Reconstruction relative à un facteur d'échelle

Position du problème et hypothèses

Sous l'hypothèse d'avoir seulement accès aux paramètres intrinsèques du système stéréo, mais aussi à $n \geq 8$ points mis en correspondance, calculer la localisation 3D des points à partir de la projection p_l et p_r .

Comme la géométrie du système n'est pas complètement connu, il demeurera une ambiguïté sur la localisation des points 3D.

En effet, puisque nous ne connaissons pas la ligne de base du système, on conçoit intuitivement que l'on ne pourra obtenir l'échelle véritable de la scène observée.

En conséquence, la reconstruction sera unique à un facteur d'échelle près.

Ce facteur pourra en revanche être déterminé, par la connaissance de la distance entre deux points dans la scène.

Revenons à la matrice Essentielle. Cette matrice E n'est connue qu'à un facteur d'échelle près. Il s'agit donc de choisir une normalisation efficace de cette dernière.

On a :
$$E^T E = S^T R^T R S = S^T S$$

$$\text{soit } E^T E = \begin{pmatrix} T_y^2 + T_z^2 & -T_x T_y & -T_x T_z \\ -T_x T_y & T_x^2 + T_z^2 & -T_z T_y \\ -T_x T_z & -T_z T_y & T_y^2 + T_x^2 \end{pmatrix} 77$$

$$\text{Donc } \text{Tr}(E^T E) = 2\|T\|^2$$

$$\text{Soit } N = \sqrt{\text{Tr}(E^T E) / 2}$$

Ainsi, en divisant les entrées de la matrice essentielle par la constante de normalisation N , on obtient une matrice essentielle dans laquelle le vecteur de translation T est normalisé et unitaire :

$$\hat{T} = \frac{T}{\|T\|}$$

C'est une façon de revenir à une géométrie du système stéréo générique dans laquelle la ligne de base est fixée unitaire.

Soit la nouvelle
matrice
essentielle
normalisée \hat{E}
telle que :

$$\hat{E}^T \hat{E} = \begin{pmatrix} 1 - \hat{T}_x^2 & -\hat{T}_x \hat{T}_y & -\hat{T}_x \hat{T}_z \\ -\hat{T}_x \hat{T}_y & 1 - \hat{T}_y^2 & -\hat{T}_z \hat{T}_y \\ -\hat{T}_x \hat{T}_z & -\hat{T}_z \hat{T}_y & 1 - \hat{T}_z^2 \end{pmatrix}$$

Il faut alors constater que cette normalisation nous rapproche de la véritable matrice essentielle $E=RS$ car $E_{\text{true}} = +/- \hat{E}$

On doit alors constater qu'on peut retrouver $\hat{T} = T / \|T\|$ à partir de n'importe quelle colonne de la matrice $\hat{E}^T \hat{E}$ à un signe global près en raison des termes quadratiques.

Alors, soit $w_i = \hat{E}_i \times \hat{T}$ pour $i = 1, 2, 3$ et E_i les trois lignes de la matrice essentielle normalisée \hat{E} vue comme des vecteurs 3D.

Si R_i sont les lignes de la matrice rotation R toujours vues comme des vecteurs 3D, on peut montrer que :

$$R_i = w_i + w_j \times w_k$$

où le triplet (i, j, k) décrit toutes les permutations cycliques de $(1, 2, 3)$.

Etant données les indéterminations sur les signe de \hat{E} et \hat{T} , on obtient quatre estimées différentes de la paire (\hat{T}, R) .

Pour chaque paire (\hat{T}, R) on va calculer la composante Z_1 de chaque pint dans le référentiel caméra gauche.

On a :

$$p_r = \frac{f R(P_l - \hat{T})}{R_3^T (P_l - \hat{T})}$$

$$P_r = R(P_l - T)$$

$$p_l = \frac{f_l}{Z_l} P_l$$

$$p_r = \frac{f_r}{Z_r} P_r$$

d'où :

$$x_r = \frac{f R_1^T (P_l - \hat{T})}{R_3^T (P_l - \hat{T})}$$

donc :

$$Z_1 = f \frac{(f R_1 - x_r R_3)^T \hat{T}}{(f R_1 - x_r R_3)^T p_l}$$

Il s'ensuit P_1 et P_r . Et seule une des 4 configurations (\hat{T}, R) fournit des coordonnées géométriquement consistantes pour lesquelles Z_1 et Z_r sont positives pour tout point.

Algo RECONSTRUCTION_EUCLID

INPUT :

- un ensemble de points mis en correspondance en coordonnées caméra; soit p_l et p_r une paire générique.
- Une estimée de la matrice essentielle normalisée \hat{E}

5. Retrouver $\hat{T} = \frac{T}{\|T\|}$
7. Construire le vecteur w et calculer les lignes de la matrice R
8. Reconstruire les coordonnées Z_l et Z_r de chaque point
9. Si les signes de Z_l et Z_r des points reconstruits sont :
 - a) Tous les deux négatifs pour quelques points, changer le signe de \hat{T} et revenir à l'étape 3.
 - b) L'un négatif et l'autre positif pour quelques points, changer le signe de chaque entrée de \hat{E} et revenir à l'étape 2.
 - c) Tous les deux positifs pour chaque point, sortir.

