

République Algérienne Démocratique et Populaire  
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique



**Université Batna 2**  
**Faculté de Technologie**  
**Département d'Électronique**



## **Mémoire**

**Présenté pour l'obtention du diplôme de  
MAGISTER en Électronique**

### **OPTION**

Contrôle

**Par**

**BRIOUA Abdelkrim**

### **Thème**

---

**Reconstitution d'une scène 3D à partir d'indices visuels 2D**

---

Soutenu devant le jury composé de :

Dr. SAIDI Lamir	Prof.	Université Batna 2	Président
Dr. BOUGUECHAL Nour-Eddine	Prof.	Université Batna 2	Rapporteur
Dr. BOUOUDEN Sofiane	M.C.A.	Université de Khenchela	Examineur
Dr. AISSI Salim	M.C.A.	Université de Batna	Examineur

# Sommaire

Liste des Figures .....	i
-------------------------	---

<b>Introduction générale .....</b>	<b>1</b>
------------------------------------	----------

<b>Chapitre I : Détermination du mouvement d'un objet 3D.....</b>	<b>4</b>
---	----------

I.1. Equations de base du mouvement 3D .....	5
I.2. Les différentes formes de représentation de la rotation.....	5
I.2.1. Paramétrisation par les angles d'Euler .....	7
I.2.2. Le vecteur de rotation r (formule de Rodrigues) .....	8
I.3. Relation entre coordonnées 3D (scène) et 2D(image) .....	9
I.3.1.Coordonnées Homogènes .....	9
I.3.2.Relation entre coordonnées homogènes et applications affines.....	10
I.3.3.Projections 3D-2D.....	11
I.3.3.1.Projection parallèle .....	11
I.3.3.2.Projection perspective.....	12
I.3.3.3.Caméra trou d'épingle (pinhole).....	12
I.4. Relation entre mouvement 3D et mouvement 2D.....	13
I.5. Calibration de la caméra .....	14
I.5.1.Introduction.....	14
I.5.2. Modélisation de la caméra .....	15
I.5.3.Modèle interne de la caméra.....	16
I.5.4.Projection perspective sur le plan image.....	17
I.5.5.Projection perspective sur le plan d'affichage .....	17
I.5.6.Modèle externe de la caméra.....	18
I.5.7.Modèle globale de la caméra .....	19
I.6. Conclusion.....	20

<b>Chapitre II : Reconstruction et localisation d'objets 3D .....</b>	<b>21</b>
II.1.Introduction .....	23
II.2.Méthodes monoculaire de reconstitution d'objets 3D .....	23
II.2.1.Shape from shading .....	24
II.2.2.Shape from texture.....	24
II.2.3.Shape from focus/defocus.....	25
II.2.4. Caméra 2.5D à temps de vol .....	26
II.3.Localisation caméra/objet .....	27
II.3.1.Correspondance de droites .....	28
II.3.2.Correspondance de points .....	30
II.4.Quelque méthodes de localisation d'objet en mouvement.....	31
II.4.1. Méthode de Dhome.....	31
II.4.1.1.Détermination de la rotation .....	31
II.4.1.2.Détermination de la translation.....	33
II.4.2.Méthode de Horaud .....	34
II.4.2.1.Aspect mathématique de la méthode .....	34
II.4.2.2.Translation .....	36
II.4.3.Méthode de Fang.....	36
II.4.3.1.Aspect mathématique de la méthode .....	37
II.4.3.2. Discussion .....	39
II.5. Conclusion .....	40
<b>Chapitre III : Manipulation pratique .....</b>	<b>41</b>
III.1.Introduction .....	42
III.2. Problème de la perspective inverse .....	43
III.3.Détermination de la rotation .....	44
III.4.Détermination de la translation.....	46
III.5.Algorithme de simulation .....	48
III.6. Résultats obtenus .....	48
III.7.Conditions de la manipulation .....	51
III.8. Conclusion .....	52
<b>Conclusion générale.....</b>	<b>53</b>
<b>Bibliographie.....</b>	<b>55</b>

## Liste des Figures

---

<b>Figure I.1 :</b>	<i>Equations du mouvement 3D</i>
<b>Figure I.2 :</b>	<i>Les 3 angles de rotation du mouvement 3D</i>
<b>Figure I.3 :</b>	<i>Rotation autour de l'axe Z</i>
<b>Figure I.4 :</b>	<i>Les trois angles d'Euler</i>
<b>Figure I.5 :</b>	<i>Formule de Rodrigues</i>
<b>Figure I.6 :</b>	<i>Projection d'un Point 3D sur un plan</i>
<b>Figure I.7 :</b>	<i>Projection perspective</i>
<b>Figure I.8 :</b>	<i>Calcul de la profondeur d (la focale)</i>
<b>Figure I.9:</b>	<i>Calibrer une caméra consiste à trouver sa fonction de transfert</i>
<b>Figure I.10:</b>	<i>Coordonnées d'un point dans les repères <math>R_c</math>, <math>R_s</math> et <math>R_i</math></i>
<b>Figure I.11:</b>	<i>Passage repère <math>R_w</math> (monde) au plan <math>R_c</math> ( caméra)</i>
<b>Figure I.12 :</b>	<i>Modèle complet de la calibration de la caméra</i>

---

<b>FigureII.1 :</b>	<i>Exemple de reconstruction par le shape from shading</i>
<b>FigureII.2 :</b>	<i>Exemple shape from texture</i>
<b>FigureII.3 :</b>	<i>Shape from défocus</i>
<b>FigureII.4 :</b>	<i>Image de distance d'un visage capturé avec une caméra temps de vol</i>
<b>FigureII.5:</b>	<i>Correspondance entre 3,4,5 et 6 points est 3 droites</i>
<b>FigureII.6 :</b>	<i>La droite appartenant à l'objet, sa projection sur le plan imageet le centre de projection F appartiennent au même plan</i>
<b>FigureII.7 :</b>	<i>Conditions de la méthode de Dhome</i>
<b>FigureII.8 :</b>	<i>Translation selon Dhome</i>
<b>FigureII.9 :</b>	<i>Conditions de la méthode de Horaud</i>
<b>FigureII.10:</b>	<i>Conditions de la méthode de Fang</i>

---

<b>FigureIII.1:</b>	<i>Problème de la perspective inverse</i>
<b>FigureIII.2:</b>	<i>Problème de la rotation</i>
<b>FigureIII.3:</b>	<i>Problème de la translation</i>
<b>FigureIII.4 :</b>	<i>Erreur sur l'estimation de l'angle de rotation suivant l'axe X</i>
<b>FigureIII.5 :</b>	<i>Erreur sur l'estimation de l'angle de rotation suivant l'axe Y</i>
<b>FigureIII.6 :</b>	<i>Erreur sur l'estimation de l'angle de rotation suivant l'axe Z</i>
<b>FigureIII.7:</b>	<i>La translation est parfaitement estimée</i>

## **Introduction générale**

Dès 1950, des études sur la perception visuelle de l'environnement furent publiées. Durant les années 60 des recherches furent menées principalement aux USA, sur le traitement par ordinateur des images scientifiques provenant de sondes spatiales, de satellites et d'observations biomédicales.

La naissance de la vision par machine ou par ordinateur, d'après [BES88], était une conséquence directe de l'évolution et de la croissance des capacités de calcul des ordinateurs.

L'étude de ce domaine tourne autour de quatre axes qui sont :

- 1- Saisie, digitalisation et traitement de l'image.
- 2- Reconnaissance sémantique et interprétation de l'image plane.
- 3- Vision tridimensionnelle, modélisation de l'objet et de l'espace.
- 4- C.A.O., infographie et synthèse d'images.

Aujourd'hui la vision par ordinateur est devenue une discipline qui s'intéresse au traitement de l'information. On peut l'assimiler à un système qui a pour entrée une séquence d'images. Ce système traite, cette entrée, en lui apportant un certain nombre de connaissances qui vont permettre d'avoir à la sortie une description en termes de conception 3D et de relation entre les objets et la scène où ils évoluent.

Son objectif est de remplacer la vision humaine pour rendre les machines plus indépendantes de l'homme. Ainsi elles auront vocation à le suppléer quand sa présence n'est pas indispensable ou nuisible pour sa santé (travail dans des milieux nocifs).

Pour atteindre cet objectif et pour assurer une reconnaissance sémantique, c'est-à-dire la compréhension de la scène, deux types de stratégies sont mises en jeu. La première qui est appelée stratégie ascendante tente de construire à partir de l'information sensorielle une représentation la plus abstraite possible (par exemple ensemble de primitives géométriques 3D). La deuxième stratégie, connue sous le nom de stratégie descendante, déduit à partir de l'ensemble d'objets connus par le système une description compatible avec les primitives extraites de l'image. Il est alors possible de mettre en correspondance la représentation

extraite de l'image avec les descriptions des objets afin de décrire les données sensorielles en termes de ces objets [HOR87].

Ces stratégies et connaissances sont traduites par des formalismes mathématiques et les algorithmes qui leurs correspondent sont construits. Les performances de ces derniers doivent répondre à la qualité exigée de ces systèmes désignés ces derniers temps comme des systèmes de vision artificielle.

### ***Contexte de l'étude :***

Une des tendances récentes qui a pris un essor considérable dans le domaine de la vision par ordinateur consiste à prendre en compte une séquence d'images plutôt qu'une seule image comme dans le cas de la vision statique. Cette approche est connue sous le nom de vision dynamique [XIA87].

En effet, l'estimation des paramètres du mouvement 3D d'un corps rigide en utilisant une séquence s'avère très intéressante, dans la mesure, où elle nous permet d'obtenir des équations du mouvement qui peuvent être résolues par des algorithmes, qui sont plus faciles que ceux utilisés dans le cas de la vision statique.

Ce travail portera sur la détermination des paramètres du mouvement 3D d'un objet, supposé rigide, à partir d'une séquence d'images prise par une caméra fixe.

Dans le premier chapitre on présentera les équations de base du mouvement 3D, le modèle de la caméra retenu pour notre étude et les différentes transformations à déterminer pour établir le modèle global de la caméra.

Le second chapitre sera un résumé des différentes méthodes vues au cours de notre recherche bibliographique, notamment celles qui se basent sur la vision monoculaire telles que les méthodes se basant sur le changement d'éclairage "shape from shading", celles utilisant le changement de la texture, "shape from texture", ou encore celles utilisant le calcul de la profondeur de la scène "depth from defocus" ou de la forme "shape from Defocus" et les méthodes utilisant les caméras 2.5D à temps de vol. Les méthodes de Dhome, Horaud et Fang seront présentées. Les hypothèses faites par chaque méthode, les équations mathématiques ainsi que les algorithmes utilisés seront expliqués. Enfin les limites, pour chaque méthode, seront également mises en discussion.

Le dernier chapitre décrira la méthode qui a été développée en simulation, les courbes de résultats obtenues ainsi que la manipulation utilisée pour la calibration de la caméra.

Une conclusion générale résumera la contribution de ce mémoire.

## Chapitre I

# Détermination du mouvement d'un objet 3D

- I.1. Equations de base du mouvement 3D
- I.2. Les différentes formes de représentation de la rotation
  - I.2.1. Paramétrisation par les angles d'Euler
  - I.2.2. Le vecteur de rotation  $r$  ( formule de Rodrigues)
- I.3. Relation entre coordonnées 3D(scène) et 2D(image)
  - I.3.1. Coordonnées homogènes
  - I.3.2. Relation entre coordonnées homogènes et application affines
  - I.3.3. Projections 3D-2D
    - I.3.3.1. Projection parallèle
    - I.3.3.2. Projection perspective
    - I.3.3.3. Caméra trou d'épingle (Pinhole)
- I.4. Relation entre mouvement 3D et Mouvement 2D
- I.5. Calibration de la caméra
  - I.5.1. Introduction
  - I.5.2. Modélisation de la caméra
  - I.5.3. Modèle interne de la caméra
  - I.5.4. Projection perspective sur le plan image
  - I.5.5. Projection perspective sur le plan d'affichage
  - I.5.6. Modèle externe de la caméra
  - I.5.7. Modèle global de la caméra
- I.6. Conclusion

## I.1. Equation de base du mouvement 3D

Tout mouvement d'un corps rigide peut se décomposer d'une manière unique comme le produit d'une rotation  $\mathbf{R}$  autour d'un axe passant par l'origine suivi d'une translation  $\mathbf{T}$ [AYA89].

On obtient ainsi l'image  $M'$  d'un point  $M$  quelconque, du corps, après mouvement par la formule suivante (Figure I.1)

$$\overrightarrow{OM'} = R \cdot \overrightarrow{OM}$$

où  $O$  et l'origine du repère

$R$  est une matrice  $3 \times 3$

$T$  vecteur de translation  $3 \times 1$ .

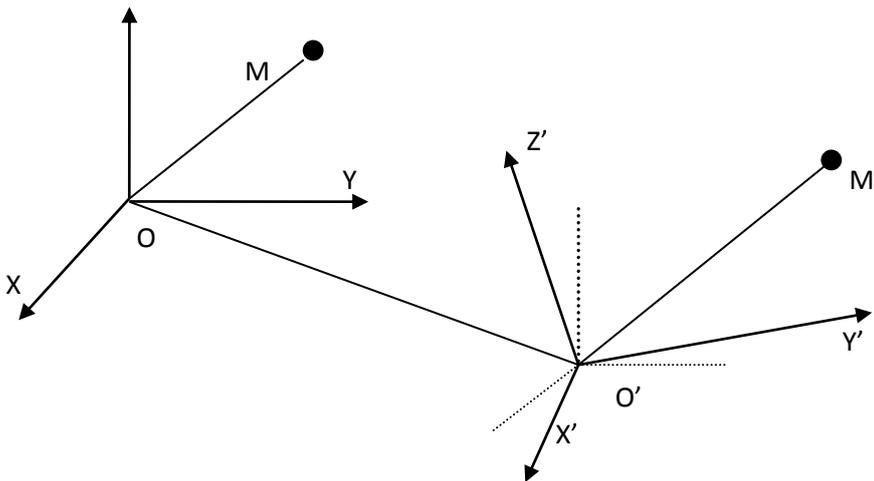


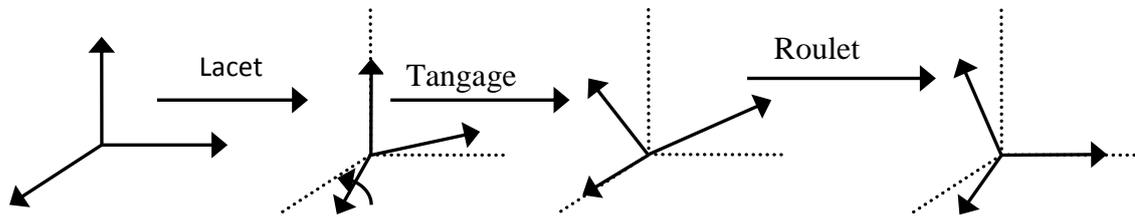
Figure I.1 : Equations du mouvement 3D

Ce qui revient à faire intervenir 6 paramètres, qui sont les trois coordonnées décrivant la position du centre de masse de l'objet (ou d'un point quelconque lui appartenant) et les trois angles de rotation.

## I.2. Les différentes formes de représentation de la rotation

Une manière courante de représenter une rotation  $R$  consiste à la décomposer en trois rotations élémentaires autour des axes  $X, Y, Z$  d'un repère fixe prédéfini. Si on note

par  $(\varphi, \theta, \Psi)$  les angles de rotation autour des axes X, Y et Z respectifs, le mouvement global de rotation sera une composition de ces trois rotations. (FigureI.2)



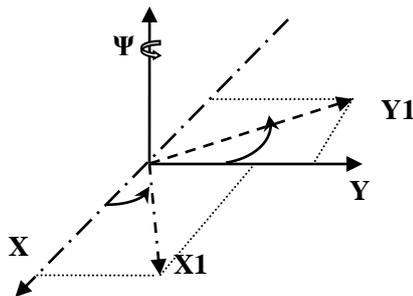
FigureI.2 :Les 3 angles de rotation du mouvement 3D

Pour la première rotation (FigureI.3) autour de l'axe Z, on peut écrire les coordonnées X', Y' et Z' coordonnées du point P', image du point P(X, Y, Z), après cette rotation suivant l'expression suivante :

$$\begin{cases} X' = X_1 \cdot \cos\Psi - Y_1 \cdot \sin\Psi \\ Y' = X_1 \cdot \sin\Psi + Y_1 \cdot \cos\Psi \\ Z = Z_1 \end{cases}$$

ou sous forme de matrice

$$\begin{bmatrix} X' \\ Y' \\ Z' \end{bmatrix} = \begin{bmatrix} \cos\Psi & -\sin\Psi & 1 \\ \sin\Psi & \cos\Psi & 0 \\ 0 & 0 & 0 \end{bmatrix} \cdot \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} \quad (I.1)$$



FigureI.3 : Rotation autour de l'axe Z

Et pour chaque rotation autour de l'axe X ou Y, on peut décrire mathématiquement cette dernière de la même façon que pour la rotation autour de Z.



