



VISION PAR ORDINATEUR

Philippe Marthon

Edition 1999

Sommaire

Chapitre 1	Introduction	
Section 1	Interprétation	7
1	Méthode ascendante	7
2	Méthode descendante	8
Chapitre 2	Analyse des images	
Section 1	Calcul d'attributs des pixels	11
1	Histogramme du premier ordre	11
2	Histogrammes du deuxième ordre	12
3	Gradient	13
4	Contour	14
5	Attributs tridimensionnels	18
Section 2	Segmentation et regroupement	18
1	Segmentation	18
2	Regroupement	19
Chapitre 3	Extraction des caractéristiques géométriques d'une figure	
Section 1	Image numérique	21
1	Obtention	21
2	Inconvénients	22
Section 2	Projections	22
Section 3	Moments géométriques centrés	22
Section 4	Descripteurs de Fourier	23
1	Descripteur de Fourier de coordonnées	23
2	Descripteur de Fourier invariant par rotation	23
Section 5	Codages statistiques	24
Section 6	Zones discriminantes	24
Chapitre 4	Extraction de la structure d'une figure	
Section 1	Squelette	27
Section 2	Chaîne	27
Section 3	Graphe de propriétés et de relations	29
1	Graphe de régions adjacentes	29
2	Généralisation	29
Section 4	Exercice	30
Chapitre 5	Modélisation géométrique et calibrage d'une caméra	
Section 1	Modélisation	33
1	Caméra en trou d'épingle	33

Section 2	Modélisation en trou d'épingle	33
2	Passage à l'image numérique	35
3	Passage au repère externe	37
4	Extension projective	38
5	Interprétation des lignes de \tilde{C} et détermination du centre optique	38
6	Modèle inverse	39
Section 3	Calibrage géométrique	39
1	Méthode linéaire de calibrage	40
Chapitre 6	Photométrie et stéréo	
Section 1	Définitions fondamentales	43
1	Eclairement	43
2	Intensité	43
3	Luminance	44
4	44
Section 2	Relation entre luminance et éclairement	44
Section 3	Réflectance	47
1	Repère local	47
2	Fonction de réflectance bidirectionnelle	48
3	Eclairement d'une surface par une source étendue	48
4	Eclairement d'une surface par une source ponctuelle . .	50
5	Surfaces typiques	50
Section 4	Modèle photométrique et carte de réflectance	53
1	Surface de Lambert blanche éclairée par une source ponctuelle	54
2	Surface de Minnaert	54
Section 5	Stéréophotométrie	54
Section 6	Exercices	55
1	Surfaces de Minnaert	55
1	55
2	55
2	Stéréovision mono-image	55
3	Stéréophotométrie avec une surface de Lambert	57
Chapitre 7	Vision stéréoscopique binoculaire	
Section 1	Principe	59
Section 2	Positionnement des caméras	59
Section 3	Mise en correspondance	60
1	Points et pixels homologues	60
2	Contraintes	62
3	Critère de vraisemblance	65
4	Stéréovision humaine	65

Section 4	Reconstruction 3D en l'absence d'information de position des caméras	66
1	Equations dans les systèmes de coordonnées normalisés	66
2	Application à la reconstruction à un facteur d'échelle près d'une scène 3D prise sous deux points de vue distincts .	67
Section 5	Proposition d'une méthode	70
2	Résultats	71
Chapitre 8	Reconnaissance des objets	73
Section 1	Définition	73
Section 2	Exemples	73
1	Lecture optique	73
2	Vision par ordinateur	73
Section 3	Deux grandes approches	73
1	Approche cognitive	73
2	Approche empirique	74
Chapitre 9	Théorie de Bayes	77
Section 1	Risque	77
Section 2	Règle de décision de Bayes	77
1	Cas de deux classes	77
2	Probabilité d'erreur minimale	78
Section 3	Exercice	79
1	79
2	79

Chapitre 1 Introduction

Section 1 Interprétation

L'interprétation des images a pour objet de fournir des informations sur les objets qui sont présents dans les images. Ces informations portent sur la nature physique de ces objets, leurs formes géométriques, leurs positions et leurs mouvements dans l'espace, leurs natures.

On peut distinguer deux grands types de méthodes d'interprétation:

- a. les méthodes *ascendantes*. Elles commencent toutes par une *analyse* des images.
- b. les méthodes *descendantes*. Elles se caractérisent par la prise en compte d'informations a priori sur la scène observée et les différents objets susceptibles d'y être présents.

On admet généralement que la vision humaine est à la fois ascendante et descendante et que l'aspect ascendant y est prépondérant.

1 Méthode ascendante

Traçons les différentes étapes:

- a. Calcul des attributs des pixels de l'image par analyse radiométrique de leurs voisinages.
- b. Segmentation et regroupement : l'image est découpée en régions censées correspondre à des images de parties significatives des objets.
- c. Mesures géométriques sur les régions,
- d. Extraction de la structure de l'image
- e. Interprétation tridimensionnelle et reconnaissance des objets

David Marr a proposé un modèle ascendant de la vision humaine dont nous nous sommes inspirés dans le modèle ci-après:

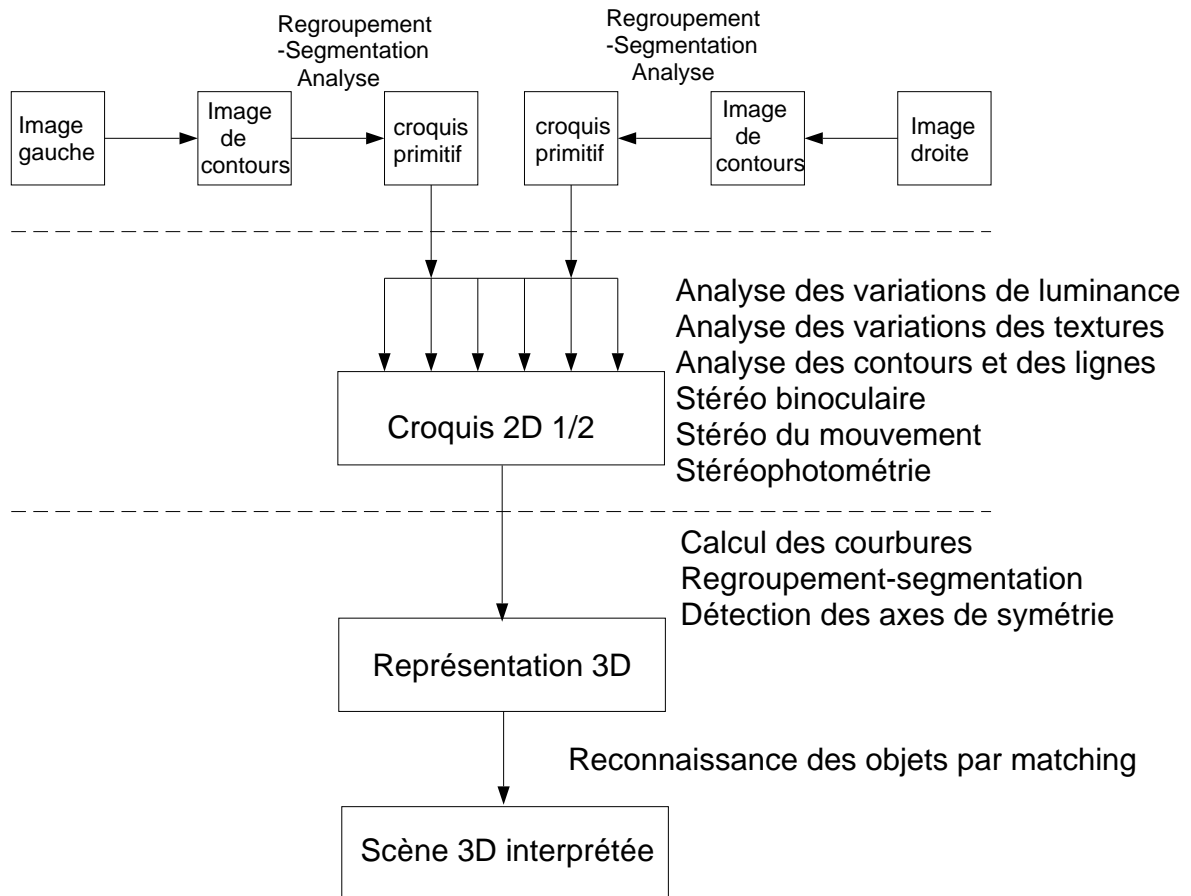


Figure 1.1 Modèle de Marr. On remarque 3 niveaux de traitement : les traitements de bas niveau (traitement des images), les traitements de moyen niveau, les traitements de haut niveau.

2 Méthode descendante

Elle intervient lorsque l'on recherche dans l'image la présence d'un objet particulier.



Figure 1.2 Qui suis-je ?

Chapitre 2 Analyse des images

Section 1 Calcul d'attributs des pixels

Chaque pixel (i,j) possède un attribut de base qui est son niveau de gris $im(i,j)$. On peut aussi calculer d'autres attributs en analysant la répartition des niveaux de gris dans un voisinage de ce pixel.

1 Histogramme du premier ordre

1.1 Définition

Un histogramme h_{ij} du premier ordre est une estimation de la distribution de probabilité du niveau de gris du pixel (i,j) où L est le nombre de niveaux de gris possibles

$$h_{ij}(x) = P\{im(i,j) = x\} \quad 0 \leq x \leq L - 1 \quad (2.1)$$

(par exemple $L = 256$).

Dans le cas où l'image est stationnaire au sens strict, chaque pixel possède la même distribution des niveaux de gris et l'on peut écrire:

$$h_{ij} = h \quad \forall i, j \quad (2.2)$$

Si l'image est ergodique, cette distribution peut s'estimer à partir des fréquences d'apparition des niveaux de gris dans l'image:

$$h(x) \simeq \frac{n(x)}{N^2} \quad (2.3)$$

où $n(x)$ désigne le nombre de pixels possédant le niveau de gris x :

$$n(x) = \text{card} \{ (i,j) \mid im(i,j) = x \} \quad (2.4)$$

et N^2 le nombre total de pixels de l'image.

1.2 Histogramme d'un voisinage

Considérons un voisinage \mathcal{V} du pixel (i,j) . D'après ce qui précède on définit l'histogramme h de \mathcal{V} comme la fonction:

$$h(x) = \frac{n(x)}{\text{card}(\mathcal{V})} \quad \forall x = 0, \dots, L - 1 \quad (2.5)$$

où $n(x)$ désigne le nombre de pixels de \mathcal{V} possédant le niveau de gris x .

1.3 Exemples

Les attributs seront tirés des caractéristiques de l'histogramme h du voisinage \mathcal{V} , définie par la formule (2.5). Parmi les principaux:

a. Moyenne

$$\bar{x} = \sum_{x=0}^{L-1} x h(x) \quad (2.6)$$

b. Attributs tirés des moments centrés

$$\mu_k = \sum_{x=0}^{L-1} (x - \bar{x})^k h(x) \quad (2.7)$$

Dans cette famille, les plus utilisés sont:

- I. la variance μ_2
- II. l' écart-type $\sigma = \sqrt{\mu_2}$
- III. le biais $\frac{\mu_3}{\sigma^3}$
- IV. le kurtosis $\frac{\mu_4}{\mu_2^2} - 3$ (le kurtosis est nul pour toute gaussienne de moyenne nulle)

c. Entropie

$$E = - \sum_{x=0}^{L-1} h(x) \log(h(x)) \quad (2.8)$$

L'entropie est maximale quand tous les niveaux de gris sont équirépartis dans le voisinage \mathcal{V} i.e. quand $h(x) = \frac{1}{L} \quad \forall x = 0 \dots L-1$

2 Histogrammes du deuxième ordre

2.1 Définition

Un histogramme $h_{i1j1i2j2}$ du deuxième ordre est une estimation de la distribution conjointe des niveaux de gris des deux pixels $(i1,j1)$ et $(i2,j2)$ où L est le nombre de

$$h_{i1j1i2j2}(x_1, x_2) = P\{im(i_1, j_1) = x_1, im(i_2, j_2) = x_2\} \quad \begin{matrix} 0 \leq x_1 \leq L-1 \\ 0 \leq x_2 \leq L-1 \end{matrix} \quad (2.9)$$

niveaux de gris possibles.

Dans le cas où l'image est stationnaire au sens strict, $h_{i1j1i2j2}$ ne dépend que de $i = i1 - i2$ et de $j = j1 - j2$. On peut donc écrire:

$$h_{i1j1i2j2} = h_{ij} \quad \forall i1, j1, i2, j2 \quad (2.10)$$

Si l'image est ergodique, cette distribution peut s'estimer comme suit:

$$h_{ij}(x_1, x_2) \simeq \frac{n(x_1, x_2, i, j)}{N^2} \quad (2.11)$$

$$n(x_1, x_2, i, j) = \text{card} \{ \{ (i_1, j_1), (i_1 + i, j_1 + j) \} \mid \text{im}(i_1, j_1) = x_1, \text{im}(i_1 + i, j_1 + j) = x_2 \} \} \quad (2.12)$$

2.2 Histogramme d'un voisinage

Considérons un voisinage \mathcal{V} du pixel (i_l, j_l) . D'après ce qui précède on définit l'histogramme h_{ij} de \mathcal{V} comme la fonction:

$$h_{ij}(x_1, x_2) = \frac{n(x_1, x_2, i, j)}{\text{card}(\mathcal{V})} \quad (2.13)$$

2.3 Exemples

Les attributs seront tirés des caractéristiques des histogrammes h_{ij} du voisinage \mathcal{V} du pixel (i_l, j_l) définis par la formule (2.13). Parmi les principaux:

a. l'autocovariance où la moyenne \bar{x} est estimée grâce à la formule (2.6)

$$\kappa_{ij} = \sum_{x_1=0}^{L-1} \sum_{x_2=0}^{L-1} (x_1 - \bar{x})(x_2 - \bar{x}) h_{ij}(x_1, x_2) \quad (2.14)$$

b. l'inertie

$$I_{ij} = \sum_{x_1=0}^{L-1} \sum_{x_2=0}^{L-1} (x_1 - x_2)^2 h_{ij}(x_1, x_2) \quad (2.15)$$

3 Gradient

En chaque pixel (i, j) de l'image supposée dérivable, on peut calculer le gradient de l'image qui est un vecteur à deux composantes :

$$\nabla \text{im}(i, j) = \left(\frac{\partial \text{im}}{\partial x}(i, j), \frac{\partial \text{im}}{\partial y}(i, j) \right) \quad (2.16)$$

Lorsqu'on dispose d'une image numérique, on peut estimer le gradient grâce à deux masques, ce qui revient à calculer deux sommes pondérées des niveaux de gris présents dans un voisinage de ce pixel.

Exemple : masques de Sobel

$$\begin{bmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{bmatrix} \text{ et } \begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix} \text{ approchent } \frac{\partial im}{\partial x} \text{ et } \frac{\partial im}{\partial y} \text{ à un facteur multiplicatif près.}$$

Ainsi:

$$(im(i+1, j-1) + 2im(i+1, j) + im(i+1, j+1)) - (im(i-1, j-1) + 2im(i-1, j) + im(i-1, j+1)) \quad (2.17)$$

approche à un facteur multiplicatif près $\frac{\partial im}{\partial x}(i, j)$

$$\text{On remarquera que: } \begin{bmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} * \begin{bmatrix} -1 & -1 \\ +1 & +1 \end{bmatrix}$$

Ainsi l'application des masques de Sobel revient-elle à moyenner l'image au moyen du masque 2x2 $\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$ puis à calculer les dérivées directionnelles horizontale et verticale de l'image moyennée en utilisant les deux masques de différence 2x2 $\begin{bmatrix} -1 & -1 \\ +1 & +1 \end{bmatrix}$ et $\begin{bmatrix} -1 & +1 \\ -1 & +1 \end{bmatrix}$

Le moyennage de l'image permet de lutter contre le bruit.

Un autre attribut intéressant est le module du gradient

$$\| \nabla im(i, j) \| = \left(\left(\frac{\partial im}{\partial x}(i, j) \right)^2 + \left(\frac{\partial im}{\partial y}(i, j) \right)^2 \right)^{\frac{1}{2}} \quad (2.18)$$



Figure 2.1 Image des modules des gradients calculés à partir de l'opérateur de Sobel. L'image initiale est la composante verte du mandrill

4 Contour

Un pixel de contour — un *edgel* — est un pixel dont la position marque un changement significatif d'intensité dans l'image.

4.1 Origine physique des contours

Un contour dans une image est le résultat de l'une des quatre causes physiques suivantes:

- variation de la physique de la surface de l'objet qui se traduit par une variation de l'albédo : l'albédo est la fraction de l'énergie du rayonnement incident réfléchi par la surface.
- discontinuité d'orientation ou arête
- variation de l'orientation d'une surface brillante induisant un reflet
- variation de l'éclairement de la surface (ombre)

Remarquons qu'une discontinuité de profondeur seule ne peut provoquer de contour. Mais il se trouve qu'une discontinuité de profondeur est généralement accompagnée d'une discontinuité d'orientation et/ou d'albédo. Donc généralement une discontinuité de profondeur s'accompagne d'un contour.

Enfin un contour peut provenir d'un bruit de numérisation voire d'une poussière sur la lentille de la caméra.

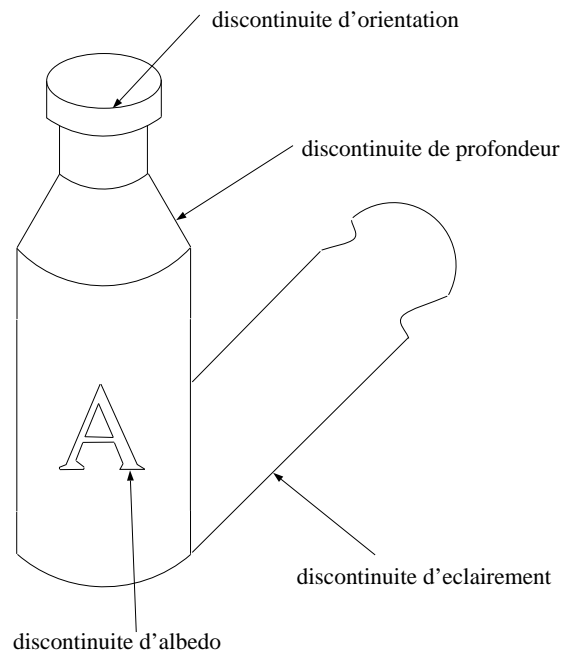


Figure 2.2 Origine physique des contours

4.2 Détection des contours

En première approximation, un pixel de contour correspond à un maximum local du module du gradient dans la direction du gradient. Autre définition équivalente: il correspond à un passage par zéro de la dérivée directionnelle seconde de l'image dans la direction du gradient.

Détecteur de Marr-Hildreth

Il est basé sur le laplacien de gaussienne.

Il se décompose en quatre étapes:

- Régularisation de l'image par une gaussienne

- b. Calcul du laplacien de l'image régularisée
- c. Détermination des passages par zéro
- d. Elimination des faux contours

Les deux premières étapes peuvent se regrouper en une seule opération de convolution:

$$\Delta(g * im) = (\Delta g) * im \quad (2.19)$$

où :

$$\Delta g(x, y) = \frac{1}{\pi \sigma^4} \left(\frac{x^2 + y^2}{2\sigma^2} - 1 \right) e^{-\frac{x^2 + y^2}{2\sigma^2}} \quad (2.20)$$

Δg est donc invariant par rotation puisqu'il ne dépend plus que de $r^2 = x^2 + y^2$

Détecteur de Shen-Castan

L'image est lissée par une fonction exponentielle séparable:

$$f(x, y) = \frac{\beta^2}{4} e^{-\beta(|x|+|y|)} \quad (2.21)$$

On applique ensuite un opérateur différentiel du 1er ou du 2ème ordre à l'image lissée et on détecte les contours en vertu de la définition précédente...

Détecteur du Bassin Versant

On met ici l'accent sur la détection des contours proprement dite.

Cet algorithme de détection comporte deux étapes:

- a. On calcule le module du gradient en chaque pixel de l'image, par exemple en appliquant l'opérateur de Sobel. On interprète alors les valeurs du module du gradient comme des altitudes et l'image des modules du gradient comme une surface. ***Les contours de l'image correspondront alors aux lignes de crête de cette surface.***
- b. Pour détecter ces lignes de crête, nous appliquons l'algorithme dit du Bassin Versant. Pour ce faire, on simule une inondation de la surface. Ses vallées vont être noyées et des bassins vont se former. On remplit ces bassins jusqu'à une hauteur prédéterminée ou seuil S. A la fin de cette opération, seuls les pics les plus hauts émergent, séparant l'eau en plusieurs bassins. Le fait de remplir les vallées jusqu'à la hauteur S nous permet de faire disparaître les sommets les plus bas, c'est-à-dire ceux qui ne représentent pas des contours significatifs. On attribue ensuite à chaque bassin un identificateur unique, par exemple un numéro. La figure (2.4) montre les bassins formés après l'étape d'initialisation de l'algorithme avec un seuil S = 4.