OUTPUT : l'ensemble de points 3D reconstruits (à un facteur d'échelle près)

Reconstruction relative à une Transformation Projective

Position du problème et hypothèses

Sous l'hypothèse d'avoir seulement accès à $n \geq 8$ points mis en correspondance (et donc les épipoles e_l et e_r), calculer la localisation des points 3D à partir de leurs projections p_l et p_r .

Il est intéressant de noter que dans ce cas particulier l'exactitude de la reconstruction est seulement fonction de l'exactitude de l'algorithme qui calcule les disparités, et plus du tout par les erreurs en calibration.

Démarche :

4. Déterminer les matrices de projection :

Le mapping de 5 points choisis arbitrairement dans la scène dans la base projective standard de \mathcal{P}^3 , et en utilisant les épipoles, la matrice de projection de chaque caméra peut être explicitement retrouvée à une transformation projective près. En l'occurrence, celle qui associe la base standard aux 5 points sélectionnés, qui est inconnue puisque nous ne connaissons pas la localisation des 5 points 3D dans le référentiel caméra.

6. Calculer la reconstruction projective :

Une fois les matrices de projection obtenues, la localisation 3D d'un point quelconque de l'espace est obtenue par triangulation dans l'espace projectif.

Déterminer les matrices de projection :

A partir de maintenant, P est un point de l'espace projectif et possède en conséquence 4 coordonnées, et sa projection p est un point du plan projectif et possède ainsi 3 coordonnées. On rappelle que ces coordonnées sont dites homogènes. De plus, p désigne la projection gauche et p' la projection droite.

L'espace 3D ordinaire (euclidien) est à présent vu comme un sous-ensemble de l'espace projectif \mathcal{P}^3 . Et chaque plan image comme un sous-ensemble de l'espace projectif \mathcal{P}^2 . Ces variétés projectives ont des comportements plus proches des variétés vectorielles qu'affines.

On devra donc voir le point 3D $[X, Y, Z]^T$ de \mathcal{R}^3 comme le point $[X, Y, Z, 1]^T$ de \mathcal{P}^3 et le point $[x, y]^T$ de \mathcal{R}^2 comme le point $[x, y, 1]^T$ de \mathcal{P}^2 .

Soient O et O' les centres de projections.

Les espaces projectifs

Un point d'un espace projectif à n dimensions, \mathbb{P}^n , est représenté par un vecteur de dimension $n+1$ de coordonnées $x=[x_1, \dots, x_{n+1}]^T$, où $\exists i / x_i \neq 0$
 x est le vecteur de coordonnées et x_i une coordonnée homogène projective.

2 vecteurs $x=[x_1, \dots, x_{n+1}]^T$ et $y=[y_1, \dots, y_{n+1}]^T$ représentent le même point ssi $\exists \lambda \neq 0 / x_i = \lambda y_i \forall i \in [1, n]$

Ainsi la correspondance entre points et vecteurs de coordonnées n'est pas bijective. Ce qui rend l'application de l'algèbre linéaire à la géométrie projective un peu plus compliquée.

Collinéation :

Soit une transformation linéaire d'un espace projectif représentée par la matrice $A \in M_{(n+1) \times (n+1)} / \det(A) \neq 0$ représentant la collinéation de \mathcal{P}^n dans \mathcal{P}^n .

L'ensemble des collinéations est un groupe. Ce groupe est appelé le groupe projectif.

La matrice associée à une collinéation est définie à un facteur multiplicatif non nul près et est notée :

$$\rho y = Ax \quad \text{ou} \quad x \bar{\wedge} y$$

Base projective :

Une base projective est un ensemble de $n+2$ points de \mathcal{P}^n tels que aucun sous-ensemble de $n+1$ points parmi eux ne soient linéairement dépendant.

L'ensemble $e_i = [0, \dots, 1, \dots, 0]^T$, où la coordonnée homogène 1 est en $i^{\text{ème}}$ position, pour $i \in [1, n+1]$ et $e_{n+2} = [1, 1, \dots, 1]^T$ est la base projective standard de \mathcal{P}^n .

Tout point x de \mathcal{P}^n peut être décrit comme la combinaison linéaire des $n+1$ points de la base standard.

$$x = \sum_{i=1}^{n+1} x_i e_i$$

Propriété 1 :

Soient x_1, x_2, \dots, x_{n+2} une base projective. \exists des matrices non singulières $A / Ae_i = \lambda_i x_i$ pour $i=1, \dots, n+2$ et $\lambda_i \neq 0$.

Toute matrice vérifiant cette propriété diffère de A seulement par un facteur d'échelle.

Propriété 2 :

Si x_1, x_2, \dots, x_{n+2} et y_1, y_2, \dots, y_{n+2} sont 2 ensembles de $n+2$ vecteurs coordonnées formant 2 bases projectives, alors \exists une matrice $P \in M_{(n+1) \times (n+1)}$ non singulière / $Px_i = \rho_i y_i$ pour $i=1, \dots, n+2$ et où les ρ_i sont des scalaires.

La matrice P est déterminée uniquement à un facteur d'échelle près.

Donc une collinéation est entièrement définie par $n+2$ points se correspondant.

