

Stéréo et auto-calibration

Diane Lingrand

`lingrand@essi.fr`

Cours destiné aux élèves de dernière année de cycle ingénieur
version provisoire 2001
mises à jour <http://www.essi.fr/~lingrand/>

Table des matières

1	Introduction	2
2	Mise en correspondance	5
2.1	Entités géométriques	5
2.2	Une méthode de mise en correspondance de points : IMAGE_MATCHING . .	5
2.2.1	Détection de points d'intérêt	6
2.2.2	Corrélation	6
2.2.3	Relaxation	7
3	Géométrie épipolaire	9
3.1	Définition - Principe	9
3.2	Contraintes épipolaires	9
3.3	Géométrie standard	9
3.3.1	Un exemple d'utilisation de la disparité	12
3.3.2	Rectification	13
4	La matrice fondamentale et les homographies	15
4.1	La matrice fondamentale	15
4.1.1	cas de la projection perspective	15
4.1.2	Retour sur la géométrie épipolaire	16
4.2	Les homographies	16
4.3	Relation entre matrice fondamentale et homographie	17
4.4	Un exemple d'utilisation des homographies <i>par Le Barbare</i>	17
4.5	Retour à la mise en correspondance	18
4.5.1	Moindre médiane des carrés	18
4.5.2	Utilisation de la géométrie épipolaire	19
5	Auto-calibration	21
5.1	Qu'est-ce que l'Auto-calibration ?	21
5.2	Cas des paramètres intrinsèques constants	21
5.3	Hypothèses sur le mouvement	21
5.3.1	cas de la rotation pure <i>D'après un article de Richard Hartley "Self-Calibration from Multiple Views with Rotating Camera" dans ECCV'94</i>	22
6	Plus de 2 vues	23
6.1	Géométrie trifocale	23
6.2	Séquences d'images	24
7	Différents types de caméras	26
8	Bibliographie	32

Chapitre 1

Introduction

La stéréoscopie est la perception tridimensionnelle de l'espace à l'aide de deux vues. Pour obtenir une vue stéréoscopique instantanée, l'homme utilise ses deux yeux. Cependant, un seul œil ou une seule caméra suffit à obtenir deux vues à condition que au moins la scène ou le capteur soit en mouvement.

Les figures suivantes présentent des stéréoscopes datant respectivement de 1860 et 1940 :

Brewster stereoscope

England 1860



Format: 70 x 130 mm

Lens: 80 mm

Viewmaster stereoscope

USA 1940 - 1998



Format: 12 x 13 mm

Lens: 50 mm

Afin de fournir à un robot une vision stéréoscopique instantanée, on peut le munir de deux caméras montées sur un axe et dont l'angle entre les axes optiques de deux caméras, appelé aussi vergence, est réglable (voir figure 1.1) ou bien, utiliser d'autres mécanismes, comme par exemple un système de miroir couplé à une unique caméra (voir figures 1.2 et 1.3) et permettant d'obtenir une image composée de deux vues différentes d'une scène.



FIG. 1.1 – Tête stéréoscopique

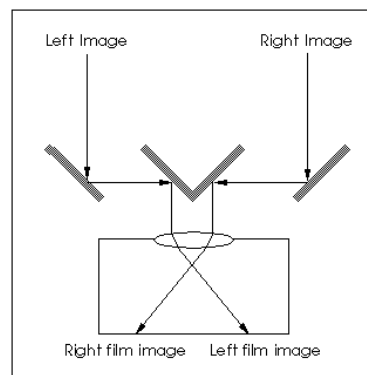


FIG. 1.2 – Principe des miroirs stéréo



FIG. 1.3 – Utilisation de miroirs et casque stéréoscopique

Chapitre 2

Mise en correspondance

La mise en correspondance consiste à trouver, pour un élément ou entité géométrique résultant de la projection d'un objet de l'espace dans une image, le projeté de ce même objet dans l'autre image, appelé correspondant.

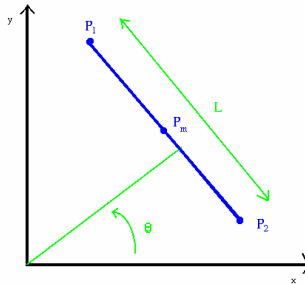
2.1 Entités géométriques

Les entités géométriques auxquelles nous nous intéressons dans les images peuvent être de plusieurs natures. De part la description des images numériques sous forme de tableaux de pixels, on pense naturellement à des points qui sont définis par leur coordonnées et dont la propriété est le niveau de gris ou, pour des images couleurs, des coordonnées chromatiques.

Dans des scènes comportant de nombreux objets géométriques, comme par exemple des scènes d'intérieur, des bâtiments, on va s'intéresser à des droites, ou plutôt, des segments de droites. Ceux-ci peuvent être définis de plusieurs façons :

- 2 points situés chacun à une extrémité
- 1 point situé à une extrémité, une longueur et une orientation
- 1 point milieu, une longueur et une orientation

Les définitions comportant des points situés aux extrémités sont bien moins précises que la dernière méthode, plus généralement adoptée.



Dans des scènes plus générales, on va vouloir mettre en correspondance des courbes. Celles-ci peuvent être paramétrées sous forme de courbes de Bézier ou de B-splines.

2.2 Une méthode de mise en correspondance de points : IMAGE_MATCHING

<http://www.inria.fr/robotvis/personnel/zhang/software.html>

- estimation des correspondances par une méthode classique
- détection de coins points d'intérêts
- corrélation : on obtient des "candidats" à la correspondance
- relaxation
- raffinement par minimisation : Méthode de la moindre médiane des carrés *Least Median Square*

2.2.1 Détection de points d'intérêt

Il existe deux grandes familles de méthodes :

- les méthodes fonctionnant en 2 étapes : détermination de contours chaînés puis détermination des points du contour de courbure maximale
- les méthodes opérant directement sur l'intensité de l'image (gradient et courbure)

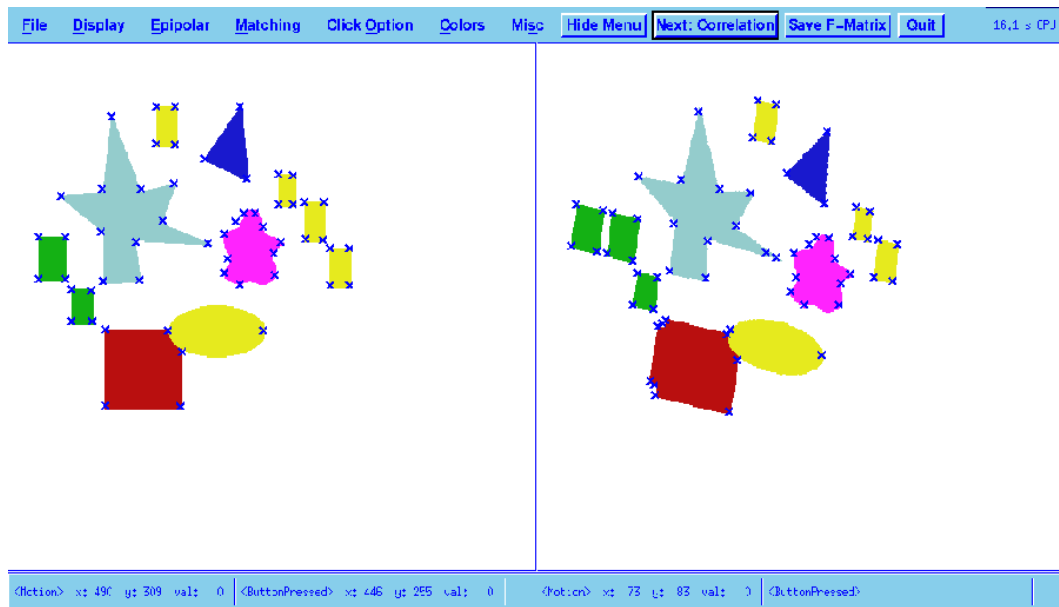
Le détecteur de points utilisé dans cette méthode est celui de Harris et Stephens (1988), basé sur l'opérateur suivant :

$$\mathcal{O} = \det((\widehat{C}(x, y) - k (\text{trace}(\widehat{C}(x, y))))^2$$

où

$$\widehat{C}(x, y) = \begin{bmatrix} \left(\frac{\partial I(x, y)}{\partial x}\right)^2 & \frac{\partial I(x, y)}{\partial x} \frac{\partial I(x, y)}{\partial y} \\ \frac{\partial I(x, y)}{\partial x} \frac{\partial I(x, y)}{\partial y} & \left(\frac{\partial I(x, y)}{\partial y}\right)^2 \end{bmatrix}$$

$I(x, y)$ représentant l'intensité en niveaux de gris du point de l'image de coordonnées (x, y) et \widehat{f} correspondant à f lissée par une gaussienne. Une valeur expérimentalement satisfaisante est $k = 0.04$ et il suffit alors de seuiller pour exhiber les coins.