Figure 2.3 Coupe d'une image de gradient

Figure 2.4 Initialisation du Bassin Versant

Une fois les bassins formés et numérotés, une étape du Bassin Versant consiste à faire monter l'eau d'un niveau supplémentaire. Certains points de la surface vont alors être inondés, et les lacs vont s'étaler.

En grossissant, des bassins qui étaient jusqu'alors isolés vont se toucher. Ainsi, si un point est le lieu de réunion de deux bassins portant des numéros distincts, ce point est alors marqué définitivement comme frontière. Il fera partie des contours de l'image résultat.

La figure (2.5) montre l'état de l'algorithme après un remplissage jusqu'au niveau 5. Les bassins 1 et 2 se sont touchés ainsi que les bassins 3 et 4. Les frontières sont apparues aux points de contact. La figure (2.6) présente le résultat de l'algorithme après le septième niveau de remplissage. Les bassins 2 et 3 se sont touchés et ont formé une nouvelle frontière.

L'exécution se termine lorsque tous les points de la surface sont immergés. A ce stade, les points marqués comme frontière entre les bassins constituent les contours de l'image initiale. La figure (2.7) montre l'état des bassins à la fin du calcul.

On remarquera qu'un tel algorithme fournit toujours des contours fermés squelettisés (d'épaisseur 1)

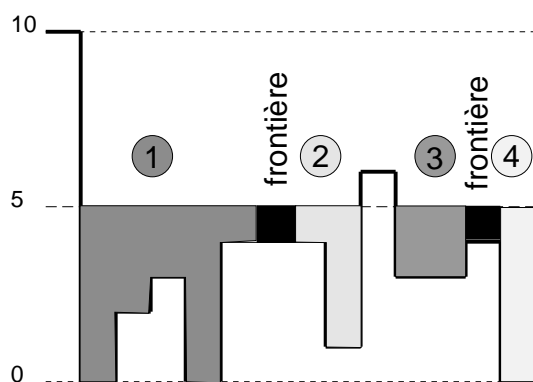


Figure 2.5. Remplissage au niveau 5

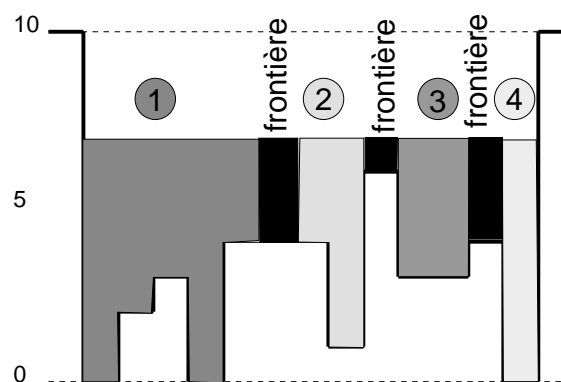


Figure 2.6 Remplissage au niveau 7

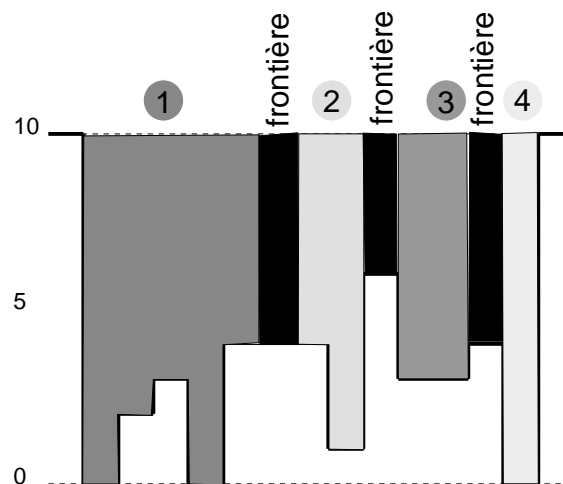


Figure 2.7 Fin du Bassin Versant



Figure 2.8 Image des modules du gradient

Figure 2.9 Image des contours obtenue en appliquant le détecteur du Bassin Versant avec un seuil $S = 24$

5 Attributs tridimensionnels

Chaque pixel est l'image d'un élément de surface — le **surfel** — qui a réfléchi la lumière. On peut donc estimer l'orientation et la profondeur de ce surfel par rapport à la caméra. On obtient ainsi deux nouveaux attributs. La technique la plus classique pour ce faire est la **stéréovision binoculaire** dont le principe de base est la **mise en correspondance** de deux images de la scène vue sous deux angles différents. Nous détaillerons cette technique au chapitre 7..

Section 2 Segmentation et regroupement

1 Segmentation

Une segmentation d'image est une partition de cette image en sous-images ou "régions", connexes, "homogènes" et différentes de ses voisines.

Le but poursuivi est de décrire l'image en des régions qui correspondent aux images des objets ou des parties d'objets présents dans la scène d'où est issue l'image.

Formellement, un ensemble de régions $\{ R_1, R_2, \dots, R_n \}$ est une segmentation de l'image im si et seulement si:

- a. $\bigcup_{i=1}^n R_i = im$
- b. $R_i \cap R_j = \emptyset \quad \forall i, j \quad i \neq j$
- c. R_i est connexe $\forall i$
- d. il existe un prédicat P |
 - I. $P(R_i) = \text{vrai} \quad \forall i$
 - II. $P(R_i \cup R_j) = \text{faux} \quad \forall i, j \quad i \neq j, R_i \text{ et } R_j \text{ étant deux régions adjacentes}$

Le résultat d'une segmentation peut être codée par une image de "labels", chaque pixel portant le numéro de la région à laquelle elle appartient.

2 Regroupement

En général, les différentes régions d'une segmentation ne correspondent pas aux images des objets (resp. des parties significatives d'objets) de la scène mais seulement à des sous-images de ces objets (resp. de leurs parties significatives). Il est donc nécessaire d'opérer des regroupements avant d'aborder l'étape de reconnaissance.

Les lois de regroupement opérés par le système visuel humain ont été essentiellement découvertes par des psychophysiciens regroupés dans l'*école de la Gestalt* (1930). Ces regroupements obéissent essentiellement à des critères de ressemblance et de proximité. Les groupes une fois créés sont à nouveau susceptibles d'être regroupés. On obtient ainsi une description hiérarchisée de l'image.

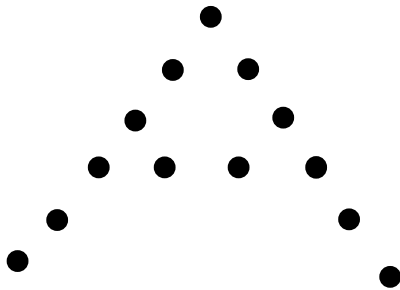


Figure 2.10

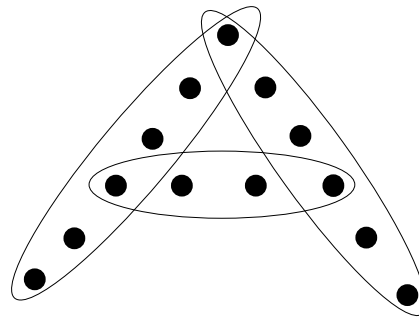


Figure 2.11

Chapitre 3 Extraction des caractéristiques géométriques d'une figure

Section 1 Image numérique

1 Obtention

Pour obtenir une telle représentation, il faut:

- a. Acquérir une image numérique de la forme en question.
Rappelons qu'on doit réaliser les fonctions suivantes:
 - I. transduction : transformation de la grandeur physique — en général l'intensité lumineuse — en un signal susceptible d'être analysé, au moyen d'un capteur.
 - II. échantillonnage de ce signal
 - III. quantification
 - IV. stockage en mémoire.

On dispose alors d'un tableau bidimensionnel de nombres entiers dont chaque élément est un *pixel* ou un *pel* (contraction de picture element).

- b. Séparer la forme du fond.
Concrètement, on doit discerner chaque pixel de la forme de chaque pixel du fond. Par exemple:
 - I. $\text{pixel} = -1$ si le pixel appartient au fond
 - II. $\text{pixel} = \text{"niveau de gris"} (\geq 0)$ si le pixel appartient à la forme
- c. Séparer éventuellement les différentes formes entre elles.
Ce sera le cas lorsque des caractères alphanumériques se toucheront.
- d. Calculer la représentation vectorielle de l'image numérique ainsi obtenue pour obtenir le vecteur des caractéristiques.

En pratique, il arrive souvent que seules les propriétés géométriques de la forme soient significatives. On peut alors représenter la forme par une image binaire sur le principe suivant:

- a. $\text{pixel} = 0$ si le pixel appartient au fond
- b. $\text{pixel} = 1$ le pixel appartient à la forme ou, plus précisément coupe la forme.

En fin de compte, chaque caractéristique correspond à chaque pixel et la mesure reflète:

- a. L'appartenance du pixel à la forme ou au fond.
- b. Eventuellement, l'intensité lumineuse de la forme mesurée en ce pixel.

2 Inconvénients

On vient de représenter l'image d'une forme plane par un vecteur de \mathbb{R}^Q où Q est le nombre de points d'échantillonnage ou de pixels de l'image. Les inconvénients de cette représentation proviennent de:

- La dimension très élevée du vecteur des caractéristiques.
- Sa sensibilité aux transformations géométriques et aux déformations, notamment aux translations.

Pour résoudre ce dernier inconvénient, on effectuera une **normalisation**. Ainsi, pour obtenir une invariance par translation, on pourra positionner le centre de gravité de la forme au centre de l'image.

Section 2 Projections

On considère un axe Δ de direction quelconque. On comptabilise alors pour chaque point M de Δ , d'abscisse x , le nombre $f(x)$ de pixels de la forme qui se projettent orthogonalement en M ; $f(x)$, pour $x_{\min} \leq x \leq x_{\max}$ constitue le vecteur de caractéristiques.

Pour détecter plus précisément les allongements de la forme qui sont perpendiculaires à Δ , on ne fera qu'une projection partielle qui arrêtera le cumul dès la rencontre d'un point du fond (méthode de Pakker).

Section 3 Moments géométriques centrés

Supposons que la forme F contienne N pixels de coordonnées (x_k, y_k) $k = 1, \dots, N$. Les coordonnées du centre de gravité de F sont:

$$x_G = \frac{1}{N} \sum_{k=1}^N x_k \quad y_G = \frac{1}{N} \sum_{k=1}^N y_k \quad (3.1)$$

Le moment centré d'ordre (p, q) est défini par:

$$M_{pq} = \sum_{k=1}^N (x_k - x_G)^p (y_k - y_G)^q i m(x_k, y_k) \quad \forall p, q \in \mathbb{N} \quad (3.2)$$

Il est invariant par translation.

Remarquons que pour une image binaire :

$$M_{pq} = \sum_{k=1}^N (x_k - x_G)^p (y_k - y_G)^q \quad \forall p, q \in \mathbb{N} \quad (3.3)$$

Les moments centrés normalisés m_{pq} définis par :

$$m_{pq} = \frac{M_{pq}}{M_{00}^\gamma} \quad \text{avec} \quad \gamma = \frac{1}{2}(p+q) + 1 \quad (3.4)$$

sont aussi invariants par changement d'échelle.

A partir des moments normalisés, Hu a défini sept coefficients qui sont également invariants par rotation. Ainsi:

$$\begin{aligned} \phi_1 &= m_{20} + m_{02} \\ \phi_2 &= (m_{20} - m_{02})^2 + 4m_{11}^2 \end{aligned} \quad (3.5)$$

Section 4 Descripteurs de Fourier

Etant donnée une fonction périodique de période L , on définit les coefficients a_n et b_n par:

$$\begin{cases} a_0 = \frac{2}{L} \int_0^L f(x) dx \\ a_n = \frac{2}{L} \int_0^L \cos\left(\frac{2\pi nx}{L}\right) f(x) dx \end{cases} \quad (3.6)$$

et

$$b_n = \frac{2}{L} \int_0^L \sin\left(\frac{2\pi nx}{L}\right) f(x) dx \quad (3.7)$$

La fonction f , si elle est continûment dérivable, peut alors être recalculée par:

$$f(x) = \frac{a_0}{2} + \sum_{n=1}^{+\infty} a_n \cos\left(\frac{2\pi nx}{L}\right) + \sum_{n=1}^{+\infty} b_n \sin\left(\frac{2\pi nx}{L}\right) \quad (3.8)$$

En pratique, on associe au contour de la forme une ou plusieurs fonctions périodiques et on calcule les M premiers couples de descripteurs de Fourier. L'erreur due à cette troncature correspond aux harmoniques élevées de f .

1 Descripteur de Fourier de coordonnées

La solution la plus simple est de prendre deux fonctions $f(x)$ et $g(x)$ respectivement égales aux deux coordonnées cartésiennes du point de contour d'abscisse curviligne x . On concatènera ensuite les deux vecteurs obtenus à partir des équations (3.6) et (3.7).

Lorsque le point origine $x = 0$ de la courbe ne peut pas être défini de manière non-ambiguë, il pourra se produire un décalage cyclique sur les valeurs de $f(x)$. Pour s'en préserver, il suffira de remplacer les termes a_n et b_n par le module du nombre complexe $c_n = a_n + i b_n$ qui est invariant dans tout décalage cyclique de $f(x)$.

Pour obtenir une invariance par translation, on devra fixer l'origine du repère par rapport à la forme, en général au centre de gravité de celle-ci.

L'invariance au changement d'échelle sera obtenue en normalisant la longueur du contour, par exemple en faisant $L = 1$.

2 Descripteur de Fourier invariant par rotation

Le descripteur de Fourier de coordonnées n'est pas invariant par rotation. Pour obtenir une telle propriété il faut adopter une fonction f invariante par rotation. Par exemple, si $M(x)$ désigne le point du contour de la forme d'abscisse curviligne x , on pourra prendre

- $f(x)$ = distance de $M(x)$ au centre de gravité de la forme
- $f(x)$ = courbure du contour en $M(x)$
- $f(x)$ = angle entre la tangente en $M(x)$ et l'axe principal d'inertie de la forme.

Section 5 Codages statistiques

On génère aléatoirement des droites dans l'image et on mesure la longueur l_F de l'intersection de la forme F avec la droite. Le vecteur de caractéristiques est alors un échantillonnage de la distribution de probabilité de la variable aléatoire L_F .

$$f(x) = P\{L_F \leq x\} \quad (3.9)$$

Pour que les résultats soient invariants par translation et rotation de la figure, il est nécessaire de choisir avec soin le générateur aléatoire (cf le paradoxe de Bertrand). On peut montrer qu'il n'y a qu'une seule façon de procéder:

- On définit chaque droite par la distance ρ de l'origine du repère-image à cette droite et par son angle θ avec l'axe des x .
- On définit le domaine D de tous les couples (ρ, θ) qui correspondent à des droites qui coupent l'image. D est donc un domaine borné.
- On tire aléatoirement (ρ, θ) selon une distribution uniforme sur D . On calcule la longueur de l'intersection de la droite $d(\rho, \theta)$ avec F .

Section 6 Zones discriminantes

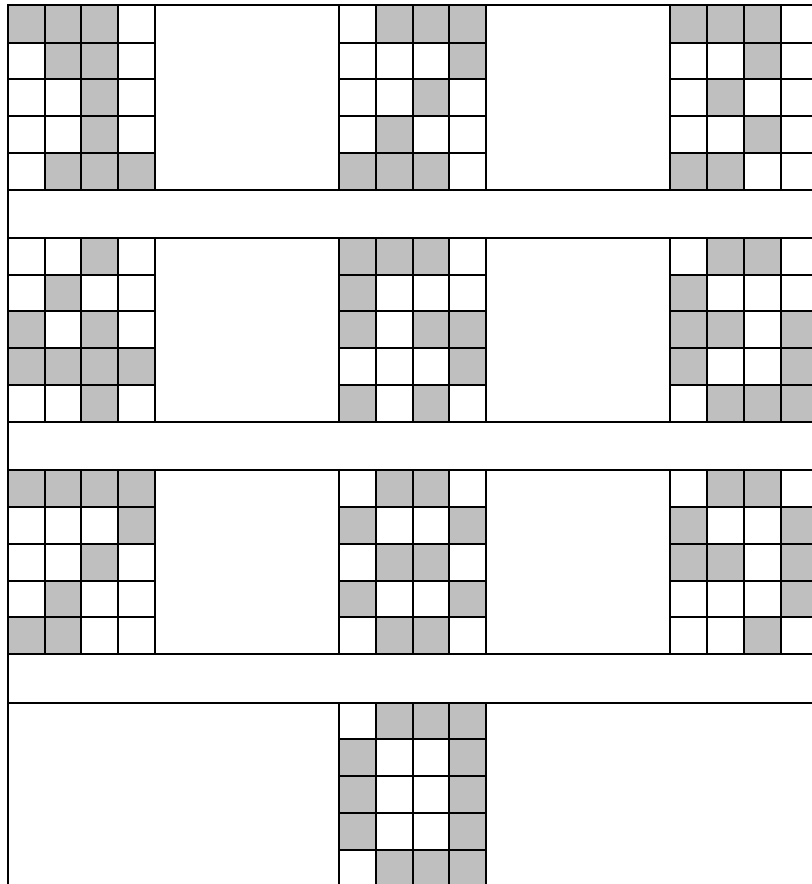


Figure 3.1 Ensemble des modèles

4	9	10	3
5	2	2	5
5	4	6	4
4	3	3	6
4	7	8	3

Figure 3.2 Histogramme

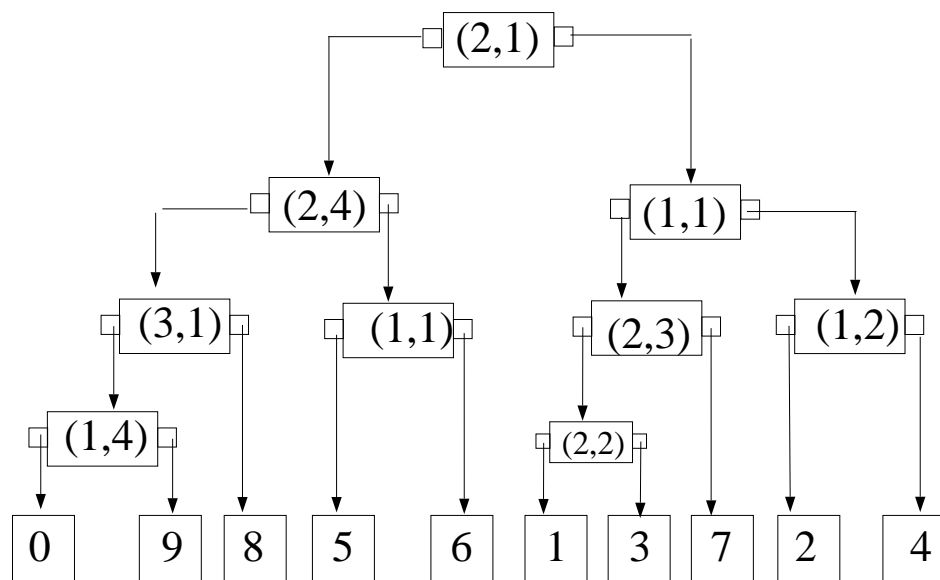


Figure 3.3 Arbre de reconnaissance

Chapitre 4 Extraction de la structure d'une figure

Section 1 Squelette

Considérons la distance d'un point x de la forme F au fond $B = \mathbb{C}F$:

$$d(x, B) = \inf\{d(x, y), y \in B\} \quad (4.1)$$

L'ensemble des points du squelette de F est l'ensemble des points $x \in F$ pour lesquels il existe au moins deux points, $y, z \in B$ vérifiant:

$$d(x, B) = d(x, y) = d(x, z) \quad (4.2)$$

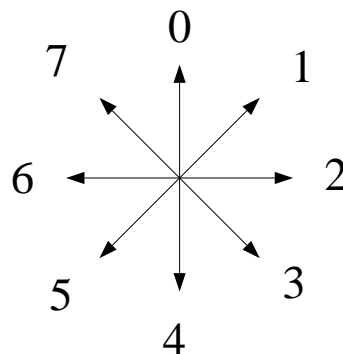
Le squelette peut donner lieu à des branches “parasites” — comme c'est le cas pour un rectangle —. Pour les éliminer différents algorithmes ont été proposés.

Une première solution consiste à opérer par “pelages successifs” en utilisant un critère local mesurant l'intériorité de chaque point de la forme.