Théorème 1 :

L'espace \mathbb{P}^n est topologiquement équivalent à la sphère unité S^n de \mathbb{R}^{n+1} dans laquelle nous identifions les points antipodes.

La droite projective

L'espace \mathcal{P}^1 est appelée droite projective.

Sa base projective standard est $e_1=[1,0]T$, $e_2=[0,1]T$ et $e_3=[1,1]T$.

Un point de la droite peut s'écrire :

$$x = x_1 \vec{e}_1 + x_2 \vec{e}_2$$

avec au moins un $x_i \neq 0$.

Soit un sous-ensemble de $\mathcal{P}^1 / x_2 \neq 0$

C'est équivalent à exclure le point e_1 .

Ces points exprimés en coordonnées homogènes peuvent s'écrire $x = \alpha e_1 + e_2$ avec $-\infty \leq \alpha \leq \infty$ et $\alpha = x_1/x_2$.

Le paramètre α est appelé le paramètre projectif du point.

Le point représenté par e_2 a un paramètre projectif nul.

Le point à l'infini :

Le point représenté par e_1 est appelé le point à l'infini de la droite \mathbb{P}^1 . Il est défini par l'équation linéaire $x_2=0$.

Si on pense que la droite projective contient la traditionnelle droite affine par la correspondance $\alpha \leftrightarrow \alpha e_1 + e_2$, alors le paramètre projectif α du point nous donne une correspondance biunivoque entre les droites affines et les droites projectives pour toute valeur de $\alpha \neq \infty$.

Les valeurs $\alpha = \pm\infty$ correspondent au point e_1 qui est en dehors de la droite affine mais est à la limite des points de la droite affine avec de grandes valeurs de α .

C'est une interprétation très utile de la relation entre droites affine et projective.

Collinéation de \mathcal{P}^1 :

Une collinéation de \mathcal{P}^1 est définie à un facteur près par la matrice de rang 2 :

$$P = \begin{bmatrix} r & s \\ t & u \end{bmatrix}$$

Si un point à un paramètre projectif α , le point transformé à un paramètre projectif α' défini par :

$$\alpha' = \frac{r\alpha + s}{t\alpha + u}$$

Le birapport de 4 points est défini par :

$$\{a, b; c, d\} = \frac{\alpha_a - \alpha_c}{\alpha_a - \alpha_d} \cdot \frac{\alpha_b - \alpha_c}{\alpha_b - \alpha_d}$$

Il est invariant pour le groupe des collinéations de \mathcal{P}^1 et en particulier il est indépendant du choix des coordonnées dans ${}^{\mathcal{P}^1}$.

On rappelle qu'une image est modélisé par le plan projectif \mathcal{P}^2 .

Points et droites :

Un point de \mathcal{P}^2 est défini par 3 nombres non tous nuls (x_1, x_2, x_3) . Ils forment un vecteur coordonnée \mathbf{x} défini à un facteur d'échelle près.

Dans \mathcal{P}^2 , il y a des objets autres que des points comme les droites. Une droite est aussi définie par un triplet de nombres (u_1, u_2, u_3) non tous nuls.

Ils forment un vecteur de coordonnées \mathbf{u} défini à un facteur d'échelle près. L'équation de la droite dans la base projective standard (e_1, e_2, e_3) de \mathcal{P}^2 est :

$$\sum_{i=1}^3 u_i x_i = 0$$

La plan projectif

On rappelle d'après ce qui précède qu'il n'y a aucune distinction entre les points et les droites du plan projectif \mathbb{P}^2 .
Pensez à la transformée de Hough.

Proposition :

Une droite passant par 2 points m_1 et m_2 est représentée par

$$m_1 \wedge m_2 = 0$$

$$\left(\vec{x} = \alpha \vec{m}_1 + \beta \vec{m}_2 \rightarrow \begin{vmatrix} \vec{x} & \vec{m}_1 & \vec{m}_2 \end{vmatrix} = 0 \rightarrow \vec{x} \cdot (\vec{m}_1 \wedge \vec{m}_2) = 0 \right)$$

Une Transformation Projective est une transformation linéaire entre espaces projectifs. En vision par ordinateur, il y a deux classes importantes de telles transformations :

- Les transformations linéaires inversibles des espaces \mathbb{P}^n , $n=1,2,3$, dans eux-mêmes.
- Les transformations entre \mathbb{P}^3 et \mathbb{P}^2 , qui modélise la formation de l'image.

Théorème

Une transformation projective de \mathbb{P}^n dans lui-même est complètement déterminée par son action sur $n+2$ points.

En particulier, une transformation projective du plan de \mathbb{P}^2 dans \mathbb{P}^2 est entièrement déterminée par l'image de 4 points parmi lesquels aucun triplet n'est collinéaire.

Comme une transformation projective est une transformation linéaire inversible, on peut la représenter dans le cadre de l'algèbre linéaire :

$$Tp' = p$$

Comme p et p' sont connus à un facteur près, les entrées de T sont aussi connues à un facteur près.

On veut montrer que T peut être déterminée en fonction des quatre points $p'_i = [x'_i, y'_i, z'_i]^T$ images de $p_1 = [1, 0, 0]^T$, $p_2 = [0, 1, 0]^T$, $p_3 = [0, 0, 1]^T$ et $p_4 = [1, 1, 1]^T$ respectivement.