### I.2.1. Paramétrisation par les angles d'Euler

Les angles d'Euler nous permettent la représentation de l'orientation d'un solide par rapport à un repère fixe. Le vecteur de rotation instantané du solide, indispensable à l'étude de sa cinématique est construit en utilisant ces trois angles. L'objectif est de passer du référentiel fixe OXYZ (le repère de navigation) au référentiel lié au solide O'X'Y'Z' (le repère mobile) par trois rotations successives :

- La précession : première rotation autour de l'axe OZ du repère fixe.
- La nutation : seconde rotation autour de l'axe Y nouvellement créé.
- La rotation propre, dernière rotation autour de l'axe X, créée suite aux deux premières rotations.

Pour chaque rotation dans le sens direct, les matrices équivalentes sont définies :

- Précession  $\Psi$  autour de l'axe Z: 
$$\begin{bmatrix} \cos \Psi & -\sin \Psi & 0 \\ \sin \Psi & \cos \Psi & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

- La nutation  $\Theta$  autour de l'axe Y: 
$$\begin{bmatrix} \cos \Theta & 0 & \sin \Theta \\ 0 & 1 & 0 \\ -\sin \Theta & 0 & \cos \Theta \end{bmatrix}$$

- Rotation propre  $\alpha$  autour de X: 
$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos \alpha & \sin \alpha \\ 0 & -\sin \alpha & \cos \alpha \end{bmatrix}$$

Le mouvement de rotation d'un solide en 3 dimensions est donc, une combinaison de 3 rotations individuelles. (Figure I.4).

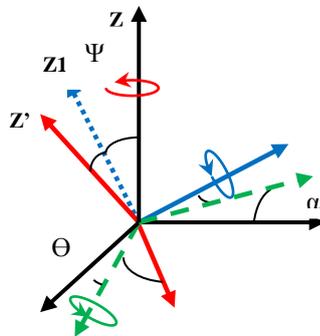


Figure 1.4 : Les trois angles d'Euler

Dans ce cas la matrice de rotation  $R(\alpha, \theta, \Psi)$  s'écrit alors :

$$\begin{bmatrix} \cos(\Psi) * \cos(\theta) & \sin(\Psi) * \cos(\theta) & \sin(\theta) \\ \cos(\Psi) * \sin(\theta) * \sin(\alpha) - \sin(\theta) \cos(\alpha) & \sin(\Psi) * \sin(\theta) * \sin(\alpha) + \cos(\Psi) \cos(\alpha) & \cos(\Psi) * \sin(\alpha) \\ \cos(\Psi) * \sin(\theta) * \cos(\alpha) - \sin(\Psi) \sin(\alpha) & \cos(\Psi) * \sin(\theta) * \cos(\alpha) - \cos(\Psi) \sin(\alpha) & \cos(\theta) * \cos(\Psi) \end{bmatrix}$$

### I.2.2. Vecteur de rotation $\mathbf{r}$ (Formule de Rodrigues)

Une rotation selon Euler est le résultat de trois rotations autour des trois axes X, Y et Z. Olinde Rodrigues lui propose un théorème fondamental où il arrive à démontrer que tout déplacement d'un corps rigide peut être effectué par une rotation autour d'un axe combinée à une translation le long de ce même axe.

Le problème revient alors à une représentation des déplacements par un vissage (la rotation) plus une translation suivant l'axe de rotation.

On remarque, par ailleurs, que :

- Toute rotation vectorielle  $\mathbf{R}$  a un axe invariant, le vecteur normé, noté  $\mathbf{n}$ .
- Les vecteurs colinéaires à  $\mathbf{n}$  sont invariants par  $\mathbf{R}$ .
- Les vecteurs appartenant au plan orthogonal du vecteur  $\mathbf{n}$  subissent une rotation  $\varnothing$  autour de l'axe défini par la direction de  $\mathbf{n}$ .

Pour un vecteur  $V$  qui subit une rotation  $R$ , il se transforme en  $V'$  et peut s'écrire en notation matricielle:

$$V' = R.V \tag{I.2}$$

En notation vectorielle ce vecteur  $V$  qui subit une rotation d'angle  $\varnothing$  d'axe  $\mathbf{n}$  ou  $(\mathbf{n} * \mathbf{n} = 1)$  est noté  $V'$  et il est défini par :

$$V' = V + \sin\varnothing \mathbf{n} \wedge V + (1 - \cos\varnothing) \mathbf{n} \wedge (\mathbf{n} \wedge V) \tag{I.3}$$

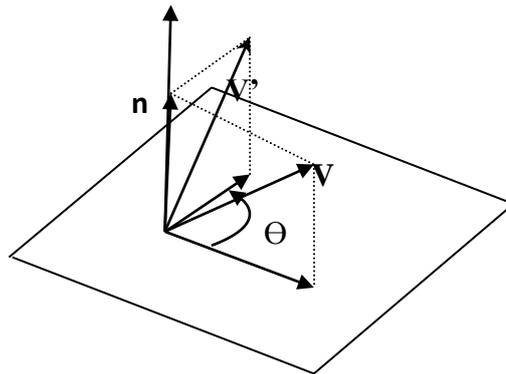


Figure1.5 : Formule de Rodrigues

On peut remarquer que le produit vectoriel de  $\mathbf{n} \wedge \mathbf{v}$  est traduit par la matrice suivante :

$$\mathbf{n} \wedge \mathbf{v} = \begin{bmatrix} -n_z \cdot v_y + n_y \cdot v_z \\ n_z \cdot v_x - n_x \cdot v_z \\ -n_y \cdot v_x + n_x \cdot v_y \end{bmatrix} = \begin{bmatrix} 0 & -n_z & n_y \\ n_z & 0 & n_x \\ -n_y & n_x & 0 \end{bmatrix} \cdot \begin{bmatrix} v_x \\ v_y \\ v_z \end{bmatrix}$$

$X(\mathbf{n})$  est la matrice antisymétrique associée au vecteur  $\mathbf{n}$  avec :

$$X(\mathbf{n}) = \begin{bmatrix} 0 & -n_z & n_y \\ n_z & 0 & n_x \\ -n_y & n_x & 0 \end{bmatrix}$$

D'où la forme matricielle de la formule de Rodrigues :

$$V' = (I + \sin\phi \cdot X(\mathbf{n}) + (1 + \cos\phi)X(\mathbf{n})^2) \cdot V$$

Donc la matrice de rotation s'écrit comme suit:

$$R(\mathbf{n}, \phi) = I + \sin\phi \cdot X(\mathbf{n}) + (1 + \cos\phi)X(\mathbf{n})^2 \quad (\text{I.4})$$

### I.3.Relation entre coordonnées 3D (scène)et 2D (image)

#### I.3.1.Coordonnées Homogènes

On définit un système de coordonnées différent pour représenter les points dans l'espace 3D. Pour le faire une quatrième composante non nulle, notée  $W$ , est rajoutée aux trois autres composantes  $(x,y,z)$ .

Un point  $M$  de l'espace 3D sera représenté par un quadruplet  $M(x,y,z,W)$  qui est appelé coordonnées Homogène du point  $M$ . Ces coordonnées ne sont pas une représentation unique, mais un quadruplet  $(x,y,z,W)$  et  $(x',y',z',W')$  représentent le même point si ces quadruplets sont multiples l'un de l'autre. Ce point  $M$  a pour coordonnées cartésiennes,  $(x,y,z)$  en coordonnées homogènes, il peut être représenté par  $(x,y,z,1)$  avec  $w=1$ .

Le passage des coordonnées d'un point  $M$  représenté par  $(x,y,z,W)$  en coordonnées homogènes, au coordonnées cartésiennes s'obtient par  $M=(x/W,y/W,z/W)$ .

L'intérêt de cette représentation par des coordonnées homogènes est de pouvoir représenter toute application affine par une seule matrice, comprenant la partie linéaire et la translation.

### I.3.2.Relation entre coordonnées homogènes et applications affines

Pour effectuer une translation  $T$  défini par  $(x_v, y_v, z_v,1)$  en coordonnées homogènes, il suffit de faire une multiplication matricielle :

$$T * \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & x_v \\ 0 & 1 & 0 & y_v \\ 0 & 0 & 1 & z_v \end{bmatrix} * \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} = \begin{bmatrix} x + x_v \\ y + y_v \\ z + z_v \\ 1 \end{bmatrix}$$

Pour appliquer une application linéaire  $M$ , dont la matrice en coordonnées cartésienne est:

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}$$

Il suffit de faire en coordonnées homogènes la multiplication matricielle par la matrice.

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} & 0 \\ a_{21} & a_{22} & a_{23} & 0 \\ a_{31} & a_{32} & a_{33} & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

Ainsi n'importe quelle application affine, composée d'une application linéaire et d'une translation, peut être obtenue en coordonnées homogènes par la multiplication par une matrice de ce type:

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} & x_v \\ a_{21} & a_{22} & a_{23} & y_v \\ a_{31} & a_{32} & a_{33} & z_v \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

La matrice en coordonnées homogènes  $M_{fg}$  composée des deux applications affines  $f$  et  $g$  est obtenu en effectuant le produit des matrices associées aux applications affines en coordonnées homogènes  $M_f M_g$ .

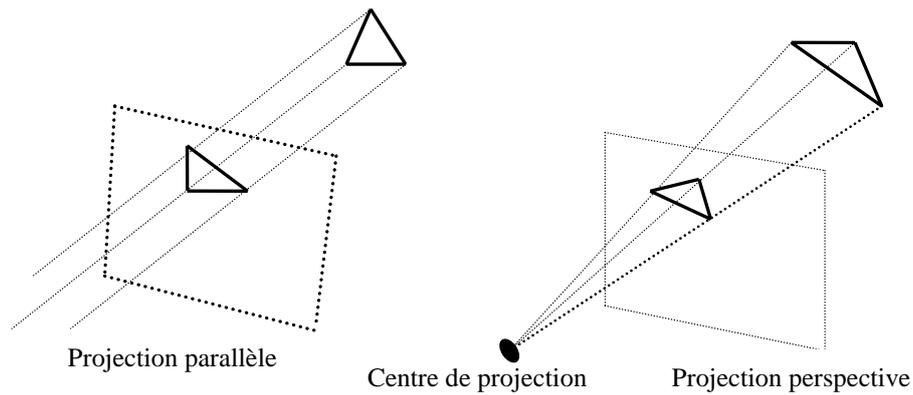
### I.3.3. Projections 3D-2D

Pour avoir une information sur le mouvement d'un objet dans la scène 3D, à partir des observations du mouvement 2D réalisées dans le plan image, il faut préalablement connaître la transformation opérée par la caméra, lors de l'acquisition de l'image.

Deux méthodes sont utilisées pour projeter des données 3D sur un plan dans le but de visualiser ces données. La projection perspective et la projection parallèle.

#### I.3.3.1. Projection parallèle

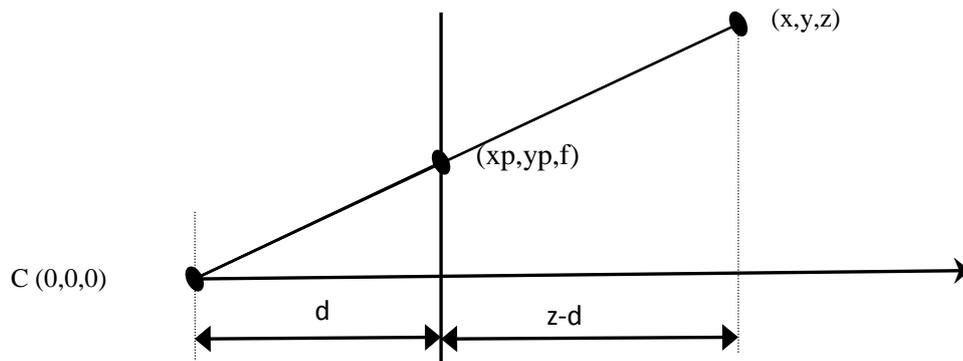
La projection parallèle est la projection sur un plan parallèlement à une direction choisie. (Voir Figure I.6). Pour réaliser cette projection on se place dans un repère  $(O, i, j, k)$  tel que le plan de projection soit le plan  $Z=0$  et la direction de la projection soit parallèle au vecteur  $k$ . Pour obtenir le projeté d'un point  $M$  de coordonnées  $(x, y, z)$ , il suffit de négliger la coordonnée  $z$  : Le projeté de  $M$  noté  $M_p$  aura pour coordonnées :  $x_p=x$ ,  $y_p=y$ ,  $z_p=0$ . La formulation de cette projection est donc particulièrement simple mais elle est très peu réaliste.



Figurel.6 : Projection d'un Point 3D sur un plan

### I.3.3.2.Projection perspective

La projection perspective, qui est souvent utilisée comme modèle de la caméra pinhole,est une projection très particulière, on suppose que le plan sur lequel on projette comme perpendiculaire au troisième axe de coordonnées  $O_z$ . Ceci correspondra au point de vue très particulier d'un observateur placé sur l'origine  $O$  et regardant dans la direction de l'axe des  $Z$ . Par des changements de repères appropriés, on ramène le cas d'un point de vue quelconque de l'observateur au cas particulier montré par la Figurel.7



Figurel.7 : Projection perspective

La transformation permettant de relier un point  $M$  de coordonnées  $(x,y,z)$  à son projeté  $M_p$  de coordonnées  $(x_p,y_p,z_p)$  est la suivante :

$$\frac{x}{x_p} = \frac{z}{d} ; \quad \frac{y}{y_p} = \frac{z}{d}$$

$$x_p = \frac{x.d}{z} \quad ; \quad y_p = \frac{y.d}{z} \quad ; \quad z_p = d \quad (I.5)$$

### I.3.3.3. Caméra trou d'épingle (pinhole)

Le principe d'une caméra "Pinhole" (trou d'épingle) est celui d'un observateur regardant une scène à travers un trou d'épingle d'où son appellation. Ce dernier est le centre de projection. La caméra est placée sur le point  $O_c=(0,0,0)$ , la direction de visée étant l'axe des  $Z_c$ . Suivant l'équation I.5 nous avons les coordonnées du projeté d'un point en fonction du paramètre  $d$  donnant l'équation  $z_p = d$  du plan de projection.

Sauf que dans le cas, d'une application graphique 3D, le paramètre  $d$  n'est pas donné directement, mais doit être déterminé à partir de l'angle d'ouverture  $\theta_c$  de la caméra et la largeur  $dimx$  de l'image.

L'angle d'ouverture  $\theta_c$  est l'angle sous lequel l'observateur placé en  $O_c(0,0,0)$  doit voir le rectangle de largeur  $dimx$  constitué par l'image, ce rectangle étant placé dans le plan " $z-d$ " comme sur la Figure I.7 centré au point M (Figure I.8)

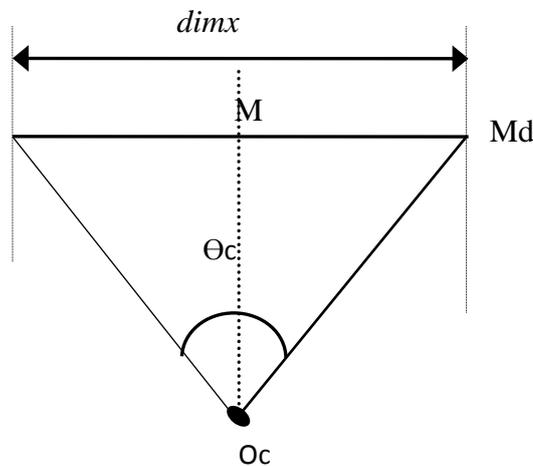


Figure I.8 : Calcul de la profondeur  $d$  (la focale)

Soit  $M_d(dimx/2,0,d)$  le point de l'image situé le plus à droite par rapport à l'observateur.

Le triangle OcMMd est droit en M on peut écrire alors que :

$$d = \frac{MMd}{tg(\frac{\theta_c}{2})} = \frac{dimx}{2tg(\frac{\theta_c}{2})} \quad (I.6)$$

## I.4.Relation entre mouvement 3D et mouvement 2D

La modélisation du mouvement utilise les équations générales de la cinématique. Le mouvement de l'objet est décomposé en un mouvement de translation défini par le vecteur ( $T$ ) et celui d'une rotation définie par ( $R$ ).