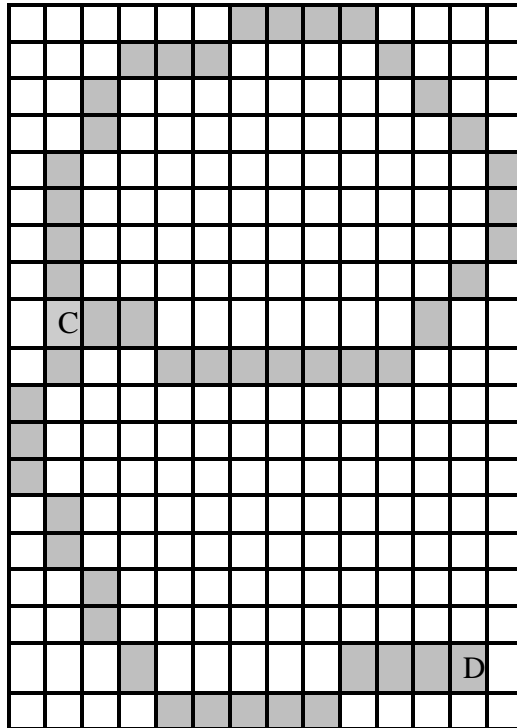
Section 2 Chaîne

En effectuant un suivi des points du squelette, on peut construire une chaîne des directions élémentaires — encore appelée chaîne de Freeman —.

Nous adopterons le codage des directions élémentaires suivant:



Considérons l'exemple suivant:



En partant du point-carrefour C on extrait les deux chaînes de Freeman suivantes:
 0000101221222333344555666666766
 454434343322221222

On peut en déduire un graphe de courbes dont chaque primitive est une courbe sans changement de convexité (ou sans point d'inflexion).

Les noeuds de ce graphe sont:

- les carrefours
- les extrémités
- les points d'inflexion
- les points de discontinuité d'orientation

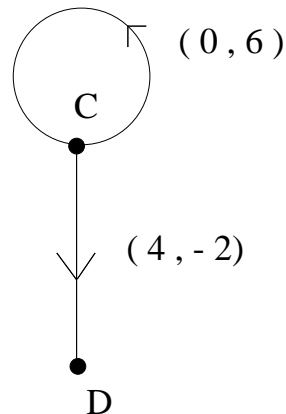
Un arc entre deux noeuds A et B indique la présence d'une courbe sans changement de convexité entre les deux points du squelette correspondant à ces noeuds. Ces arcs sont valués par:

- la direction initiale de la tangente
- l'amplitude de la boucle.

Dans l'exemple précédent, l'analyse de ces deux chaînes fait apparaître deux courbes sans changement de convexité.

- La première correspond à la boucle fermée du "e". La direction initiale de sa tangente est 0 et son amplitude +6 (270 degrés).
- La seconde correspond à la branche inférieure du "e". La direction initiale de sa tangente est 4 et son amplitude -2 (-90 degrés).

Le graphe de courbes est:



Section 3 Graphe de propriétés et de relations

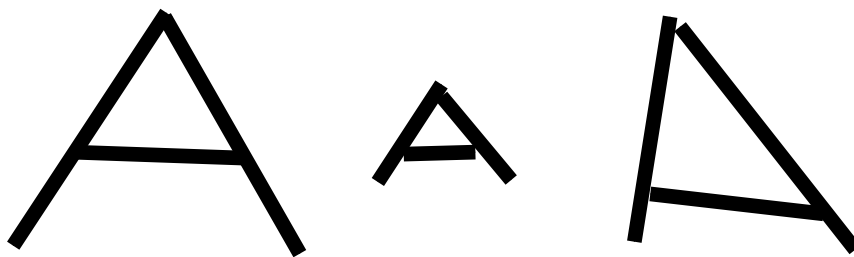
1 Graphe de régions adjacentes

A partir d'une image de labels, il est aisé d'établir le graphe des régions adjacentes dont les noeuds identifient les régions et les arcs, les adjacences entre les régions.

2 Généralisation

Plus généralement, la structure de l'image sera codée par un (ou plusieurs) graphe dont les noeuds représentent des primitives et les arcs différentes relations géométriques entre ces primitives. De plus, on associera à chaque noeud un ensemble de propriétés caractérisant géométriquement et radiométriquement la primitive de ce noeud.

Prenons pour exemple l' *image d'un A majuscule* . Une définition (intuitive) pourrait être: "deux segments de droite obliques se rejoignant vers le haut et reliés par un segment horizontal". Cette définition rend bien compte d'une certaine catégorie de A :



Notant O1, O2 deux primitives obliques et H1 une primitive horizontale, la définition précédente est équivalente au graphe suivant:

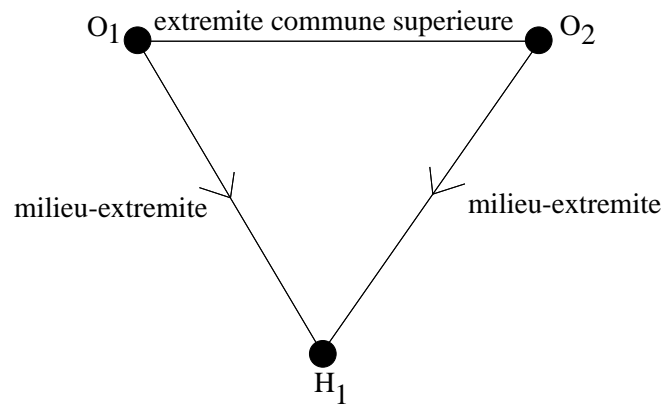
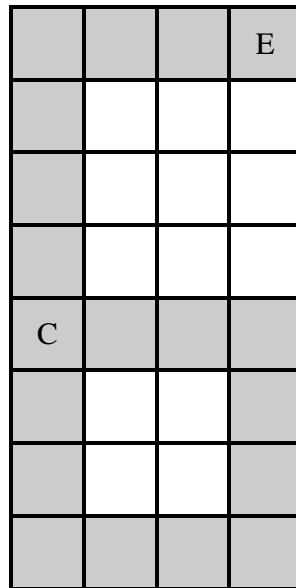


Figure 4.1 Représentation d'un A majuscule

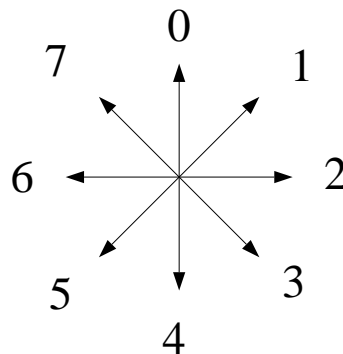
Section 4 Exercice

Soit la forme plane suivante:



1. Extraire les chaînes de Freeman de cette forme plane. Pour ce faire, on respectera les indications suivantes:

- I. On adoptera le codage des directions élémentaires suivant:



- II. L'ordre de consultation des pixels voisins non encore parcourus sera celui donné par le codage ci-dessus.

III. On partira du point carrefour C et on terminera chaque chaîne lorsqu'on rencontrera à nouveau un carrefour (éventuellement le même) ou une extrémité.

2. Regrouper les directions élémentaires en (macro) directions.
3. Soit x et y deux directions. Soit $d = y - x$. On considère la fonction f suivante:

$$\begin{aligned} \text{si } |d| \leq 4 & \quad f(x, y) = d \\ \text{si } d > 4 & \quad f(x, y) = d - 8 \\ \text{si } d < -4 & \quad f(x, y) = d + 8 \end{aligned} \quad (4.3)$$

A quoi sert une telle fonction ?

4. Soit abc une chaîne de trois directions successives. A partir de f , établir un test indiquant s'il se produit ou non un changement de convexité dans cette chaîne. Application aux chaînes 424 , 105 , 670.
5. Soit une chaîne de n directions successives a_1, a_2, \dots, a_n . Donner brièvement un algorithme permettant de segmenter cette chaîne en courbes simples ne possédant pas de changement de convexité. Application à l'exemple.
6. Utilisant la fonction f , donner une méthode permettant d'extraire de chaque courbe simple, la direction initiale de la tangente et l'amplitude de la courbe. Application à l'exemple.
7. Extraire des calculs précédents le graphe de courbes de la forme de l'exemple. Que deviendrait ce graphe si la forme était renversée ?
8. Le graphe de courbes est-il
 - I. invariant par translation ?
 - II. invariant par changement d'échelle ?
 - III. invariant par rotation ?

Chapitre 5 Modélisation géométrique et calibrage d'une caméra

Section 1 Modélisation

1 Caméra en trou d'épingle

Si l'on perce un écran opaque avec une aiguille et qu'on dispose cet écran devant un mur, on réalise ainsi la caméra la plus simple possible. Un objet éclairé placé devant l'écran fournira une image inversée sur le mur; cette image est la projection perspective ou centrale de l'objet dont le centre de projection est le trou de l'épingle — “pinhole” en anglais — (cf figure 5.1).

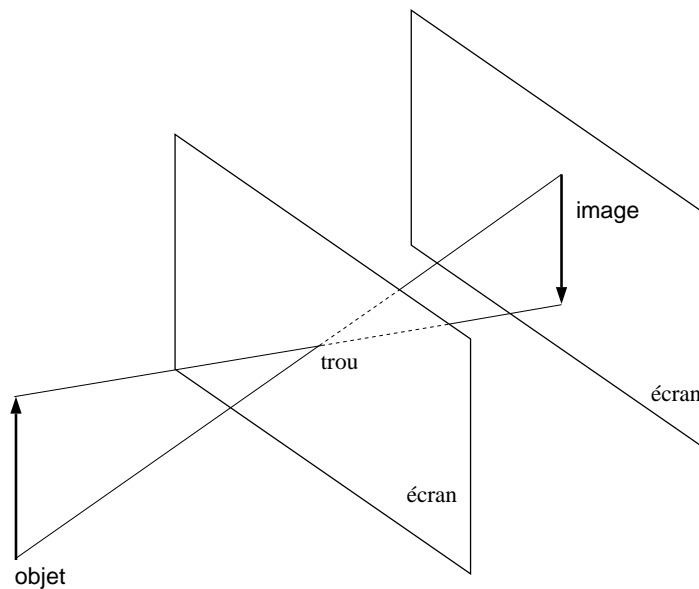


Figure 5.1 Caméra en trou d'épingle

Section 2 Modélisation en trou d'épingle

Le modèle perspectif de la caméra en trou d'épingle est parfaitement défini une fois connue la distance du trou au mur-écran. On désignera par longueur focale une telle distance et on la notera f ; elle est souvent exprimée en millimètre.

On désignera aussi par:

- centre optique L , le centre de projection, ici le trou d'épingle
- axe optique, la droite perpendiculaire au plan image IM — ici, le mur-écran — passant par le centre optique L .

Un point M de l'espace 3D se projette dans le plan image en m le long d'une droite passant par L et M (cf fig. 5.2)

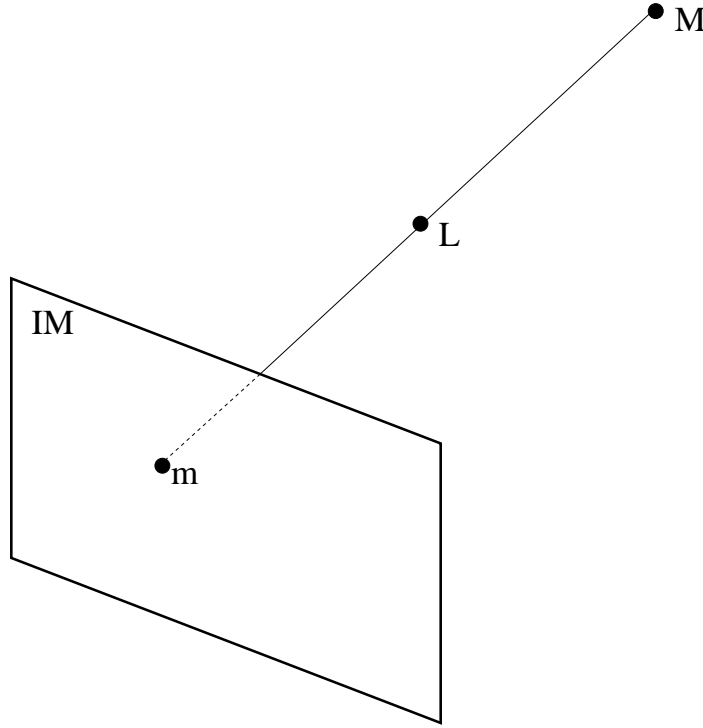


Figure 5.2 Modèle de la caméra trou d'épingle

Les points L , m et M étant alignés :

$$\exists \lambda < 0 \mid \vec{Lm} = \lambda \vec{LM} \quad (5.1)$$

Choisissons un repère orthonormé direct $(Lx_c y_c z_c)$ attaché à la caméra. Plaçons son origine en L, le plan $Lx_c y_c$ parallèle au plan image et l'axe des z_c confondu avec l'axe optique tourné vers l'objet.

Soit (x_c, y_c, z_c) les coordonnées de M dans ce repère. D'après (5.1), m a comme coordonnées $(\lambda x_c, \lambda y_c, \lambda z_c = -f)$. Les coordonnées de m sont donc:

$$\begin{aligned} x'_c &= -\frac{f x_c}{z_c} \\ y'_c &= -\frac{f y_c}{z_c} \\ z'_c &= -f \end{aligned} \quad (5.2)$$

Adoptant les coordonnées homogènes, on peut écrire cette transformation sous la forme matricielle suivante:

$$\begin{bmatrix} wx'_c \\ wy'_c \\ wz'_c \\ w \end{bmatrix} = \begin{bmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & f & 0 \\ 0 & 0 & -1 & 0 \end{bmatrix} \begin{bmatrix} x_c \\ y_c \\ z_c \\ 1 \end{bmatrix} \quad (5.3)$$

Pour éviter le signe négatif de -1 , on peut très bien supposer que l'image se forme en avant et non en arrière du centre optique (cf fig. 5.3). On a maintenant :

$$\begin{bmatrix} wx'_c \\ wy'_c \\ w \end{bmatrix} = \tilde{P} \begin{bmatrix} x_c \\ y_c \\ z_c \\ 1 \end{bmatrix} \quad (5.4)$$

$$\text{avec } \tilde{P} = \begin{bmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}$$

2 Passage à l'image numérique

Le dispositif de lecture de la caméra doit échantillonner l'image en pixels. Ces pixels sont en général rectangulaires et alignés sur des lignes et des colonnes orthogonales. Rappelons qu'un pixel est repéré par son numéro de ligne et par son numéro de colonne, la première ligne se situant en haut de l'image numérique et la première colonne à gauche. Aussi l'origine Ω du repère image (Ωuv) est-elle située dans le coin supérieur gauche de l'image. Pour repérer le point image m en unités pixels dans le repère image (Ωuv), on doit connaître les quatre autres paramètres suivants:

- e_u : nombre de pixels par unité de mesure (mm par exemple) sur une colonne; e_u^{-1} représente donc la dimension verticale d'un pixel
- e_v : nombre de pixels par unité de mesure sur une ligne; e_v^{-1} représente donc la dimension horizontale d'un pixel
- (p_u, p_v) : coordonnées en pixels dans (Ωuv) , de la projection orthogonale P de L sur le plan image IM. P est appelé le **point principal**.

Pour ce faire, précisons le repère de la caméra ($Lx_c y_c z_c$) en prenant l'axe Lx_c parallèle aux colonnes de pixels donc à l'axe Ωu , l'axe Ly_c dirigé suivant le vecteur $\vec{Ly}_c = \vec{Lz}_c \wedge \vec{Lx}_c$: cet axe est donc parallèle aux lignes de pixels, donc à l'axe Ωv sous réserve que les lignes soient perpendiculaires aux colonnes, ce qui est le plus souvent vérifié.

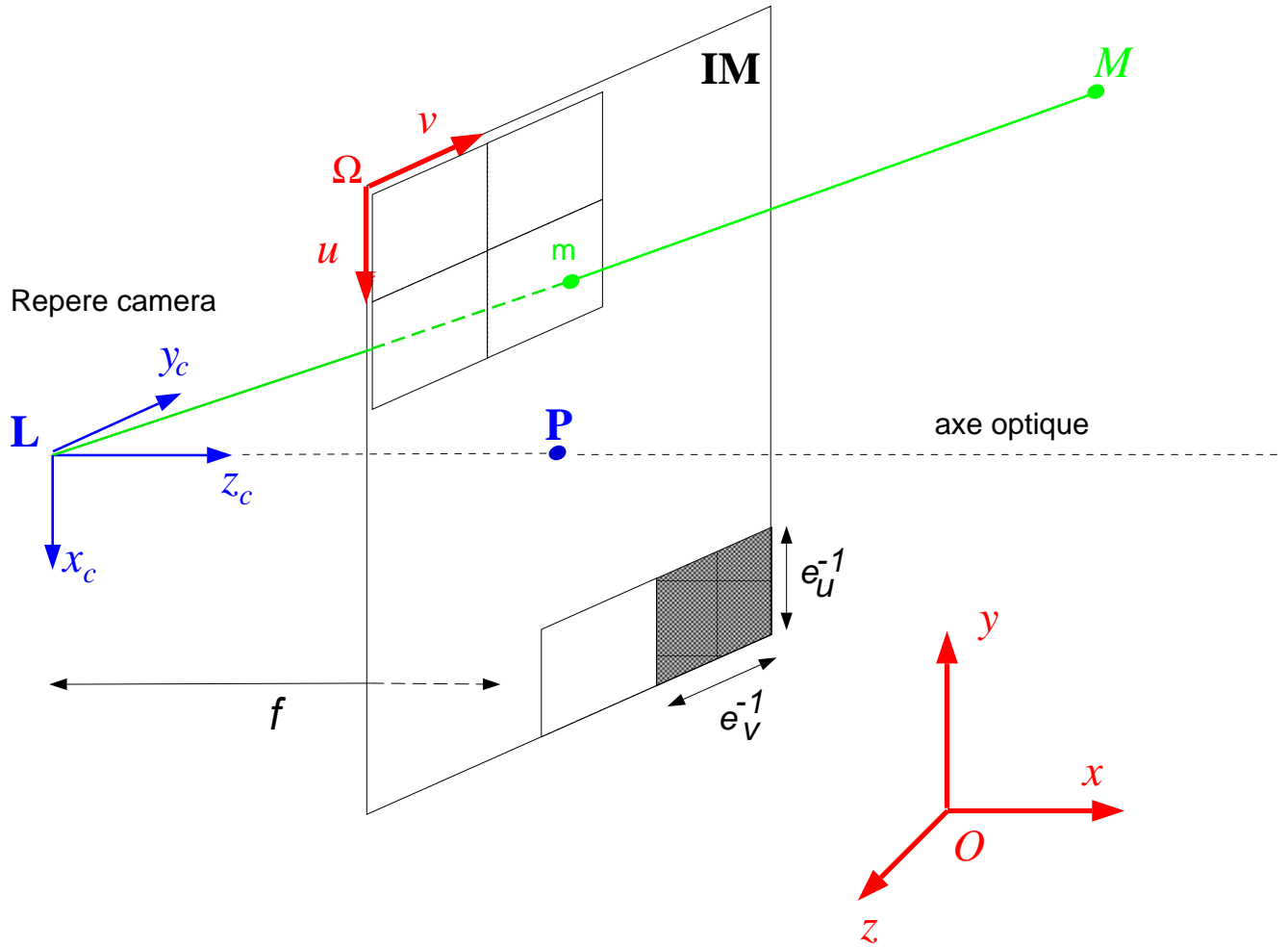


Figure 5.3 Modèle pinhole avant

Soit (u, v) les coordonnées en pixels de m . Puisque

$$\vec{\Omega m} = \vec{\Omega P} + \vec{Pm} = \vec{\Omega P} + \vec{Lm} - \vec{LP} \quad (5.5)$$

en exprimant les coordonnées de ces vecteurs dans le repère image, il vient:

$$\begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} p_u \\ p_v \end{bmatrix} + \begin{bmatrix} \frac{x'_c}{e_u^{-1}} \\ \frac{y'_c}{e_v^{-1}} \end{bmatrix} = \begin{bmatrix} p_u \\ p_v \end{bmatrix} + \begin{bmatrix} e_u & 0 \\ 0 & e_v \end{bmatrix} \begin{bmatrix} x'_c \\ y'_c \end{bmatrix} \quad (5.6)$$

Exprimé en coordonnées homogènes:

$$\begin{bmatrix} wu \\ wv \\ w \end{bmatrix} = \tilde{H} \begin{bmatrix} wx'_c \\ wy'_c \\ w \end{bmatrix} \quad (5.2)$$

où $\tilde{H} = \begin{bmatrix} e_u & 0 & p_u \\ 0 & e_v & p_v \\ 0 & 0 & 1 \end{bmatrix}$

Utilisant (5.4), il vient:

$$\begin{bmatrix} wu \\ wv \\ w \end{bmatrix} = \tilde{H} \tilde{P} \begin{bmatrix} x_c \\ y_c \\ z_c \\ 1 \end{bmatrix} \quad (5.3)$$

$$\text{avec } \tilde{H}\tilde{P} = \begin{bmatrix} fe_u & 0 & p_u & 0 \\ 0 & fe_v & p_v & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}$$

Ces équations fournissent la transformation qui à un point M de la scène repéré dans le repère caméra fait correspondre son image m repéré dans le repère de l'image: elles représentent le modèle géométrique direct de la caméra exprimé dans son repère interne ($Lx_c y_c z_c$). les quatre paramètres qui interviennent $f e_u$, $f e_v$, p_u , p_v sont appelés paramètre internes (ou intrinsèques) de la caméra en trou d'épingle.