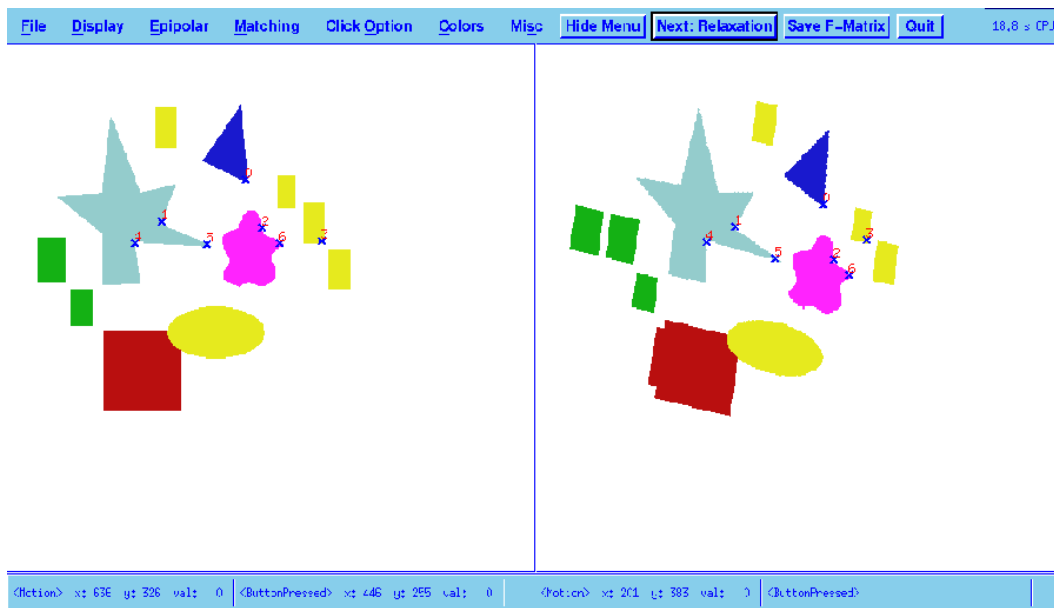
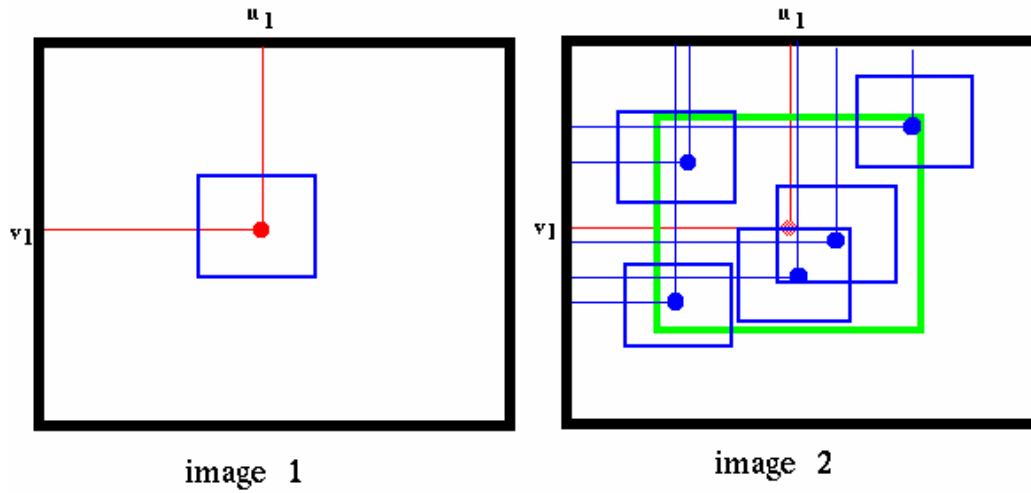


2.2.2 Corrélation

Considérant un point \mathbf{m}_1 de l'image 1, on recherche son correspondant dans l'image 2 à l'intérieur d'un espace de recherche centré sur la position de \mathbf{m}_1 . La taille de l'espace est déterminée par une estimée de la disparité entre les deux images. Par exemple, pour une séquence vidéo, l'espace de recherche sera bien plus petit que pour des vues stéréoscopiques.

Pour chaque point \mathbf{n}_i de l'espace de recherche, on effectue la corrélation entre une fenêtre de l'image 1 centrée sur \mathbf{m}_1 et la fenêtre de l'image 2 centrée sur \mathbf{n}_i . On obtient alors, pour chaque point \mathbf{n}_i un score de corrélation entre -1 et +1.

On fixe alors un seuil au dessus duquel les points seront comme des candidats à l'appariement. Ainsi, chaque point de l'image 1 possède un ensemble de points candidats dans l'image 2 tandis que, inversement, chaque point de l'image 2 possède un ensemble de points candidats dans l'image 1.



2.2.3 Relaxation

Afin de départager les différents points candidats à l'appariement pour un seul point, on va effectuer une opération de relaxation dont l'idée est de propager des contraintes telles que l'unicité ou la continuité sur les appariements .

Cette méthode est basée sur une fonction d'énergie :

$$\mathcal{E} = \sum_i S_M(\mathbf{m}_{1i}, \mathbf{m}_{2i})$$

dépendant d'une force associée à une paire de correspondances définie par :

$$S_M(\mathbf{m}_{1i}, \mathbf{m}_{2i}) = \sum_{(\mathbf{n}_1, \mathbf{n}_2) \in \mathcal{V}(\mathbf{m}_{1i}, \mathbf{m}_{2i})} \delta((\mathbf{m}_{1i}, \mathbf{m}_{2i}), (\mathbf{n}_1, \mathbf{n}_2))$$

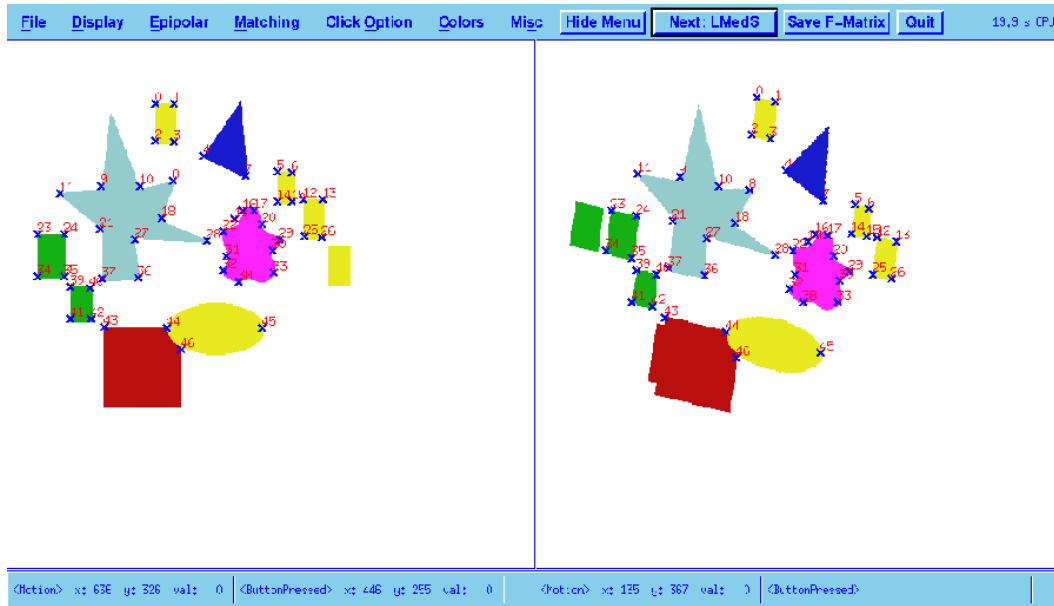
en fonction d'une force de similarité δ

$$\delta((\mathbf{m}_{1i}, \mathbf{m}_{2i}), (\mathbf{n}_1, \mathbf{n}_2)) = \begin{cases} \frac{1}{1 + \frac{\|\mathbf{m}_{1i} - \mathbf{n}_1\| + \|\mathbf{m}_{2i} - \mathbf{n}_2\|}{2}} & \text{si } \frac{1}{k_0} < \frac{\|\mathbf{m}_{2i} - \mathbf{n}_2\|}{\|\mathbf{m}_{1i} - \mathbf{n}_1\|} < k_0 \text{ et } |(\mathbf{m}_{1i} \vec{\mathbf{n}}_1, \mathbf{m}_{2i} \vec{\mathbf{n}}_2)| < \theta_0 \\ 0 & \text{sinon} \end{cases}$$

Cette force de similarité peut être comprise comme une limitation de la rotation et de l'homothétie.

On va donc minimiser cette énergie par un processus itératif jusqu'à convergence :

- calcul des forces pour chaque couple candidat
- minimisation de la fonction d'énergie
- réactualisation des correspondances
- et on recommence jusqu'à convergence de l'énergie



A ce stade, on dispose d'une estimation des appariements. L'étape suivante consiste à utiliser la géométrie épipolaire comme contrainte supplémentaire, comme nous allons l'étudier dans la section suivante.

Chapitre 3

Géométrie épipolaire

3.1 Définition - Principe

On note (voir 3.1 pour plus de clarté) \mathbf{C}_1 et \mathbf{C}_2 les centres optiques des caméras 1 et 2, \mathbf{m}_1 et \mathbf{m}_2 les images de \mathbf{M} . L'image de la droite $(\mathbf{C}_1\mathbf{M})$ (resp. $(\mathbf{C}_2\mathbf{M})$) par la caméra 2 (resp. 1) est la droite épipolaire $(\mathbf{e}_2\mathbf{m}_2)$ (resp. $(\mathbf{e}_1\mathbf{m}_1)$) dont \mathbf{e}_2 (resp. \mathbf{e}_1), appelé épipôle, est le point d'intersection avec l'axe $(\mathbf{C}_1\mathbf{C}_2)$. Si on fait varier le point \mathbf{M} , les droites épipolaires correspondantes vont varier, mais les épipôles \mathbf{e}_1 et \mathbf{e}_2 restent fixes (voir 3.2); ce sont les images respectives des centres optiques \mathbf{C}_2 et \mathbf{C}_1 par les caméras 1 et 2.

3.2 Contraintes épipolaires

La géométrie épipolaire implique naturellement certaines contraintes :

contrainte d'unicité : chaque point de l'image admet au plus un correspondant dans la seconde image (et vice-versa), voir 3.3. Cette contrainte est respectée lorsque la scène est opaque (il n'y a pas de transparence).

contrainte de continuité : on suppose la profondeur de la scène continue (3.4). En pratique, il s'agira d'une continuité par morceaux car des ruptures ont lieu lors de l'occlusion d'un objet par un autre.

contrainte d'ordre : l'ordre des points image est conservé d'une image à l'autre (3.5). Cette contrainte est respectée uniquement lors de faibles variations de profondeur par rapport à l'écart entre les caméras.