On suppose entre deux instants successifs  $t_k$  et  $t_{k+1}$  un objet a subi un mouvement composé de Rotation  $R$  et de translation  $T$ , on peut écrire alors l'équation suivante entre les coordonnées D'un point  $P(X_w, Y_w, Z_w)$  à l'instant  $t_k$  et  $t_{k+1}$  de la manière suivante :

$$P_{k+1} = R.P_k + T \quad (I.7)$$

où  $P_k$  : coordonnées du point  $P$  à l'instant  $t=t_k$

$P_{k+1}$ :coordonnées du point  $P$  à l'instant  $t=t_{k+1}$

Après projection des coordonnées 3D, cette équation mettra en évidence d'une part le mouvement projeté et d'autre part le mouvement réel et la structure de l'objet.

De ce fait on peut dire que l'estimation du mouvement des objets et celle de leurs structures sont étroitement liées, ceci explique l'usage fréquent du terme "structure from motion" pour désigner les algorithmes d'analyse du mouvement 3D à partir d'une séquence monoculaire [DUG92].

Pour répondre à ce problème il faut résoudre des équations générales et difficiles à manipuler ceci amène souvent les chercheurs à considérer des hypothèses sur le mouvement pour simplifier ces équations et permettre leur résolution.

## I.5.Calibration de la caméra

### I.5.1.Introduction

La transformation géométrique permettant d'établir la relation mathématique entre les coordonnées des points 3D de la scène observée, et les coordonnées 2D de leur projection

dans l'image désigne le calibrage de la caméra. Etablir cette relation mathématique constitue le point initial pour plusieurs applications de la vision artificielle, comme par exemple la reconnaissance d'objets, le contrôle dimensionnel de pièces, la reconstruction de l'environnement pour la navigation d'un robot mobile, etc.



Figurel.9 : Calibrer une caméra consiste à trouver sa fonction de transfert

Cette opération de calibrage de caméra est particulièrement importante lorsque l'on doit obtenir, à partir des images acquises, des informations métriques en vue d'application de mesures dimensionnelles ou de reconstitution de scène. Calibrer une caméra, revient à choisir un modèle de caméra à priori et déterminer ensuite les paramètres de ce modèle.

Le modèle sténopé (*pinhole* en anglais) modélise une caméra par une projection perspective. Ce modèle transforme un point 3D de l'espace  $M$  en un point-image  $m$ . Cette transformation est le résultat de trois transformations élémentaires successives.

- La première transformation entre le repère du monde et celui de la caméra.
- La deuxième entre le repère caméra et le repère image (plan rétinien).
- La troisième entre le repère capteur et le repère image.

### 1.5.2. Modélisation de la caméra

Le modèle géométrique sténopé, dans notre cas, est constitué d'un plan, appelé plan rétinien, dans lequel l'image se forme à l'aide d'une projection perspective. Un point  $O_c$  est placé à une distance fixe du plan rétinien. Ce point est appelé le centre optique ou le point focal. Il est choisi de sorte que les rayons de lumière reflétés par les objets passent par lui, formant une image perspective de la scène dans le plan rétinien. A ce moment, deux considérations doivent être prises en compte. La première est que le point focal peut être placé devant ou derrière le plan rétinien. Dans le cas où le centre optique est placé devant le

plan rétinien, l'image obtenue est une projection inverse de la scène. Dans l'autre cas, l'image obtenue est une projection de la scène.

Le modèle de la caméra retenu pour cette étude est celui de la projection centrale dans lequel les rayons visuels passent par le centre de la lentille et sont rectilignes. (Figure I.10)

- La distance focale  $f$ , correspond à la distance (en mm) qui sépare le plan rétinien (plan image) du point focal  $O_c$ .
- Les paramètres de conversion  $e_u$  et  $e_v$ , représentent l'ajustement horizontal et vertical permettant le passage du repère du plan rétinien (exprimé en mm) au repère image (exprimé en pixel).
- La position  $(u_0, v_0)$  du point principal (le centre de l'image) définie comme étant le projeté du centre optique  $O_c$  sur le plan image, exprimée en pixel.

Il existe deux ensembles de paramètres. Le premier ensemble des paramètres intrinsèques modélise la géométrie interne et les caractéristiques optiques de la caméra. Le second ensemble contient les paramètres extrinsèques qui lient le repère de la caméra au repère associé à la scène où se trouvent les objets de référence utiles pour la calibration.

### I.5.3. Modèle interne de la caméra

Pour exprimer le modèle interne de la caméra, il faut établir les équations qui nous permettent de passer du repère lié à la caméra  $R_c(X_c, Y_c, Z_c)$  au repère lié au système d'affichage de l'image  $R(u, v)$ . Pour faire ce passage nous devons introduire les paramètres suivants:

- $e_u$ : le facteur d'échelle horizontal (pixels/mm),
- $e_v$ : le facteur d'échelle vertical (pixels/mm).
- $u_0, v_0$  et  $w_0$  sont les coordonnées de  $O_c$  dans le repère image (mesurées en pixel).

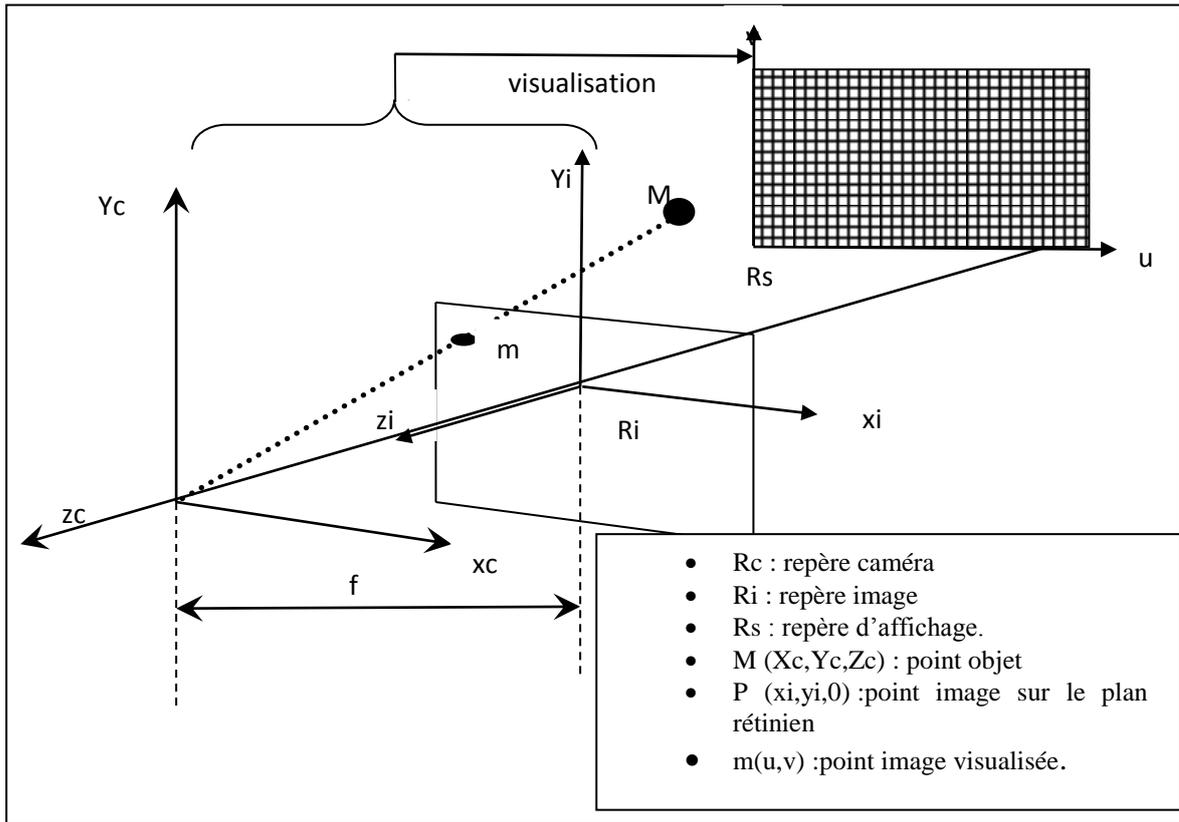


Figure.10 : Coordonnées d'un point dans les repères  $R_c$ ,  $R_s$  et  $R_i$

### I.5.4. Projection perspective sur le plan image

Le passage du repère  $R_c$  au repère  $R_i$  (repère du plan image) est une simple projection perspective d'un point sur le plan placé à  $Z_c=f$ ,  $f$  étant la focale de la caméra (Figure I.9).

Ces deux repères sont orthonormés et ont leurs axes des  $z$  confondus.

On peut écrire :

$$\frac{X_c}{X_i} = \frac{-Z_c}{f} \quad ; \quad \frac{Y_c}{Y_i} = \frac{Z_c}{f} \quad (I.8)$$

Qui peut s'écrire en utilisant les coordonnées homogènes:

$$\begin{bmatrix} s \cdot X_i \\ s \cdot Y_i \\ s \end{bmatrix} = \begin{bmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \cdot \begin{bmatrix} X_c \\ Y_c \\ Z_c \\ 1 \end{bmatrix} \text{ avec } s=Z_c$$

### I.5.5. Projection dans le plan d'affichage

Cette transformation consiste en une homothétie suivant chaque  $X_i$  et  $Y_i$ , puis en un changement d'origine, qui se traduit par les relations suivantes :

$$\begin{cases} u = e_u \cdot X_i + g_u \\ v = e_v \cdot Y_i + g_v \end{cases}$$

où  $e_u$  et  $e_v$  : Facteur d'échelle en pixels/mm

$g_u$  et  $g_v$  : Translations suivant les axes  $u$  et  $v$  en pixels.

Et on peut l'écrire sous forme matricielle :

$$\begin{bmatrix} s \cdot u \\ s \cdot v \\ s \end{bmatrix} = \begin{bmatrix} e_u & 0 & g_u \\ 0 & e_v & g_v \\ 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} s \cdot X_i \\ s \cdot Y_i \\ s \end{bmatrix} \quad (\text{I.9})$$

La matrice de passage du repère  $R_c$  au repère  $R_s$ , s'appelle : Modèle interne de la caméra elle est notée **Mint (3\*4)** et s'écrit :

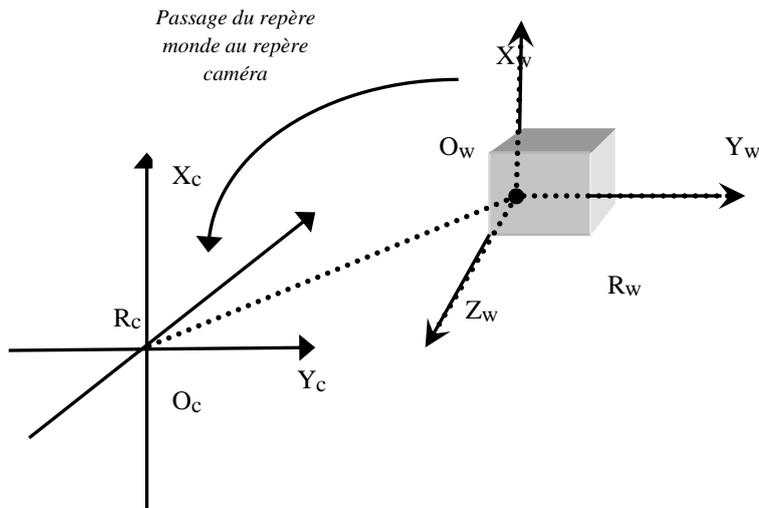
$$\begin{bmatrix} s \cdot u \\ s \cdot v \\ s \end{bmatrix} = \begin{bmatrix} e_u & 0 & g_u \\ 0 & e_v & g_v \\ 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} f & 0 & g_u & 0 \\ 0 & f & g_v & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} X_c \\ Y_c \\ Z_c \\ 1 \end{bmatrix} \quad (\text{I.10})$$

avec toujours  $s=Z_c$  ; et  $f, e_u, e_v, g_u$  et  $g_v$  sont les paramètres internes de la caméra.

### I.5.6. Modèle externe de la caméra

Lors d'une prise de vue réelle, les coordonnées des objets de la scène sont relatives à un repère absolu lié à la scène. (Figure I.11)

Avant d'appliquer le modèle interne de la caméra pour reconstituer l'image, il faut exprimer les coordonnées absolues  $R_w(X_w, Y_w, Z_w)$  dans le repère de la caméra  $R_c(X_c, Y_c, Z_c)$ .



Figurel.11 : Passage repère  $R_w$  (monde) au repère  $R_c$  (caméra)

Le passage du repère  $R_c$  au repère  $R_w$  se fait par une translation du centre  $O_c$  en  $O_w$  suivant le vecteur  $(O_x, O_y, O_z)/R_c$  et trois rotations :

- Une rotation d'angle  $\Theta$  autour de l'axe  $(OZ_w)$
- Une rotation d'angle  $\Omega$  autour de l'axe  $(OY_w)$
- Une rotation d'angle  $\Psi$  autour de l'axe  $(OX_w)$

La matrice de transformation permettant le passage, notée  $T_{cw}(4 \times 4)$ , est appelée modèle externe de la caméra et s'écrit :

$$\begin{bmatrix} X_c \\ Y_c \\ Z_c \\ 1 \end{bmatrix} = \begin{bmatrix} R_{11}R_{12}R_{13}O_x & & & \\ R_{11}R_{12}R_{13}O_y & & & \\ R_{11}R_{12}R_{13}O_z & & & \\ 0 & 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix} \quad (\text{I.11})$$

où  $(X_w, Y_w, Z_w)$  sont les coordonnées d'un point objet M dans le repère  $(R_w)$

et  $(X_c, Y_c, Z_c)$  sont les coordonnées de ce même point dans le repère  $(R_c)$

avec :

$$R_{11} = \cos(\varphi_3) \cdot \cos(\varphi_2)$$

$$R_{12} = \sin(\varphi_3) \cdot \cos(\varphi_2)$$

$$R_{13} = \sin(\varphi_2)$$

$$R_{21} = \cos(\varphi_3) \cdot \sin(\varphi_2) \cdot \sin(\varphi_1) - \sin(\varphi_2) \cos(\varphi_1)$$

$$R_{22} = \sin(\varphi_3) \cdot \sin(\varphi_2) \cdot \sin(\varphi_1) - \cos(\varphi_3) \cos(\varphi_1)$$

$$R_{23} = \cos(\varphi_3) \cdot \sin(\varphi_1)$$

$$R_{31} = \cos(\varphi_3) \cdot \sin(\varphi_2) \cdot \cos(\varphi_1) - \sin(\varphi_3) \cdot \sin(\varphi_1)$$

$$R_{32} = \cos(\varphi_3) \cdot \sin(\varphi_2) \cdot \cos(\varphi_1) - \cos(\varphi_3) \cdot \sin(\varphi_1)$$

$$R_{33} = \cos(\varphi_2) \cdot \cos(\varphi_3)$$

$(O_x, O_y, O_z)$  coordonnées de l'origine ' $O_w$ ' dans le repère  $(R_c)$

tel que :

$$O_x = (l_x \cdot R_{11} + l_y \cdot R_{12} + l_z \cdot R_{13})$$

$$O_y = (l_x \cdot R_{21} + l_y \cdot R_{22} + l_z \cdot R_{23})$$

$$O_z = (l_x \cdot R_{31} + l_y \cdot R_{32} + l_z \cdot R_{33})$$

### I.5.7. Modèle global de la caméra

La matrice qui détermine la projection d'un point défini dans le repère  $(R_w)$  sur le repère  $(R_s)$  est appelée "modèle global de la caméra". Cette matrice notée  $\mathbf{H}$ , de dimension 3x4 est déduite directement des équations I.10 et I.11. Elle est donnée par :

$$\begin{bmatrix} s \cdot u \\ s \cdot v \\ s \end{bmatrix} = \begin{bmatrix} e_u & 0 & g_u \\ 0 & e_v & g_v \\ 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} f & 0 & 0 \\ 0 & f & 0 \\ 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} R_{11} R_{12} R_{13} O_x \\ R_{21} R_{22} R_{23} O_y \\ R_{31} R_{32} R_{33} O_z \\ 0 & 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} X_0 \\ Y_0 \\ Z_0 \\ 1 \end{bmatrix}$$

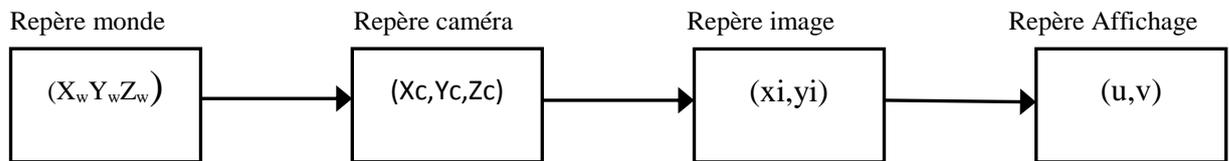
Qui après multiplication donne :

$$\begin{bmatrix} s \cdot u \\ s \cdot v \\ s \end{bmatrix} = \begin{bmatrix} H_{11} H_{12} H_{13} H_{14} \\ H_{21} H_{22} H_{23} H_{24} \\ R_{31} R_{32} R_{33} H_{34} \end{bmatrix} \cdot \begin{bmatrix} X_0 \\ Y_0 \\ Z_0 \\ 1 \end{bmatrix} = \mathbf{H} \cdot \begin{bmatrix} X_0 \\ Y_0 \\ Z_0 \\ 1 \end{bmatrix} \quad (\text{I.12})$$

$$\begin{aligned}
 \text{où} \quad H_{11} &= -R_{11} \cdot f \cdot e_u + R_{31} \cdot g_v & H_{23} &= -R_{23} \cdot f \cdot e_u + R_{33} \cdot g_v \\
 H_{12} &= -R_{12} \cdot f \cdot e_u + R_{32} \cdot g_v & H_{24} &= -O_y \cdot f \cdot e_u + O_z \cdot g_v \\
 H_{13} &= -R_{13} \cdot f \cdot e_u + R_{33} \cdot g_v & H_{31} &= -R_{31} \\
 H_{14} &= -O_x \cdot f \cdot e_u + O_z \cdot g_v & H_{32} &= -R_{32} \\
 H_{21} &= -R_{21} \cdot f \cdot e_u + R_{31} \cdot g_v & H_{33} &= -R_{33} \\
 H_{22} &= -R_{22} \cdot f \cdot e_u + R_{32} \cdot g_v & H_{34} &= O_z
 \end{aligned}$$

Cette matrice est souvent normalisée en divisant ses coefficients  $H_{ij}$  par une constante non nulle.