3 Passage au repère externe

Dans de nombreuses applications, le repère "utile" est extérieur à la caméra. C'est le cas dans une application de saisie d'objets par un robot sous contrôle visuel où le repère utile est alors le repère du robot. C'est encore le cas dans la navigation d'un robot qui doit se situer à l'intérieur de la scène qu'il traverse : le repère utile est alors celui de la scène. Un point M de la scène doit donc être repéré dans le repère externe "utile" (Oxyz). Pour établir le modèle direct nous avons besoin de connaître:

- les coordonnées (l_x , l_y , l_z) en unités de mesure (mm par exemple) du centre optique L dans (Oxyz)
- la matrice de rotation R qui permet le passage du repère caméra au repère externe.

(l_x , l_y , l_z) et R^T fournissent donc la position et l'orientation de la caméra dans le repère externe. R étant définie par un axe de rotation Δ et un angle de rotation Φ , les paramètres (l_x , l_y , l_z , Δ , Φ) constituent les paramètres externes ou extrinsèques de la caméra. Puisque Δ est défini par deux paramètres, ils sont au nombre de six.

Déterminons à présent le modèle direct.

Soit (x,y,z) les coordonnées de M dans le repère externe. Puisque

$$L\vec{M} = L\vec{O} + O\vec{M} \quad (5.4)$$

les coordonnées de $L\vec{M}$ dans le repère externe sont ($x-l_x$, $y-l_y$, $z-l_z$) et (x_c, y_c, z_c) dans le repère caméra. On a donc:

$$\begin{bmatrix} x_c \\ y_c \\ z_c \end{bmatrix} = R \begin{bmatrix} x - l_x \\ y - l_y \\ z - l_z \end{bmatrix} = R \begin{bmatrix} 1 & 0 & 0 & -l_x \\ 0 & 1 & 0 & -l_y \\ 0 & 0 & 1 & -l_z \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} \quad (5.5)$$

soit encore:

$$\begin{bmatrix} x_c \\ y_c \\ z_c \\ 1 \end{bmatrix} = \begin{bmatrix} R & 0_3 \\ 0_3 & 1 \end{bmatrix} \begin{bmatrix} I & -l \\ 0_3 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} \quad (5.6)$$

Ainsi:

$$\begin{bmatrix} x_c \\ y_c \\ z_c \\ 1 \end{bmatrix} = \tilde{K} \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} \quad (5.7)$$

avec

$$\tilde{K} = \begin{bmatrix} R & t \\ 0_3^T & 1 \end{bmatrix} \quad (5.8)$$

$$\text{et } t = \begin{pmatrix} t_x \\ t_y \\ t_z \end{pmatrix} = -R \begin{pmatrix} l_x \\ l_y \\ l_z \end{pmatrix}$$

Utilisant (5.3) et (5.7), il vient:

$$\begin{bmatrix} wu \\ wv \\ w \end{bmatrix} = \tilde{C} \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} \quad (5.9)$$

où $\tilde{C} = \tilde{H}\tilde{P}\tilde{K}$ représente le modèle géométrique direct de la caméra qui est donc linéaire en coordonnées homogènes.

$$\tilde{C} = \begin{bmatrix} \alpha_u r_1^T + p_u r_3^T & \alpha_u t_x + p_u t_z \\ \alpha_v r_2^T + p_v r_3^T & \alpha_v t_y + p_v t_z \\ r_3^T & t_z \end{bmatrix} \quad (5.10)$$

où $r_i^T = (r_{i1}, r_{i2}, r_{i3})$ $i = 1, 2, 3$ est le i -ème vecteur ligne de R , $l = (l_x, l_y, l_z)^T$, $t_x = -r_1^T l$, $t_y = -r_2^T l$, $t_z = -r_3^T l$.

Remarque: le rang de \tilde{C} est égal à 3, puisque α_u et α_v étant non nuls, $\alpha_u r_1^T + p_u r_3^T$, $\alpha_v r_2^T + p_v r_3^T$ et r_3^T sont linéairement indépendants.

4 Extension projective

On peut étendre la tranformation du monde 3D à l'image 2D réalisée par une caméra en trou d'épingle aux espaces projectifs P^3 et P^2 qui les englobent et qui contiennent aussi des "points à l'infini".

En effet, soit (X, Y, Z, T) les coordonnées projectives de $\tilde{M} \in P^3$ dans la base projective issue du repère $(Oxyz)$ et (U, V, W) celles de son image $\tilde{m} \in P^2$ dans la base projective issue du repère (Ωuv) , d'après (5.9), on a

$$\begin{bmatrix} U \\ V \\ W \end{bmatrix} = \tilde{C} \begin{bmatrix} X \\ Y \\ Z \\ T \end{bmatrix} \quad (5.11)$$

5 Interprétation des lignes de \tilde{C} et détermination du centre optique

Soit \tilde{c}_i^T le i -ème vecteur ligne de \tilde{C} $i=1,2,3$.

$\{\tilde{M} \mid \tilde{c}_1^T \tilde{M} = 0\} = \{\tilde{M} \mid U = 0\}$ est le plan Π_1 contenant les droites $L\Omega$ et $\Omega\vec{v}$

$\{\tilde{M} \mid \tilde{c}_2^T \tilde{M} = 0\} = \{\tilde{M} \mid V = 0\}$ est le plan Π_2 contenant les droites $L\Omega$ et $\Omega\vec{u}$

$\{\tilde{M} \mid \tilde{c}_3^T \tilde{M} = 0\} = \{\tilde{M} \mid W = 0\} = \{\tilde{M} \mid Z_c = 0\}$ est le plan focal Π_3 , parallèle au plan image et passant par le centre optique L .

On a donc : $\Pi_1 \cap \Pi_2 = \text{droite } L\Omega$ et $\Pi_1 \cap \Pi_2 \cap \Pi_3 = \{L\} = \left\{ \widetilde{M} \mid \widetilde{C}\widetilde{M} = 0 \right\}$ et on en déduit que:

$$\widetilde{C} \begin{pmatrix} l_x \\ l_y \\ l_z \\ 1 \end{pmatrix} = 0 \quad (5.12)$$

Soit C la matrice carrée (3,3) formée à partir des trois premières colonnes de \widetilde{C} et \widetilde{c}^4 la quatrième colonne de \widetilde{C} : $\widetilde{C} = [C \ \widetilde{c}^4]$

L'équation (5.12) se réécrit:

$$C \begin{pmatrix} l_x \\ l_y \\ l_z \end{pmatrix} + \widetilde{c}^4 = 0 \quad (5.13)$$

C étant inversible:

$$\begin{pmatrix} l_x \\ l_y \\ l_z \end{pmatrix} = -C^{-1}\widetilde{c}^4 \quad (5.14)$$

Cette équation fournit les coordonnées du centre optique dans le repère utile.

6 Modèle inverse

Le rayon de projection associée à m est la droite qui passe par L et m . Cette droite contient aussi son point à l'infini \widetilde{D} dont les coordonnées projectives sont $(D, 0)^T$. Ce point a pour image m donc: $\widetilde{m} = \widetilde{C}\widetilde{D} = CD$ et $D = C^{-1}\widetilde{m}$. D fournit la direction du rayon de projection. Un point courant M du rayon de projection s'écrit:

$$M = L + \lambda \vec{Lm} \quad \lambda > 0 \quad (5.15)$$

On en déduit l'équation du rayon de projection:

$$M = C^{-1}(-\widetilde{c}^4 + \mu \widetilde{m}) \quad \mu > 0 \text{ ou } \mu < 0 \quad (5.16)$$

Grâce au modèle inverse, un point image m étant donné, on peut déterminer grâce à (5.16) l'équation de la demi-droite de projection qui contient tous les points 3D M qui se projettent en m .

Section 3 Calibrage géométrique

Calibrer géométriquement "fortement" une caméra revient à déterminer son modèle géométrique direct. Pour une caméra en trou d'épingle, ceci revient à déterminer la matrice \widetilde{C} . Dans certaines applications (navigation de robots), il sera aussi utile de déterminer les paramètres internes et externes de la caméra.

En pratique, pour calibrer, on a besoin d'une mire de calibrage et de points de calibrage matérialisés sur cette mire. Plusieurs solutions sont envisageables: en général,

les points de calibrage sont les noeuds d'une grille ou les centres de disques disposés régulièrement sur un dièdre (solution INRIA) ou à l'intérieur d'un parallélépipède (solution N7). On déterminera alors avec la plus grande précision possible les coordonnées 3D des N points de calibrage M_i ainsi que les coordonnées 2D de leurs images. Ces mesures permettront d'obtenir \tilde{C} .

Meilleur sera le calibrage, meilleure aussi sera la correspondance entre les objets et leurs images, et en fin de compte, meilleure sera l'estimation de l'emplacement de ces objets à partir de leurs images.

Dans une application robotique de manipulation d'objets ou de navigation, la localisation des objets doit souvent être suffisamment précise et en conséquence le calibrage des caméras du robot doit faire l'objet d'un soin particulier.

1 Méthode linéaire de calibrage

Cette méthode s'appuie sur le théorème suivant:

Soit $\tilde{C} = \begin{bmatrix} c_1^T & c_{14} \\ c_2^T & c_{24} \\ c_3^T & c_{34} \end{bmatrix}$ une matrice (3,4) de rang 3.

Alors \tilde{C} modélise une caméra en trou d'épingle si et seulement si:

$$\begin{aligned} \|c_3\| &= 1 \\ (c_1 \wedge c_3 \mid c_2 \wedge c_3) &= 0 \end{aligned} \quad (5.17)$$

Remarque

Soit θ l'angle entre les lignes et les colonnes de l'image ($\theta \in]0, \pi[$) alors

$$\cos \theta = - \frac{(c_1 \wedge c_3 \mid c_2 \wedge c_3)}{\|c_1 \wedge c_3\| \|c_2 \wedge c_3\|} \quad (5.18)$$

Donc (5.17) $\Leftrightarrow \theta = \frac{\pi}{2} \Leftrightarrow$ lignes et colonnes sont orthogonales.

Soit M_i de coordonnées (x_i, y_i, z_i) les points de calibrage et soit m_i leurs images de coordonnées (u_i, v_i)

$$\tilde{C}M_i = m_i \Leftrightarrow \begin{cases} \frac{c_1^T M_i + c_{14}}{c_3^T M_i + c_{34}} = u_i \\ \frac{c_2^T M_i + c_{24}}{c_3^T M_i + c_{34}} = v_i \end{cases} \quad (5.19)$$

soit encore:

$$\tilde{C}M_i = m_i \Leftrightarrow \begin{cases} c_1^T M_i - u_i c_3^T M_i + c_{14} - u_i c_{34} = 0 \\ c_2^T M_i - v_i c_3^T M_i + c_{24} - v_i c_{34} = 0 \end{cases} \quad (5.20)$$

qui sont deux équations linéaires homogènes par rapport aux inconnues c_{ij}

Pour N points de calibrage, on obtient donc 2N équations linéaires homogènes:

$$Ac = 0 \quad (5.21)$$

où $A(2N, 12)$ dépend des coordonnées 3D des M_i et des coordonnées 2D des m_i et $c = [c_1^T, c_{14}, c_2^T, c_{24}, c_3^T, c_{34}]^T$: c est donc défini à un facteur d'échelle près.

Nous avons besoin de savoir maintenant à quelle condition, il y a une solution unique non nulle (à un facteur d'échelle près) au système linéaire (5.21). Puisque $rg(A) + \dim(Ker A) = 12$, cette condition est équivalente à $rg(A) = 11$. Le théorème montre qu'en général cette condition est remplie si $N \geq 6$ et si les points de calibrage ne sont pas coplanaires.

- a. Si $N \geq 6$, en général, $rg(A) = 11$
 b. Si $N \geq 6$ et si les points sont coplanaires, en général $rg(A) = 8$

La méthode de résolution de (5.21) consiste à résoudre l'un des deux problèmes d'optimisation avec contraintes suivant:

$$(P_1) \begin{cases} \text{Min } \|Ac\|^2 \\ \|c_3\|^2 = 1 \end{cases} \quad (5.22)$$

ou bien

$$(P_2) \begin{cases} \text{Min } \|Ac\|^2 \\ \|c_3\|^2 = 1 \\ (c_1 \wedge c_3 \mid c_2 \wedge c_3) = 0 \end{cases} \quad (5.23)$$

D'après les deux théorèmes précédents, si $N \geq 6$, (P_1) et (P_2) ont (en général) une solution unique, les points de calibrage ne devant pas être coplanaires.

La solution de (P_1) peut être explicitement calculée. Pour le montrer, nous avons besoin du théorème suivant:

Soit

$$(P) \begin{cases} \text{Min } \|Ax\|^2 \\ \|Cx\|^2 = 1 \end{cases} \quad (5.24)$$

où A est une matrice (N,M) et C est une matrice (P,N) qui échantillonne certaines coordonnées du vecteur x . Soit $z = Cx$ et y le $(M-P)$ vecteur composé des coordonnées restantes de x . Soit B $(N,M-P)$ et D (N,P) les sous-matrices de A correspondant à y et z : $Ax = By + Dz$

alors

posant $E = D^T D - D^T B (B^T B)^{-1} B^T D$, E est une matrice semi-définie positive — $(Ex \mid x) \geq 0 \quad \forall x$ — et toute solution de (P) \bar{x} est telle que:

- a. \bar{z} est un vecteur propre unitaire associé à la plus petite valeur propre de E .
 b. $\bar{y} = -(B^T B)^{-1} B^T D \bar{z}$

Appliquons ce théorème à (P_1) . $z = c_3$ est le vecteur obtenu en extrayant les 9-ème, 10-ème et 11-ème coordonnée de c . B est donc la $(2N,9)$ sous-matrice de A coorespondant à ses huit premières et à sa dernière colonne, D est la $(2N,3)$ sous-matrice de A correspondant aux trois autres colonnes. c_3 est donc un vecteur propre unitaire de la matrice $(3,3)$ symétrique définie positive $D^T (I - B(B^T B)^{-1} B^T) D$ associé à sa plus petite valeur propre. Ce calcul nécessite d'inverser la matrice $(9,9)$ $B^T B$. Les neuf autres coordonnées de c sont égales à $\bar{y} = -(B^T B)^{-1} B^T D c_3$

On remarquera que cette méthode de calibrage aboutit à des calculs très simples.

Chapitre 6 Photométrie et stéréo

Dans ce chapitre, nous nous intéressons aux relations qui existent entre la distribution spatiale des niveaux de gris de l'image d'un objet et la physique et la géométrie de la surface de cet objet.

Section 1 Définitions fondamentales

1 Eclairement

L'**éclairement énergétique** ou **irradiance** E en un point d'une surface est la valeur du rapport de la puissance ou flux énergétique à l'aire réceptrice. Elle s'exprime en watts par mètre carré (W m^{-2}).

$$E = \frac{dP}{dS} \quad (6.1)$$

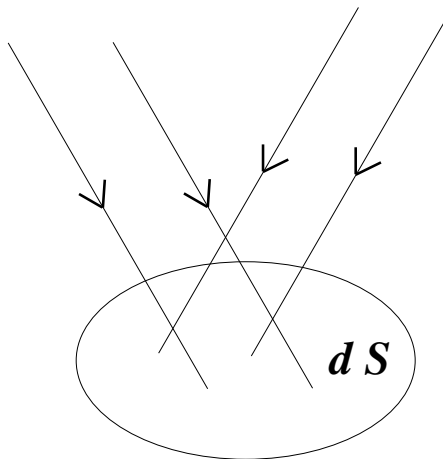


Figure 6.1 Eclairement

2 Intensité

L' **intensité lumineuse** I d'une source ponctuelle O dans une direction Δ est égale au rapport de la puissance dP émise par la source à l'intérieur d'un cône infiniment petit de sommet O contenant Δ , à l'angle solide $d\Omega$ de ce cône.

$$I = \frac{dP}{d\Omega} \quad (6.2)$$

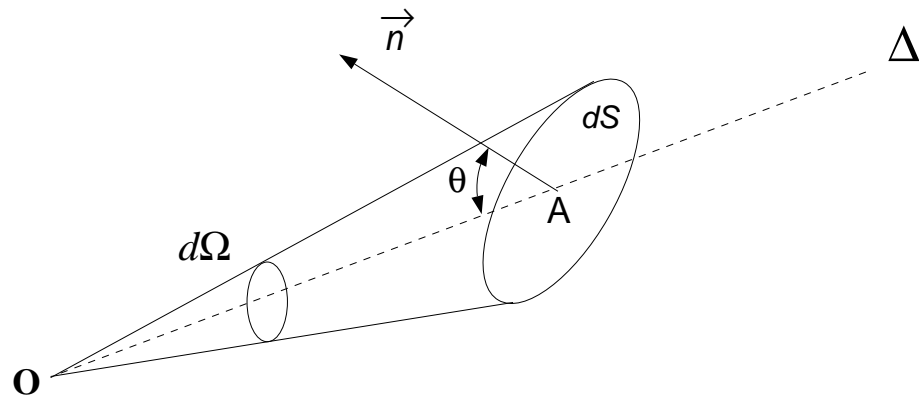


Figure 6.2 Intensité

3 Luminance

La **luminance énergétique** ou **radiance** L dans une direction Δ d'un élément de surface entourant un point O est égale au rapport de l'intensité dI de l'élément dans la direction Δ à sa surface raccourcie dans cette direction $d\Sigma \cos \alpha$. Elle s'exprime en watts par mètre carré par stéradian ($\text{W m}^{-2} \text{sr}^{-1}$)

$$L = \frac{dI}{d\Sigma \cos \alpha} \quad (6.3)$$

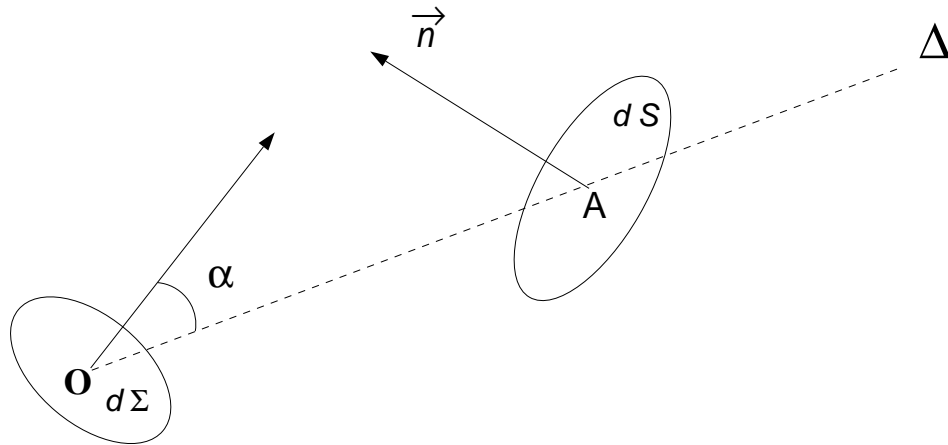


Figure 6.3 Luminance

4

Section 2 Relation entre luminance et éclairement

L'éclairement dE de l'élément de surface A d'aire dS par l'élément de surface O d'aire $d\Sigma$ vaut d'après 6.1 et 6.2:

$$dE = \frac{dI d\Omega}{dS} \quad (6.4)$$

Or

$$d\Omega = \frac{dS \cos \theta}{z^2} \quad (6.5)$$

où z représente la distance de O à A (voir la figure 6.2) et

$$dI = L d\Sigma \cos \alpha \quad (6.6)$$

d'après 6.3. Reportant 6.5 et 6.6 dans 6.4, il vient:

$$dE = \frac{L d\Sigma \cos \alpha \cos \theta}{z^2} \quad (6.7)$$

Or

$$d\psi = \frac{d\Sigma \cos \alpha}{z^2} \quad (6.8)$$

n'est autre que l'angle solide sous lequel $d\Sigma$ est vu de A (voir la figure 6.4).