On a :
$$T^{-1}p_1 = p'_1$$

Donc la première colonne de T^{-1} peut s'écrire $\lambda[x'_1, y'_1, z'_1]^T$ avec λ indéterminé

Les transformations projectives de \mathcal{P}^2 dans \mathcal{P}^2

Et de fait :

$$T^{-1} = \begin{pmatrix} \lambda x'_1 & \mu x'_2 & \nu x'_3 \\ \lambda y'_1 & \mu y'_2 & \nu y'_3 \\ \lambda z'_1 & \mu z'_2 & \nu z'_3 \end{pmatrix}$$

avec λ, μ et ν indéterminés.

Comme il n'y a pas trois points collinéaires parmi les p'_i ,
 Test bien inversible et on peut résoudre le système linéaire
 suivant à trois équations et trois inconnues λ, μ et ν
 relativement au facteur inconnu ρ correspondant au point

$$p'_4 : \quad \lambda x'_1 + \mu x'_2 + \nu x'_3 = x'_4$$

$$\lambda y'_1 + \mu y'_2 + \nu y'_3 = y'_4$$

$$\lambda z'_1 + \mu z'_2 + \nu z'_3 = z'_4$$

Ainsi les 9 entrées de T seront connues à un facteur ρ près.

Soient P^1, \dots, P^n les points de \mathcal{P}^3 à reconstruire à partir de leurs projections p_1, \dots, p_n et p'_1, \dots, p'_n . Et supposons que parmi les cinq premiers P^i (P^1, P^2, \dots, P^5), il n'y en ait pas 3 parmi eux collinéaires ou 4 coplanaires.

On va d'abord montrer que si on choisit (P^1, P^2, \dots, P^5) comme la base standard projective de \mathcal{P}^3 , chaque matrice de projection peut être déterminée à un facteur projectif près qui dépend de la localisation des épipoles.

Puisqu'une transformation projective de l'espace \mathcal{P}^3 est entièrement déterminée par la données de l'image par cette transformation de 5 points, on peut construire une transformation projective qui envoie P^1, P^2, \dots, P^5 dans la base projective standard de \mathcal{P}^3 , $P^1=[1,0,0,0]^T$, $P^2=[0,1,0,0]^T$, $P^3=[0,0,1,0]^T$, $P^4=[0,0,0,1]^T$ et $P^5=[1,1,1,1]^T$

Pour le point image correspondant p_i dans la caméra gauche, on peut écrire $MP^i = \rho_i p_i$

Où M est la matrice de projection gauche et $\rho_i \neq 0$.

De façon similaire, puisqu'une transformation projective du plan \mathcal{P}^2 est entièrement déterminée par la données de l'image par cette transformation de 4 points, on peut aussi construire une transformation projective qui envoie les 4 premiers points p_i dans la base projective standard de \mathcal{P}^2 , $p_1 = [1, 0, 0]^T$, $p_2 = [0, 1, 0]^T$, $p_3 = [0, 0, 1]^T$ et $p_4 = [1, 1, 1]^T$.

Dans tout ce qui suit, on suppose que les coordonnées du cinquième point p_5 , de l'épipôle et de tout autre point p_i peuvent être obtenues en appliquant cette transformation aux anciennes coordonnées.

Grâce à ces transformations de base, on peut écrire M très simplement :

$$M = \begin{bmatrix} \rho_1 & 0 & 0 & \rho_4 \\ 0 & \rho_2 & 0 & \rho_4 \\ 0 & 0 & \rho_3 & \rho_4 \end{bmatrix}$$

Soit $[\alpha, \beta, \gamma]^T$ les coordonnées de p_5 dans la base standard, on peut alors éliminer ρ_1 , ρ_2 et ρ_3 de l'expression de M :

$$M = \begin{bmatrix} \alpha \rho_5 - \rho_4 & 0 & 0 & \rho_4 \\ 0 & \beta \rho_5 - \rho_4 & 0 & \rho_4 \\ 0 & 0 & \gamma \rho_5 - \rho_4 & \rho_4 \end{bmatrix}$$

Comme une matrice de projection est définie à un facteur d'échelle près, on peut diviser par ρ_4

$$M = \begin{bmatrix} \alpha x - 1 & 0 & 0 & 1 \\ 0 & \beta x - 1 & 0 & 1 \\ 0 & 0 & \gamma x - 1 & 1 \end{bmatrix} \text{ avec } x = \frac{\rho_5}{\rho_4}$$

Ainsi, la matrice de projection M de la caméra gauche a été déterminée à un paramètre projectif près x .