3.3 Géométrie standard

Définition

Il s'agit d'un cas particulier pour lequel les lignes épipolaires sont parallèles et horizontales; les épipôles sont donc situés à l'infini (voir figure 3.6).

Cette géométrie possède l'avantage de simplifier beaucoup la mise en correspondance puisque le correspondant d'un point d'une image se trouve dans la même ligne dans l'autre image (voir figure 3.7).

Disparité

Dans ce cas précis de géométrie standard, en notant par D la distance entre les deux caméras, par Z la profondeur d'un point \mathbf{M} de la scène, par f la focale commune des caméras

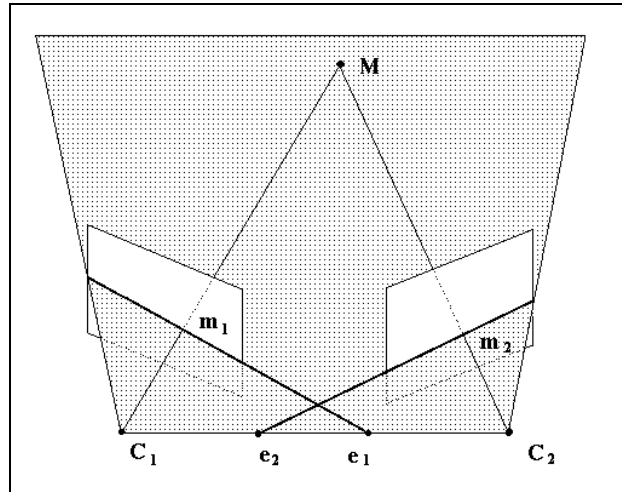


FIG. 3.1 – Géométrie épipolaire : Notations

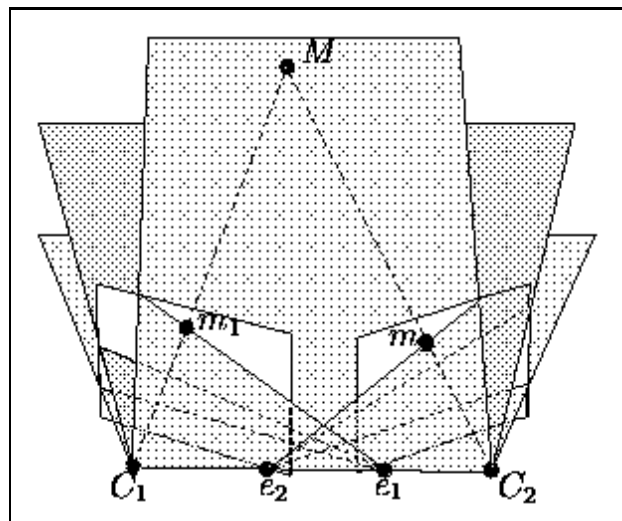


FIG. 3.2 – Principe de la géométrie épipolaire

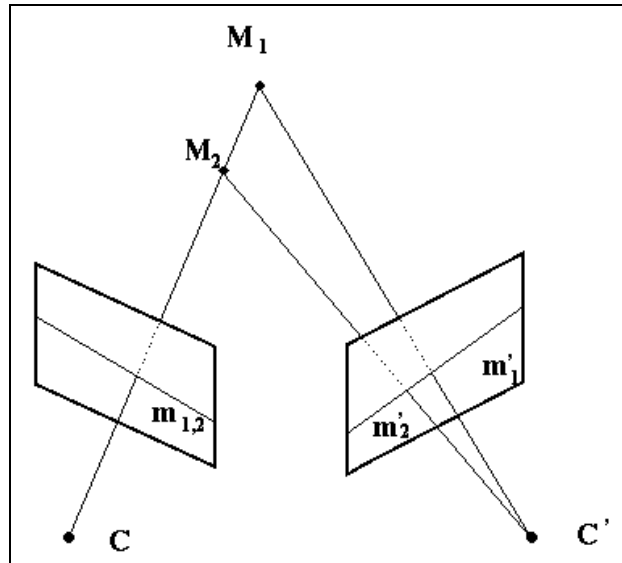


FIG. 3.3 – Violation de la contrainte d'unicité : le point \mathbf{m} de la première image admet deux points \mathbf{m}'_1 et \mathbf{m}'_2 pour correspondants dans la deuxième image, ceci par transparence.

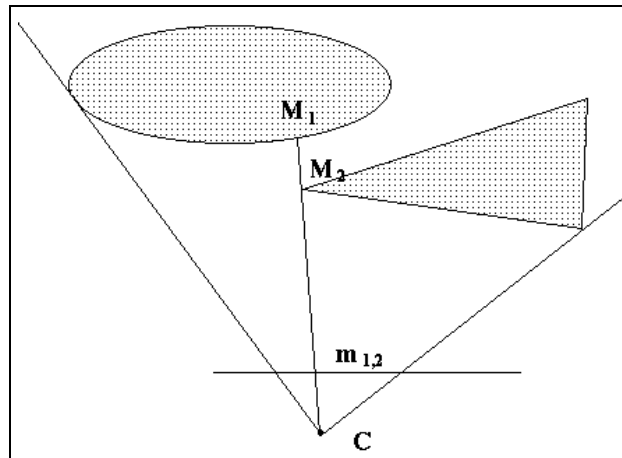


FIG. 3.4 – Violation de la contrainte de continuité : il y a une discontinuité de profondeur en M_1 et M_2 car il y a occlusion d'un objet par un autre.

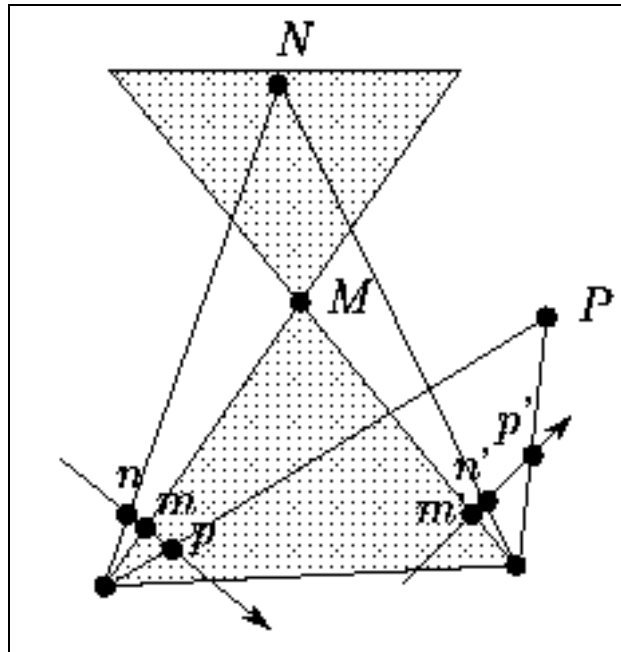


FIG. 3.5 – Violation de la contrainte d'ordre : sur la première image, \mathbf{m} est entre \mathbf{n} et \mathbf{p} , sur la seconde, \mathbf{n}' est entre \mathbf{m}' et \mathbf{p}' . La différence de profondeur des points \mathbf{M} , \mathbf{N} et \mathbf{P} est importante par rapport à l'écart des centres optiques des caméras.

et par d la différence entre les deux abscisses des projetés de \mathbf{M} (voir figure 3.8), on obtient la relation suivante :

$$d = u_2 - u_1 = D \frac{f}{Z}$$

Ce qui permet, pour chaque couple de projeté apparié, de retrouver la profondeur. La disparité a pour propriétés d'être continue par morceaux, monotone et d'avoir un gradient borné.

3.3.1 Un exemple d'utilisation de la disparité

Dans cet exemple, pour chaque point de l'image, on a calculé la disparité et construit ainsi l'image de disparité ou carte dense de disparité (voir figure 3.9). Ainsi, pour chaque point de l'image, on peut déterminer la profondeur et reconstruire le modèle 3D.

Cette méthode possède des avantages :

- carte dense de disparité
- algorithme simple
- parallélisable

et des inconvénients :

- coûteux
- étape de rectification
- précision sur d
- problèmes avec des motifs répétitifs et sur des occlusions

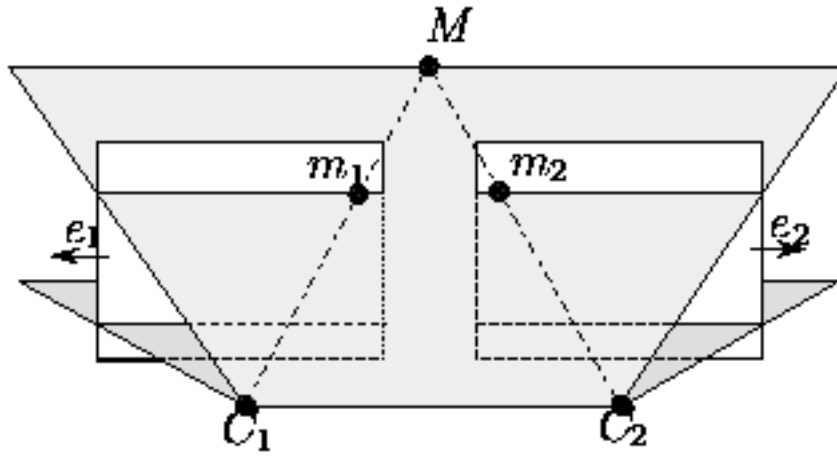


FIG. 3.6 – Caméras en position de géométrie standard

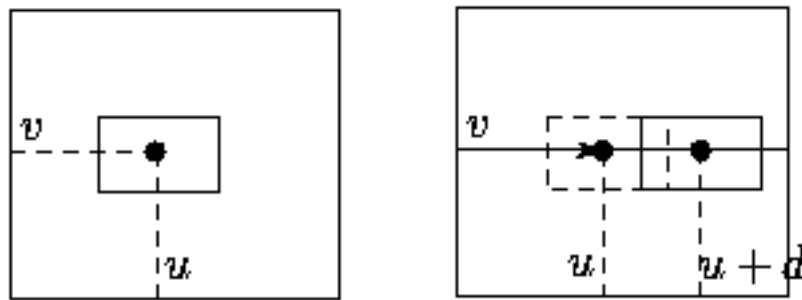


FIG. 3.7 – Recherche d'une correspondance dans le cas de la géométrie standard

3.3.2 Rectification

Cependant, sauf éventuellement avec des systèmes robotiques de précision, les paires d'images ne sont pas en configuration de géométrie standard. On peut cependant effectuer des transformations sur ces images afin de se trouver en position de géométrie standard.