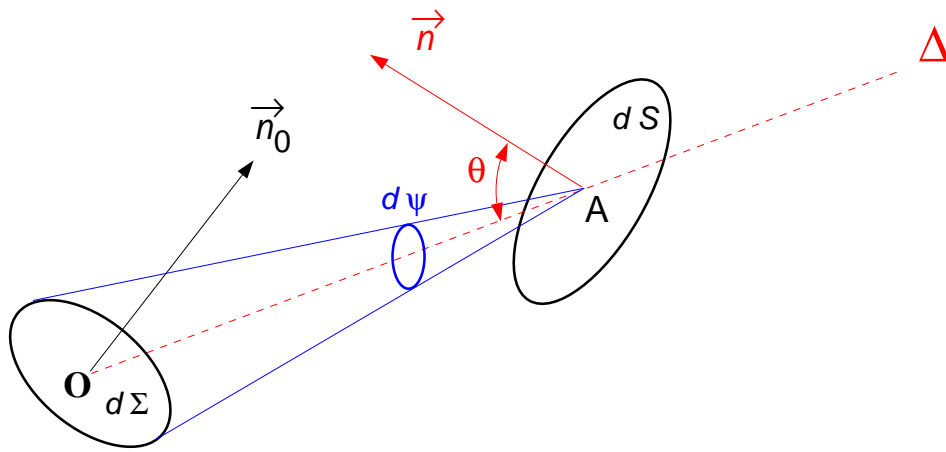


Figure 6.4

Finalement :

$$dE = L \cos \theta d\psi \quad (6.9)$$

Nous pouvons maintenant établir la relation qui existe entre la luminance d'un point de la surface d'un objet et l'éclairement de l'image de ce point. Considérons une lentille de diamètre d à une distance f du plan image. Soit un élément de surface de l'objet d'aire dob dont l'image a une aire dim . Soit α l'angle que fait le rayon joignant l'élément dob au centre de la lentille avec l'axe optique (voir figure 6.5). On a alors la relation:

$$E = L \frac{\pi}{4} \left(\frac{d}{f} \right)^2 \cos^4 \alpha \quad (6.10)$$

Il y a donc proportionnalité entre éclaïrement et luminance.

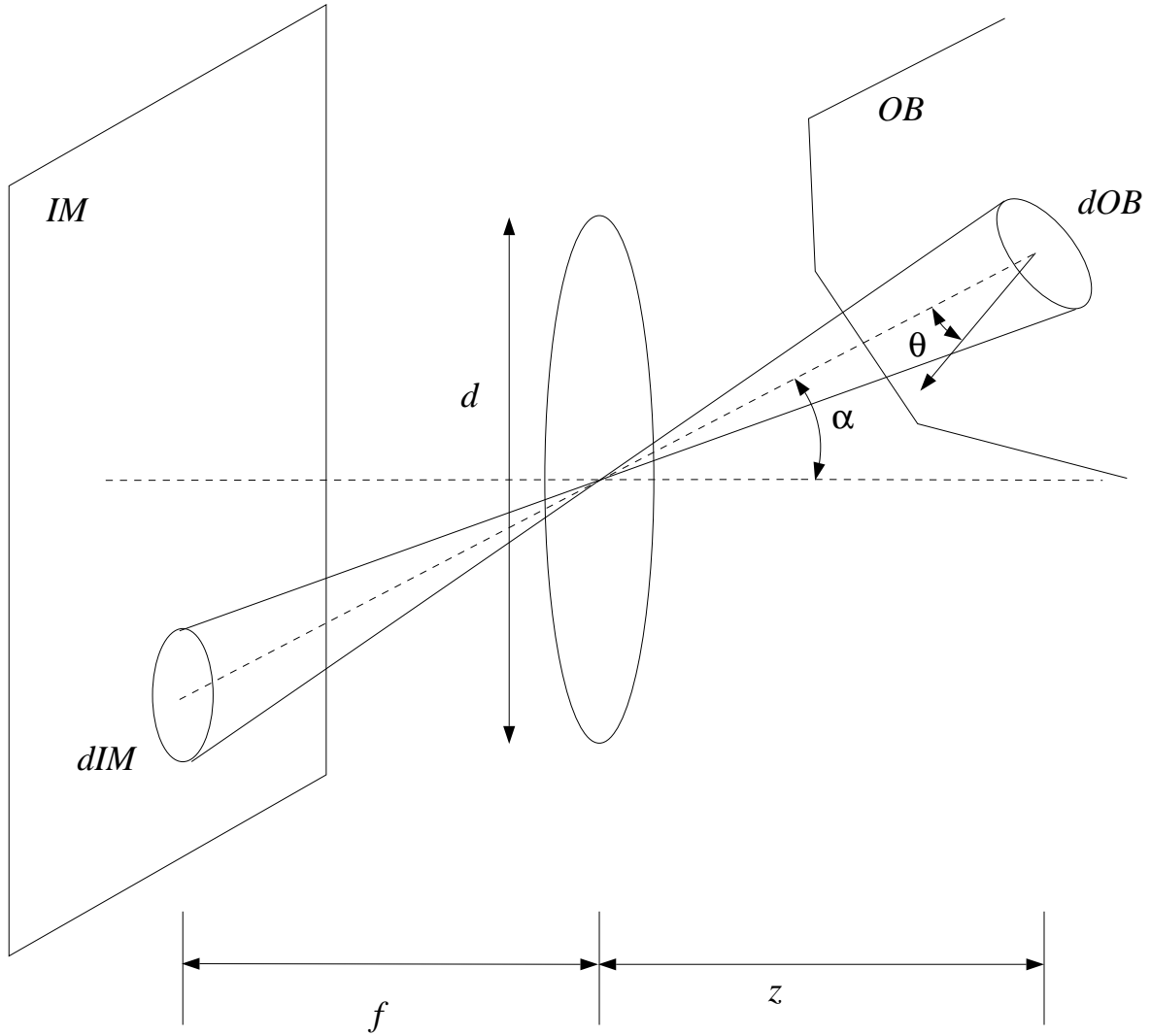


Figure 6.5

Démonstration

$$E = \frac{dP}{dIM} \quad (6.11)$$

où dP est la puissance issue de dOB et traversant la lentille. Donc :

$$dP = dI \Omega = L dOB \cos \theta \Omega \quad (6.12)$$

où Ω est l'angle solide sous lequel on voit la lentille à partir de dOB . Or

$$\Omega = \frac{\pi}{4} \frac{d^2}{\left(\frac{z}{\cos \alpha}\right)^2} \cos \alpha = \frac{\pi}{4} \frac{d^2}{z^2} \cos^3 \alpha \quad (6.13)$$

D'où :

$$E = L \frac{dOB}{dIM} \frac{\pi}{4} \left(\frac{d}{z}\right)^2 \cos^3 \alpha \cos \theta \quad (6.14)$$

Comme l'angle solide sous lequel on voit dIM du centre de la lentille est égal à celui sous lequel on voit dOB du centre de la lentille :

$$\frac{dIM \cos \alpha}{\left(\frac{f}{\cos \alpha}\right)^2} = \frac{dOB \cos \theta}{\left(\frac{z}{\cos \alpha}\right)^2} \quad (6.15)$$

On en déduit:

$$\frac{dOB}{dIM} = \frac{\cos \alpha}{\cos \theta} \left(\frac{z}{f}\right)^2 \quad (6.16)$$

D'où:

$$E = L \frac{\pi}{4} \left(\frac{d}{f}\right)^2 \cos^4 \alpha \quad (6.17)$$

■

Effet de **vignettage**: cet effet provoque un affaiblissement de l'éclairement lorsqu'on s'éloigne du centre de l'image. Il est d'autant plus important que le diaphragme est ouvert.

Section 3 Réflectance

1 Repère local

La direction d'un rayon lumineux frappant un élément de surface ou réfléchi par lui, peut être déterminé dans un repère local à cet élément grâce à l'angle polaire $\theta \in [0, \frac{\pi}{2}]$ et à l'azimut $\phi \in [-\pi, \pi]$.

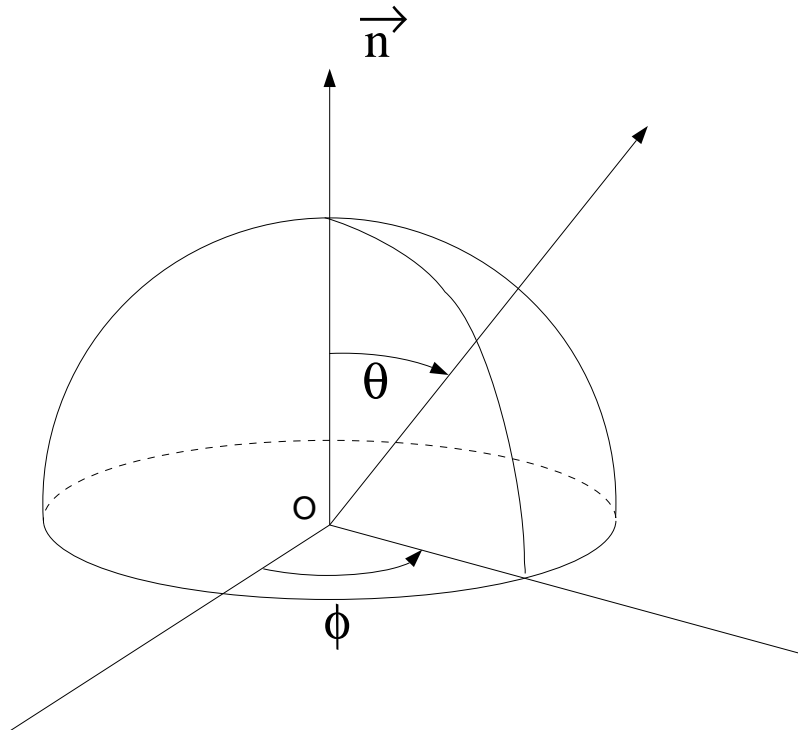


Figure 6.6

2 Fonction de réflectance bidirectionnelle

Soit:

- $dE(\theta_i, \phi_i)$ l'éclairement d'un élément de surface en provenance d'une direction (θ_i, ϕ_i)
- $dL(\theta_e, \phi_e)$ la luminance de ce même élément de surface dans la direction (θ_e, ϕ_e)

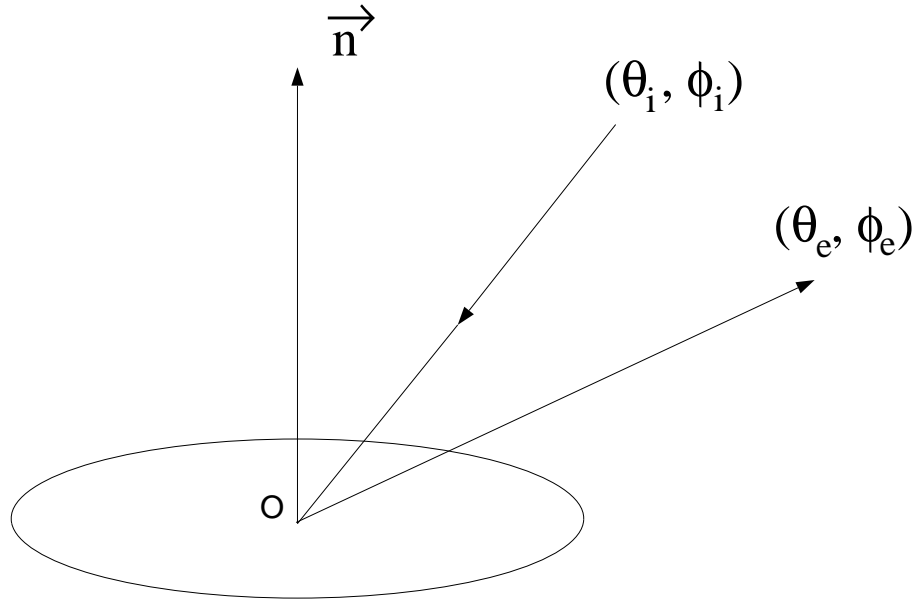


Figure 6.7

On définit alors la fonction de réflectance bidirectionnelle f par :

$$f(\theta_i, \phi_i, \theta_e, \phi_e) = \frac{dL(\theta_e, \phi_e)}{dE(\theta_i, \phi_i)} \quad (6.18)$$

Il y a une contrainte sur la forme d'une fonction de réflectance bidirectionnelle, celle donnée par la condition de réciprocité d'Helmholtz:

$$f(\theta_i, \phi_i, \theta_e, \phi_e) = f(\theta_e, \phi_e, \theta_i, \phi_i) \quad (6.19)$$

En effet lorsque deux surfaces sont en équilibre thermique, le rayonnement de l'une sur l'autre est contrebalancé par le rayonnement se propageant en sens opposé. Si ce n'était pas le cas, l'une se réchaufferait tandis que l'autre se refroidirait, ce qui violerait la seconde loi de la thermodynamique.

3 Eclairement d'une surface par une source étendue

Soit $E(\theta_i, \phi_i)$ la luminance dans la direction (θ_i, ϕ_i) de l'élément de la source de dimensions $d\theta_i$ et $d\phi_i$ et d'aire $\sin \theta_i d\theta_i d\phi_i$. Soit dE l'éclairement de cet élément sur un élément de surface O (cf figure 6.8).

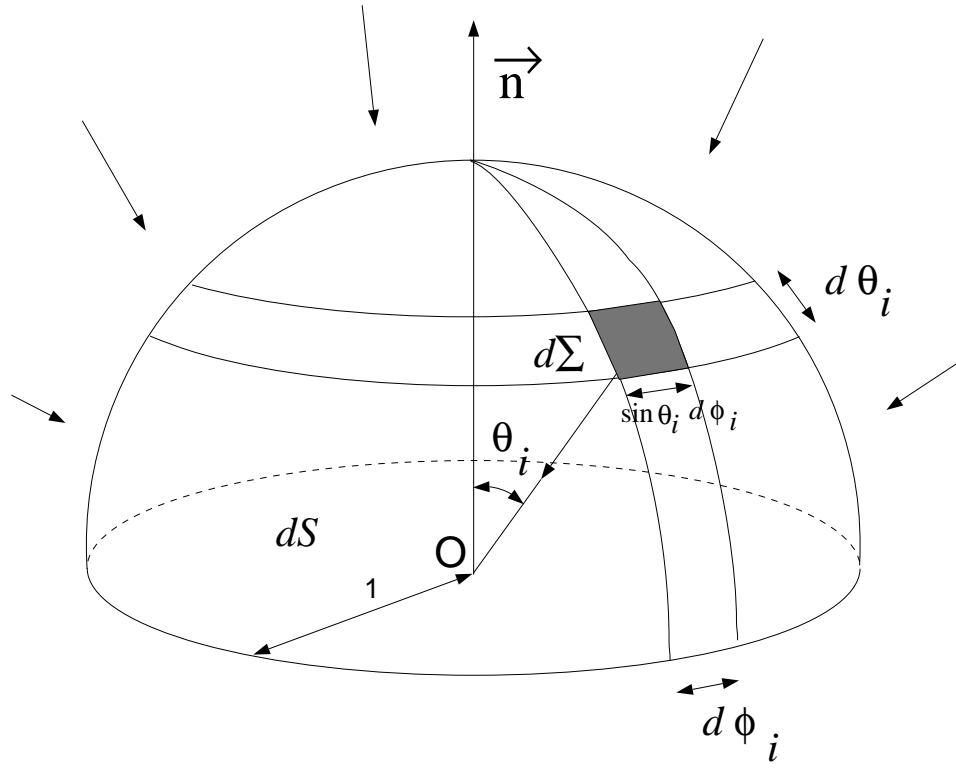


Figure 6.8

D'après (6.9) :

$$dE = E(\theta_i, \phi_i) \cos \theta_i d\psi_i \quad (6.20)$$

où $d\psi_i$ est l'angle solide du cône de sommet O sous-tendu par l'élément de la source (θ_i, ϕ_i) , donc égal à l'aire de cet élément:

$$d\psi_i = \sin \theta_i d\theta_i d\phi_i \quad (6.21)$$

On en déduit:

$$dE = E(\theta_i, \phi_i) \sin \theta_i \cos \theta_i d\theta_i d\phi_i \quad (6.22)$$

et l'éclairement de l'élément de surface par la source étendue est:

$$E_{tot} = \int dE = \int_{-\pi}^{+\pi} \int_0^{\pi/2} E(\theta_i, \phi_i) \sin \theta_i \cos \theta_i d\theta_i d\phi_i \quad (6.23)$$

On peut maintenant calculer la luminance L de cet élément de surface :

$$dL = f dE \quad (6.24)$$

et :

$$L(\theta_e, \phi_e) = \int dL = \int_{-\pi}^{+\pi} \int_0^{\pi/2} f(\theta_i, \phi_i, \theta_e, \phi_e) E(\theta_i, \phi_i) \sin \theta_i \cos \theta_i d\theta_i d\phi_i \quad (6.25)$$

4 Eclairage d'une surface par une source ponctuelle

Considérons une source ponctuelle dans la direction (θ_s, ϕ_s) et d'intensité E . Soit $E(\theta_i, \phi_i)$ la luminance de cette source dans la direction (θ_i, ϕ_i) . On a nécessairement:

$$E(\theta_i, \phi_i) = K \delta(\theta_i - \theta_s) \delta(\phi_i - \phi_s) \quad (6.26)$$

Déterminons la constante K :

$$E = \int_{-\pi}^{+\pi} \int_0^{\pi/2} E(\theta_i, \phi_i) \sin \theta_i d\theta_i d\phi_i = K \sin \theta_s \quad (6.27)$$

D'où :

$$E(\theta_i, \phi_i) = E \frac{\delta(\theta_i - \theta_s) \delta(\phi_i - \phi_s)}{\sin \theta_s} \quad (6.28)$$

et l'éclairage de l'élément de surface par la source ponctuelle est:

$$E_{tot} = \int_{-\pi}^{+\pi} \int_0^{\pi/2} E(\theta_i, \phi_i) \sin \theta_i \cos \theta_i d\theta_i d\phi_i = E \cos \theta_s \quad (6.29)$$

5 Surfaces typiques

Etudions maintenant la réflectance de quelques surfaces typiques.

5.1 Surface de Lambert blanche

Une surface de Lambert est une surface dont la luminance reste la même dans toutes les directions, à éclairage constant.

Une surface est dite blanche si elle réfléchit toute la lumière qu'elle reçoit (pas d'absorption)

Déterminons la fonction de réflectance bidirectionnelle d'une telle surface. Pour ce faire, éclairons-la par une source ponctuelle de direction (θ_s, ϕ_s) et d'intensité E .

La première propriété s'écrit:

$$L(\theta_e, \phi_e) = f(\theta_s, \phi_s, \theta_e, \phi_e) E \cos \theta_s = \text{constante} \quad \forall (\theta_s, \phi_s, \theta_e, \phi_e) \quad (6.30)$$

pourvu que l'éclairage $E \cos \theta_s$ reste constant. On en déduit que la fonction de réflectance est constante:

$$f(\theta_s, \phi_s, \theta_e, \phi_e) = \text{constante} = K \quad \forall (\theta_s, \phi_s, \theta_e, \phi_e) \quad (6.31)$$

Pour déterminer la constante, on fait appel à la deuxième propriété, à savoir que la puissance réémise par l'élément de surface est égale à la puissance reçue.

$$dP_i = dP_e \quad (6.32)$$

La puissance reçue est :

$$dP_i = E \cos \theta_s d\Sigma \quad (6.33)$$

où $d\Sigma$ est l'aire de l'élément de surface

Pour calculer la puissance réémise, considérons l'hémisphère reposant sur cet élément de surface. L'éclairement du point (θ_e, ϕ_e) de cet hémisphère est:

$$dE = L(\theta_e, \phi_e) \cos \theta_e d\Sigma \quad (6.34)$$

La puissance reçue par l'élément de dimensions $d\theta_e$ et $d\phi_e$ et d'aire $\sin \theta_e d\theta_e d\phi_e$ est donc:

$$d^2 P_e = L(\theta_e, \phi_e) \cos \theta_e \sin \theta_e d\theta_e d\phi_e d\Sigma \quad (6.35)$$

La puissance réémise par l'élément de surface est donc:

$$dP_e = \int_{-\pi}^{+\pi} \int_0^{\pi/2} L(\theta_e, \phi_e) \cos \theta_e \sin \theta_e d\theta_e d\phi_e d\Sigma \quad (6.36)$$

avec

$$L(\theta_e, \phi_e) = K E \cos \theta_s \quad \forall \theta_e, \phi_e \quad (6.37)$$

Appliquant (6.32), on déduit:

$$E \cos \theta_s d\Sigma = \int_{-\pi}^{+\pi} \int_0^{\pi/2} K E \cos \theta_s \cos \theta_e \sin \theta_e d\theta_e d\phi_e d\Sigma \quad (6.38)$$

soit après simplification:

$$2\pi K \int_0^{\pi/2} \sin \theta_e \cos \theta_e d\theta_e = 1 \quad (6.39)$$

Puisque $2 \sin \theta_e \cos \theta_e = \sin 2\theta_e$, il vient:

$$\pi K = 1 \quad (6.40)$$

On en déduit qu'une surface est lambertienne si et seulement si sa fonction de réflectance vérifie:

$$f(\theta_i, \phi_i, \theta_e, \phi_e) = \frac{1}{\pi} \quad (6.41)$$

Dans le cas où la surface est grise et non plus blanche (cas général), la fonction de réflectance s'écrit:

$$f(\theta_i, \phi_i, \theta_e, \phi_e) = \frac{\rho}{\pi} \quad (6.42)$$

où ρ ($0 \leq \rho \leq 1$) est l'albédo, c'est-à-dire la fraction de l'énergie réfléchie.