Le modèle complet de la caméra peut être illustré par le schéma suivant :



Figurel.12 : Modèle complet de la calibration de la caméra.

Les paramètres de la calibration ou les relations géométriques permettant le passage d'un repère à un autre sont :

- Soit calculés.
- Soit fournis par le système de prise de vue lors de l'acquisition des images.

Dans ce cas la partie concernant la calibration de la caméra est réalisée en utilisant le module caméra calibration Toolbox du logiciel Matlab.

L'ensemble des matrices de calibration ainsi que toutes les matrices permettant de passer d'un repère à un autre seront données en annexe.

## **I.6. Conclusion**

Ce chapitre a été dédié à la détermination du mouvement d'un objet 3D. Pour cela, toutes les équations de base explicitant un tel mouvement ont été explicitées. Aussi, la mise en place d'une caméra pour estimer et calculer une telle opération a nécessité de faire appel à la relation d'une image 2D et une scène 3D. Dans ce cas, les mécanismes de projection utiles pour appréhender cette démarche ont été clairement développés.

Par ailleurs, la modélisation de la caméra a fait également l'objet de notre étude. Nous avons donné tous les modèles nécessaires pour calibrer cet outil afin d'estimer la détermination de mouvement.

## Chapitre II

# Reconstruction et localisation d'objets 3D

II.1.Introduction

II.2.Méthodes monoculaires de reconstruction d'objet 3D

II.2.1.Shape from shading

II.2.2.Shape from texture

II.2.3.Shape from focus / defocus

II.2.4.Caméra 2,5D temps de vol

II.3.Localisation caméra/objet.

II.3.1.Correspondance de droites

II.3.2.Correspondance de points

II.4.Quelques méthodes de localisation d'objets en mouvement.

II.4.1.Méthode de Dhome [DHO89]

II.4.1.1.Détermination de la rotation.

II.4.1.2.Détermination de la translation.

II.4.2.Méthode de Horaud [HOR87]

II.4.2.1.Aspect mathématique de la méthode

II.4.2.2.Translation

II.4.3.Méthode de Fang [FAN83]

II.4.3.1.Aspect mathématique de la méthode

II.4.3.2.Discussion

II.5. Conclusion

## II.1.Introduction

Les résultats donnés par les chercheurs dans le domaine de la détermination de mouvement d'objet 3D à partir d'images sont très variés. Certains sont qualitatifs, c'est-à-dire décrivent la scène en utilisant des prédicats "à gauche de", "à droite de", "au-dessus de", "au-dessous de" ... ; ce qui permet de donner les positions relatives des objets afin d'en extraire les caractéristiques telles que l'occlusion, la profondeur relative des objets, les orientations des surfaces et le mouvement relatif des objets.

Mais la plupart du temps, on a plus d'intérêt à exploiter les résultats quantitatifs en estimant les paramètres du mouvement d'un objet, en calculant la profondeur d'un pixel spécifique ou en reconstruisant les formes des surface d'un objet [XIA 87].

Pour obtenir des solutions optimales, la recherche dans l'espace paramétrique d'ordre 3 ou 4 est indispensable. En réalité, l'estimation des paramètres 3D est un problème dont la solution n'est pas unique et demande un temps de calcul considérable.

Cette contrainte a poussé beaucoup d'auteurs à trouver des solutions partielles en limitant les possibilités du mouvement. Ainsi Williams [WIL 80] a obtenu des résultats numériques en limitant les possibilités du mouvement de son objet à une rotation autour de l'axe vertical.

Pour l'étude que nous allons mener, nous nous intéressons seulement aux méthodes qui déterminent les paramètres du mouvement à partir de leurs primitives.

## II.2.Méthodes monoculaires de reconstruction d'objets 3D

Les méthodes de reconstruction 3D se basant sur une seule caméra utilisent généralement certaines caractéristiques de l'image pour aboutir à l'information de la profondeur. Cette information caractéristique qui peut nous aider à estimer la profondeur peut être l'éclaircissement "*Shape from shading*", la déformation de la texture "*shape from texture*", la variation des paramètres de mise au point de l'objet "*Shape from focus/defocus*", la détection des points de fuites [GMMB00] ou aussi des méthodes par temps de vol de lumière (caméra à temps de vol).

Toutes ces méthodes s'appuient sur certaines hypothèses fortes qui permettent d'estimer la profondeur ; ces informations sont prises lors de la prise d'image.

### II.2.1. Shape from shading

Cette approche se base sur l'exploitation de la forme de l'objet éclairé, à partir des variations graduelles de l'éclairage. Dans une image au niveau de gris, on essaye de déterminer la position de la source de lumière ainsi que la surface de l'objet pour chaque pixel de l'image.

Cette méthode, même dans l'hypothèse d'un modèle d'éclairage lambertien (luminance énergétique est la même dans toutes les directions), demeure difficile à implémenter.

Si la surface de la forme est décrite par la normale, cela revient à résoudre pour chaque pixel une équation linéaire à trois inconnues. Ce problème passe par la résolution d'une équation non linéaire à deux inconnues dans le cas où la surface est décrite en termes de gradient. Ainsi pour alléger les algorithmes de résolution de cette approche on a souvent recours à définir des contraintes supplémentaires [PF05][MBB07].

Ces méthodes nécessitent des temps de calcul très élevés, ce qui représente un handicap pour leur utilisation sur des systèmes temps réel.

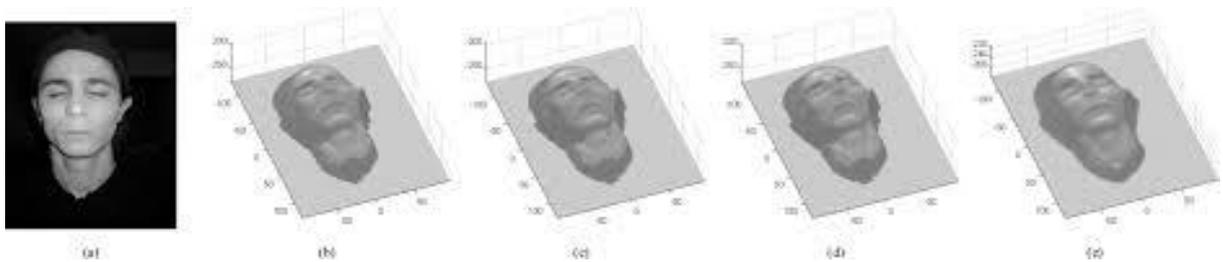


Figure II.1 : Exemple de reconstruction par la méthode "shape from shading" [INRIA]

### II.2.2. Shape from texture

Dans ce type d'approche, on estime la forme des objets en fonction des variations observées dans la texture [AS88][GAR92][LH05]. Les objets à reconstruire doivent avoir une texture soit naturelle soit artificielle.

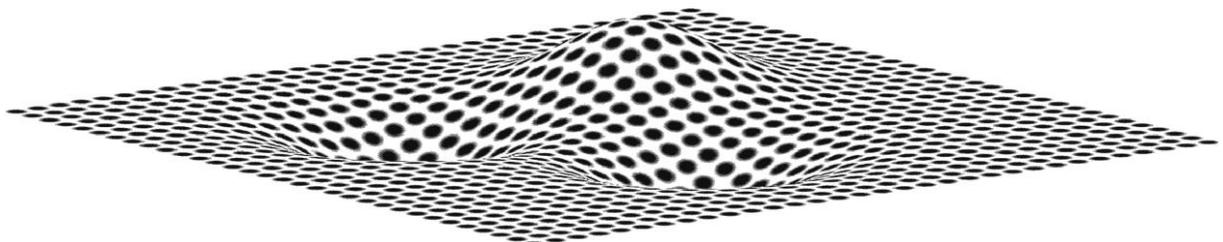


Figure II.2 : Exemple shape from texture journal of vision avril 2006

### II.2.3. Shape from focus/defocus

Pour estimer la profondeur "Depth from defocus" ou la forme "shape from defocus", il est nécessaire d'avoir plusieurs images prises depuis une caméra fixe avec des valeurs de mise au point différentes. En se mettant dans des hypothèses particulières sur la caméra, on ramène le problème à une inversion de l'équation d'intégration qui décrit le processus de création de l'image. Cette méthode s'appuie sur le fait que l'image de la scène, sur le plan image, dépend des deux paramètres de la région qui sont la radiance et la forme. La valeur de chaque pixel est le résultat de l'intégration d'une radiance inconnue par un noyau de convolution inconnu qui dépend aussi de la forme de la scène. Connaissant la valeur du pixel, le problème revient à estimer la valeur de la radiance et du noyau de convolution.

On se met dans l'hypothèse qui considère la surface de la scène localement plane et en même temps parallèle au plan de la caméra (hypothèse équifocal) [AFM98][DW90]. Plusieurs travaux ont été menés et qui ont démontré que cette approche reste très sensible au bruit et au manque de texture des objets à reconstruire.

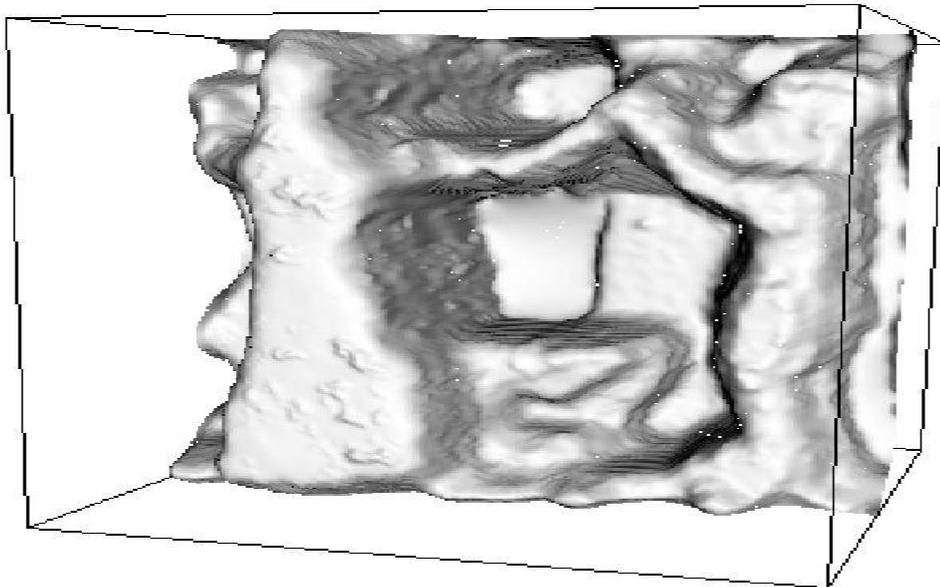


Figure II.3: "Shape from defocus" algorithm (joint work with H.Jin).

#### II.2.4. Caméra 2,5D à temps de vol

Les caméras à temps de vol ont fait leur apparition au début des années 2000. Leur principe de fonctionnement est similaire à celui des radars, la seule différence provient du fait que l'onde envoyée, par la caméra, n'est pas acoustique mais lumineuse.

Cette caméra est composée d'une source de lumière infrarouge, d'amplitude modulée, et d'un capteur qui mesure l'intensité rétrodiffusée de cette lumière infrarouge.

Deux technologies sont utilisées pour le calcul du temps de vol :

- La première mesure le temps que met le signal pour effectuer le trajet entre l'objet et la caméra, se basant sur des impulsions (pulsed time of flight TOF).
- La deuxième, et qui est la plus utilisée, se base sur des ondes de modulation continues (continuous wave CW). Elle mesure le déphasage entre le signal émis et celui réfléchi par démodulation synchrone du signal réfléchi.

Ce type de caméra donne une estimation directe de la profondeur de la scène, mais présente des limites par rapport à sa faible résolution. De plus ces caméras ne fournissent qu'une information 2.5D pour passer à une information 3D ; il est alors nécessaire d'utiliser plus d'une ; ce qui augmente considérablement le coût de ce genre de solution. En plus, il faut un algorithme adéquat comme celui proposé par [BG08].



Figure II.4 : Image de distance d'un visage capturé avec une caméra temps de vol

### II.3. Localisation caméra/objet

La détermination de la position d'un objet et de son orientation par rapport à une caméra est identique au problème de la détermination des paramètres extrinsèques d'une caméra lorsque les paramètres intrinsèques de cette dernière sont connus.

Le problème qu'on rencontre, dans ce cas, est souvent dû au fait de ne pas disposer d'assez de points de correspondance pour nous permettre d'avoir une approche linéaire.

Localiser un objet en mouvement à partir des coordonnées de points ou droites lui appartenant, dans le repère monde et leurs projections 2D décrites dans le repère caméra, revient à la détermination de son mouvement (rotation et translation).

Les solutions pour la résolution de ce problème sont regroupées en deux catégories [HOR 87]

1. Solutions analytiques : Cette méthode est utilisée dans les cas où on ne dispose pas d'assez de points de correspondances (nombre compris entre 3 et 6). La solution consiste à résoudre une équation non linéaire avec plusieurs solutions. Pour 3 droites qui peuvent être construites à partir de 3,4,5 ou 6 points (Figure II.5), des solutions sont proposées.

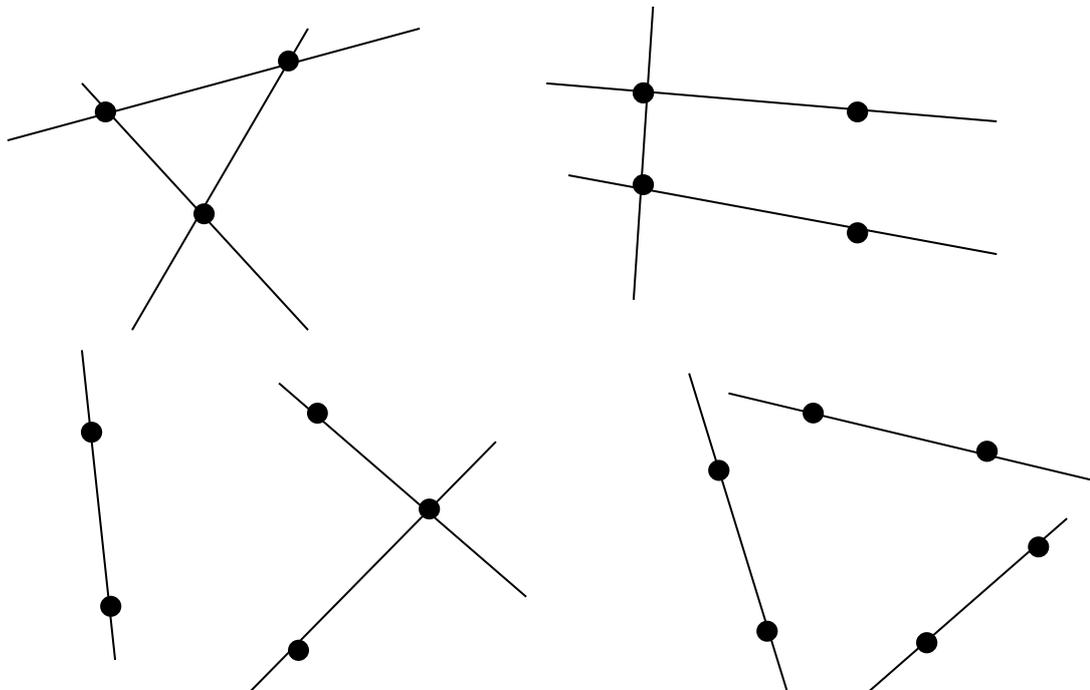


Figure II.5 : Correspondance entre 3,4,5 et 6 points et 3 droites [HOR 87]

2. Solutions Numériques : lorsque le nombre de points de correspondance est très important, les solutions analytiques ne sont plus efficaces du fait que le nombre d'équations est supérieur au nombre d'inconnues. Dans ce cas, des solutions numériques sont nécessaires.

