Chapitre 6

Vision stéréoscopique

6.1. Introduction

Un des objectifs de la vision par ordinateur est de reconstruire la structure tri-dimensionnelle (3D) de l'espace à partir d'une ou plusieurs images. La vision stéréoscopique utilise deux images prises avec deux caméras. Connaissant le modèle de projection de chaque caméra et la relation spatiale entre les deux caméras, il s'agit de calculer les coordonnées 3D d'un point à partir de ses deux projections dans les deux images.

Les problèmes d'estimation du modèle de projection et d'estimation de la relation spatiale entre les deux caméras ont été abordés au chapitre 5. Dans ce chapitre nous allons nous concentrer sur les problèmes d'appariement et de reconstruction.

Le problème d'appariement ou de mise en correspondance consiste à trouver un ensemble de couples ou de paires, chaque couple étant formé d'un point d'une image apparié avec un point de l'autre image. Ce problème a une nature combinatoire et la seule façon de réduire cette combinatoire est de définir une mesure de ressemblance entre un point d'une image et un point de l'autre image et de mettre en œuvre des contraintes qui ont trait à la géométrie du capteur stéréoscopique et à la structure de la scène et des objets observés.

Le problème de reconstruction est directement lié au problème de calibration du capteur stéréo qui a été abordé au chapitre 5.

Considérons par exemple la paire d'images stéréoscopiques de la figure 6.1. Chacune de ces deux images représente une projection différente de la même scène 3D. Les propriétés métriques ne se conservant pas par projection il n'est pas facile de comparer les éléments d'une image à ceux de l'autre image afin de décider si on peut les mettre en correspondance ou non. Il faut également tenir compte du phénomène d'occultation : un élément d'une image peut ne pas être vu dans l'autre image (et *vice versa*) soit par ce qu'il est caché par un autre élément, soit parce qu'il est dans le champ de vue d'une caméra et hors du champ de vue de l'autre caméra, soit enfin à cause de phénomènes physiques tels que les ombres et les reflets qui dépendent des positions relatives de la source, de l'objet observé et de la caméra.

Figure 6.1. *Une paire d'images stéréo. Cet exemple illustre les difficultés de mise en correspondance et notamment le phénomène d'occultation.*

La démarche calculatoire sera la suivante :

- établir un ensemble d'appariements entre les éléments des deux images. Chacun de ces appariements forme une mise en correspondance *locale* ;
- utiliser des contraintes afin de réduire l'espace de recherche d'une mise en correspondance *globale* ;
- mettre en œuvre une méthode de mise en correspondance globale.

Une mise en correspondance locale est tout simplement un appariement qui satisfait une mesure de ressemblance et qui est cohérent avec les contraintes découlant de la calibration du capteur. Une mise en correspondance globale est un ensemble d'appariements qui satisfont à d'autres contraintes qui seront explicitées dans ce chapitre à la section 6.5.

Dans la section suivante nous allons discuter du choix quant aux éléments image à mettre en correspondance. Ce choix dépend d'un certain nombre de critères qui seront explicités. Ensuite nous allons consacrer une section à la géométrie de la stéréo et à la reconstruction. Cette section reprend la formulation du chapitre 5 en introduisant une méthode permettant le calcul de la géométrie épipolaire sans passer par la calibration des deux caméras.

Nous allons ensuite détailler dans la section 6.4. toutes les contraintes qui peuvent réduire la taille de l'espace de recherche d'une mise en correspondance globale.

Finalement nous allons étudier quelques méthodes de mise en correspondance ainsi que quelques cas particuliers tels que la vision trinoculaire (stéréo avec trois caméras) et la détection d'obstacles avec un système stéréo.

6.2. Primitives stéréoscopiques

Le choix des éléments (ou primitives) image à mettre en correspondance est crucial. La primitive “idéale” est telle que (i) ces propriétés intrinsèques permettent une mesure de ressemblance fortement discriminante, (ii) elle permette la mise en œuvre efficace de contraintes stéréoscopique et (iii) la reconstruction 3D soit possible. Les techniques de segmentation d'images que nous avons étudiées permettent l'extraction d'une gamme variée de primitives telles que :

• *des points* (pixels, points d'intérêt, éléments de contour, points caractéristiques le long d'un contour, jonctions, etc.) ;

• *des segments* (segments de droite, arcs de cercle, portions de conique, etc.) ;

• *des régions.*

Sauf quelques cas particuliers, les régions sont mal adaptées à la vision stéréo et cela pour plusieurs raisons. Une région se trouvant sur la surface d'un objet donnera naissance, dans les deux images, à deux régions de taille et de forme différentes, ceci étant dû au phénomène de distorsion projective. On ne peut donc pas comparer deux régions dans deux images différentes provenant de la même région 3D, figure 6.2. Par ailleurs, en l'absence d'un appariement terme à terme entre des points à l'intérieur des deux régions, on ne pourra pas décider s'il s'agit d'une surface 3D gauche ou planaire, figure 6.2.