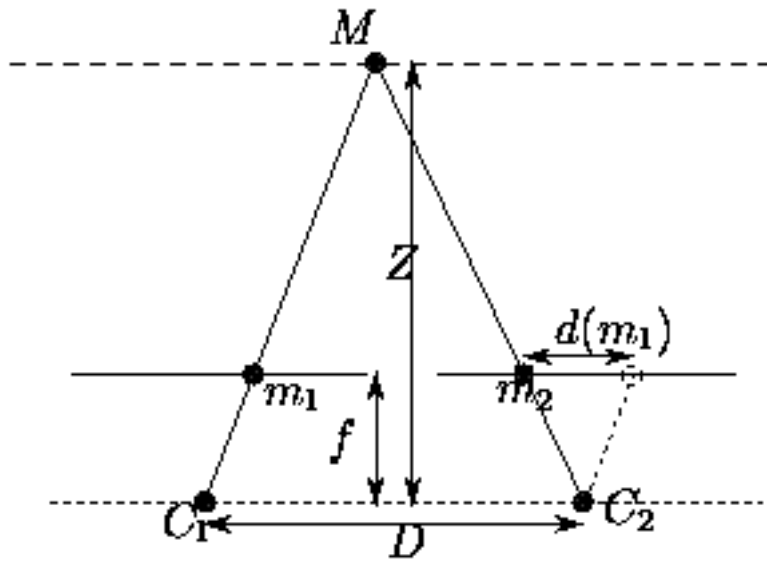


FIG. 3.8 – Disparité

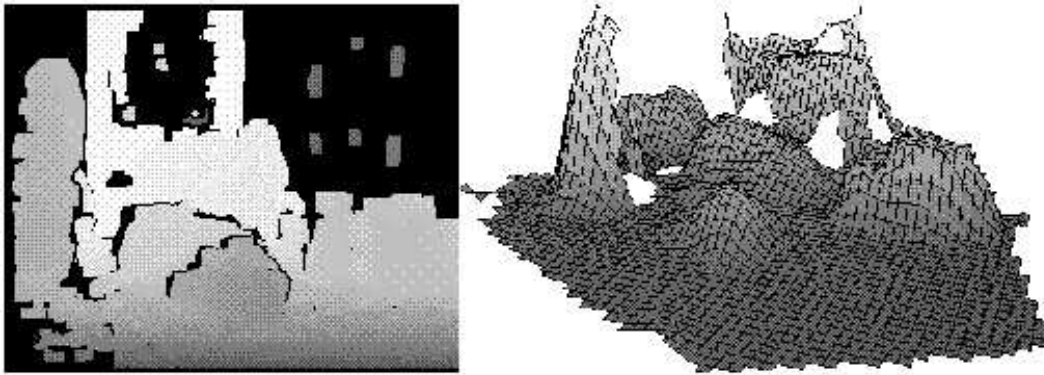


FIG. 3.9 – Carte de disparité et reconstruction 3d correspondante

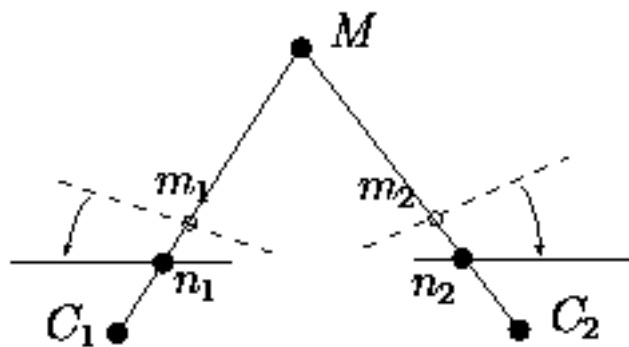


FIG. 3.10 – Rectification de géométrie

Chapitre 4

La matrice fondamentale et les homographies

Après avoir étudié la disparité, nous allons maintenant nous intéresser à la matrice fondamentale qui établit une relation entre les points des images sans dépendre de la profondeur comme c'était le cas dans le chapitre précédent.

4.1 La matrice fondamentale

4.1.1 cas de la projection perspective

Dans ce paragraphe, on notera par $\mathbf{M}_1 = [X_1 \ Y_1 \ Z_1]^T$ un point de l'espace qui se projette en $\mathbf{m}_1 = [u_1 \ v_1 \ 1]^T$ (on prend la partie (3x3) non nulle de la matrice de projection) selon :

$$\kappa_1 \mathbf{m}_1 = \mathbf{A}_1 \mathbf{M}_1 \quad (4.1)$$

Après une rotation \mathbf{R} et une translation \mathbf{t} , il atteint une position $\mathbf{M}_2 = [X_2 \ Y_2 \ Z_2]^T$ et se projette en $\mathbf{m}_2 = [u_2 \ v_2 \ 1]^T$:

$$\mathbf{M}_2 = \mathbf{R} \mathbf{M}_1 + \mathbf{t} \quad (4.2)$$

et

$$\kappa_2 \mathbf{m}_2 = \mathbf{A}_2 \mathbf{M}_2 \quad (4.3)$$

En combinant (4.2) et (4.3), on obtient :

$$\kappa_2 \mathbf{m}_2 = \mathbf{A}_2 \mathbf{R} \mathbf{M}_1 + \mathbf{A}_2 \mathbf{t}$$

\mathbf{M}_1 est remplacé par sa valeur en (4.1) ce qui fournit :

$$\kappa_2 \mathbf{m}_2 = \kappa_1 \underbrace{\mathbf{A}_2 \mathbf{R} \mathbf{A}_1^{-1}}_{\mathbf{H}_\infty} \mathbf{m}_1 + \underbrace{\mathbf{A}_2 \mathbf{t}}_{\mathbf{e}} \quad (4.4)$$

On remarque l'apparition dans cette équation de l'homographie du plan à l'infini (voir paragraphe ??) \mathbf{H}_∞ ainsi que de l'épipôle \mathbf{e} défini précédemment.

Il suffit alors de prendre le produit vectoriel de $\mathbf{A}_2 \mathbf{t}$ par l'équation 4.4 et d'en prendre ensuite la produit scalaire par \mathbf{m}_2 pour obtenir :

$$0 = \mathbf{m}_2^T (\mathbf{A}_2 \tilde{\mathbf{t}}) \mathbf{A}_2 \mathbf{R} \mathbf{A}_1^{-1} \mathbf{m}_1$$

La matrice fondamentale, dans le cas de la projection perspective, s'exprime alors par :

$$\mathbf{F} = (\mathbf{A}_2 \tilde{\mathbf{t}}) \mathbf{A}_2 \mathbf{R} \mathbf{A}_1^{-1}$$

ce qui est une expression qui ne dépend pas des coordonnées des points de l'image, contrairement aux coefficients de la disparité vectorielle.

On peut remarquer dans ce cas que la matrice fondamentale ne permettra jamais de retrouver la norme de la translation; celle-ci ne dépend donc que de deux paramètres.

4.1.2 Retour sur la géométrie épipolaire

La relation fondamentale $\mathbf{m}_2^T \mathbf{F} \mathbf{m}_1 = 0$ exprime le fait que le point \mathbf{m}_2 appartient à la droite épipolaire $\mathbf{F} \mathbf{m}_1$ de même que, en prenant la transposée, le point \mathbf{m}_1 appartient à la droite $\mathbf{F}^T \mathbf{m}_2$. Toutes les droites épipolaires se coupent en un point appelé épipole.

4.2 Les homographies

Dans le cas où la projection est perspective et la translation \mathbf{t} nulle, l'expression de la matrice fondamentale vue précédemment s'annule et on ne peut plus parler alors de matrice fondamentale. Par contre, si on prend les parties (3x3) non nulles de \mathbf{A}_1 et \mathbf{A}_2 , alors, les équations :

$$\begin{cases} Z_1 \mathbf{m}_1 &= \mathbf{A}_1 \mathbf{M}_1 \\ Z_2 \mathbf{m}_2 &= \mathbf{A}_2 \mathbf{M}_2 \\ \mathbf{M}_2 &= \mathbf{R} \mathbf{M}_1 \end{cases}$$

fournissent la relation suivante :

$$\mathbf{m}_2 = \frac{Z_1}{Z_2} \mathbf{A}_2 \mathbf{R} \mathbf{A}_1^{-1} \mathbf{m}_1$$

ce qui constitue une homographie entre les deux images : $\kappa \mathbf{m}_1 = \mathbf{H}_\infty \mathbf{m}_2$ où $\mathbf{H}_\infty = \mathbf{A}_2 \mathbf{R} \mathbf{A}_1^{-1}$.