Des méthodes basées sur des hypothèses particulières sur le mouvement de l'objet, sont développées par un certain nombre de chercheurs.

Ainsi une solution linéaire a été présentée par Ganapathy [GAN 84], la rotation est définie par une matrice 3X3. Les résultats obtenus sont très sensibles au bruit.

Une autre méthode itérative était proposée par Liu [LIU90] où la rotation est représentée par les angles d'Euler. La correspondance est faite par rapport aux droites et non plus par rapport aux points. Le mouvement est décomposé en une rotation et une translation. Une fois la rotation déterminée la translation devient un problème linéaire. Cette méthode d'après ses auteurs présente des limites dès que les angles de la rotation dépassent les 30°[LIU90].

J.-Q.Fang et T.S.Hang ont obtenu aussi des résultats numériques sur les paramètres du mouvement en supposant que les angles de rotation d'un objet étaient très faibles [FAN 83].

Dhome a mis en œuvre une méthode qui ne pose aucune condition sur les paramètres du mouvement mais plutôt sur le choix des repères de travail et des images. Cette méthode aboutit à la solution d'une équation non linéaire à huit inconnues [DHO89].

Une autre méthode qui ne conditionne pas les paramètres du mouvement est celle de Horaud, elle nécessite l'utilisation de la projection de trois droites concourantes. Pour la détermination des paramètres du mouvement, il résout une équation non linéaire en utilisant une méthode graphique [HOR 87].

Ces trois dernières méthodes feront l'objet d'une étude détaillée dans notre étude.

### II.3.1. Correspondance de droites

On considère un point  $\mathbf{m}$  de l'image ; il a pour coordonnées  $u, v$  dans le repère lié à l'image.  $X_c, Y_c$  les coordonnées de ce même point dans le repère lié à la caméra.

On a 
$$X_c = (u - u_0)/\alpha_u$$

$$Y_c = (v - v_0)/\alpha_v$$

où  $u_0, v_0, \alpha_u, \alpha_v$  sont les paramètres intrinsèques de la caméra.

$X, Y$  et  $Z$  les coordonnées d'un point 3D  $\mathbf{M}$ , de l'objet, dans le repère caméra.

$$X_c = X/Z$$

$$Y_c = Y/Z$$

On supposera que  $m$  appartient à une droite de l'image dont l'équation s'écrit :

$$aX_c + bY_c + c = 0$$

$$aX + bY + cZ = n \cdot \overrightarrow{FM} = 0 \quad (\text{II.1})$$

Avec  $n$  la normale au plan de projection comme le montre la figure suivante :

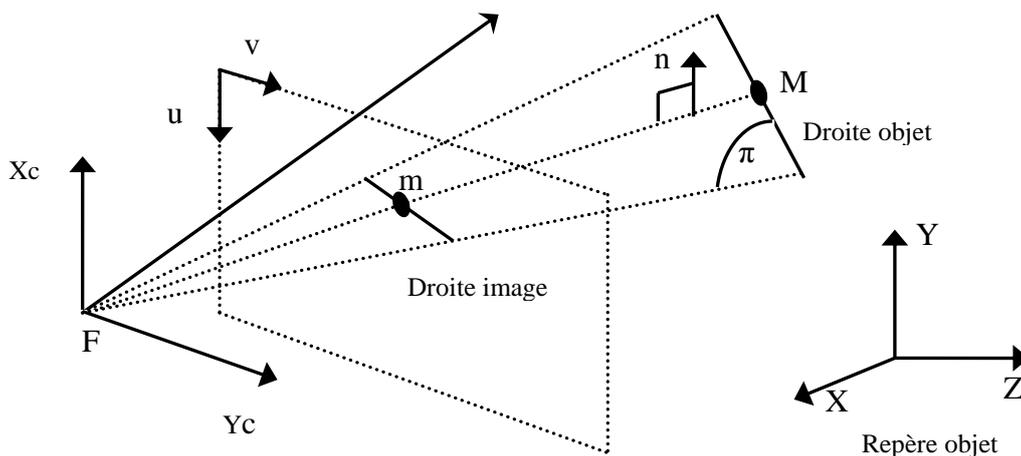


Figure 11.6 : La droite appartenant à l'objet, sa projection sur le plan image et le centre de projection  $F$  appartiennent au même plan [HOR 87]

Considérons maintenant la droite de l'objet décrite par le point  $M$ , dans le repère lié à l'objet ; cette droite peut être définie par son vecteur directeur  $v$  (unitaire) et  $p$  son vecteur position.

$$\overrightarrow{FM} = p' + \mu v'$$

avec

$$p' = R \cdot p + T$$

$$v' = R \cdot v$$

Où  $R$  et  $T$  sont la rotation et la translation que subit le repère lié à l'objet et celui de la caméra.

La contrainte de correspondance exprime le fait que la droite de l'objet appartient au plan défini par l'équation II.1 ; d'où

$$n'.(R.v) = 0$$

$$n.(R.p + T) = 0$$

Par conséquent, pour chaque correspondance droite-objet avec une droite-image on a deux contraintes. Le nombre de paramètres inconnus à déterminer est de 6 (trois pour la rotation et trois pour la translation).

Comme le montre la figure II.5, 3 droites peuvent être construites par 3,4,5 ou 6 points. Il nous faut alors au moins 3 correspondances droite/droite ou point/point pour utiliser cette approche.

### II.3.2. Correspondance de points

Comme pour le cas de la droite, si on considère un point  $\mathbf{M}_i$ , de l'objet, ses coordonnées sont  $(X_i, Y_i, Z_i)$  dans le repère lié à l'objet,  $(X_{ci}, Y_{ci})$  les coordonnées du point  $\mathbf{m}_i$ , qui est la projection de  $\mathbf{M}_i$  dans le repère caméra.

La colinéarité des points  $\mathbf{M}_i$ ,  $\mathbf{m}_i$  et  $F$  (le centre de projection) s'exprime de la façon suivante :

$$X_{ci} = \frac{r_{11}.X_i + r_{12}.Y_i + r_{13}.Z_i + T_1}{r_{31}.X_i + r_{32}.Y_i + r_{33}.Z_i + T_3} \quad (\text{II.2})$$

$$Y_{ci} = \frac{r_{21}.X_i + r_{22}.Y_i + r_{23}.Z_i + T_2}{r_{31}.X_i + r_{32}.Y_i + r_{33}.Z_i + T_3} \quad (\text{II.3})$$

où  $r_{ij}$  est un élément de la matrice de rotation  $R$  ;  $T_i$  est une composante du vecteur de translation  $T$ . Ces équations peuvent s'écrire aussi sous la forme matricielle suivante

$$[1 \quad 0 \quad -X_{ci}] \cdot R \cdot \begin{bmatrix} X_i \\ Y_i \\ Z_i \end{bmatrix} + [1 \quad 0 \quad -X_{ci}] \cdot \begin{bmatrix} T_1 \\ T_2 \\ T_3 \end{bmatrix} = 0 \quad (\text{II.4})$$

$$[1 \quad 0 \quad -Y_{ci}] \cdot R \cdot \begin{bmatrix} X_i \\ Y_i \\ Z_i \end{bmatrix} + [1 \quad 0 \quad -Y_{ci}] \cdot \begin{bmatrix} T_1 \\ T_2 \\ T_3 \end{bmatrix} = 0 \quad (\text{II.5})$$

Cette approche en principe peut fonctionner avec seulement trois points.

## II.4. Quelques méthodes de localisation d'objet en mouvement

### II.4.1. Méthode de Dhome [DHO89]

La méthode de Dhome nécessite l'utilisation d'au moins trois droites **3D**, qui doivent vérifier la condition suivante :

- Leurs projections sur l'image **2D** ne doivent pas être confondues.

Pour déterminer la position de l'objet après son mouvement, on considère le problème de la rotation seule puis on détermine celui de la translation.

#### II.4.1.1. Détermination de la rotation

La rotation sera définie par l'estimation des trois angles de la rotation suivant les trois axes. Pour le faire et pour simplifier le problème, on se place dans des repères particuliers (figure II.7)

- L'axe  $X_c$  de  $R_c$  est porté par la droite 2D  $l_i$
- Le plan  $X_c O_c Z_c$  est le plan passant par  $O_c$  et  $l_i$  (Figure II.5).

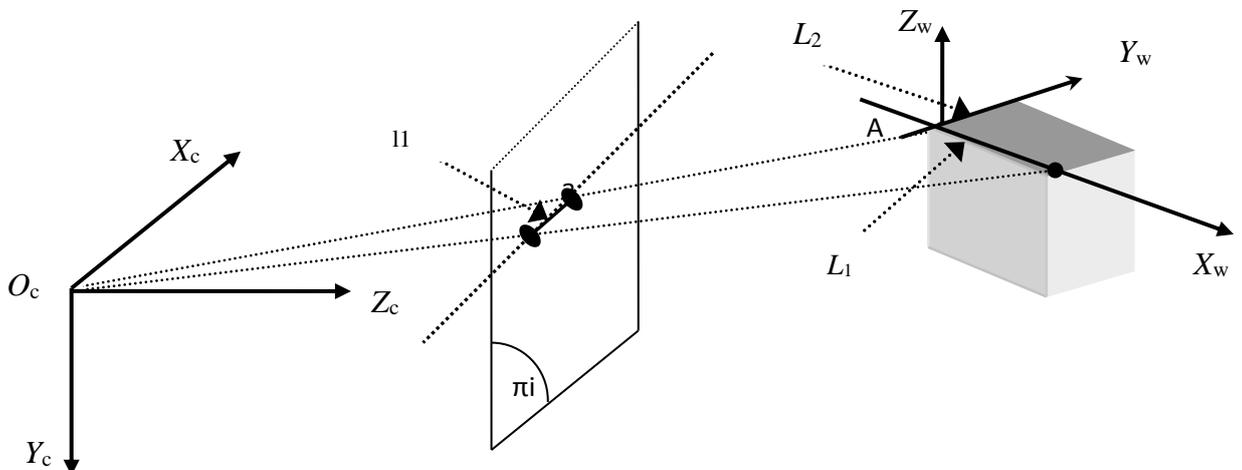


Figure II.7: Conditions de la méthode de Dhome

Le repère  $R_w$ , lié à l'objet en mouvement, sera défini par :

- porté par la droite  $\mathbf{3D} L_i$  qui correspond à  $l_i$ .
- Le plan  $X_oAY$  contenant la droite  $L_2$

De cette façon la matrice de changement de repère  $R_w$  à  $R_c$  peut être définie par deux angles uniquement :

1. L'angle  $\alpha$  qui correspond à la rotation autour de l'axe  $Z_w$  de  $R_w$  permettant d'amener l'axe  $X_w$  parallèle à  $l_i$ .
2. L'angle  $\beta$  qui correspond à la rotation autour du nouvel axe  $X_w$  faisant correspondre les axes  $Y$  et  $Z$  des repères  $R_c$  et  $R_w$ .

Le vecteur directeur de la droite  $l_i$ , noté  $ab$ , aura pour coordonnées  $(a_i, b_i, 0)$ .

Le plan d'interprétation de  $l_i$  est le plan passant par  $l_i$  et  $O_c$  ;  $i=[1..3]$ .

Le vecteur  $\mathbf{N}$  normal au plan d'interprétation de  $l_i$  est donc de la forme  $(b_i, -a_i, c_i)$  avec  $i=[1..3]$

On mesure  $a_i$  et  $b_i$  dans  $R_c$ .  $L_i$  une droite du plan d'interprétation de  $l_i$  est donc orthogonale au vecteur  $N_i$  normal à ce plan. La méthode consiste à exprimer cette propriété pour chacune des droites considérées.

Ainsi on aura à résoudre le système d'équations suivant :

$$R\alpha\beta \cdot [A_i \ B_i \ C_i]^t \cdot [b_i \ -a_i \ c_i] = 0 \quad (\text{II.6})$$

où :

1.  $(A_i, B_i, C_i)$  sont les coordonnées du vecteur directeur de la droite  $\mathbf{3D} L_i$ .
2.  $(b_i, -a_i, c_i)$  sont les coordonnées du vecteur normal au plan d'interprétation de  $l_i$ .
3.  $R\alpha\beta$  est la matrice de rotation qui permet le passage de  $R_c$  à  $R_w$ .

Ce système traduit le fait que chaque droite  $L_i$  fait partie du plan passant par  $O_c$  et  $l_i$  dont  $(b_i - a_i c_i)$  est la normale.

$R\alpha\beta^*(A_i, B_i, C_i)$  représente les coordonnées du vecteur directeur de la droite  $L_i$  exprimée dans le repère  $R_c$ . Résoudre cette équation revient à résoudre une équation à deux variables en sinus et cosinus qui s'avère très difficile ; pour cela on procède à des changements de variables pour aboutir à une équation polynomiale d'ordre huit dans le cas général. Cette

équation est d'ordre quatre lorsque les droites traitées sont concourantes ou lorsqu'elles sont coplanaires.

#### II.4.1.2. Détermination de la translation

Soit  $(\pi_i)$ , plan d'interprétation du segment  $[ab]$ , c'est-à-dire le plan défini par le segment  $[ab]$  et le centre optique de la caméra  $O_c$ .

Soit  $N_i$  la normale au plan  $(\pi_i)$  et  $P_i$  un point du segment  $[AB]$  de l'espace.

On sait que pour tout  $i=1..3$ .

$$O_c P_i \cdot N_i = 0 \quad (\text{II.7})$$

(du fait que  $O_c P_i$  appartient au plan  $(\pi_i)$ ).

Or, si on considère  $P_i'$ , le point correspondant au point  $P_i$  dans le repère  $R_w$ , on aura :

$$O_c P_i \cdot N_i = (O_c P_i' + P_i' P_i) \cdot N_i \quad (\text{II.8})$$

Avec  $P_i' P_i = T$  vecteur de translation ; on peut écrire alors  $(O_c P_i' + T) \cdot N_i = O_c P_i \cdot N_i = 0$

$$T \cdot N_i = -O_c P_i' \cdot N_i \quad (\text{II.9})$$

On connaît les coordonnées de  $O_c P_i'$  et  $N_i$ . Il est donc possible d'obtenir à partir de trois segments trois équations à trois inconnues qui sont les coordonnées du vecteur de translation  $T$ . (Figure II.8).

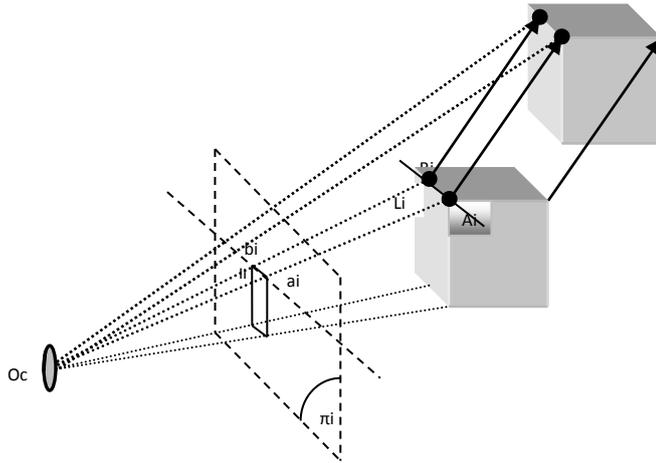


Figure II.8 : Translation selon Dhome

**Discussion :** Cette méthode présente l'avantage d'avoir un principe simple. En plus, elle ne conditionne ni la nature du mouvement de l'objet, ni son amplitude.