Eclairement par une source ponctuelle

Si on éclaire une telle surface par une source ponctuelle de direction (θ_s, ϕ_s) et de luminance E , alors la luminance de cette surface vérifie:

$$L(\theta_e, \phi_e) = \frac{1}{\pi} E \cos \theta_s \quad (6.43)$$

La luminance est donc indépendante de la direction d'observation.

Eclairement par une source étendue

Eclairons maintenant notre surface de Lambert par un ciel uniforme de luminance E . Appliquant (6.25)

$$L(\theta_e, \phi_e) = \int_{-\pi}^{+\pi} \int_0^{\pi/2} \frac{E}{\pi} \sin \theta_i \cos \theta_i d\theta_i d\phi_i \quad (6.44)$$

soit encore:

$$L(\theta_e, \phi_e) = E \quad (6.45)$$

La luminance de la surface de Lambert est donc la même que celle du ciel.

5.2 Surface spéculaire ou miroir

Une surface est dite spéculaire lorsqu'elle reflète toute la lumière venant de la direction (θ_i, ϕ_i) dans la direction symétrique $(\theta_i, \phi_i + \pi)$

Sa fonction de réflectance vérifie donc:

$$f(\theta_i, \phi_i, \theta_e, \phi_e) = k(\theta_i, \phi_i, \theta_e, \phi_e) \delta(\theta_e - \theta_i) \delta(\phi_e - \phi_i - \pi) \quad (6.46)$$

Ecrivant que la puissance réémise par l'élément de surface est égale à la puissance reçue et appliquant les relations (6.33) et (6.36), il vient:

$$E \cos \theta_i = \int_{-\pi}^{+\pi} \int_0^{\pi/2} L(\theta_e, \phi_e) \sin \theta_e \cos \theta_e d\theta_e d\phi_e \quad (6.47)$$

D'après (6.46) :

$$L(\theta_e, \phi_e) = k(\theta_e, \phi_e, \theta_i, \phi_i) \delta(\theta_e - \theta_i) \delta(\phi_e - \phi_i - \pi) E \cos \theta_i \quad (6.48)$$

On en déduit:

$$\int_{-\pi}^{+\pi} \int_0^{\pi/2} k(\theta_e, \phi_e, \theta_i, \phi_i) \delta(\theta_e - \theta_i) \delta(\phi_e - \phi_i - \pi) \sin \theta_e \cos \theta_e d\theta_e d\phi_e = 1 \quad (6.49)$$

soit encore:

$$k(\theta_e, \phi_e, \theta_i, \phi_i) \sin \theta_i \cos \theta_i = 1 \quad (6.50)$$

On en déduit qu'une surface est spéculaire si et seulement si sa fonction de réflectance vérifie:

$$f(\theta_i, \phi_i, \theta_e, \phi_e) = \frac{\delta(\theta_e - \theta_i) \delta(\phi_e - \phi_i - \pi)}{\sin \theta_i \cos \theta_i} \quad (6.51)$$

D'après (6.25), sa luminance vaut:

$$L(\theta_e, \phi_e) = \int_{-\pi}^{+\pi} \int_0^{\pi/2} \frac{\delta(\theta_e - \theta_i) \delta(\phi_e - \phi_i - \pi)}{\sin \theta_i \cos \theta_i} E(\theta_i, \phi_i) \sin \theta_i \cos \theta_i d\theta_i d\phi_i \quad (6.52)$$

D'où :

$$L(\theta_e, \phi_e) = E(\theta_e, \phi_e - \pi) \quad (6.53)$$

ce qui était attendu.

Section 4 Modèle photométrique et carte de réflectance

Il s'agit maintenant de mettre en évidence la relation entre l'éclairement d'un point image m et l'orientation de la surface au point M dont l'image est m . Cette relation sera utile pour retrouver la géométrie 3D de l'objet à partir de son image: c'est la base du "shape from shading" — restitution de la forme à partir des niveaux de gris —.

Modèle photométrique

La relation (6.10) permet de relier la mesure de l'éclairement d'un point de l'image $m(x,y)$ à la luminance du point de la surface M qui s'est imagé en m . Grâce au calibrage géométrique, nous pouvons évaluer:

$$im_{\alpha}(x, y) = \frac{im(x, y)}{\cos^4 \alpha} \quad (6.54)$$

Un calibrage photométrique est maintenant nécessaire: il consiste à mesurer l'éclairement $im_{\alpha}(x, y)$ — le niveau de gris — provoqué par une source dont on connaît l'orientation.

On pose maintenant $\widehat{im}(x, y) = \frac{im_{\alpha}(x, y)}{im_{\alpha}(x_0, y_0)}$ qu'on appellera "niveau de gris calibré".

Le modèle photométrique s'écrit alors: $\widehat{im}(x, y) = \frac{L(\theta_e, \phi_e)}{L(\theta_e^0, \phi_e^0)}$

Carte de réflectance

Considérons le repère camera ($Lx_c y_c z_c$) et exprimons dans ce repère les coordonnées d'un vecteur normal en M à la surface de l'objet et dirigé vers la caméra (i.e. vers l'extérieur de l'objet). L'équation de la surface peut s'écrire:

$$z_c = F(x_c, y_c) \quad (6.55)$$

et $\vec{n} = \left(\frac{\partial F}{\partial x_c}(x_c, y_c), \frac{\partial F}{\partial y_c}(x_c, y_c), -1 \right)$ est une normale à la surface de l'objet dirigé vers l'extérieur.

On pose:

$$\begin{aligned} p &= \frac{\partial F}{\partial x_c}(x_c, y_c) \\ q &= \frac{\partial F}{\partial y_c}(x_c, y_c) \end{aligned} \quad (6.56)$$

On appelle alors carte de réflectance la fonction R qui à (p, q) associe le niveau de gris calibré du point éclairé par M :

$$(p, q) \rightarrow R(p, q) = \widehat{im}(x, y) = \frac{L(\theta_e, \phi_e)}{L(\theta_e^0, \phi_e^0)} \quad (6.57)$$

Etant donné un niveau de gris, elle fournit toutes les orientations de la surface compatibles avec ce niveau de gris.

Exemples

Surface de Lambert blanche éclairée par une source ponctuelle La source étant “ponctuelle” se situe à une grande distance de l’objet vis-à-vis de ses dimensions. Soit $\vec{n}_s = (p_s, q_s, 1)$ la direction commune des rayons lumineux qui atteignent l’objet exprimée dans le repère de la caméra. Soit E l’intensité de cette source.

La luminance d’un point de la surface de l’objet est donc:

$$L(\theta_e, \phi_e) = \frac{1}{\pi} E \cos \theta_s \quad (6.58)$$

où θ_s est l’angle que la normale à la surface en ce point $\vec{n} = (p, q, 1)$ fait avec \vec{n}_s

Supposons que le surfel ayant servi au calibrage photométrique soit orienté dans la direction de la source: $\theta_s^0 = 0$. On a donc:

$$L(\theta_e^0, \phi_e^0) = \frac{1}{\pi} E \quad (6.59)$$

D’où:

$$R(p, q) = \cos \theta_s \quad (6.60)$$

Or

$$\cos \theta_s = \frac{(\vec{n} | \vec{n}_s)}{\|\vec{n}\| \|\vec{n}_s\|} = \frac{p_s p + q_s q + 1}{\sqrt{p_s^2 + q_s^2 + 1} \sqrt{p^2 + q^2 + 1}} \quad (6.61)$$

On en déduit que:

$$R(p, q) = \frac{p_s p + q_s q + 1}{\sqrt{p_s^2 + q_s^2 + 1} \sqrt{p^2 + q^2 + 1}} \quad (6.62)$$

Surface de Minnaert La fonction de réflectance d’une telle surface s’écrit:

$$f(\theta_i, \phi_i, \theta_e, \phi_e) = \frac{k+1}{2\pi} (\cos \theta_i \cos \theta_e)^{k-1} \quad (6.63)$$

avec $0 \leq k \leq 1$. Pour $k = \frac{1}{2}$, f modélise la surface des mers de la lune. Se reporter à l’exercice 1.

Section 5 Stéréophotométrie

Si on produit deux éclairagements différents de la scène à mesurer et si l’on acquiert les deux images correspondantes, on obtient pour chaque pixel, deux équations à deux inconnues (d’après (6.57):

$$\begin{aligned} im_1(x, y) &= R_1(p, q) \\ im_2(x, y) &= R_2(p, q) \end{aligned} \quad (6.64)$$

Dans certains cas, par exemple, si les deux équations sont linéaires et indépendantes, on trouvera une solution unique (\bar{p}, \bar{q}) qui donnera l’orientation du surfel. Dans d’autres cas, un troisième éclairage sera nécessaire pour obtenir une solution unique.

Section 6 Exercices

1 Surfaces de Minnaert

Considérons les surfaces dites de Minnaert dont les fonctions de réflectance s'écrivent:

$$f(\theta_i, \phi_i, \theta_e, \phi_e) = \frac{k+1}{2\pi} (\cos \theta_i \cos \theta_e)^{k-1} \quad (6.65)$$

avec $0 \leq k \leq 1$

1.1

On éclaire cette surface par un ciel hémisphérique de luminance uniforme E_0 :

$$E(\theta_i, \phi_i) = E_0 \quad \forall \theta_i \in \left[0, \frac{\pi}{2}\right], \forall \phi_i \in [-\pi, +\pi] \quad (6.66)$$

Calculer la luminance de cette surface $L(\theta_e, \phi_e) \quad \forall \theta_e \in \left[0, \frac{\pi}{2}\right], \forall \phi_e \in [-\pi, +\pi]$

1.2

On l'éclaire maintenant par une source ponctuelle de luminance E_0 dans la direction (θ_s, ϕ_s) :

$$E(\theta_i, \phi_i) = \frac{E_0 \delta(\theta_i - \theta_s) \delta(\phi_i - \phi_s)}{\sin \theta_s} \quad \forall \theta_i \in \left[0, \frac{\pi}{2}\right], \forall \phi_i \in [-\pi, +\pi] \quad (6.67)$$

Calculer $L(\theta_e, \phi_e) \quad \forall \theta_e \in \left[0, \frac{\pi}{2}\right], \forall \phi_e \in [-\pi, +\pi]$

2. On suppose maintenant $k = \frac{1}{2}$

a. Calculer la carte de réflectance $R(p, q)$

b. Soit un pixel (x_0, y_0) de niveau de gris $im(x_0, y_0) = a$; soit le surfel M dont l'image est m .

Quelle est l'équation nécessairement vérifiée par le gradient (p, q) de M ? Quel est le lieu géométrique des gradients possibles de M ?

2 Stéréovision mono-image

On considère une surface S éclairée et observée par une caméra. On supposera que le modèle géométrique de cette caméra est une projection orthogonale. On suppose que la carte de réflectivité de S s'écrit:

$$R(p, q) = ap + bq \quad (6.68)$$

dans un repère $Oxyz$ lié à la caméra; a et b sont deux constantes positives ou nulles

On considère l'angle α_0 défini par:

$$\cos \alpha_0 = \frac{a}{\sqrt{a^2 + b^2}} \quad ; \quad \sin \alpha_0 = \frac{b}{\sqrt{a^2 + b^2}} \quad (6.69)$$

et dans le plan Oxy, la demi-droite (D_0) passant par O et de vecteur directeur $(\cos \alpha_0, \sin \alpha_0)$.

On suppose aussi connues:

a. l'image de la surface S :

$$\widehat{im}(x, y) = \lambda |x| + \mu |y| \quad (6.70)$$

où λ et μ sont deux constantes positives ou nulles.

b. la profondeur $|z_0|$ du surfel M_0 situé sur l'axe optique de la caméra (donc de coordonnées $(0, 0, z_0)$ dans le repère de la caméra Oxyz)

On demande de calculer la profondeur $|z|$ d'un surfel M qui se projette orthogonalement sur (D_0) en $m(t \cos \alpha_0, t \sin \alpha_0)$ où $t \geq 0$.

Conseil : si dans Oxyz la surface s'écrit:

$$z + f(x, y) = 0 \quad (6.71)$$

on dérivera la fonction g suivante:

$$\begin{aligned} t \rightarrow g(t) &= f(t \cos \alpha_0, t \sin \alpha_0) = -z \\ \mathbb{R}^+ &\rightarrow \mathbb{R} \end{aligned} \quad (6.72)$$

3 Stéréophotométrie avec une surface de Lambert

Considérons une surface de Lambert éclairée par une source éloignée. La direction de cette source étant (p_s, q_s, I) dans le repère classique lié à la caméra, on sait que l'éclairement $im(x,y)$ — le “niveau de gris” — du pixel (x,y) est liée à la normale (p,q,I) de l'élément de surface qui l'éclaire par la relation (formule (6.57) du photocopié):

$$im(x,y) = \frac{p_s p + q_s q + 1}{\sqrt{p^2 + q^2 + 1} \sqrt{p_s^2 + q_s^2 + 1}} \quad (6.73)$$

3.1

On éclaire successivement cette surface par deux sources ponctuelles éloignées de directions (p_1, q_1, I) et (p_2, q_2, I) . Quelles sont les deux équations vérifiées par la normale (p,q,I) d'un élément de surface qui réfléchit la lumière sur le pixel (x,y) ?

3.2

Montrer que les solutions de ces deux équations appartiennent nécessairement à une droite que l'on précisera.

3.3

En déduire le nombre maximal de solutions à ces deux équations.

3.4

L'équation (6.73) peut encore s'écrire:

$$im(x,y) = (\vec{n} \mid \vec{s}) \quad (6.74)$$

où \vec{n} est le vecteur unitaire normal à la surface, \vec{s} le vecteur unitaire dans la direction de la source lumineuse et $(. \mid .)$ le produit scalaire euclidien.

A partir de la relation (6.74) réécrire les deux équations mentionnées dans la question 3.1.

3.5

En déduire les coordonnées de \vec{n} dans la base $(\vec{s}_1, \vec{s}_2, \vec{s}_1 \wedge \vec{s}_2)$

3.6

Reprendre la question 3.3. En déduire aussi une condition nécessaire et suffisante pour que la solution à ces deux équations soit unique.

3.7

Quelle nouvelle expérience faudrait-il faire pour obtenir une solution unique dans tous les cas de figure ?

Chapitre 7 Vision stéréoscopique binoculaire

Section 1 Principe

Le principe de base de la stéréovision binoculaire est la ***mise en correspondance*** ou ***appariement*** de deux images d'une même scène observée sous deux points de vue différents.

En effet, supposons qu'on ait repéré deux points m_g et m_d qui soient les deux images d'un même point M de la scène 3D, il est alors possible par triangulation de déterminer la position de ce point M . Dans le cas classique où l'on modélise les deux caméras par une projection perspective, M est à l'intersection des deux droites de projection associées aux deux images m_g et m_d . Les équations de ces deux droites dans un repère quelconque peuvent être déterminées grâce au calibrage géométrique des deux caméras.

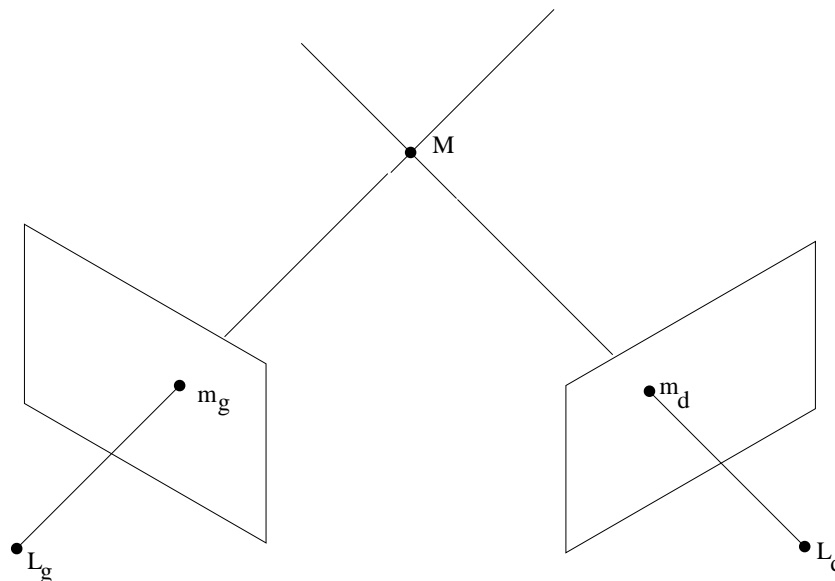


Figure 7.1 Triangulation

Il y a donc trois problèmes à résoudre:

- le positionnement des caméras
- la mise en correspondance des images
- la triangulation

Section 2 Positionnement des caméras

Les relations géométriques entre le capteur binoculaire et la scène influent sur la qualité de la reconstruction. Intuitivement, les deux caméras doivent viser la scène à reconstruire de façon à maximiser la surface totale visible des deux caméras.. En déterminant un point de fixation — intersection des axes optiques — qui est au milieu de la scène, on accroît généralement la surface de la scène visible des deux caméras et on réduit la zone de recherche des points homologues.

D'un point de vue théorique, un tel positionnement permet de définir un critère de vraisemblance d'appariement plus robuste.

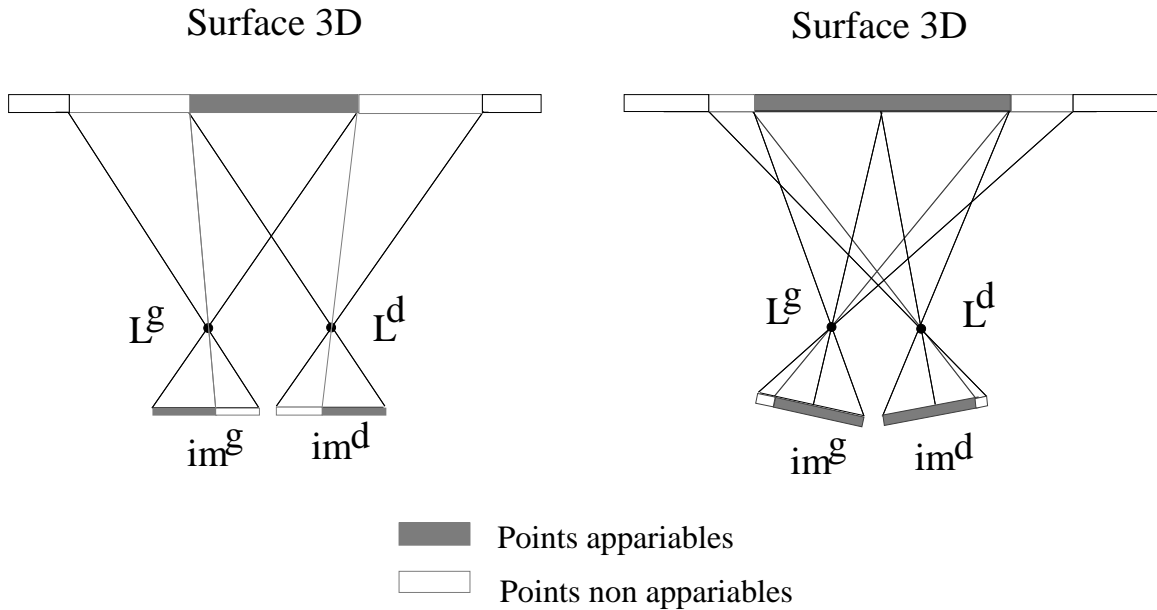


Figure 7.2 Mouvement de convergence des caméras vers la scène

Section 3 Mise en correspondance

1 Points et pixels homologues

On dira que deux points image sont *homologues* s'ils sont les images d'un même point 3D de la scène. La mise en correspondance de deux images consiste à déterminer pour chaque point image (gauche ou droit) s'il possède ou non un homologue et si oui lequel.

Remarque: au lieu de mettre en correspondance les points qui sont en nombre infini, il est plus commode de mettre en correspondance les pixels qui eux sont en nombre fini. Deux pixels seront dits **homologues** si l'homologue du point central de l'un appartient à l'autre et vice-versa. On remarquera que la relation "homologue" ainsi définie est bien symétrique.