D'autre part, si on considère une scène plane où le plan a pour normale \mathbf{n} et distance par rapport à l'origine d , on obtient :

$$[\mathbf{n}^T \mathbf{M}_1 = d] \cap [\mathbf{M}_2 = \mathbf{R} \mathbf{M}_1 + \mathbf{t}] \implies \mathbf{M}_2 = \left(\mathbf{R} + \frac{\mathbf{t} \mathbf{n}^T}{d} \right) \mathbf{M}_1$$

Le cas de la rotation pure correspond à ce cas en faisant tendre $\frac{1}{d}$ vers 0. Inversément, ce cas correspond à celui de la rotation pure en prenant $\mathbf{R}' = \mathbf{R} + \frac{\mathbf{t} \mathbf{n}^T}{d}$. Ainsi, il y a équivalence entre mouvement sans translation et mouvement d'un plan. Si on se place dans le cas d'une projection perspective, il existe une relation homographique entre les deux images :

$$\lambda \mathbf{m}_2 = \mathbf{H} \mathbf{m}_1$$

avec :

$$\mathbf{H} = \mathbf{A}_2 \left(\mathbf{R} + \frac{\mathbf{t} \mathbf{n}^T}{d} \right) \mathbf{A}_1^{-1}$$

Lorsque d tend vers l'infini, ce qui correspond à des points situés sur un plan à l'infini, le terme $\frac{\mathbf{t} \mathbf{n}^T}{d}$ tend vers 0 et l'expression de l'homographie \mathbf{H} tend vers l'expression correspondant au cas de la rotation pure :

$$\mathbf{H} \xrightarrow{d \rightarrow \infty} \mathbf{A}_2 \mathbf{R} \mathbf{A}_1^{-1} = \mathbf{H}_\infty$$

\mathbf{H}_∞ représente ainsi l'homographie du plan à l'infini.

4.3 Relation entre matrice fondamentale et homographie

Réécrivons les équations obtenues précédemment :

$$Z_2 \mathbf{m}_2 = Z_1 \mathbf{A}_2 \mathbf{R} \mathbf{A}_1^{-1} \mathbf{m}_1 + \mathbf{A}_2 \mathbf{t}$$

$$\mathbf{F} = (\tilde{\mathbf{A}}_2 \mathbf{t}) \mathbf{A}_2 \mathbf{R} \mathbf{A}_1^{-1}$$

$$\mathbf{H} = \mathbf{A}_2 \mathbf{R} \mathbf{A}_1^{-1} + \mathbf{A}_2 \mathbf{t} \frac{\mathbf{n}^T}{d} \mathbf{A}_1^{-1}$$

On peut alors définir plusieurs entités :

Homographie des points à l'infini :

$$\mathbf{H}_\infty = \mathbf{A}_2 \mathbf{R} \mathbf{A}_1^{-1}$$

foyer d'expansion :

$$\mathbf{e}' = \mathbf{A}_2 \mathbf{t}$$

matrice fondamentale : $\mathbf{F} = \tilde{\mathbf{e}}' \mathbf{H}_\infty$

homographie : $\mathbf{H} = \mathbf{H}_\infty + \mathbf{e}' \frac{\mathbf{n}^T}{d} \mathbf{A}_1^{-1}$

4.4 Un exemple d'utilisation des homographies *par Le Barbare*

On veut retrouver les structures planes d'une scène. Pour cela, on va tirer aléatoirement des sous-ensembles de points et déterminer de façon robuste des relations homographiques. On fusionne ensuite les sous-ensembles ayant même matrice homographique jusqu'à obtenir des ensembles de points appartenant à différents plans de la scène. Dans cet exemple, trois plans ont été détectés : le plan du tableau et deux des faces du carton.

Si l'on observe les résultats numériques :

– matrice fondamentale

$$\mathbf{F} = \begin{pmatrix} 0 & 4.76495e-06 & -0.00220561 \\ -4.76495e-06 & 0 & -0.707103 \\ 0.00220561 & 0.707103 & 0 \end{pmatrix}$$

– Collinéations

$$\mathbf{H}_1 = \begin{pmatrix} 0.00764348 & 3.92544e-05 & -0.99992 \\ -4.9159e-06 & 0.00709485 & -0.000297399 \\ 0 & 0 & 0.00708294 \end{pmatrix}$$

$$\mathbf{H}_2 = \begin{pmatrix} 0.013372 & 3.25211e-05 & -0.999683 \\ 0.013372 & 3.25211e-05 & -0.999683 \\ -1.72255e-06 & 0.0143671 & 0.00651458 \\ 0 & 0 & 0.0143814 \end{pmatrix}$$

$$\mathbf{H}_3 = \begin{pmatrix} 0.0209325 & -5.31735e-05 & -0.99922 \\ -2.68491e-05 & 0.0208353 & 0.0158142 \\ 0 & 0 & 0.0209152 \end{pmatrix}$$

on observe que \mathbf{H}_1 et \mathbf{H}_2 ont des normales opposées en x ce qui correspond effectivement aux faces orthogonales du carton et que \mathbf{H}_3 a une composante en x faible ce qui correspond au plan du tableau, approximativement fronto-parallèle.

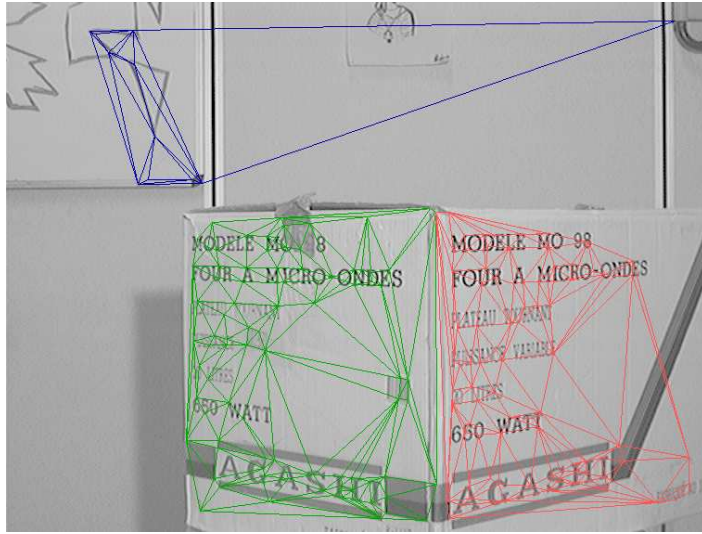


FIG. 4.1 – Segmentation d’une scène en structures planes

4.5 Retour à la mise en correspondance

On va maintenant utiliser la relation fondamentale reliant les points images. Pour cela, on va estimer la matrice fondamentale à partir des correspondances estimées. Parmi les points candidats, certains sont erronés, d’autres sont bruités. On va donc utiliser une méthode robuste permettant de calculer cette matrice fondamentale à partir des données non aberrantes.

4.5.1 Moindre médiane des carrés

Cette méthode utilise la technique d’échantillonnage aléatoire : au lieu d’estimer les paramètres pour chaque sous-ensemble de données, on ne les estime que sur certains sous-ensembles, tirés aléatoirement, ceci afin de réduire la complexité de l’algorithme. On minimise ici non pas la somme des carrés des distances mais la médiane de ces mêmes carrés, ce qui permet d’avoir jusqu’à 50% de données erronées. L’algorithme est le suivant :

1. On réalise aléatoirement m sous-ensembles de données de taille minimale (la plus petite taille qui permette d’estimer les paramètres).
2. Pour chaque sous-ensemble j , on calcule les paramètres $A_j = (a_i)_{1 \leq i \leq m}$ (il peut exister plusieurs solutions).
3. Pour chaque solution, on calcule la médiane du résidu au carré M_j sur toutes les données (voir figure 4.2).
4. On conserve la solution pour laquelle M_j est minimale.
5. On affine finalement le résultat : on compense la non-optimalité de la méthode au bruit gaussien en utilisant l’estimée de la déviation standard robuste : $\hat{\sigma} = 1,4826 [1 + \frac{5}{(n-p)}] \sqrt{M_j}$. On élimine ainsi les données aberrantes que l’on qualifie ainsi si, r_k étant le résidu de la donnée k , : $r_k^2 \geq (2,5 \hat{\sigma})^2$.

Pour estimer le nombre m de sous-échantillons de taille p que l’on doit tirer à la première étape, on calcule la probabilité qu’au moins un des sous-échantillons fournisse une bonne