Néanmoins, on lui fait le reproche d'aboutir à une équation d'un degré élevé qui nous donne généralement une solution qui n'est pas unique. Pour résoudre ce problème, il faut augmenter le nombre de segments utilisés ce qui augmente considérablement la complexité de l'algorithme.

#### II.4.2. Méthode de Horaud [HOR87]

Cette deuxième méthode est basée sur l'étude de l'image de trièdres formés par les coins de l'objet. La condition d'utilisation de cette approche est donc le fait que l'image contient au moins la projection de trois droites concourantes dans l'espace. En effet, si un point d'une image n'est pas un indice robuste, l'intersection de trois droites l'est.

A partir de l'appariement d'un groupe de trois droites de l'image à un trièdre de l'objet, il est possible d'obtenir mathématiquement toutes les rotations possibles dans l'espace qui produisent la projection 2D observée.

##### II.4.2.1. Aspect mathématique de la méthode

Chaque trièdre de l'objet est modélisé par trois angles qui le définissent dans l'espace. Ces angles sont les suivants :

- $\varphi$  : L'angle formé par les deux premiers vecteurs du trièdre.
- $\theta, \Psi$  : les deux coordonnées angulaires du troisième vecteur du trièdre dans la sphère de Gauss dont l'axe principal est le premier vecteur, et dont le plan xOy est le plan défini par les deux premiers vecteurs  $L_1, L_1 \wedge L_2$ .
- On se place donc dans le repère dont les vecteurs sont  $L_1, L_1 \wedge L_2$  et  $L_1 \wedge L_2 \wedge L_3$  comme le montre la Figure II.9.

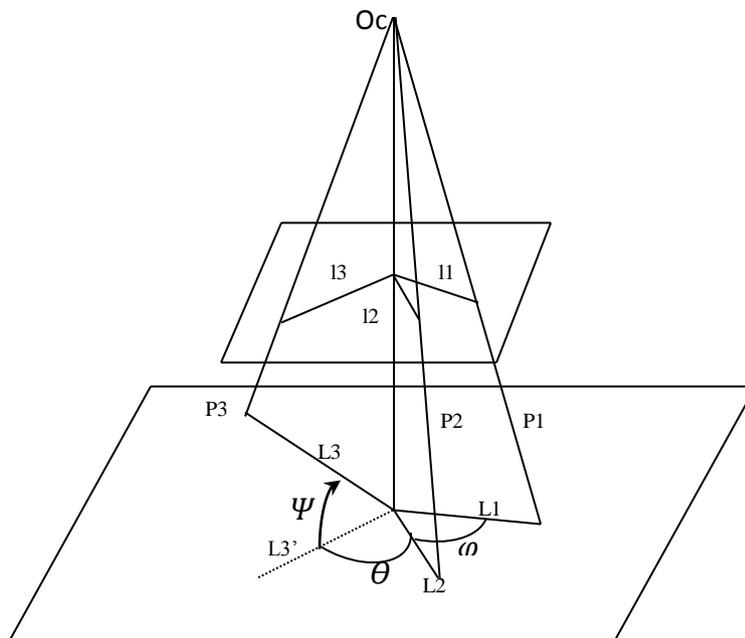


Figure II.9 : Conditions de la méthode de Horaud

La méthode se base sur la détermination du vecteur  $S$  normal au plan formé par les deux premiers vecteurs  $L_1$  et  $L_2$  qui définit à lui seul l'orientation de l'objet en mouvement.

Soit  $P_i$ , la normale au plan d'interprétation de la droite  $L_i$ .

On a les équations suivantes :

$$L_1 = S \wedge P_1 \quad \text{et} \quad L_2 = S \wedge P_2$$

On sait aussi que  $L_1 \cdot L_2 = |L_1| \cdot |L_2| \cdot \cos \varphi$

Donc on peut écrire :

$$(S \wedge P_1) \cdot (S \wedge P_2) = |S \wedge P_1| \cdot |S \wedge P_2| \cdot \cos\varphi$$

Si on exprime le vecteur  $L_3$  dans la base formée par les vecteurs  $S, L_1$  et  $(S \wedge P_1)$  on aura

$$L_3 = S \cdot \sin\Psi + L_1 \cdot \cos\theta \cdot \cos\Psi + (S \wedge P_1) + L_1 \cdot \sin\theta \cdot \cos\Psi$$

Sachant que  $L_3 \cdot P_3 = 0$  on obtient donc l'équation suivante :

$$(S \cdot P_3) \cdot |S \wedge P_3| \cdot \sin\Psi + S \cdot (P_1 \wedge P_3) \cdot \cos\theta \cdot \cos\Psi + (S \cdot P_1) \cdot (S \cdot P_3) \cdot \cos\theta \cdot \cos\Psi = (P_1 \cdot P_3) \cdot \cos\theta \cdot \cos\Psi$$

dont l'inconnue est  $S$  et où  $\Psi$  et les  $P_i$  sont connus.

Il est possible de résoudre cette équation par une méthode graphique en définissant le vecteur  $S(\alpha, \beta)$  dans la sphère de Gauss et en faisant varier les angles  $\alpha$  et  $\beta$  le caractérisant.

#### II.4.2.2. Translation

Pour déterminer la translation, on utilise une approche voisine de celle de Dhome. Horaud exploite le fait d'avoir un segment 3D qui passe par le plan d'interprétation ou segment 2D (plan passant par le centre optique  $O_c$  et le segment  $[ab]$ ). De la même manière que dans la méthode de Dhome, on obtient trois équations qui nous permettent de déterminer le vecteur de translation  $T$ .

#### Discussion

Pour beaucoup d'objets, les nombreuses symétries qui apparaissent permettent de mettre en évidence des trièdres semblables. Des optimisations importantes propres à cette méthode sont alors exploitables. Ceci peut être utilisé, par exemple, dans le cas de reconnaissance de parallélépipèdes rectangles dont les angles sont tous formés de trièdres droits. Dans ce cas, une partie des résultats des calculs nécessaires à l'application de cette méthode peut être réutilisée pour tous les angles semblables. Par contre, la mise en place de structures de données particulières devient alors nécessaire.

#### II.4.3. Méthode de Fang [Fan83]

Cette méthode est conçue pour l'estimation des mouvements de faible amplitude. Elle se base sur l'utilisation d'une séquence d'images de la scène de l'objet en mouvement.

En effet, l'utilisation de deux images successives dans une séquence nous permet de supposer que le mouvement qui s'effectue entre instants est très faible si la vitesse de l'objet n'est pas très importante.

Pour calculer les paramètres du mouvement, il faut connaître les coordonnées de cinq points ou plus et leurs correspondants dans deux images successives.

### II.4.3.1. Aspect mathématique de la méthode

La géométrie du problème est représentée sur la figure II.10.

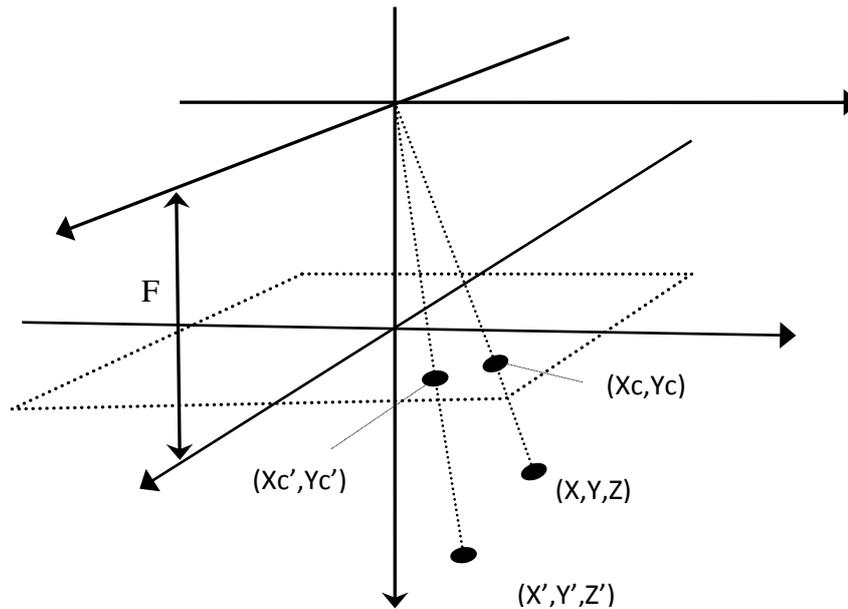


Figure II.10 : Conditions de la méthode de Fang

On note :

- $(X, Y, Z)$  les coordonnées d'un point de l'espace **3D**.
- $(x, y)$  les coordonnées d'un point de l'image.

On considère deux images successives, extraites d'une séquence d'images de la scène contenant l'objet en mouvement.

Soit :  $(X, Y, Z)$  coordonnées 3D du point  $P$  à l'instant  $t_1$

$(X', Y', Z')$  coordonnées 3D du point  $P$  à l'instant  $t_2$

$(x, y)$  coordonnées de la projection du point  $P$  à l'instant  $t_1$

$(x', y')$  coordonnées de la projection du point  $P$  à l'instant  $t_2$

On définit :

$$\Delta x = x' - x \quad (\text{II.10})$$

$$\Delta y = y' - y \quad (\text{II.11})$$

L'équation du mouvement 3D d'un corps rigide est la suivante :

$$\begin{bmatrix} X' \\ Y' \\ Z' \end{bmatrix} = R \cdot \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} + \begin{bmatrix} \Delta X \\ \Delta Y \\ \Delta Z \end{bmatrix}$$

où  $(\Delta X, \Delta Y, \Delta Z)$  est le vecteur de translation et  $R$  la matrice de rotation (3x3)

Donnée par

$$\begin{bmatrix} R_{11} & R_{12} & R_{13} \\ R_{21} & R_{22} & R_{23} \\ R_{31} & R_{32} & R_{33} \end{bmatrix}$$

avec :

$$\begin{aligned} R_{11} &= n_1^2 + (1 - n_1^2) \cos \theta \\ R_{12} &= n_1 \cdot n_2 + (1 - \cos \theta) - n_3 \cdot \sin \theta \\ R_{13} &= n_1 \cdot n_3 + (1 - \cos \theta) - n_2 \cdot \sin \theta \\ R_{21} &= n_2^2 + (1 - n_2^2) \cos \theta \\ R_{22} &= n_1 \cdot n_2 + (1 - \cos \theta) - n_3 \cdot \sin \theta \\ R_{23} &= n_1 \cdot n_3 + (1 - \cos \theta) - n_2 \cdot \sin \theta \\ R_{31} &= n_2^2 + (1 - n_2^2) \cos \theta \\ R_{32} &= n_1 \cdot n_2 + (1 - \cos \theta) - n_3 \cdot \sin \theta \\ R_{33} &= n_1 \cdot n_3 + (1 - \cos \theta) - n_2 \cdot \sin \theta \end{aligned}$$

Avec  $n_1, n_2, n_3$  cosinus directeurs de l'axe de rotation.

Si on se met dans l'hypothèse du faible mouvement la matrice de rotation aura pour composantes :

$$R_{11} = R_{22} = R_{33} = 1$$

$$R_{12} = -R_{21} = -n_3 \cdot \theta$$

$$R_{23} = -R_{32} = -n_1 \cdot \theta$$

La relation qui existe entre les coordonnées 3D de chaque point et celles de sa projection est donnée par :

$$X=f \cdot x/z \quad \text{et} \quad Y=f \cdot y/z$$

L'équation de base du mouvement est donnée par :

$$\frac{\Delta X' - x - \Delta X}{\Delta X(\varphi_1 \cdot x + \varphi_1 \cdot y + f) + (f \cdot y \cdot \varphi_3 - (\varphi_2 + x^2)) \cdot \varphi_2 + x \cdot y \cdot \varphi_1} = \frac{\Delta Y' - y - \Delta Y}{\Delta Y(\varphi_1 \cdot x + \varphi_1 \cdot y + f) + (f \cdot y \cdot \varphi_3 - (\varphi_2 + x^2)) \cdot \varphi_2 + x \cdot y \cdot \varphi_1}$$

Pour résoudre cette équation on passe à une représentation en coordonnées homogènes, après quoi on utilise des méthodes de résolution de systèmes non linéaires.

#### II.4.3.2. Discussion

Les résultats numériques obtenus par cette méthode sont très sensibles au nombre et au choix des points. Il a été démontré mathématiquement que neuf points choisis de telle manière qu'ils n'appartiennent pas au même plan donnent une solution unique. Dans la pratique suivant les résultats donnés par l'auteur six points suffisent généralement pour trouver une solution unique. Cinq donnent une solution qui n'est pas unique.

Le second aspect de la sensibilité lié au choix des points concerne la distance séparant l'objet du plan image. Plus la distance augmente plus l'erreur augmente.

Le troisième aspect concerne l'amplitude des angles de rotation. En effet, les résultats qui ont été fournis montrent que l'erreur augmente très vite lorsque l'angle de rotation dépasse quelques degrés.

## II.5. Conclusion

Dans ce chapitre nous avons explicité les différentes méthodes de reconstruction et de localisation d'objets 3D. Toutes ces méthodes sont basées sur la vision monoculaire. Ainsi, nous avons rappelé les méthodes les plus intéressantes et les plus utilisées en vision ; on peut citer la méthode basée sur le changement d'éclairage "shape from shading", la méthode basée sur le changement de la texture, "shape from texture", la méthode basée sur calcul de la profondeur de la scène "depth from defocus" ou de la forme "shape from Defocus" et les méthodes utilisant les caméras 2.5D à temps de vol. A cet effet, nous avons présenté des méthodes pratiques notamment celle de Dhome, Horaud et Fang en formulant leurs aspects théorique et pratique.

## **Chapitre III**

# **Manipulation pratique**

III.1.Introduction

III.2.Problème de la perspective inverse

III.3.Détermination de la rotation

III.4.Détermination de la translation

III.5.Algorithme de Simulation

III.6.Résultats obtenus

III.7.Manipulation réalisée

III.8.Conclusion

### III.1.Introduction

Dans ce chapitre, il s'agira d'élaborer un algorithme qui sera en mesure d'estimer la position 3D d'un objet mobile, en utilisant des séquences d'images prises par une caméra fixe pendant le mouvement de cet objet.

Pour réaliser cet algorithme on se met dans les hypothèses suivantes :

- La caméra est fixe.
- Le mouvement de l'objet entre deux instants successifs est suffisamment faible.
- L'objet en mouvement est rigide.

Le principe de ce travail se base essentiellement sur l'exploitation des travaux de [DHO 89] et celui de [FAN 83] pour l'estimation du mouvement 3D.

Il s'agira donc, de déterminer les paramètres du mouvement 3D, c'est-à-dire les trois angles de rotation autour des axes et les trois composantes du vecteur de translation ce qui ne permettra d'avoir la position de notre objet.

Pour le faire, il est nécessaire de:

- Connaître la position initiale 3D de l'objet.
- Avoir une image qui contient au moins la position de trois segments de l'objet mobile.

Le repère lié à la caméra sera noté  $R_c$ , le repère lié à l'objet lui sera noté  $R_o$  et celui lié au plan image sera noté  $R_m$ .

Le modèle de la caméra retenu est celui de la projection centrale, ainsi la relation entre les coordonnées 3D et les coordonnées 2D est celle donnée par l'équation suivante :

$$p_i(x_i, y_i) = (f \cdot X_c, f \cdot Y_c) \quad (f: \text{la focale})$$

Comme dans le cas de la méthode de Dhome, on décompose le mouvement de l'objet en deux :

- Une rotation  $R(\varphi_1, \varphi_2, \varphi_3)$  tel que  $\varphi_1, \varphi_2, \varphi_3$  représentent les angles de rotation suivant les axes  $X, Y, Z$  respectivement.
- Un translation  $T(\Delta X, \Delta Y, \Delta Z)$

On commencera par la détermination de la rotation  $R(\varphi_1, \varphi_2, \varphi_3)$  ensuite on détermine la translation  $T(\Delta X, \Delta Y, \Delta Z)$ .