Pour déterminer x , il est utile de relier les entrées de M aux coordonnées du centre de projection O de la caméra gauche. Ceci peut être réalisé en observant que M modélise une projection perspective de centre de projection O . Donc, M projète tous les points de \mathcal{P}^3 en un point de \mathcal{P}^2 , à l'exception du point O . Comme M est de rang 3, le noyau de M est non trivial et O est donné par :

$$MO = 0$$

soit

$$O = \left[\frac{1}{1 - \alpha x}, \frac{1}{1 - \beta x}, \frac{1}{1 - \gamma x}, 1 \right]^T$$

De façon symétrique pour la caméra droite,

$$M' = \begin{bmatrix} \alpha' x' - 1 & 0 & 0 & 1 \\ 0 & \beta' x' - 1 & 0 & 1 \\ 0 & 0 & \gamma' x' - 1 & 1 \end{bmatrix} \text{ avec } x' = \frac{\rho'_5}{\rho'_4}$$

et

$$O' = \left[\frac{1}{1 - \alpha' x'}, \frac{1}{1 - \beta' x'}, \frac{1}{1 - \gamma' x'}, 1 \right]^T$$

Comme la localisation des épipoles est connue, x et x' peuvent être déterminés par :

$$MO' = \sigma e$$

et

$$M'O = \sigma' e' \quad \text{avec} \quad \sigma \neq 0 \text{ et } \sigma' \neq 0$$

D'où

$$\begin{bmatrix} \alpha & -\alpha' & \alpha' e_x \\ \beta & -\beta' & \beta' e_y \\ \gamma & -\gamma' & \gamma' e_z \end{bmatrix} \begin{pmatrix} x \\ x' \\ \sigma x' \end{pmatrix} = \begin{pmatrix} \sigma e_x \\ \sigma e_y \\ \sigma e_z \end{pmatrix}$$

Si on essaye de résoudre ce système vu comme un système linéaire en les trois inconnues x , x' et σ , on a

$$\sigma x' = \sigma \frac{e^T (p_5 \times p_5')}{v^T (p_5 \times p_5')}$$

$$\text{avec } v = (\alpha' e_x, \beta' e_y, \gamma' e_z)$$

Comme e , p_5 , p_5' et v sont connus, on a bien x' .
De même,

$$x = \frac{e'^T (p_5 \times p_5')}{v'^T (p_5 \times p_5')}$$

$$\text{avec } v' = (\alpha e'_x, \beta e'_y, \gamma e'_z)$$

Connaissant x et x' on a complètement déterminé les matrices de projection ainsi que les centres de projection ! ¹⁰⁷

Calculer la reconstruction projective :

A partir de maintenant, on est paré pour reconstruire n'importe quel point de \mathcal{P}^3 connaissant ses points images correspondants $p=[p_x, p_y, p_z]^T$ et $p'=[p'_x, p'_y, p'_z]^T$.

La reconstruction sera unique à une transformation projective fixée par le choix de P^1, \dots, P^5 comme la base standard projective de \mathcal{P}^3 mais inconnue.

Il faut observer que la droite projective l définie par :

$$\lambda O + \mu [O_x p_x, O_y p_y, O_z p_z, 0]^T$$

avec $(\lambda, \mu) \in \mathbb{R}^2, (\lambda, \mu) \neq (0, 0)$

passe par O ($\lambda=1$ et $\mu=0$) et par p car :

$$M \begin{pmatrix} O_x p_x \\ O_y p_y \\ O_z p_z \\ 0 \end{pmatrix} = p$$

De façon similaire, il faut observer que la droite projective l' définie par :

$$\lambda' O' + \mu' [O'_x p'_x, O'_y p'_y, O'_z p'_z, 0]^T$$

avec $(\lambda', \mu') \in \mathbb{R}^2, (\lambda', \mu') \neq (0, 0)$

passe par O' ($\lambda'=1$ et $\mu'=0$) et par p' .

Le point projectif P peut être vu comme l'intersection des deux droites projectives l et l' .

Ce qui revient à déterminer la solution non-triviale du système linéaire homogène suivant :

$$\begin{bmatrix} O_x & O_x p_x & -O'_x & -O'_x p'_x \\ O_y & O_y p_y & -O'_y & -O'_y p'_y \\ O_z & O_z p_z & -O'_z & -O'_z p'_z \\ 1 & 0 & -1 & 0 \end{bmatrix} \begin{pmatrix} \lambda \\ \mu \\ \lambda' \\ \mu' \end{pmatrix} = 0$$

Algo UNCAL_STEREO

INPUT :

- $n \geq 5$ paires de points se correspondant, p_i et p'_i , images de n points P^1, \dots, P^n . On suppose que parmi les 5 premiers (P^1, P^2, \dots, P^5), il n'y en a pas 3 collinéaires ou 4 coplanaires.
- On a estimé la localisation des épipoles, e et e' .
- Soit (P^1, P^2, \dots, P^5) , la base projective standard de \mathcal{P}^3 .

2. Déterminer les transformations projectives planaires T et T' qui mappent les p_i et p'_i ($i=1, \dots, 4$) en la base projective standard de \mathcal{P}^2 sur chaque plan image. Appliquer T sur les p_i et l'épipole e , et T' sur les p'_i et l'épipole e' . Soient (α, β, γ) et $(\alpha', \beta', \gamma')$ les nouvelles coordonnées de p_5 et p'_5 .
3. Déterminer x et x' .
4. Déterminer O et O' .
5. Etant donnée une paire de points conjugués p et p' , reconstruire la localisation du point P dans la base projective standard de \mathcal{P}^3 .

OUTPUT : les coordonnées des points P^1, P^2, \dots, P^5 dans la base projective standard.