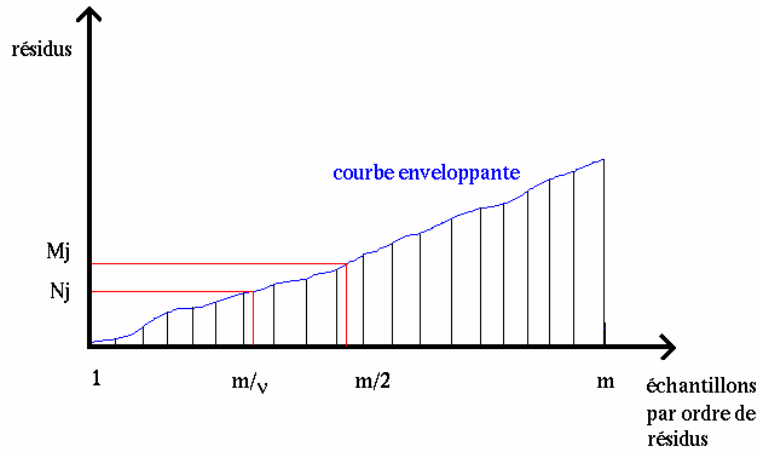
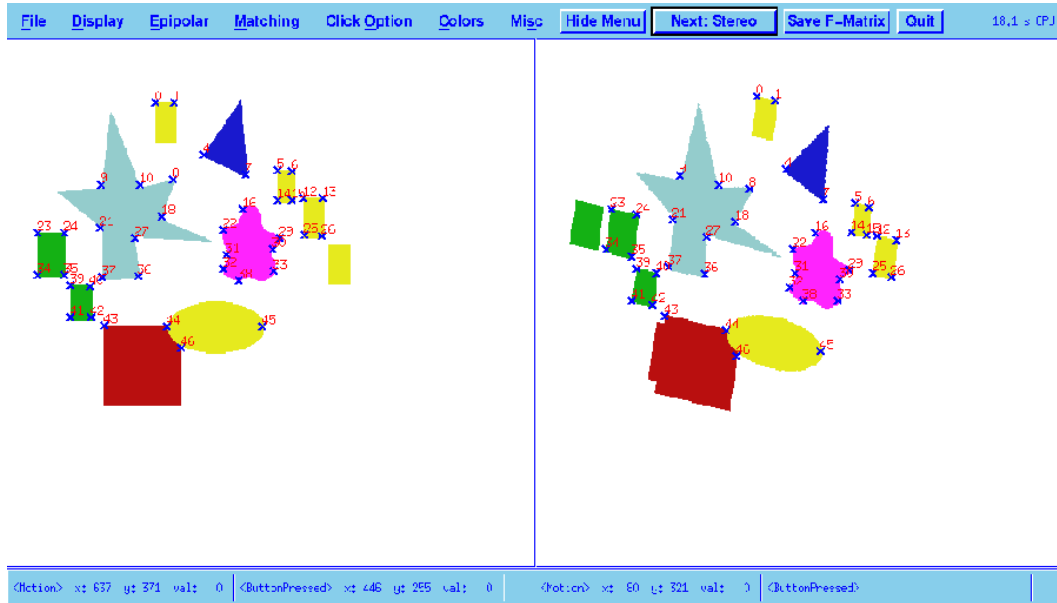


FIG. 4.2 – Détermination de la médiane

estimation. En notant ϵ la fraction de points aberrants,

$$P = 1 - [1 - (1 - \epsilon)^p]^m \implies m = \frac{\log(1 - P)}{\log(1 - (1 - \epsilon)^p)}$$

Si on pose, par exemple, $\epsilon = 40\%$, $p = 7$ et $P = 0,99$, on obtient $m = 163$.



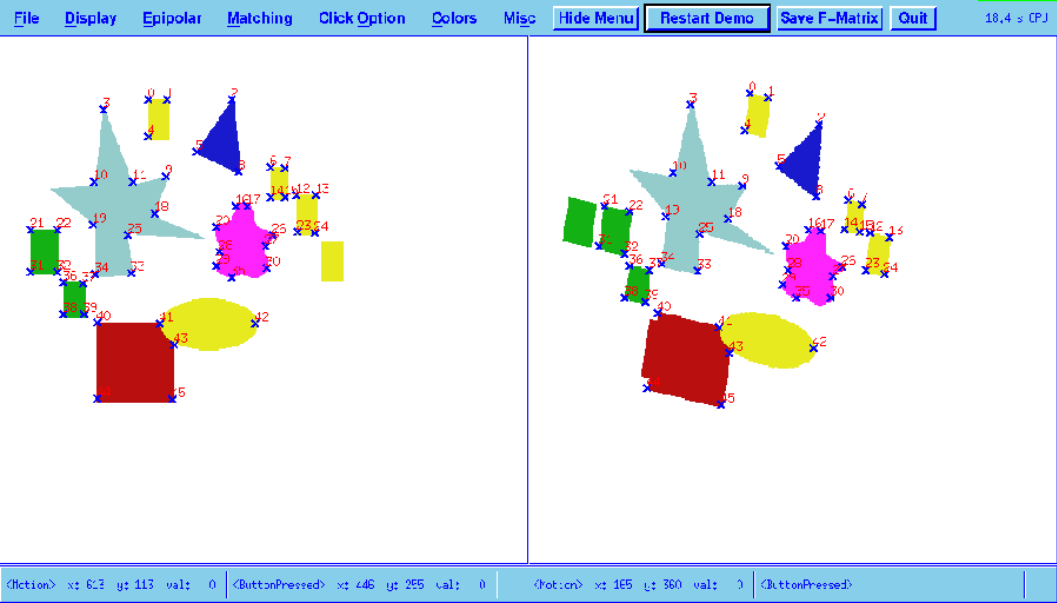
4.5.2 Utilisation de la géométrie épipolaire

Dans notre cas, les résidus correspondent aux distances des points images aux droites épipolaires :

$$\epsilon_{\mathbf{F}}^2(\mathbf{F}) = \sum_{\{\mathbf{m}_1\}} [d(\mathbf{m}_2, \mathbf{F}\mathbf{m}_1)^2 + d(\mathbf{m}_1, \mathbf{F}^T\mathbf{m}_2)^2] \quad (4.5)$$

Ainsi, après élimination des points aberrants, on obtient une meilleure estimée de la

matrice fondamentale que l'on utilise à nouveau pour réestimer les correspondances dont voici le résultat sur notre exemple de données synthétiques :



Chapitre 5

Auto-calibration

5.1 Qu'est-ce que l'Auto-calibration ?

C'est trouver un maximum de paramètres (projection, mouvement, structure) sans s'encombrer d'une mire de calibration. Mais pour cela, il va falloir faire des hypothèses.

L'hypothèse la plus connue est la plus utilisée consiste à supposer que les paramètres intrinsèques sont constants. Ce n'est pourtant pas une hypothèse réaliste lors de zoom ou changement de focus. D'autres hypothèses concernent une partie des paramètres intrinsèques ainsi que la connaissance de propriétés sur le mouvement (par exemple, pour une voiture roulant sur une route, on suppose que le mouvement est plan) ou la structure (points coplanaires, arêtes orthogonales, ...).

5.2 Cas des paramètres intrinsèques constants

Intéressons nous aux degrés de liberté. La matrice de projection, dans le cas de la projection perspective, possède de 3 à 5 paramètres, la rotation, 3 et la translation, 2 dans le cas où on utilise la matrice fondamentale. Cette dernière possède 9 coefficients mais est définie à un facteur d'échelle près et son déterminant est nul par définition ¹ ce qui fait 7 degrés de liberté pour 8 à 10 paramètres.

Dans le cas où on a 9 paramètres, on obtient ainsi 2 relations reliant les paramètres intrinsèques et qui sont connues sont le nom d'équations de Kruppa. Il suffit alors d'effectuer 2 déplacements indépendants pour déterminer l'ensemble des paramètres.

5.3 Hypothèses sur le mouvement

Les hypothèses couramment faites sur le mouvements sont les suivantes :

- translation pure
- rotation pure
- translation orthogonale à l'axe de rotation
- zoom, utilisation du focus, connaissance de la vergence

Une autre étude consiste en l'inventaire des cas particuliers.

¹A votre avis, pourquoi ?

5.3.1 cas de la rotation pure *D'après un article de Richard Hartley "Self-Calibration from Multiple Views with Rotating Camera" dans ECCV'94*

Nous sommes dans le cas d'une caméra à paramètres constants et de points 3D fixes. Comme nous l'avons vu précédemment, nous sommes en présence d'une relation homographique entre les points des deux images.

$$\left. \begin{array}{l} Z \mathbf{m} = \mathbf{A} \mathbf{R} \mathbf{M} \\ Z \mathbf{m}' = \mathbf{A} \mathbf{R}' \mathbf{M} \end{array} \right\} \Rightarrow \mathbf{m}' = \mathbf{A} \mathbf{R}' \mathbf{R}^{-1} \mathbf{A}^{-1} \mathbf{m} = \mathbf{H} \mathbf{m}$$

L'algorithme est le suivant :

1. Pour chaque image i , calculer la transformation \mathbf{H}_i t.q. $\mathbf{m}_i = \mathbf{H} \mathbf{m}_0$.
2. Trouver une matrice triangulaire supérieure \mathbf{A} t.q. $\mathbf{A}^{-1} \mathbf{P}_i \mathbf{A} = \mathbf{R}_i$, \mathbf{R}_i rotation.
3. Raffiner l'estimation de \mathbf{H}_i par la technique itérative de Levenberg-Marquardt.

Et si nous détaillons l'étape 2 de l'algorithme :

$$(\mathbf{A} \mathbf{A}^T) \mathbf{H}_i = \mathbf{H}_i (\mathbf{A} \mathbf{A}^T)$$

- trouver $\mathbf{C} = \mathbf{A} \mathbf{A}^T$
- trouver \mathbf{A} par factorisation de Choleski