### III.2. Problème de la perspective inverse

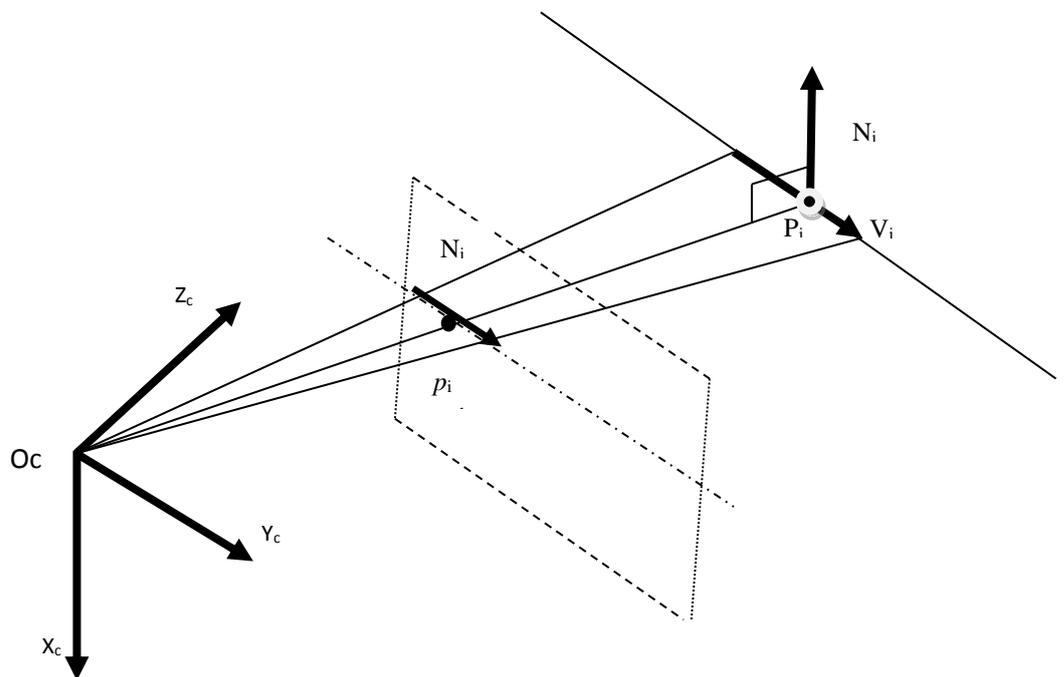


Figure III.1: Problème de la perspective inverse

Soit  $(l_i)$  une droite du plan image ; elle est caractérisée par son vecteur directeur  $v_i$  qui a pour composantes  $(a_i, b_i, 0)$ , dans le repère  $R_c$  ;  $p_i$  un point lui appartenant de coordonnées  $(x_c, y_c, f)$  dans le même repère. Cette droite correspond à la projection de la droite 3D  $(L_i)$ . Le plan qui passe par l'origine du repère  $R_c$  noté  $O_c$  et qui contient les deux droites  $(L_i)$  et sa projection  $(l_i)$  est appelé plan d'interprétation [DHO89].

Le vecteur normal à ce plan peut être calculé en effectuant le produit vectoriel du vecteur  $v_i$  avec  $O_c p_i$  défini par le centre  $O_c$  et le point  $p_i$ . (Figure III.1)

Ce vecteur sera noté  $N_i$  et aura pour composantes  $(N_{xi}, N_{yi}, N_{zi})$  qui s'écrivent :

$$\begin{bmatrix} N_{xi} \\ N_{yi} \\ N_{zi} \end{bmatrix} = \begin{bmatrix} a_i \\ b_i \\ 0 \end{bmatrix} \wedge \begin{bmatrix} x_i \\ y_i \\ f \end{bmatrix}$$

d'où:

$$\begin{bmatrix} N_{xi} \\ N_{yi} \\ N_{zi} \end{bmatrix} = \begin{bmatrix} b_i \cdot f \\ -a_i \cdot f \\ a_i \cdot y_i - b_i \cdot x_i \end{bmatrix} \quad (\text{III.1})$$

Soit  $V_i(A_i, B_i, C_i)$  le vecteur directeur de la droite 3D ( $L_i$ ).

Le produit scalaire des deux vecteurs  $V_i$  et  $N_i$  doit être nul (condition d'orthogonalité)(Figure III.1).

D'où 
$$A_i \cdot N_{xi} + B_i \cdot N_{yi} + C_i \cdot N_{zi} = 0$$

avec 
$$N_{xi} = b_i \cdot f$$

$$N_{yi} = -a_i \cdot f$$

$$N_{zi} = a_i \cdot y_i + b_i \cdot x_i$$

### III.3.Détermination de la rotation

Soient :

- un objet 3D en mouvement de rotation pure.
- ( $L_i$ ) une droite 3D de cet objet tel que  $i=1:3$  caractérisée par son vecteur directeur  $V_{i_k}$  à l'instant  $t_k$  et  $p_i$  un point lui appartenant.
- ( $l_i$ ) la projection de la droite ( $L_i$ ) qui lui correspond.
- $V_{i_{k+1}}$  les vecteurs directeurs des droites ( $l_i$ ) à l'instant  $t_{k+1}$ .(Figure III.2).

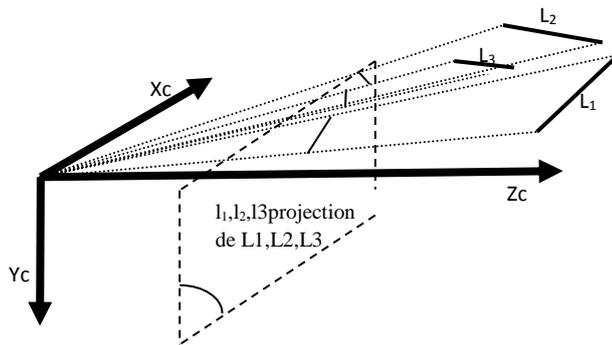


Figure III.2: Problème de la rotation

Les coordonnées du vecteur  $V_{i_{k+1}}$  peuvent être exprimées en fonction de celles de  $V_{i_k}$  par la relation suivante :

$$V_{i_{k+1}} = R(\varphi_1, \varphi_2, \varphi_3) \cdot V_{i_k}$$

Cette dernière équation est celle du mouvement d'un corps rigide.

Si on se met dans l'hypothèse du faible mouvement entre  $t_k$  et  $t_{k+1}$  la matrice de rotation se réduira à la matrice suivante:

$$\begin{bmatrix} 1 & -\varphi_3 & \varphi_2 \\ \varphi_3 & 1 & -\varphi_1 \\ -\varphi_2 & \varphi_1 & 1 \end{bmatrix}$$

où  $\varphi_1$ ,  $\varphi_2$  et  $\varphi_3$  sont les angles de rotation autour des axes  $X_c$ ,  $Y_c$  et  $Z_c$  respectivement.

En remplaçant dans l'équation par la valeur de la matrice de rotation on aura :

$$\begin{bmatrix} A_{i_{k+1}} \\ B_{i_{k+1}} \\ C_{i_{k+1}} \end{bmatrix} \begin{bmatrix} A_{i_{k+1}} \\ B_{i_{k+1}} \\ C_{i_{k+1}} \end{bmatrix} = [R(\varphi_1, \varphi_2, \varphi_3)] \cdot \begin{bmatrix} A_{i_k} \\ B_{i_k} \\ C_{i_k} \end{bmatrix} \quad (\text{III.2})$$

Soit  $N_{i_{k+1}}$  le vecteur normal au vecteur  $V_{i_{k+1}}$  qui est déterminé par l'équation III.2 et qui a pour composantes  $N_{ix_{k+1}}$ ,  $N_{iy_{k+1}}$  et  $N_{iz_{k+1}}$

Le produit scalaire des deux vecteurs  $N_{i_{k+1}}$  et  $V_{i_{k+1}}$  sera donné par l'équation suivante :



$$[R(\varphi_1, \varphi_2, \varphi_3)] \cdot \begin{bmatrix} A_{i_k} \\ B_{i_k} \\ C_{i_k} \end{bmatrix} \cdot \begin{bmatrix} N_{xi_{k+1}} \\ N_{yi_{k+1}} \\ N_{zi_{k+1}} \end{bmatrix} = 0 \quad (\text{III.3})$$

En développant cette équation on obtient :

$$\begin{aligned} (C_{i_k} \cdot N_{yi_{k+1}} - B_{i_k} \cdot N_{xi_{k+1}}) \cdot \varphi_1 - (C_{i_k} \cdot N_{xi_{k+1}} - A_{i_k} \cdot N_{zi_{k+1}}) \cdot \varphi_2 \\ - (B_{i_k} \cdot N_{xi_{k+1}} - A_{i_k} \cdot N_{yi_{k+1}}) \cdot \varphi_3 = A_{i_k} \cdot N_{xi_{k+1}} - B_{i_k} \cdot N_{yi_{k+1}} + C_{i_k} \cdot N_{zi_{k+1}} \end{aligned}$$

En prenant  $i$  qui varie de 1 à 3 on obtient trois équation à trois inconnues qu'on pourra mettre sous la forme :

$$\begin{cases} \alpha_{11} \cdot \varphi_1 + \alpha_{12} \cdot \varphi_2 + \alpha_{13} \cdot \varphi_3 = \alpha_{14} \\ \alpha_{21} \cdot \varphi_1 + \alpha_{22} \cdot \varphi_2 + \alpha_{23} \cdot \varphi_3 = \alpha_{24} \\ \alpha_{31} \cdot \varphi_1 + \alpha_{32} \cdot \varphi_2 + \alpha_{33} \cdot \varphi_3 = \alpha_{34} \end{cases}$$

Et qui pourra se résoudre soit par la méthode itérative de résolution de systèmes d'équations linéaires ou par la méthode de **Cramer**.

### III.4. Détermination de la translation

La détermination de la rotation nous a permis d'avoir l'orientation de l'objet en mouvement. Pour la localisation de ce dernier il faut déterminer sa translation. (Figure III.3).

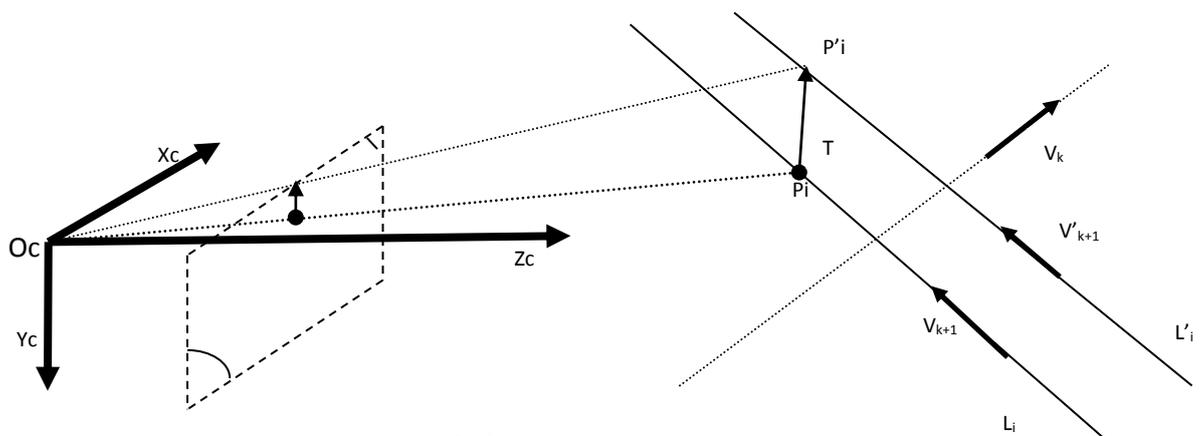


Figure III.3: Problème de la translation

- Soient  $(X_{i_{k+1}}, Y_{i_{k+1}}, Z_{i_{k+1}})$  les coordonnées du vecteur  $V_{i_k}$  après sa rotation.
- $P_i$  un point de la droite  $L_i$  de vecteur directeur  $V_{i_{k+1}} = (X_{i_{k+1}}, Y_{i_{k+1}}, Z_{i_{k+1}})$

- $P_i'$  un point de la droite  $L_i'$  de vecteur directeur  $V_{i'k+1}$ .
- $(\Delta X, \Delta Y, \Delta Z)$  les coordonnées du vecteur de translation  $T=P_i'P_i$ .

On peut écrire :

$$O_c P_i' * N_{i_{k+1}} = 0 \quad (III.4)$$

avec:

$$O_c P_i' = O_c P_i + P_i P_i' \quad (III.5)$$

D'où :

$$(O_c P_i + P_i P_i') \cdot N_{i_{k+1}} = 0 \quad (III.6)$$

Soit :

$$O_c P_i = \begin{bmatrix} P_{x_{i_{k+1}}} \\ P_{y_{i_{k+1}}} \\ P_{z_{i_{k+1}}} \end{bmatrix} \quad (III.7)$$

Donc l'équation (III.7) devient :

$$\begin{bmatrix} P_{x_{i_{k+1}}} + \Delta X \\ P_{y_{i_{k+1}}} + \Delta Y \\ P_{z_{i_{k+1}}} + \Delta Z \end{bmatrix} \cdot \begin{bmatrix} N_{xi_{k+1}} \\ N_{yi_{k+1}} \\ N_{zi_{k+1}} \end{bmatrix} = 0 \quad (III.8)$$

$$\begin{aligned} N_{xi_{k+1}} \cdot \Delta X + N_{yi_{k+1}} \cdot \Delta Y + N_{zi_{k+1}} \cdot \Delta Z = \\ -P_{x_{i_{k+1}}} \cdot N_{xi_{k+1}} - P_{y_{i_{k+1}}} \cdot N_{yi_{k+1}} - P_{z_{i_{k+1}}} \cdot N_{zi_{k+1}} \end{aligned} \quad (III.9)$$

Comme dans le cas de la rotation, si on prend trois segments, c'est-à-dire  $i$  varie de 1 à 3 on aura trois équations à trois inconnus qu'on peut résoudre par la même méthode que pour la rotation.

### III.5. Algorithme de simulation

Pour la simulation de cet algorithme établi, on utilise trois vecteurs qui sont considérés liés à l'objet fictif.

Pour chaque image, la procédure suivante est réalisée :

1. Sélectionner les trois vecteurs à suivre.
2. On fait tourner ces vecteurs d'un angle très petit de l'ordre de 1 degré suivant l'axe  $X_c, Y_c, Z_c$ .
3. Projection de ces vecteurs sur un plan rétinien fictif.
4. On calcule les normales aux plans d'interprétations des vecteurs projetés
5. Formulation de l'équation  $A * R(\varphi_1, \varphi_2, \varphi_3) = B$ .
6. Cette équation est résolue par la méthode de Cramer si le nombre des segments est égal à trois sinon on utilise la méthode des moindres carrés, ainsi on obtient les trois angles de rotation.
7. Détermination de la position des trois segments en utilisant les angles  $\varphi_1, \varphi_2, \varphi_3$ .
8.  $K=k+1$  et on revient à 2.

Pour le calcul de la translation on procède de la manière suivante :

1. On choisit 3 points qui appartiennent aux trois droites qui portent les vecteurs sélectionnés.
2. La relation entre ces points et leurs correspondants après translation est donnée par le système suivant :  $C.\Delta T=D$  où  $\Delta T$  est le vecteur de translation, Cest une matrice 3x3 et  $D$  est un vecteur 3x1.
3. Le système obtenu est résolu de la même manière que dans le cas de la rotation.

### III.6.Résultats obtenus

Pour tester cet algorithme, on a sélectionné trois vecteurs d'un polyèdre fictif qu'on fait tourner d'un angle assez faible suivant les trois axes. Ces angles sont calculés par le programme développé. On compare le résultat au résultat théorique pour trouver l'erreur commise sur la mesure de ceux-ci.