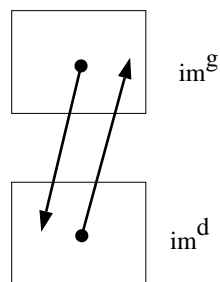


Figure 7.3 Pixels homologues

Un point image ne possède pas d'homologue si l'une des deux situations suivantes se produit:

- il est l'image d'un élément de surface 3D qui n'appartient pas au champ visuel de l'autre caméra;

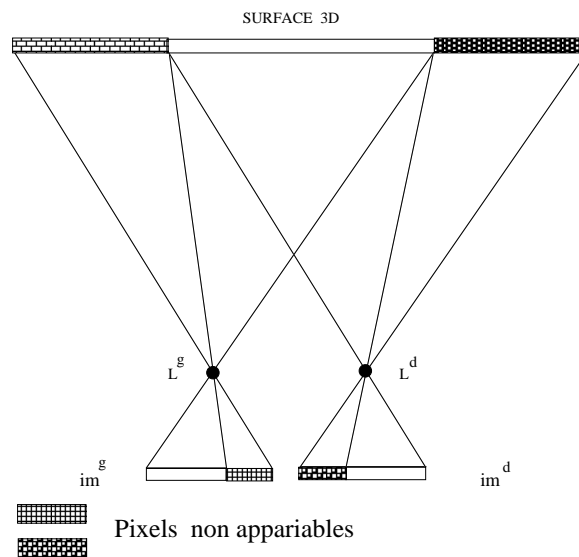


Figure 7.4 Points et pixels qui sont images d'un élément de surface 3D qui n'appartient pas au champ visuel commun aux deux caméras.

- il est l'image d'un élément de surface 3D qui appartient au champ visuel commun aux deux caméras, mais n'est visible que d'une des deux caméras: il y a donc ici un phénomène d'*occultation*;

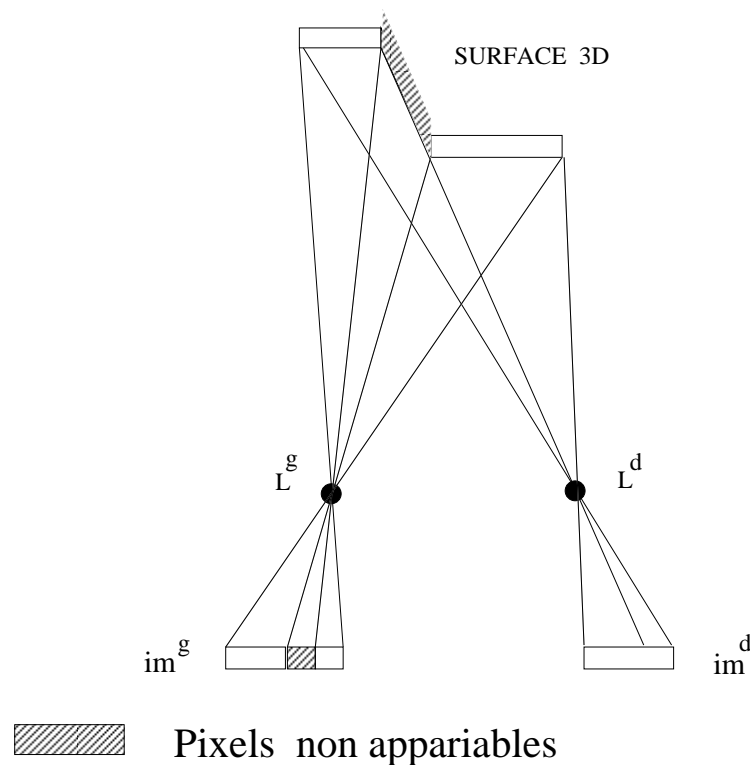


Figure 7.5 Points et pixels non appariables provenant d'un phénomène d'occultation.

Un pixel image ne possède pas d'homologue dans l'un des deux cas précédents mais aussi dans le cas où il est l'image d'un élément de surface 3D incliné par rapport à la ligne de base (la droite joignant les deux centres de projection). Dans ce cas, le nombre de pixels représentant -après projection- le même élément de surface 3D est différent d'une image à une autre: il y a un phénomène de raccourcissement.

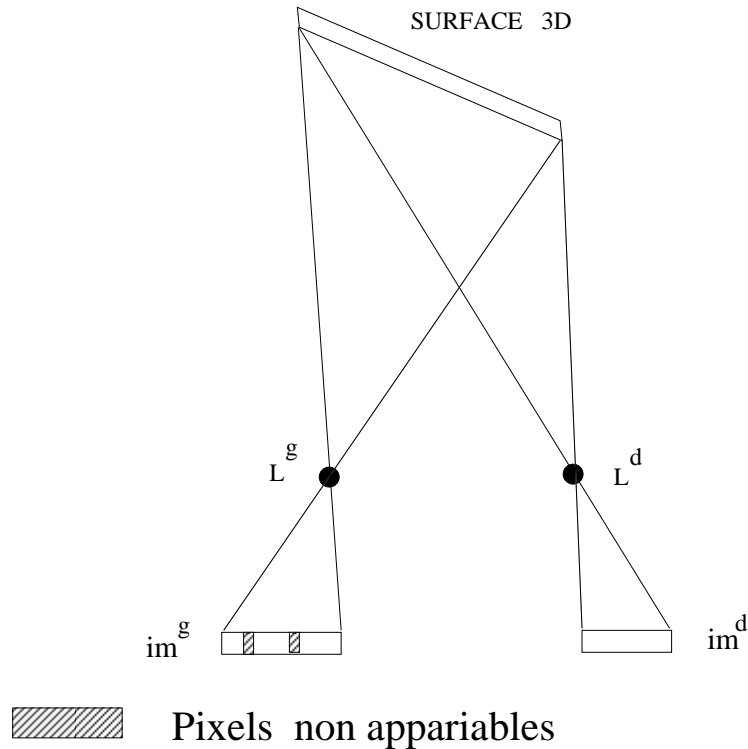


Figure 7.6 Pixels non appariables provenant d'un raccourcissement entre les images.

2 Contraintes

Nous commençons par présenter deux contraintes de base qui permettent de réduire la recherche des points ou des pixels homologues.

2.1 Contrainte d'unicité-occultation

Elle s'énonce: "Un point ou un pixel de l'image gauche possède *au plus* un homologue dans l'image droite et vice-versa".

2.2 Contrainte épipolaire

La géométrie d'un capteur binoculaire contraint la recherche des points ou pixels homologues. Modélisons notre capteur binoculaire sous la forme classique de deux projections centrales. Soit

- a. M un élément infinitésimal de surface qui réfléchit la lumière
- b. m^g son image gauche
- c. m^d son image droite
- d. L^g le centre optique de la caméra gauche
- e. L^d le centre optique de la caméra droite
- f. $e^g = im^g \cap L^g L^d$ l'épipôle gauche du capteur binoculaire

- g. $e^d = im^d \cap L^g L^d$ l'épipôle droit du capteur binoculaire
- h. ω_M^g le point de fuite associé à la direction de $m^d L^d$ (image d'un point M à l'infini)
- i. ω_M^d le point de fuite associé à la direction de $m^g L^g$;

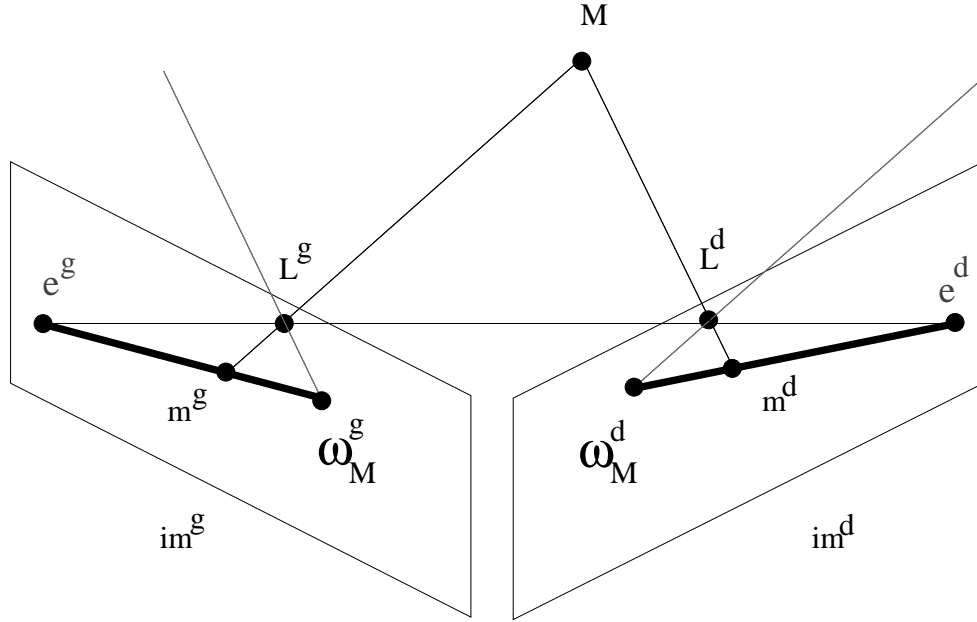


Figure 7.7 Segments épi-polaires conjugués

La contrainte épi-polaire s'énonce:

L'homologue d'un point m^g (resp. m^d) appartient nécessairement au segment épi-polaire $e^d \omega_M^d$ (resp. $e^g \omega_M^g$).

Déterminons les équations de ces segments épi-polaires en écrivant les modèles géométriques directs des deux caméras, gauche et droite:

$$\begin{aligned} \tilde{m}^g &= \tilde{C}^g \tilde{M} \\ \tilde{m}^d &= \tilde{C}^d \tilde{M} \end{aligned} \quad (7.1)$$

où $\tilde{M} \in P^3$ et $\tilde{m} \in P^2$. L^g et L^d étant les centres optiques:

$$\begin{aligned} \tilde{C}^g \tilde{L}^g &= 0 \\ \tilde{C}^d \tilde{L}^d &= 0 \end{aligned} \quad (7.2)$$

Posant: $\tilde{C}^g = [C^g \tilde{c}^g]$ et $\tilde{C}^d = [C^d \tilde{c}^d]$, il vient:

$$\tilde{L}^g = - \begin{bmatrix} -(C^g)^{-1} \tilde{c}^g \\ 1 \end{bmatrix} \quad (7.3)$$

et

$$\tilde{e}^d = \tilde{C}^d \tilde{L}^g = \tilde{C}^d \begin{bmatrix} -(C^g)^{-1} \tilde{c}^g \\ 1 \end{bmatrix} = \tilde{c}^d - C^d (C^g)^{-1} \tilde{c}^g \quad (7.4)$$

Le point à l'infini du rayon de projection $L^g m^g$, Ω_m^g a pour coordonnées projectives:

$$\tilde{\Omega}_m^g = \begin{bmatrix} (C^g)^{-1} \tilde{m}^g \\ 0 \end{bmatrix} \quad (7.5)$$

Le point ω_M^d est l'image droite de Ω_m^g . Donc:

$$\tilde{\omega}_M^d = \tilde{C}^d \tilde{\Omega}_m^g = \tilde{C}^d \begin{bmatrix} (C^g)^{-1} \tilde{m}^g \\ 0 \end{bmatrix} = C^d (C^g)^{-1} \tilde{m}^g \quad (7.6)$$

L'équation de la droite épipolaire (D_e^d) qui passe par e^d et ω_M^d est donc:

$$(D_e^d) = \left\{ \tilde{m} = \alpha \tilde{e}^d + \beta \tilde{\omega}_M^d \right\} = \left\{ \tilde{m} \mid \tilde{m}, \tilde{e}^d, \tilde{\omega}_M^d \mid = 0 \right\} = \left\{ \tilde{m} \mid \tilde{m}^T (\tilde{e}^d \wedge \tilde{\omega}_M^d) = 0 \right\} \quad (7.7)$$

Le segment épipolaire s'écrit:

$$[\tilde{\omega}_M^d, \tilde{e}^d] = \left\{ \alpha \tilde{\omega}_M^d + \beta \tilde{e}^d \right\} \quad (7.8)$$

où α et β sont de mêmes signes.

Droites épipolaires conjuguées

D'après (7.7), (D_e^d) s'identifie au vecteur $\tilde{e}^d \wedge \tilde{\omega}_M^d$. Or, si nous considérons la matrice (3,3) \tilde{E}^d telle que: $\tilde{E}^d x = \tilde{e}^d \wedge x$, \tilde{E}^d est une matrice anti-symétrique puisque, posant $\tilde{e}^d = (e_1^d, e_2^d, e_3^d)^T$,

$$\tilde{E}^d x = \begin{pmatrix} e_1^d \\ e_2^d \\ e_3^d \end{pmatrix} \wedge \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} e_2^d x_3 - e_3^d x_2 \\ e_3^d x_1 - e_1^d x_3 \\ e_1^d x_2 - e_2^d x_1 \end{pmatrix} = \begin{pmatrix} 0 & -e_3^d & e_2^d \\ e_3^d & 0 & -e_1^d \\ -e_2^d & e_1^d & 0 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \quad (7.9)$$

On en déduit que:

$$\tilde{e}^d \wedge \tilde{\omega}_M^d = \tilde{E}^d \tilde{\omega}_M^d = \tilde{E}^d C^d (C^g)^{-1} \tilde{m}^g = \tilde{F} \tilde{m}^g \quad (7.10)$$

où

$$\tilde{F} = \tilde{E}^d C^d (C^g)^{-1} \quad (7.11)$$

La matrice \tilde{F} porte le nom de **matrice fondamentale**. Elle est définie à un facteur d'échelle près et est de rang 2; en effet, les matrices C^g et C^d étant inversibles, $rg(\tilde{F}) = rg(\tilde{E}^d) = 2$.

Les droites épipolaires conjuguées, qui sont les intersections du plan épipolaire avec les plans-images gauche et droit ont donc pour équations:

$$\begin{aligned} (D_e^d) &= \left\{ \tilde{m} \mid \tilde{m}^T \tilde{F} \tilde{m}^g = 0 \right\} \\ (D_e^g) &= \left\{ \tilde{m} \mid \tilde{m}^T \tilde{F}^T \tilde{m}^d = 0 \right\} \end{aligned} \quad (7.12)$$

On en déduit aussi qu'une condition nécessaire pour que deux points m^g et m^d soient homologues est qu'ils vérifient la relation :

$$\tilde{m}^{dT} \tilde{F} \tilde{m}^g = 0 \quad (7.13)$$

3 Critère de vraisemblance

Pour résoudre complètement le problème de la mise en correspondance, on a besoin d'un critère qui mesure la vraisemblance d'une mise en correspondance. La solution sera celle qui maximise le critère.

On admet généralement que ce critère comporte deux termes :

- une mesure de **ressemblance locale** qui s'appuie sur le fait suivant: lorsqu'un élément de surface est vu des deux caméras, ses images gauche et droite sont en général semblables aussi bien géométriquement que radiométriquement.
- une mesure de **régularité** qui est nécessaire lorsque la mesure de ressemblance est insuffisante pour fournir une solution unique au problème de mise en correspondance.

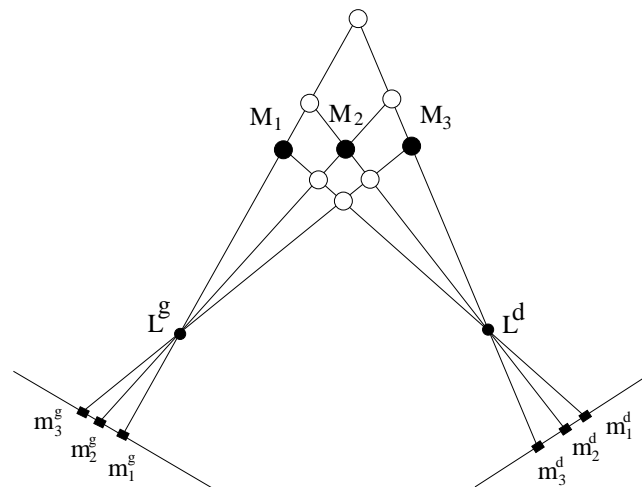


Figure 7.8 Problème des fausses cibles. Dans cette figure, chaque point d'une image a trois homologues possibles dans l'autre image. Parmi toutes les solutions envisageables une seule correspond à la réalité physique (points noirs).

4 Stéréovision humaine

L'invention des stéréogrammes de points aléatoires par Julesz suggère que le processus mis en oeuvre par le système visuel humain pour fusionner les deux images est fondamentalement un processus de "bas niveau" qui ne s'appuie ni sur une reconnaissance monoculaire, ni sur une connaissance a priori de la scène (figure 7.9).



Figure 7.9 Stéréogramme de points aléatoires. Au moyen d'un stéréoscope, on peut percevoir le relief. Les formes n'apparaissent, qu'une fois le processus de fusion accompli.

Section 4 Reconstruction 3D en l'absence d'information de position des caméras

Dans de nombreuses applications, les positions et orientations des caméras et donc leurs paramètres externes, sont inconnues. Il peut s'agir du cas où l'on dispose de deux caméras dont les positions et orientations absolues et relatives sont inconnues ou bien d'une seule caméra se déplaçant selon un mouvement inconnu (navigation robotique). Il est malgré tout possible de reconstruire à un facteur d'échelle près une scène en trois dimensions lorsqu'on possède deux images de celle-ci et au moins 8 points (voire 5 mais c'est plus difficile) points homologues

1 Equations dans les systèmes de coordonnées normalisés

On suppose simplement connus les paramètres internes des deux caméras gauche et droite ($\alpha_u = f e_u$, $\alpha_v = f e_v$, p_u , p_v)_{g,d}. On définit deux plans images normalisés situés à la distance unité (au lieu de f^g et de f^d) des deux centres optiques L^g et L^d .

On est en mesure de calculer les deux images normalisées grâce aux formules de passage:

$$\begin{aligned} u' &= \frac{u - p_u}{\alpha_u} \\ v' &= \frac{v - p_v}{\alpha_v} \end{aligned} \quad (7.14)$$

On considèrera que le repère 3D de référence est le repère caméra gauche ($L^g x_c y_c z_c$). Le repère image gauche (resp. droit) sera le repère centré de l'image normalisée gauche (resp. droite) $P^g u' v' (P^d u' v')$

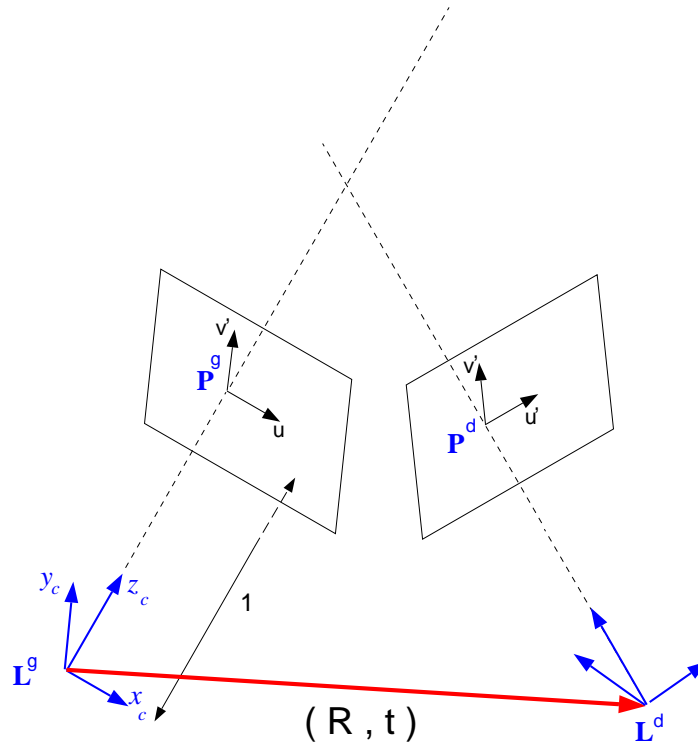


Figure 7.10

On considère aussi le déplacement $D = (R, t)$ qui fait passer du repère caméra gauche au repère caméra droit; R est la matrice de rotation gauche vers droit et $t = L^g L^d$

Le modèle direct de la caméra gauche est donc

$$\tilde{C}^g = [I_3 \ 0] \quad (7.15)$$

où I_3 est la matrice identité.