Chapitre 6

Plus de 2 vues

6.1 Géométrie trifocale

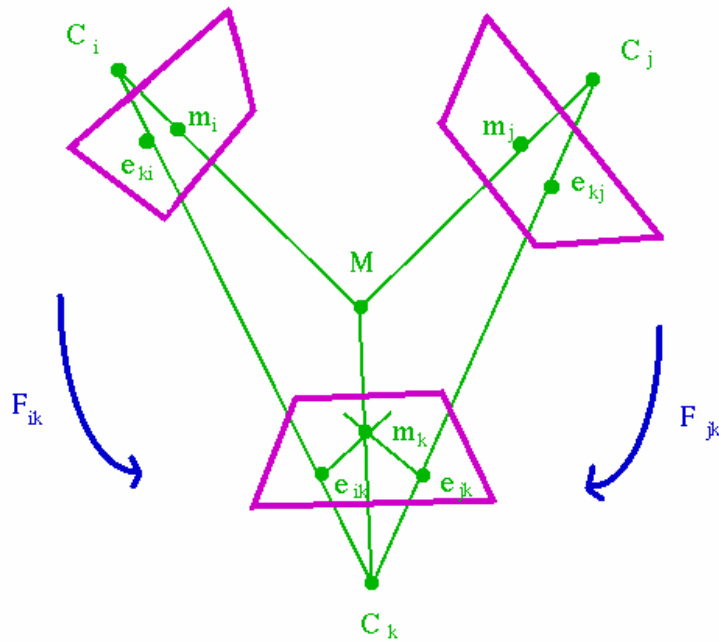


FIG. 6.1 – Géométrie épipolaire dans le cas de 3 caméras

Dans le cas de deux vues, on obtenait, sauf cas particuliers, une relation que l'on peut qualifier d'ordre 0 sur les projetés car elle ne concernait que les coordonnées des points images. Dans le cas de trois vues, on obtient évidemment des relations entre coordonnées mais également des relations entre les tangentes ainsi que les courbures des contours :

contrainte d'ordre 0 : point projeté

$$\left. \begin{array}{l} \mathbf{m}_k^T \mathbf{F}_{ik} \mathbf{m}_i = 0 \\ \mathbf{m}_k^T \mathbf{F}_{jk} \mathbf{m}_j = 0 \end{array} \right\} \Rightarrow \mathbf{m}_k = \mathbf{F}_{ik} \mathbf{m}_i \wedge \mathbf{F}_{jk} \mathbf{m}_j$$

contrainte d'ordre 1 : tangente au point projeté (pente τ , point ∞ : t)

$$\tau_k = (\mathbf{F}_{ik} \mathbf{m}_i \wedge \mathbf{F}_{jk} \mathbf{m}_j) \wedge (\mathbf{F}_{ik} \mathbf{t}_i \wedge \mathbf{F}_{jk} (\mathbf{F}_{ij} \mathbf{t}_i \wedge (\mathbf{m}_j \wedge \mathbf{t}_j)))$$

contrainte d'ordre 2 : courbure au point projeté (la relation existe mais sa formulation est compliquée)

On définit le plan trifocale comme le plan des centres des caméras $\mathcal{P}(C_i, C_j, C_k)$

Il existe, tout comme dans le cas de 2 vues, des cas particuliers :

- \mathbf{M} est dans le plan trifocal : il faut alors se référer aux homographies
- C_i, C_j et C_k alignés

L'apport d'une troisième vue permet, d'une part, de réduire les ambiguïtés en imposant une contrainte géométrique et, d'autre part, de permettre la synthèse d'une troisième vue : on connaît la position, la tangente et la courbure du point synthétique.

6.2 Séquences d'images

Au delà de 3 vues, on n'obtient pas de contrainte algébrique supplémentaire. Il est cependant clair que si plus de 3 images n'apportent pas de contraintes au niveau algébrique, l'apport de nouvelles vues comporte des avantages certains :

- les parties occultées de la scène seront évidemment d'autant plus visibles que l'on accumulera les points de vues différents
- les parties ne vérifiant pas les contraintes de la géométrie épipolaire (voir 3.2) par suite de configuration singulière seront vues correctement si l'on dispose de plusieurs vues
- la détermination de paramètres du problème lorsqu'ils sont trop nombreux pour être déterminés par deux ou trois vues (par exemple en effectuant différents mouvements de rotation)
- le raffinement numérique de l'estimation de la structure tridimensionnelle

De plus, dans un but d'application de réalité virtuelle, deux ou trois vues ne sont pas suffisantes si on désire effectuer une reconstruction sous tous les angles et s'affranchir un maximum du problème des occlusions.

Les images sont alors souvent traitées 2 par 2 voir 3 par 3, plus rarement. Quelques modèles de mouvement continu sont cependant disponibles.

Mise en correspondance de points sur une séquence d'images : Pour cela, plusieurs approches coexistent. On peut considérer simplement le problème de mise en correspondance pour des images prises deux à deux. Sachant par ailleurs que les trilinearités apportent des contraintes supplémentaires, on peut alors songer à considérer les images trois par trois. Lorsqu'il s'agit d'une séquence d'images avec forte disparité d'une image à l'autre, on propagera ainsi les correspondances. Par contre, à moins d'un mouvement de va-et-vient, chaque point n'aura des correspondants que dans quelques images. C'est un cas qui nous permettra d'avoir une vue bien plus large mais qui ne permettra pas d'obtenir plus de précision sur les paramètres de mouvement, de caméra ou de structure. Diverses techniques de fusion des données ainsi recueillies existent mais ne font pas partie du sujet abordé ici.

Si la séquence vidéo ne présente pas de forte disparité entre deux images, on va pouvoir suivre les correspondances sur un plus grand nombre de vues et des algorithmes plus simples que ceux évoqués précédemment peuvent être mis en oeuvre. Les algorithmes travaillant "au vol" sur les séquences d'images, c'est-à-dire traitant les images au fur et à mesure de leur acquisition, calculent une estimée de la position future puis, lorsque l'image suivante est disponible, réajustent au besoin leur mécanisme de précision. C'est le principe du filtre de Kalman que nombre d'auteurs ont utilisé.

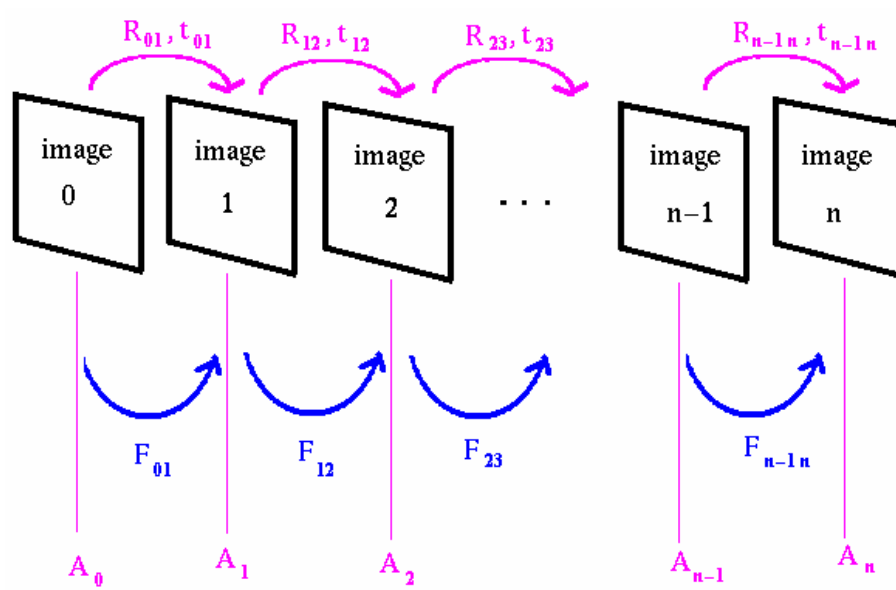


FIG. 6.2 – N vues

Chapitre 7

Différents types de caméras

Les problèmes que l'on cherche à résoudre en Vision par Ordinateur sont de nature assez diverses et d'autres alternatives sont proposées, permettant d'adapter les capteurs à la tâche de vision considérée.

Prenons l'exemple de la navigation d'un robot dans un monde comportant des obstacles. Une reconstruction 3D très précise de l'environnement n'est pas nécessaire car seule une mesure de distance des objets les plus proches est nécessaire. Par contre, une rapidité de calcul est impérative. On peut alors choisir un algorithme simple et efficace donnant une reconstruction même grossière de l'environnement.

Ces problèmes ont cependant été abordés par certains auteurs selon un axe tout à fait différent : ils ont cherché à construire une caméra adaptée à un problème donné en fournissant des données appropriées au problème considéré (voir 7.2).

les lentille demi-sphérique ont été introduites en 1986, ces lentilles offrent une vue panoramique pour une surveillance sur un grand angle (exemple au fond d'une piscine). Par contre, elles ne permettent pas la détection d'obstacle car justement, ceux-ci apparaissent en périphérie des images et la résolution est alors mauvaise.

le miroir conique a été introduit par Jarvis et Byrne en 1988 puis implémenté par Yagi et son équipe en 1994.

En exprimant les coordonnées des points 3D dans un repère cylindrique $\mathbf{M}(r, \theta, Z)$ et celles des points 2D dans un repère polaire $\mathbf{m}(\rho, \theta)$ (voir figure 7.1), on obtient les équations suivantes :

$$\begin{aligned}\rho &= Z \tan(\alpha) + Z \tan\left(\frac{\pi}{2} - 2\alpha\right) \\ &= \frac{Z}{\sin(2\alpha)}\end{aligned}$$

Pour obtenir un champ de vue le plus large possible, on prend $\alpha = \frac{\pi}{4}$ et obtient alors la relation suivante : $\rho = Z$. Par contre, la distance d'un objet à l'axe du cône n'intervient pas dans la projection.

Ce système est donc dédié à l'évitement d'obstacle verticaux posés sur le sol mais ne permet pas d'évitement d'obstacles généraux et ne se prête absolument pas à la reconstruction tridimensionnelle.

le miroir à double courbure a été introduit récemment par Southwell et Basu en 1996, ce miroir à pour but la reconstruction tridimensionnelle de scène et d'éviter de projeter les objets sur de trop petites portions d'image. Pour le moment, la qualité des images est faible et les distorsions grandes.