Une fois qu'on a trouvé une reconstruction projective de nos points, **comment revenir à des coordonnées euclidiennes ?**

Si on connaît la localisation exacte des points (P^1, P^2, \dots, P^5), dans le référentiel monde, on peut déterminer la transformation projective initialement inconnue de \mathcal{P}^3 qui mappait ces 5 points, vus comme des points de \mathcal{P}^3 , en la base projective standard de \mathcal{P}^3 .

Si on ignore cette localisation euclidienne des 5 points, il existe des algorithmes mais qui utilisent plus de 2 images.

| <i>Connaissance a priori</i> | <i>Reconstruction 3D à partir de deux vues</i> |
|---|--|
| Paramètres intrinsèques et extrinsèques | Pas d'ambiguïté (coordonnées absolues) |
| Paramètres intrinsèques seulement | A un facteur d'échelle près |
| Pas d'information | A une transformation projective près |

Vision trinoculaire

<http://www.ptgrey.com/products/triclopsSDK/index.html>

Stereo Algorithm

- An area based correlation with SAD (Sum of Absolute Differences) is done on band passed images.
- The algorithm is fairly robust and it has a number of validation steps that reduce the level of noise.
- The method requires texture and contrast to work correctly. Occlusions, repetitive features and specularities can cause problems.
- Stereo can be performed with both 2-camera and 3-camera stereo configurations.

Speed (FPS)

- This chart illustrates the speed of the Triclops™ SDK using a 640x480 B&W Digidlops® on a P4 2.4GHz PC. The Digidlops baseline is 10cm. The Bumblebee baseline is 12cm and is therefore more accurate.
- The disparity range was selected so that it corresponds to a distance range of 0.5m to infinity.
- The data was obtained using the demo program that is supplied with the Triclops SDK
- The number of disparities is proportional to the speed. Therefore, if you cut the number of disparities in half, the system will run roughly twice as fast. The speed of processing is independent of the mask size.

| Resolution | Disparity Range | Subpixel | Validation | Frames/Second |
|------------|-----------------|----------|------------------------------------|---------------|
| 160x120 | 32 | No | Texture & Uniqueness | 91 |
| 160x120 | 32 | Yes | Texture & Uniqueness | 91 |
| 160x120 | 32 | Yes | Texture, Uniqueness & Segmentation | 83 |
| 320x240 | 48 | No | Texture & Uniqueness | 31 |
| 320x240 | 48 | Yes | Texture & Uniqueness | 26 |
| 320x240 | 48 | Yes | Texture, Uniqueness & Segmentation | 25 |
| 640x480 | 96 | No | Texture & Uniqueness | 4.5 |
| 640x480 | 96 | Yes | Texture & Uniqueness | 4.3 |
| 640x480 | 96 | Yes | Texture, Uniqueness & Segmentation | 4.4 |