La matrice de passage du repère caméra gauche au repère caméra droite est en coordonnées projectives (cf 5.8) :

$$\tilde{K} = \begin{bmatrix} R & t \\ 0_3^T & 1 \end{bmatrix} \quad (7.16)$$

La matrice de passage en coordonnées projectives du repère caméra droit au repère caméra gauche est donc:

$$\tilde{K}^{-1} = \begin{bmatrix} R^T & -R^T t \\ 0_3^T & 1 \end{bmatrix} \quad (7.17)$$

Le modèle direct de la caméra droite, exprimé dans le repère caméra droite est $[I_3 \ 0]$; exprimé dans le repère caméra gauche (le repère de référence), il est donc:

$$\tilde{C}^d = [I_3 \ 0] \tilde{K}^{-1} = [I_3 \ 0] \begin{bmatrix} R^T & -R^T t \\ 0_3^T & 1 \end{bmatrix} = [R^T \ -R^T t] \quad (7.18)$$

Les coordonnées projectives des épipôles sont donc:

$$\begin{aligned} \tilde{e}^g &= \tilde{C}^g \tilde{L}^d = [I_3 \ 0] \begin{bmatrix} t \\ 1 \end{bmatrix} = t \\ \tilde{e}^d &= \tilde{C}^d \tilde{L}^g = [R^T \ -R^T t] \begin{bmatrix} 0_3 \\ 1 \end{bmatrix} = -R^T t \end{aligned} \quad (7.19)$$

La droite épipolaire associée au point de l'image gauche m^g est représentée par $\tilde{e}^d \wedge C^d (C^g)^{-1} \tilde{m}^g$. Comme ici, $C^d = R^T$ et $C^g = I_3$, elle est donc représentée par $-R^T t \wedge R^T \tilde{m}^g$ soit encore par $R^T (t \wedge \tilde{m}^g)$

La droite épipolaire associée au point de l'image droite m^d est représentée par: $\tilde{e}^g \wedge C^g (C^d)^{-1} \tilde{m}^d = t \wedge R \tilde{m}^d$

On en déduit que si les deux points m^g et m^d sont homologues alors:

$$\left(\tilde{m}^g \mid t \wedge R \tilde{m}^d \right) = 0 \quad (7.20)$$

2 Application à la reconstruction à un facteur d'échelle près d'une scène 3D prise sous deux points de vue distincts

Comme application, considérons une seule caméra se déplaçant selon un mouvement inconnu. Appliquons les résultats précédents en considérant que la caméra gauche correspond à la caméra en position 1 (ou initiale) et la caméra droite à la position 2 (ou finale).

2.1 Matrice essentielle

Soit la matrice anti-symétrique T telle que: $Tx = t \wedge x$. On a donc:

$$T = \begin{pmatrix} 0 & -t_3 & t_2 \\ t_3 & 0 & -t_1 \\ -t_2 & t_1 & 0 \end{pmatrix} \quad (7.21)$$

On appelle **matrice essentielle** la matrice

$$E = TR \quad (7.22)$$

Notant (u'_1, v'_1) et (u'_2, v'_2) les coordonnées de deux points homologues m^1 et m^2 dans les repères des images normalisées et posant $u^1 = (u'_1, v'_1, 1)^T$ et $u^2 = (u'_2, v'_2, 1)^T$, alors la relation (7.20) s'écrit:

$$(u^1 \mid Eu^2) = 0 \quad (7.23)$$

Algorithme à 8 points (Longuet-Higgins)

1ère étape: détermination de la matrice essentielle

On suppose connus 8 couples de points homologues $(u_i, v_i)_{i=1,\dots,8}$. On en déduit 8 équations linéaires homogènes:

$$(u_i^1 \mid Eu_i^2) = 0 \quad i = 1, \dots, 8 \quad (7.24)$$

soit encore

$$Ax = 0 \quad (7.25)$$

où A est une matrice de taille (8,9).

Supposons que les points M_1, M_2, \dots, M_8 , antécédents des 8 couples de points homologues, soient tels que $\text{rang}(A_8) = 8$. Notons a^i le i ème vecteur-colonne de A . Supposons, par exemple, que a^1, a^2, \dots, a^8 soient linéairement indépendants et posons:

$$A_I = [a^1, a^2, \dots, a^8] \quad (7.26)$$

et

$$x_I = (x_1, x_2, \dots, x_8)^T \quad (7.27)$$

alors d'après (7.25):

$$x_I = -x_9 A_I^{-1} a^9 \quad (7.28)$$

On a ainsi déterminé les 9 composantes de la matrice essentielle à un facteur d'échelle près.

2ème étape: détermination de la translation à un facteur d'échelle près

Soit:

- e^i le i -ème vecteur colonne de E
- r^i le i -ème vecteur colonne de R

alors:

$$e^i = t \wedge r^i \quad i = 1, 2, 3 \quad (7.29)$$

Les trois vecteurs-colonnes de E sont donc perpendiculaires à t et donc t est parallèle au produit vectoriel de deux d'entre eux:

$$t = \frac{\|t\|}{\|t_0\|} t_0 \quad (7.30)$$

avec

$$\begin{aligned} t_0 &= e^1 \wedge e^2 & \text{si } (t \mid r^3) > 0 \\ t_0 &= -e^1 \wedge e^2 & \text{si } (t \mid r^3) < 0 \end{aligned} \quad (7.31)$$

3ème étape : détermination de la rotation

On montre (exercice !) que:

$$R = \frac{1}{\|t\|^2} (co(E) - TE) \quad (7.32)$$

où $co(E) = [e^2 \wedge e^3, e^3 \wedge e^1, e^1 \wedge e^2]$ désigne la matrice des cofacteurs de E

Section 5 Proposition d'une méthode

On suppose connue la géométrie épipolaire.

1.1 Synoptique

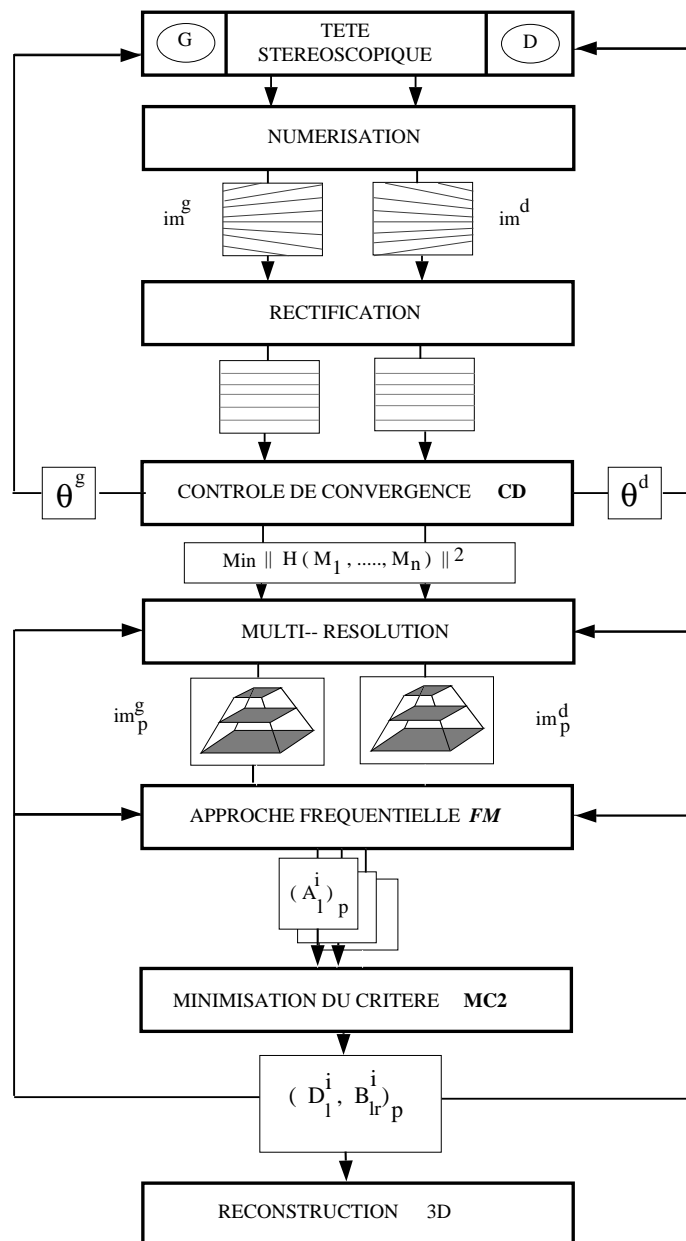


Figure 7.11 Graphe descriptif de la méthode MMT

2 Résultats

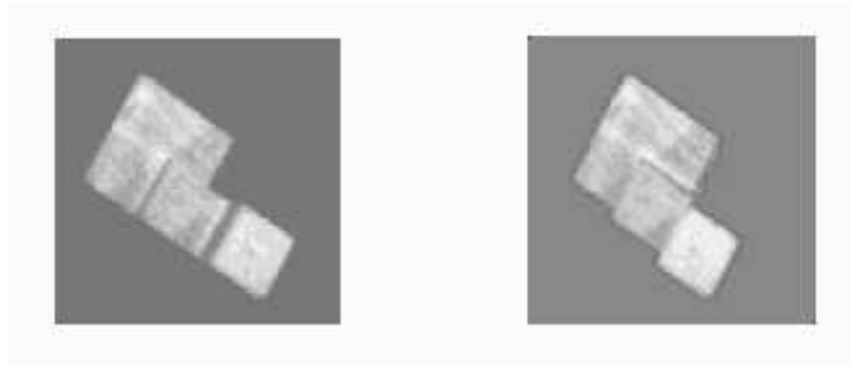


Figure 7.12 Stéréogramme d'un escalier en bois. Avec l'aide d'un stéréoscope, on peut percevoir le relief.

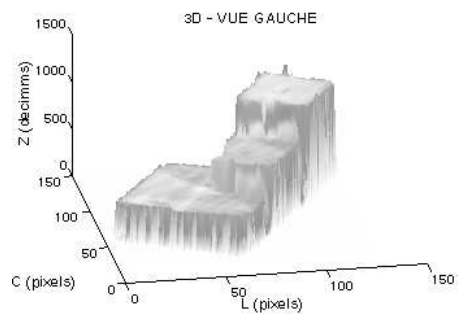


Figure 7.13 Vue gauche de la carte d'élévation de l'objet par rapport à son socle

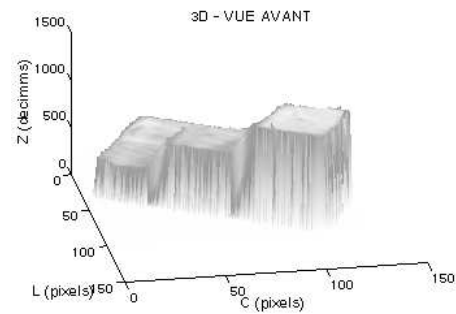


Figure 7.14 Vue avant de cette même carte d'élévation

Chapitre 8 Reconnaissance des objets

Section 1 Définition

La Reconnaissance des Formes a pour objet la conception et l'analyse des algorithmes de classification applicables à des signaux de toute nature (images, sons, ...).

Section 2 Exemples

1 Lecture optique

En entrée, l'algorithme dispose de l'image numérique d'une suite de caractères.

En sortie, il fournit une identification de chacun des caractères de la liste.

2 Vision par ordinateur

En entrée, l'algorithme dispose d'une ou de plusieurs images numériques d'une scène composée d'objets tridimensionnels.

En sortie, il fournit la liste des noms des objets visibles de cette scène.

Section 3 Deux grandes approches

1 Approche cognitive

1.1 Définition

Dans cette approche, on cherche des *définitions* des classes des formes à identifier. Ces définitions fournissent des descriptions des structures ou des configurations de formes.

L'algorithme de reconnaissance consiste alors

- a. à comparer la structure de la forme inconnue à celle qui est fournie par la définition d'une classe donnée
- b. à décider de l'appartenance ou de la non-appartenance de la forme inconnue à la classe en question selon que sa structure répond suffisamment bien ou non à la définition susdite.

1.2 Exemple

On cherche à reconnaître des *images numériques de cercles*.

Le "cognitiviste" se sert d'une définition de l'objet "idéal" cercle ; par exemple, un cercle est une surface plane limitée par une courbe dont tous les points sont à égale distance d'un même point appelé son centre.

Il en déduit une définition de l'image numérique d'un cercle: par exemple, c'est une région dont la frontière est une composante connexe de pixels de contours qui sont tous "approximativement" à égale distance d'un même pixel.

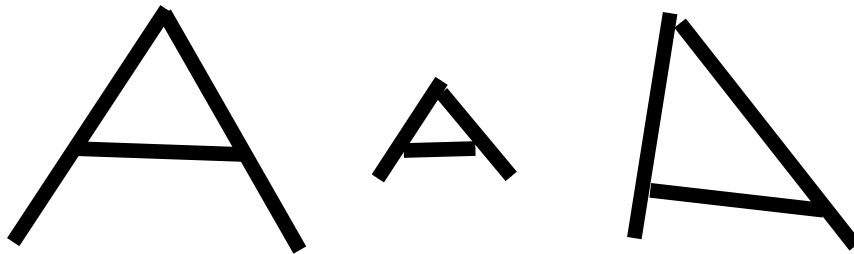
Il bâtit ensuite un algorithme de reconnaissance comprenant les étapes suivantes :

- a. calcul d'une image de contours
- b. extraction de toutes les composantes 8-connexes
- c. sélection de toutes celles qui possèdent la propriété : "tous ses pixels sont approximativement à la même distance d'un autre pixel"

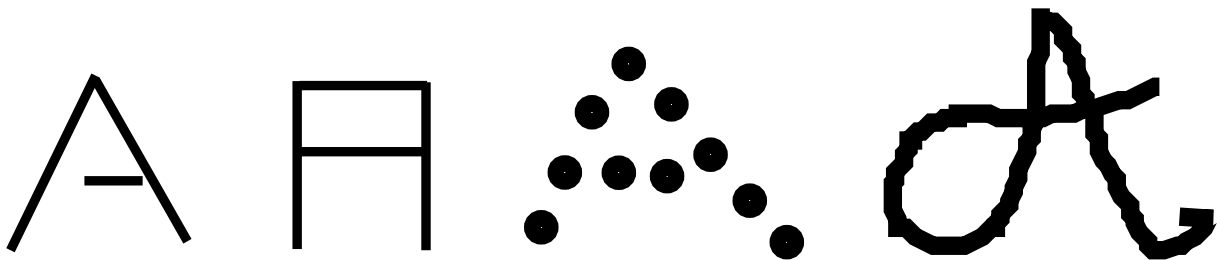
1.3 Difficulté

La difficulté essentielle de ce type d'approche est de trouver de bonnes définitions: les formes d'une même classe présentent en effet souvent de grandes différences d'aspects.

Pour le montrer, considérons l' *image d'un A majuscule* . Une définition (intuitive) pourrait être: "deux segments de droite obliques se rejoignant vers le haut et reliés par un segment horizontal". Cette définition rend bien compte d'une certaine catégorie de A :



En revanche, elle résiste mal au bruit — par exemple, lorsque le segment horizontal ne touche plus l'un ou l'autre des deux segments obliques — et ne convient pas non plus à la description d'autres types de A.



Il semble donc nécessaire d'envisager une multitude de descriptions possibles de la forme de ce caractère. Mais lesquelles ?

Deux autres problèmes se posent:

- a. isoler les caractères lorsqu'ils se touchent
- b. distinguer des caractères géométriquement proches : le contexte peut y aider

2 Approche empirique

Cette approche renonce à établir une définition et se contente de chercher des "indications" sur la classe à reconnaître. Ces indications peuvent être:

- a. des *exemples* , prototypes ou échantillons, de la classe à reconnaître
- b. des *caractéristiques* , déterministes ou statistiques, de la classe à reconnaître

Une forme inconnue fera alors partie d'une classe donnée si elle est suffisamment proche d'un exemple de cette classe ou si elle en possède les caractéristiques essentielles.

2.1 Difficulté

- a. Trouver un ensemble représentatif d'exemples de chaque classe (cf la classe A majuscule)
- b. Trouver des caractéristiques les plus pertinentes possibles
- c. Etablir de bonnes règles de classification — encore appelées règles de décision —; une méthode classique est de faire un apprentissage par l'exemple (classification dite supervisée)

Chapitre 9 Théorie de Bayes

C'est l'approche statistique fondamentale du problème de la classification des formes

Section 1 Risque

Soit :

- a. $\Omega = \{ 1, 2, \dots, L \}$ les différentes classes (“ les états de la nature”)
- b. $A = \{ \alpha_1, \dots, \alpha_a \}$ les actions possibles
- c. $\lambda_{ij} = \lambda (\alpha_i | j)$ le coût obtenu quand l'état de la nature est j et l'action entreprise α_i .

On définit le risque conditionnel $R(\alpha_i | x)$ comme étant le coût moyen obtenu quand le vecteur des caractéristiques est x et l'action entreprise α_i .

$$R(\alpha_i | x) = \sum_{j=1}^L \lambda(\alpha_i | j) P(j | x) \quad (9.1)$$

Le risque global R est alors le coût moyen associé à une règle de décision d'actions α .

$$R = \int R(\alpha(x) | x) p(x) dx \quad (9.2)$$

Section 2 Règle de décision de Bayes

C'est la règle qui minimise le risque global R . D'après la relation (9.2), l'action choisie est celle qui rend $R(\alpha_i | x)$ minimal.

On supposera dans la suite que l'action α_i correspond à la prise de décision que l'état de la nature est l'état i . λ_{ij} est alors le coût obtenu quand on décide que l'état de la nature est i alors qu'en fait, il est j .

1 Cas de deux classes

La règle de décision de Bayes C s'écrit:

$$\begin{aligned} C(x) = 1 & \Leftrightarrow R(\alpha_1 | x) < R(\alpha_2 | x) \\ C(x) = 2 & \Leftrightarrow R(\alpha_2 | x) < R(\alpha_1 | x) \end{aligned} \quad (9.3)$$

D'après (9.1) :

$$\begin{aligned} R(\alpha_1 | x) &= \lambda_{11}P(1 | x) + \lambda_{12}P(2 | x) \\ R(\alpha_2 | x) &= \lambda_{21}P(1 | x) + \lambda_{22}P(2 | x) \end{aligned} \quad (9.4)$$

On obtient:

$$C(x) = 1 \Leftrightarrow (\lambda_{21} - \lambda_{11})P(1 | x) > (\lambda_{12} - \lambda_{22})P(2 | x) \quad (9.5)$$

Mais en pratique, $P(1 | x)$ et $P(2 | x)$ sont inconnues. On doit donc appliquer la formule de Bayes:

$$P(j | x) = \frac{p(x | j) P(j)}{p(x)}$$

$$p(x) = \sum_{j=1}^2 p(x | j) P(j) \quad (9.6)$$

D'où

$$C(x) = 1 \Leftrightarrow (\lambda_{21} - \lambda_{11}) P(1) p(x | 1) > (\lambda_{12} - \lambda_{22}) P(2) p(x | 2) \quad (9.7)$$

On remarquera que $\lambda_{21} \geq \lambda_{11}$ et $\lambda_{12} \geq \lambda_{22}$.

2 Probabilité d'erreur minimale

On suppose maintenant que toutes les erreurs entraînent les mêmes pertes. On peut donc prendre:

$$\begin{aligned} \lambda_{ij} &= 0 \text{ si } i = j \\ &= 1 \text{ si } i \neq j \end{aligned} \quad (9.8)$$

Alors:

$$R(\alpha_i | x) = \sum_{j=1}^L \lambda_{ij} P(j | x) = \sum_{j \neq i} P(j | x) = P(\text{erreur} | x) \quad (9.9)$$

On déduit de (9.2) et de (9.9) que le risque global est dans ce cas égal à la probabilité d'erreur. La règle de décision de Bayes minimise donc ici la probabilité d'erreur. Comme $R(\alpha_i | x) = 1 - P(i | x)$, cette règle s'écrit:

$$C(x) = i \Leftrightarrow P(i) p(x | i) > P(j) p(x | j) \quad \forall j \neq i \quad (9.10)$$

Remarque

En pratique, le problème le plus difficile à résoudre est de calculer les densités conditionnelles $p(x | i)$

Un cas "classique" est celui où ces densités sont normales. S'il n'y a qu'une seule caractéristique, on obtient:

$$p(x | i) = \frac{1}{\sigma_i \sqrt{2\pi}} e^{-\frac{1}{2} \left(\frac{x - \mu_i}{\sigma_i} \right)^2} \quad (9.11)$$

avec

$$\mu_i = E[x | i] \quad \text{et} \quad \sigma_i^2 = E\{[(x | i) - \mu_i]^2\} \quad (9.12)$$

Le problème est donc d'estimer μ_i , σ_i^2 $\forall i = 1 \dots L$

On se rappellera que des exemples de la classe i tendent à se regrouper autour de la moyenne avec un écart type σ_i . Environ 95 % de ces exemples tomberont dans l'intervalle : $|x - \mu_i| \leq 2\sigma_i$

Section 3 Exercice

Un radar de surveillance d'un porte-avions fournit des images d'avions amis (classe 1) ou ennemis (classe 2). Ces images sont ensuite analysées pour fournir un vecteur de caractéristiques x . On suppose connues les probabilités conditionnelles:

$$p(x | 1) = p(x | \text{avion ami})$$

$$p(x | 2) = p(x | \text{avion ennemi})$$

On suppose aussi que la probabilité d'apparition d'un avion ennemi est le double de celle d'un avion ami. Notant λ_{ij} le coût obtenu quand on décide que l'avion est de classe i alors qu'en fait il est de classe j , on fait:

$$\lambda_{11} = \lambda_{22} = 0$$

$$\lambda_{12} = 3$$

$$\lambda_{21} = 1$$

1

Pourquoi a-t-on choisi $\lambda_{12} > \lambda_{21} > 0$?

2

Quelle est la règle de décision de Bayes ?