L'erreur commise sur l'angle de rotation suivant l'axe X est donnée par la figure suivante :

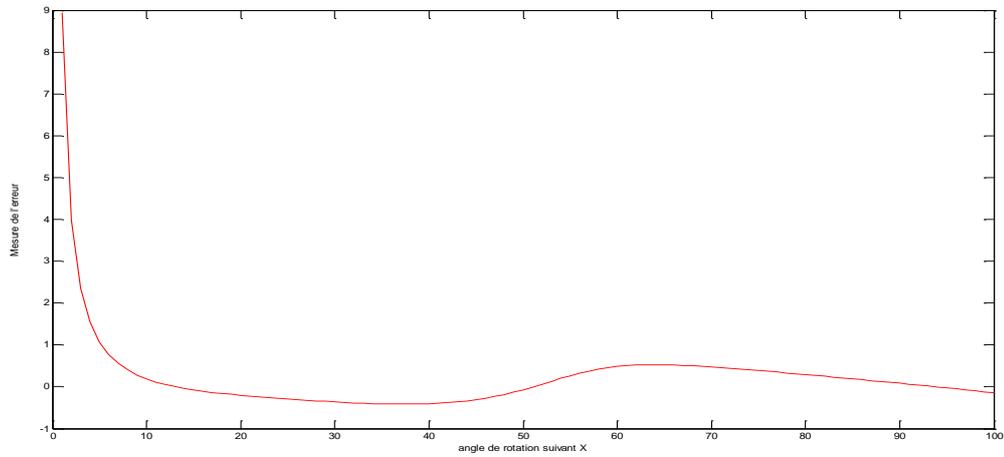


Figure III.4: Erreur sur l'estimation de l'angle de rotation suivant l'axe X

Celle commise sur l'estimation de l'angle de rotation suivant Y est donnée par la figure suivante :

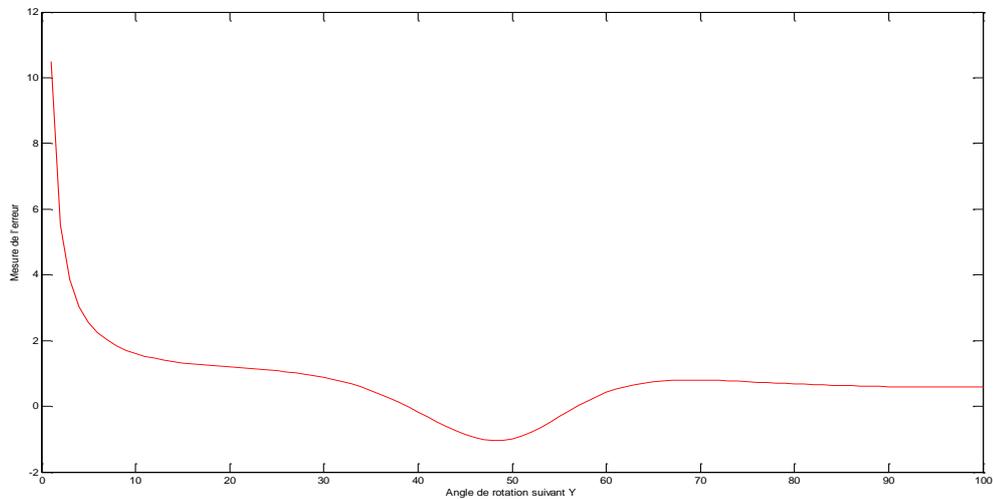


Figure III.5 : Erreur sur l'estimation de l'angle de rotation suivant l'axe Y

Enfin, l'erreur sur l'angle de rotation suivant l'axe Z est explicitée sur la figure suivante :

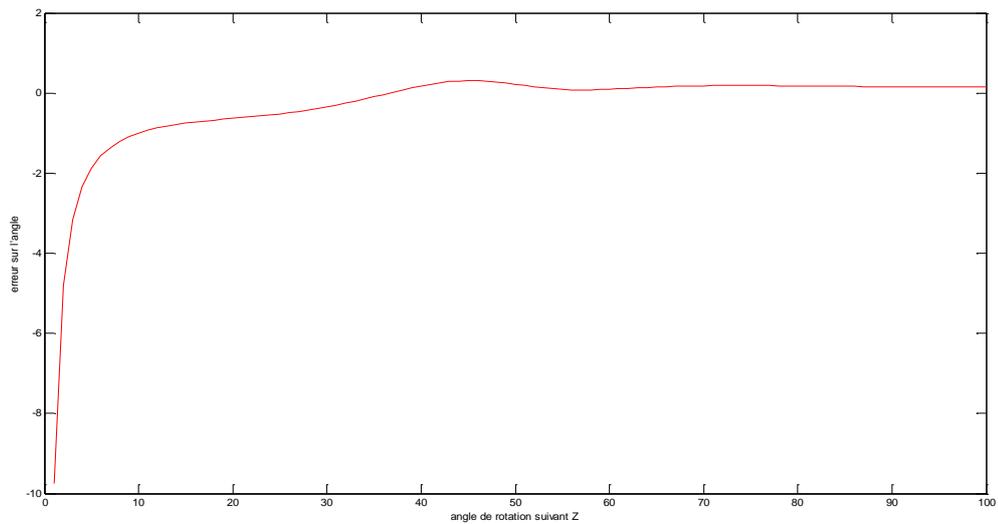


Figure III.6 : Erreur sur l'estimation de l'angle de rotation suivant l'axe Z

Le tracé des courbes représente l'évolution de l'erreur en fonction de la valeur de l'angle de rotation.

Cette erreur devient importante, au fur et à mesure que le mouvement de l'objet évolue, ce qui est dû sûrement au cumul de l'erreur. En effet, la position de l'objet après mouvement est obtenue en utilisant les angles de rotations calculés par notre algorithme. Ceux-ci étant entachés d'erreur, due à l'approximation faite sur les angles de rotation.

Pour la translation, les composantes du vecteur de translation sont parfaitement estimées. En effet, l'erreur pour les trois composantes est quasi nulle (figure III.7).

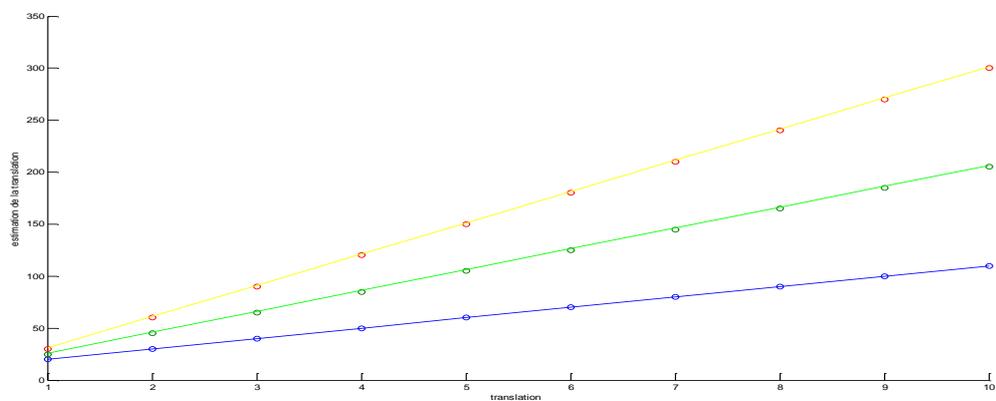


Figure III.7 : La translation est parfaitement estimée

### III.7. Conditions de la manipulation

Pour les conditions expérimentales suivantes :

- Nombre de points utilisés pour la calibration six.
- La focale  $f=0.0085$
- Matrice globale calculée par matlab

951.25	2.80	330.87	371.09
32.95	421.22	-113.78	97.94
0.33	0.01	-1.22	1
0	0	0	1

Le modèle externe de la caméra ( $T_{oc}$ ) est :

0.9649	-0.0005	0.2623	-0.905
-0.0090	0.9999	0.0034	-0.285
-0.2623	-0.0056	0.9649	1
0	0	0	1

Le modèle interne ( $M_{int}$ ) est :

-652.14	0	446.30	0
0	-329.9	109.92	0
0	0	1	0

Les paramètres de la caméra sont donnés par le constructeur.

Après calibration de la caméra, on sélectionne trois segments de l'objet mobile. Leurs positions 3D initiales étant connues dans le repère  $R_o$  (repère de l'objet), celles dans le repère  $R_c$  (repère de la caméra) sont déduites directement en utilisant la matrice  $T_{oc}$  qui permet le passage de  $R_o$  vers  $R_c$ .

Le système qui permet de faire les rotations autour de l'axe  $OX_c$  et  $OY_c$  a une précision qui n'est pas disponible.

Avant de commencer la manipulation les coordonnées des 3 vecteurs sont supposées connues dans le repère  $R_o$ . Le passage vers  $R_c$  est obtenu par la matrice  $T_{oc}$  calculée lors de la calibration.

Après chaque mouvement, les coordonnées  $u$  et  $v$  sont prises dans le repère d'affichage. Le passage vers le repère image est obtenu en utilisant la matrice interne. Les coordonnées image sont exprimées dans le repère  $R_c$ .

Les résultats obtenus seront présentés dans un tableau. Les coordonnées des trois vecteurs, appelées théoriques, sont obtenues par la matrice de rotation réelle sans approximation et celles appelées expérimentales, sont calculées à partir des angles de rotation obtenus par la méthode adoptée.

Les résultats de l'expérience montrent que l'algorithme suit bien les droites sélectionnées si on prend compte des erreurs commises lors de la manipulation. En effet, la sélection des vecteurs se fait sur l'écran de visualisation, par un opérateur humain.

### **III.8. Conclusion**

Dans ce chapitre, un algorithme d'estimation de la position d'un objet 3D a été développé. Cet algorithme s'appuie sur une image recueillie par une caméra fixe permettant de repérer le mouvement de l'objet. En utilisant une projection centrale, les relations planaires 2D – spatiales 3D ont été établies.

En décomposant le mouvement de l'objet en rotation et translation, la reconstitution de la scène est possible.

L'algorithme développé donne des résultats acceptables. On espère le développer dans des travaux futurs.

## Conclusion générale

Notre travail durant cette étude se résume en deux grandes parties :

1. La première partie était consacrée à une recherche bibliographique sur l'estimation des paramètres du mouvement d'un objet mobile, dans le but de développer une application qui nous permettra de positionner un bras de robot manipulateur dans son environnement de travail. Notre intérêt dans cette recherche s'est porté aussi sur les méthodes de reconstitution d'images statiques prises par deux caméras (cas de la stéréo vision) pour ne pas citer que celle-là.
2. Une mise en place d'un algorithme de localisation d'objet à partir d'une séquence d'images. En effet, l'utilisation d'une séquence d'images nous a permis de se mettre dans une position où on peut faire intervenir des simplifications sur le mouvement 3D afin d'aboutir à des équations de mouvement simples à résoudre.

Au début de cette recherche bibliographique on s'est intéressé aux méthodes de reconstitution et de localisation d'objets 3D en utilisant des méthodes monoculaires. On s'est intéressé plus particulièrement aux méthodes de localisation d'objets. Ainsi trois méthodes qui nous paraissaient les mieux adaptées à notre problème ont été développées.

Ces méthodes sont :

- La méthode de Dhome qui aboutit, dans la détermination du mouvement de l'objet 3D, à une équation d'ordre 8 en général difficile à résoudre.
- La méthode de Horaud qui est basée sur une résolution graphique nécessitant un temps assez long pour la résolution de ce problème.
- La méthode de Fang, utilisée pour les faibles rotations et qui nous a semblé la mieux adaptée à notre cas.

- Dans l'algorithme qu'on a utilisé, on se met dans l'hypothèse du faible mouvement ; ce qui nous a permis de simplifier la matrice de rotation considérablement et d'aboutir à une équation linéaire [Fan83]. Cette approche combinée avec une approche voisine à celle utilisée par Dhomenous permet d'établir des équations linéaires faciles à résoudre.
- Les résultats obtenus ne sont pas toujours satisfaisants et cela est dû essentiellement aux erreurs diverses introduites par notre appareillage utilisé.
- Toutefois on pense que ces résultats peuvent être améliorés par un choix judicieux des appareils ainsi qu'une augmentation du nombre de vecteurs suivis (supérieur à trois).
- La translation pure est parfaitement déterminée par le suivi des trois vecteurs.
- Dans ce travail on s'est intéressé essentiellement à l'estimation des paramètres du mouvement, au suivi des segments et à la localisation d'objets en mouvement et nous espérons qu'elle fera l'objet d'autres études dans le futur.

## Bibliographie

- [ALO 88] Y. Aloimonos, M.J.Swain. "*Shape from texture. BioCyber pp.345-360 1988 1989.*
- [AYA 89] N.Ayache." *Vision par stéréoscopique et perception multi - sensorielle; Application à la robotique mobile*", Inter Edition Sciences informatique, pp.193-229.1989.
- [BES 88] J.E.Besançon. "*Vision par ordinateur en deux et trois dimensions*". Edition Eyrolles, pp. 3-13.1988.
- [BOC 08] B.Bocquillon. "*Contributions à l'auto-calibrage des caméras :modélisations et solutions garanties par l'analyse d'intervalle*".Thèse de doctorat, Université de Toulouse,octobre 2008.
- [BOG 08] A. Bogomjakov,C. Gotsman." *Reduced depth and visual halls of complex 3d scenes*". Computer Graphics, pp. 175-182. Avril 2008.
- [BOU 99] C.Boucher, J-C. Noyer etM.Benjelloun." *Estimation des positions d'objets 3D à partir d'une séquence*".France Dix-septième colloque GRETSI, Vannes, 13-17 septembre 1999.
- [CAU 98] C.Cauchois, "*Calibration du capteur de vision omnidirectionnelle*", Rapport de Diplôme d'Etudes Approfondies. Université de Technologie de Compiègne. Juil. 1998
- [CHA 92] F.Chavand, E.Colle, M.Mallem, H. Loukil etY.Chekhar." *Assistance à la perception en téléopération et en supervisionde : procédures automatique*, Rapport ANVAR. Sept 1992.
- [DHO 89] M.Dhome,M.Richetin,J.T.Lapresteet G.Rives. "*Determination of attitude 3D objects, from a single perspective view*".IEEE transaction pattern Analysis and Machine Intelligence, Vol 11,N°12, p1265-1278, Déc. 1989.
- [DUG 92] J.L.Dugelay. "*Estimation du mouvement 3D et application à la télévision en relief*". Thèse de doctorat, Université Rennes. France, 1992.
- [FAN 83] J.Q.Fang et T.S.Huang. "*Solvingthree-dimmmensional small-rotation motion. Equations :uniqueness,algorithms,and numerical results*". Computer vision, Graphics, and ImageProcessing 26, pp 183-206, 1983.

- [GAN 84] S. Ganapathy. "*Decomposition of transformation matrices for robot vision*", Proc IEEE I Conf on Robotics and Automation pages 130-139 Atlanta Georgia, USA Mars 1984.
- [GUI 00] E. Guillou, D. Meneveau, E. Maisel, K. Bouatouch. "*Using vanishing of camera calibration and coarse 3d reconstruction from a single image*", The visual computer pages 396-410, 2000.
- [HOR 87] R. Horaud. "*New methods for matching 3D objects with single perspective views*", IEEE transaction pattern Analysis and Machine Intelligence, Vol. 09, N°3 p401-412, 1987.
- [HUA 89] T.S. Huang et C.H. Lee. "*Motion and structure from orthographic projection*". IEEE PAMI, Vol 11, N°5, pp. 536-540, 1989.
- [JER 12] T. Jerbi. "*Recalage de structures tridimensionnelles à partir d'acquisitions stéréo radiographiques basse dose. Application à l'estimation de mouvements humains*". Thèse de Doctorat, Université européenne de Bretagne. Jan. 2012.
- [LER 08] F. Lerasle., "*Perception 3D*", Rapport de recherche. Université Paul Sabatier Toulouse. 2008.
- [LIU 90] Y Liu, T.S. Huang et O. D Faugeras. "*Determination of camera location from 2-D to 3-d line and point correspondences*". IEEE Transactions on Pattern Analysis and Machine Intelligence Jan. 1990.
- [RAM 06] S. Ramlingam. Thèse de doctorat. "*Generic Imaging Models: Calibration and 3D Reconstruction Algorithms*". Institut National Polytechnique de Grenoble. Nov. 2006.
- [REK 08] W. Rekiq, D. Béréziat et S. Dubuisson. Rapport de recherche LIP6. "*Fusion de données spatiales (3D) et temporelles (2D+t) pour la reconstruction de scènes 3D+t : deux approches. Merging 3D Spatial Data and 2D+t Temporal Sequences for 3D+t Scene Reconstruction: Two Approaches*". Université Pierre et Marie Curie 2 INRIA Mai 2008.
- [WIL 80] T.D. Williams. "*Depth from camera motion in real world scene*". IEEE PAMI, Vol. 02, N°6, pp. 536-540, 1980.
- [XIA 87] T.U. Xiao Wei. "*Détection et estimation des objets mobiles dans une séquence d'images*". Thèse de doctorat, Université de Compiègne, avril 1987.
- [XIE 88] M. Xie et P. Rives. "*Un algorithme efficace pour la mise en correspondance des segments 2D dans une séquence d'images*". Rapport de recherche INRIA, 1988.

Caractéristiques caméra

Modèle	JX 660
Marque	Fujifilm
Catégorie	Compact
Nombre de Pixels	16
Capteur	CCD avec filtre
Ouverture de l'objectif	f/3.5 - 6.3
Distance focale	26 - 130 mm
Obturbateur	1/1400
Sensibilité (ISO equiv.)	100 - 3200 ISO
Zoom optique	x5
Zoom numérique	x7
Modes de flash intégré	
Flash intégré	oui
Flash externe	Non

Calibrage

Pour notre manipulation on a utilisé l'outil existant de calibration disponible dans matlab "calibration camera toolbox " qui permet la détermination des paramètres du modèle de la caméra utilisée.

La mire utilisée pour la calibration est un damier de 10x10 carrés.

La taille en millimètre de chaque carré sur l'axe x et l'axe y et 50.

Pour chaque image on détermine les quatre extrémités du damier dans le même ordre.

L'ordre est important car il définit le repère du travail.

Le nombre des images utilisées et de 8.