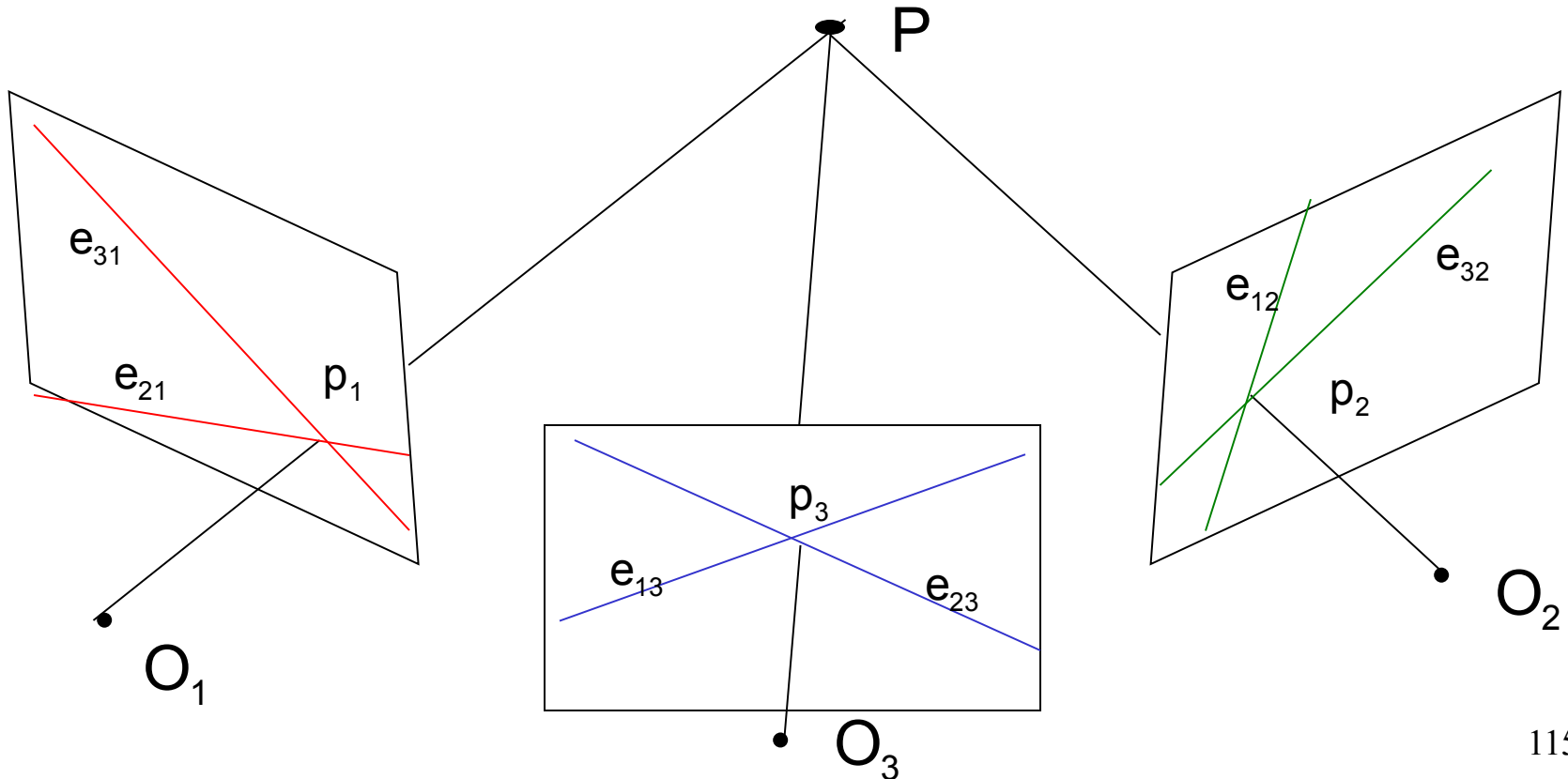


Dans ce cas, un point P de la scène se projette en trois points p_1, p_2, p_3 selon les droites passant par les centres de projection des trois caméras.

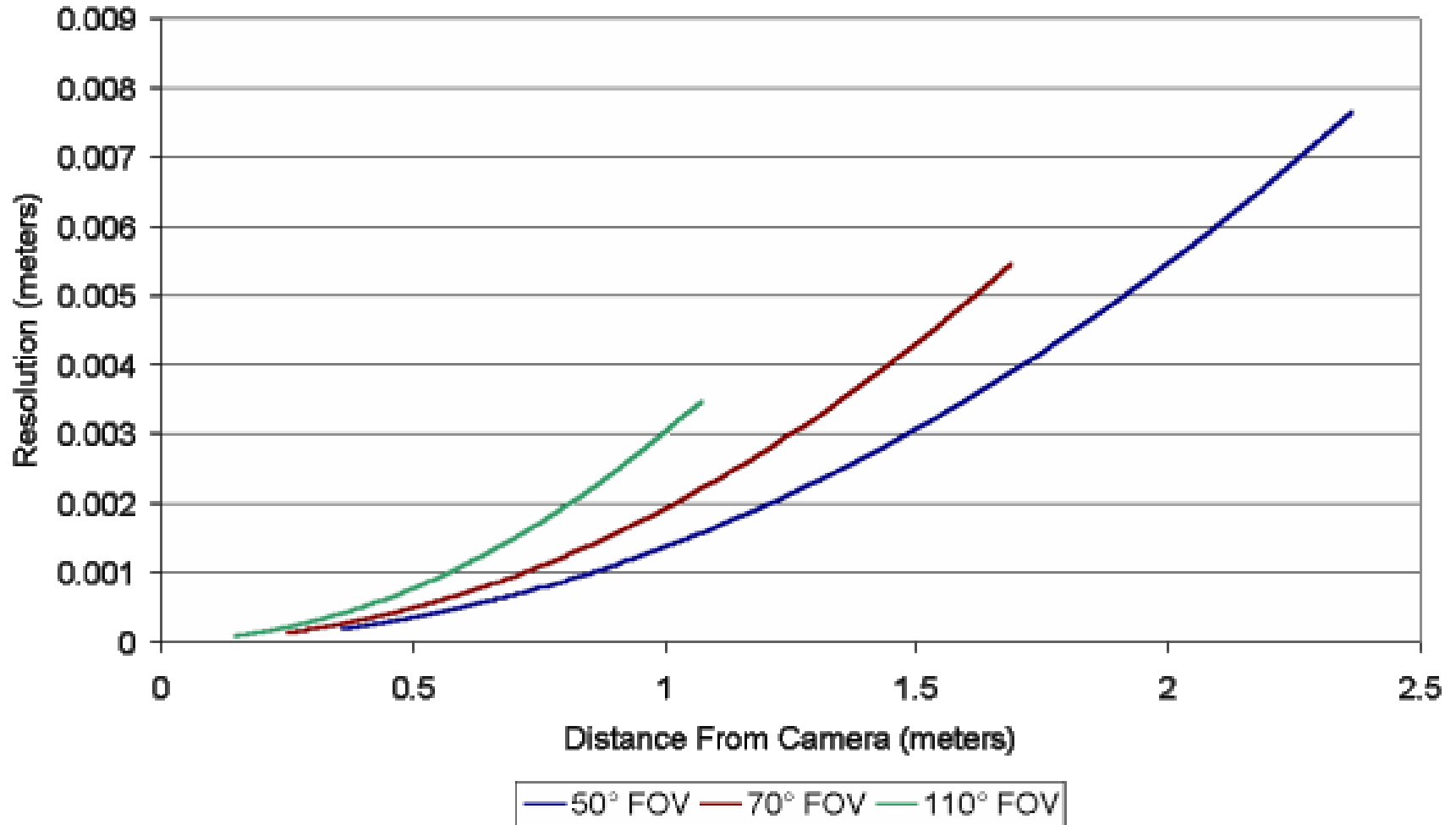
Il y a six épipolaires associées à trois caméras et aux trois projections de P : l'épipolaire e_{ij} correspond à la projection de la droite Pp_i sur l'image j . Ces 6 épipolaires forment 3 paires d'épipolaires conjuguées : l'épipolaire e_{ij} est conjuguée avec l'épipolaire e_{ji} .