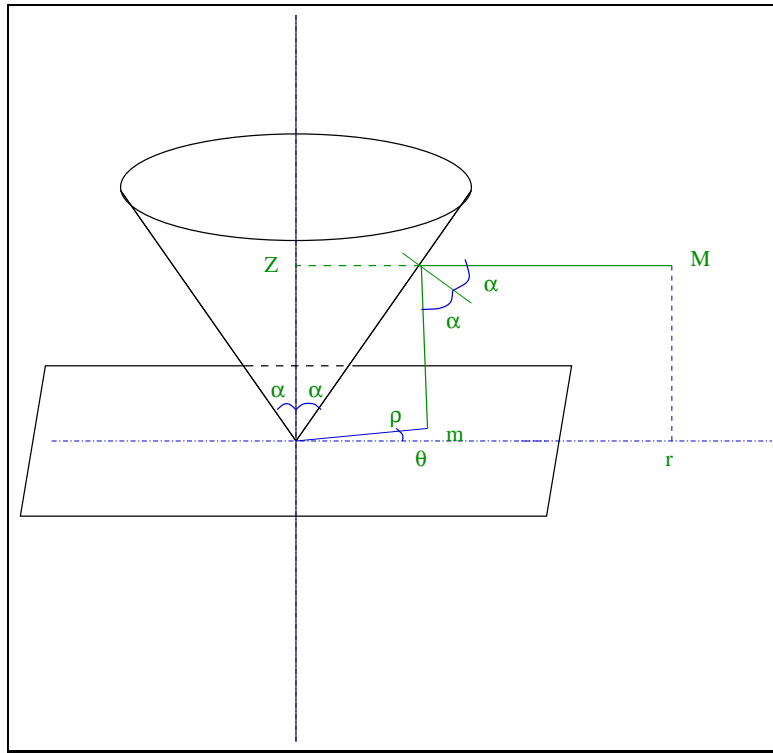


FIG. 7.1 – Miroir conique : notations et calcul des angles

le miroir parabolique a pour but la vision panoramique ainsi que l'évitement d'obstacle ou la reconstruction tridimensionnelle. Cette caméra est disponible dans le commerce (voir par exemple : <http://www.cyclovision.com/>) et fournit des images de bonne qualité.

des barettes CCD en rotation autour d'un axe commun : ce système (voir les thèses de T. Manière et R. Benosman) a été conçu récemment dans un but de reconstruction intérieure de scènes tridimensionnelles. Outre la qualité des images, ce système possède l'avantage de faire de la stéréo d'une dimension de moins que dans le cas des caméras classiques puisque les barettes CCD sont suffisamment précisément alignées pour réduire l'espace de recherche à une colonne de l'image. Le cas de la géométrie épipolaire en dimension 1 a été traité en détail dans la thèse Benosman.

Dans la suite, nous ne considérerons que les caméras ou appareils photos standards que l'on trouve dans le commerce à base d'éléments à transfert de charge (CCD).

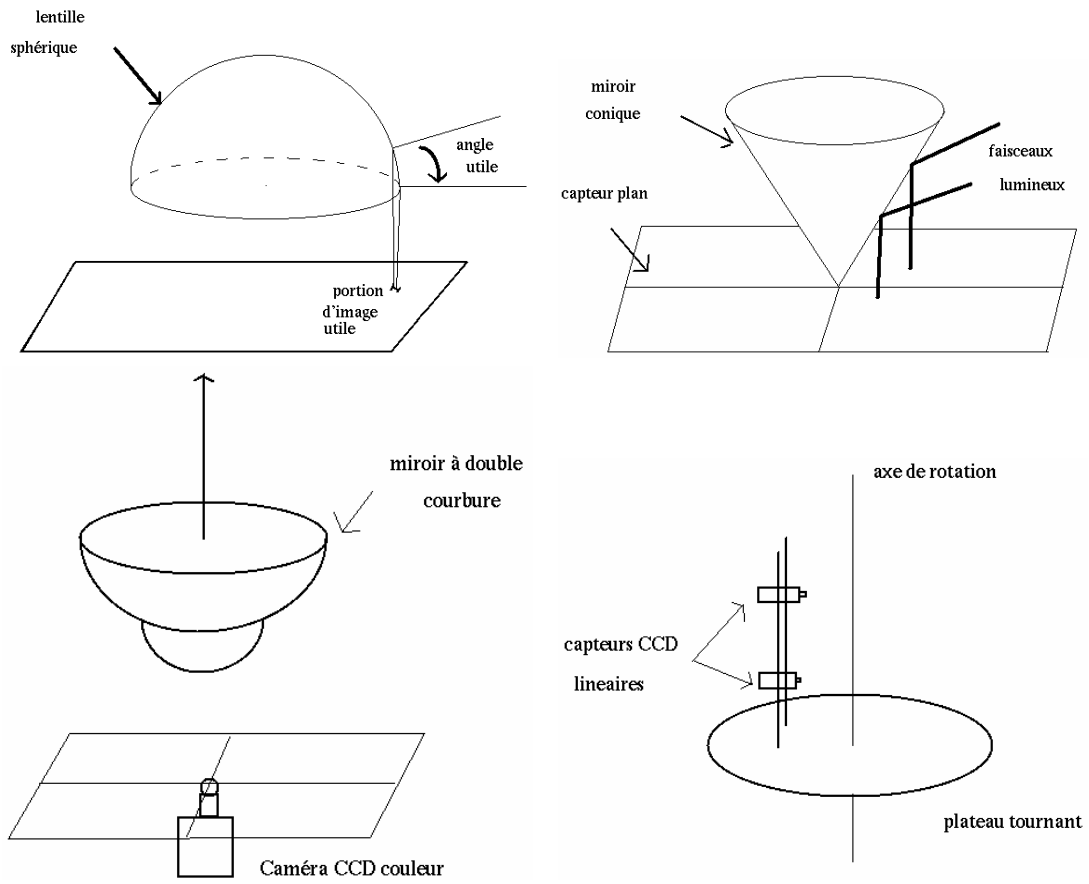


FIG. 7.2 – de gauche à droite et de haut en bas : (a) Lentille sphérique; (b) Miroir conique; (c) Miroir à double courbure; (d) Barettes CCD en rotation;

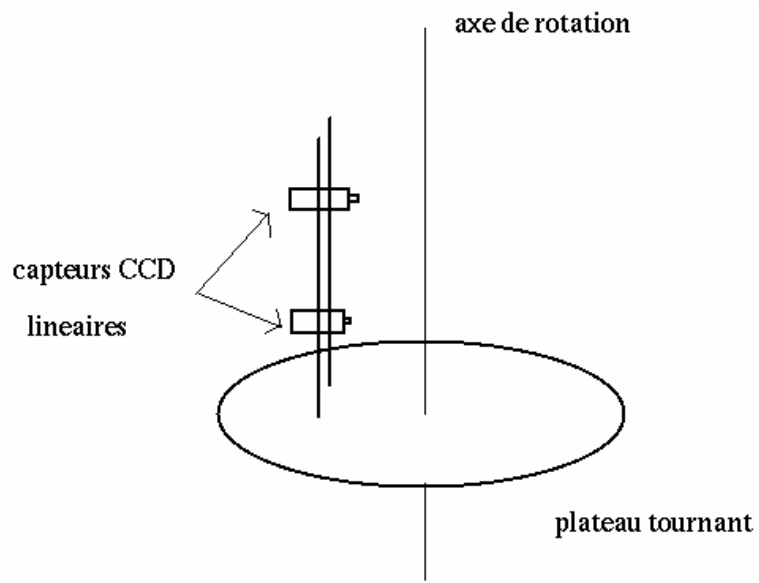


FIG. 7.3 – Le modèle cylindrique



FIG. 7.4 – Le système de vision périphérique



FIG. 7.5 – Couple d'images stéréoscopiques

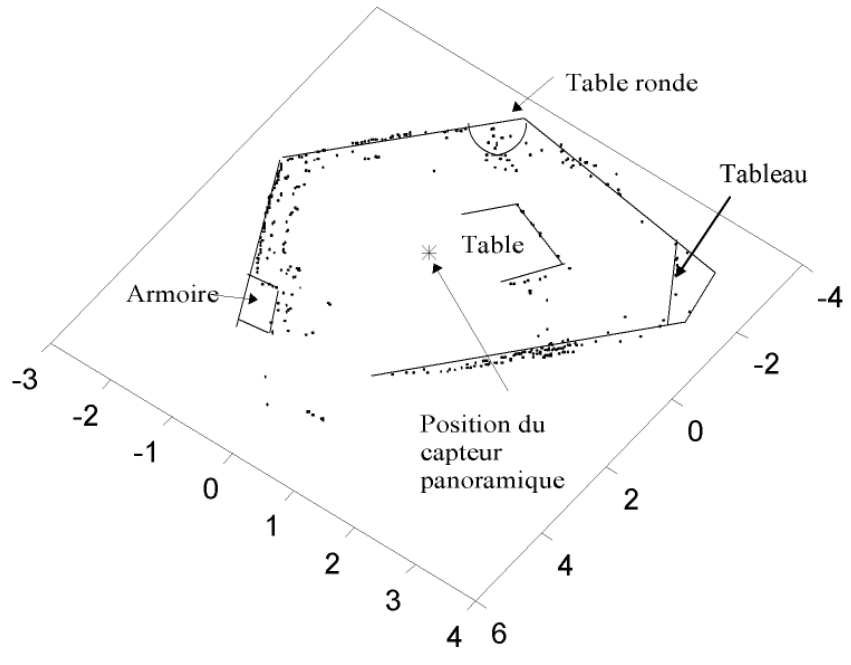


FIG. 7.6 – Reconstruction 3D

Chapitre 8

Bibliographie

Géométrie épipolaire :

- <http://www.inria.fr/robotvis/personnel/sbounou/Meta3DViewer/EpipolarGeo.html>

Stéréo :

- <http://www.inria.fr/robotvis/personnel/vthierry/Enseign/PolyStereo/stereo.ps>
- Stéréo Club Français : <http://www.cnam.fr/hebergement/scf/>
- Spécifications de l'appareil photo stéréo idéal :
http://www.cnam.fr/hebergement/scf/specif_camera.html

Auto-Calibration :

- Thèse de Reyes Enciso : *Auto-Calibration des Capteurs Visuels Actifs. Reconstruction 3D Active*, Université de Paris XI, Orsay, 1995