Lorsque le point p_1, p_2 et p_3 forment un appariement correct, le point p_i ($i \in \{1,2,3\}$) se trouve nécessairement à l'intersection des épipolaires e_{ki} et e_{ji} . Ceci nous conduit naturellement au processus d'appariement suivant :

Soit un point q_1 de l'image 1 et un point q_2 de l'image 2.
 Si l'appariement q_1/q_2 est correct, alors q_3 (leur homologue dans l'image 3) se trouve précisément à l'intersection de deux épipolaires. On peut remarquer qu'un tel système est parfaitement symétrique et qu'on peut donc vérifier dans l'image k un appariement i/j .



Graph Showing Distance From Camera vs. Resolution for Several Fields of View



Rectified Image

All images processed by the Triclops SDK are rectified to remove lens distortions and misalignments. Below is a set of images showing stereo quality before and after rectification.

Before Rectification - The left image illustrates some of the visible distortions including the curvature of the ceiling, and non-square pixels. To the right is the resultant depth image, which is clearly very poor.

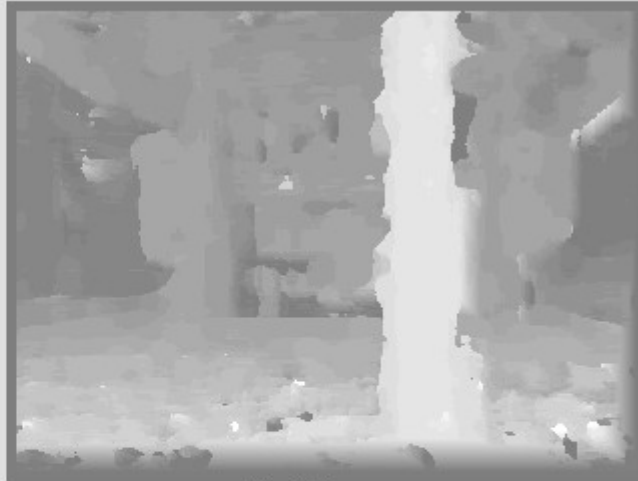


After Rectification - The stereo quality shown in the right image is noticeably better.



**Depth
Image**

In this example, the camera module has simultaneously obtained three images of the scene. While the images appear quite similar, closer inspection reveals a shift between closer objects and those that are further away. Based on the amount of shift, the system is able to determine the distance to the objects in the scene. The depth image represents closer objects with brighter shades of gray and objects further away with darker shades of gray.



Depth Image



Top Image



Left Image



Right Image

Bibliographie

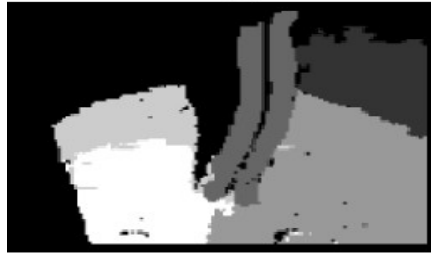
- “Multiple View Geometry”, Hartley/Zisserman 2000, Cambridge University Press, 2000. Une référence !!
- “Une introduction à la géométrie projective”, Daniel Lehmann, Ellipses
- “Algebraic projective geometry”, Semple et Kneeborne, Clarendon Press, Oxford, 1979
- “A computational theory of human stereo vision”, Proc. R. Soc. Lond. B, Vol 204, pp.301-328, 1979
- “A stereo matching algorithm with an adaptative window : theory and experiments, IEEE Trans. On PAMI, vol. 16, pp.920-932, 1994
- “Applications of Non-metric vision to some visually guided robotic tasks”, Robert, Zeller and Faugeras, Technical report 2584, INRIA, 1995.
- “From projective to euclidean reconstruction”, Devernay and Faugeras, Technical report 2725, INRIA, 1995.
- “A computer algorithm for reconstructing a scene from two projections”, Nature, Vol. 293, no. 10, pp. 133-135, 1981

Thèmes à suivre

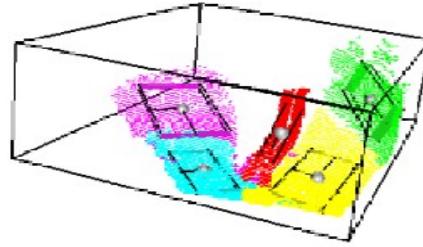
- Analyse de nuages de points 3D



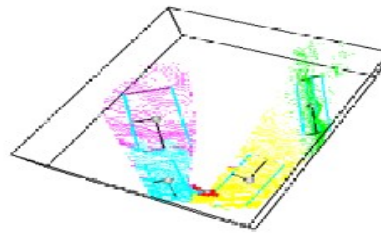
(a)



(b)



(c)



(d)

- Construction de et visualisation par maillages 3D (voir géométrie algorithmique)