Les segments de droite ont été utilisés avec quelque succès mais leur utilisation est limitée aux scènes polyédriques [11], [8], [166] et [96]. Il est à noter qu'un segment de droite a la même représentation géométrique qu'un point de

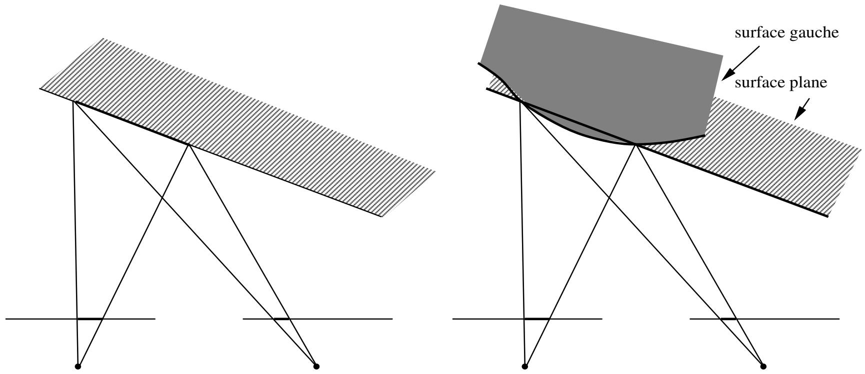

Figure 6.2. *Un objet incliné va donner naissance dans les deux images à deux régions de tailles différentes. Comment peut-on comparer ces deux régions ? (figure de gauche). En l'absence d'un appariement entre les points des deux régions on ne peut décider si cela correspond à une surface gauche ou à une surface plane (figure de droite). Ces phénomènes mettent sérieusement en échec les méthodes de mise en correspondance de régions.*

contour (position dans l'image et direction, figure 6.5.) et les mêmes contraintes stéréoscopiques s'appliquent également aux segments de droite et aux points de contour.

Bien que plus "pauvres" quant au contenu sémantique, les points s'avèrent finalement être les primitives les mieux adaptées pour la stéréo. La plupart des contraintes que nous allons étudier s'appliquent aux points, la contrainte figurale étant réservée aux points se trouvant le long d'un contour. Le seul vrai problème que pose l'utilisation des points comme primitives pour l'appariement est le choix d'une mesure de similarité (ressemblance) qui ne soit pas perturbée par la distorsion projective. A cette distorsion près, on peut comparer deux points (ou plus précisément deux régions carrées de taille égale) en utilisant une mesure de corrélation. Soit $\mathbf{p}$ un point d'une image de coordonnées $(u\ v)$ et $\mathbf{p}'$ un point de l'autre image de coordonnées $(u'\ v')$. Une mesure de similarité est le degré de corrélation entre les deux régions carrées entourant chacun de ces points :

$$c(\mathbf{p}, \mathbf{p}') = \sum_{x=-M}^{x=M} \sum_{y=-N}^{y=N} [I(u-x, v-y) - I'(u'-x, v'-y)]^2 \qquad [6.1]$$

pour une région de taille $2M \times 2N$. $I(u, v)$ est la valeur de niveau de gris au pixel de coordonnées u et v. La corrélation est nulle pour une ressemblance parfaite et strictement positive sinon. En développant l'expression de la corrélation de l'équation [6.1] et en négligeant les termes constants on obtient l'expression du produit de corrélation entre deux régions de deux images différentes centrées aux points $\mathbf{p}$ et $\mathbf{p}'$:

$$C(\mathbf{p}, \mathbf{p}') = \sum_{x=-M}^{x=M} \sum_{y=-N}^{y=N} [I(u-x, v-y) I'(u'-x, v'-y)] \qquad [6.2]$$

Le produit de corrélation est maximal pour une ressemblance parfaite. Généralement le produit de corrélation est appliqué aux images brutes (I et I') ayant éventuellement subi une opération de normalisation car la dynamique des deux caméras n'est pas forcément identique. Voir à ce propos l'étude comparative qui a été menée dans [7]. Une amélioration de la mise en correspondance par corrélation qui prend en ligne de compte une déformation affine est proposée dans [158].

Quel que soit le type de primitive choisi, une propriété importante est la fréquence de sa présence dans les deux images. La présence d'un grand nombre de primitives augmente la combinatoire mais en même temps *densifie* la reconstruction 3D qui en découle. Lorsque les primitives ne sont pas nombreuses la combinatoire est réduite mais la reconstruction sera d'autant moins bonne. Une représentation multi-résolution des images résout ce paradoxe de la façon suivante. Chaque image est décrite par un nombre croissant de primitives au fur et à mesure qu'on augmente la résolution. A faible résolution la combinatoire est réduite mais on a une reconstruction pauvre. L'appariement obtenu à faible résolution peut contraindre la combinatoire à haute résolution pour obtenir une reconstruction plus riche. Ce concept sera détaillé à la fin de ce chapitre (section 6.5.1.).

6.3. Géométrie épipolaire et reconstruction

Reprenons la présentation d'un capteur stéréo faite au chapitre 5. En particulier nous avons établi les équations [5.42] et [5.43]. Ces équations expriment la relation entre les coordonnées d'un point p$'$ de l'image de gauche, de son correspondant p dans l'image de droite, la matrice de transformation gauche/droite et la profondeur Z du point P de la scène qui se projette en p et p$'$.

En éliminant Z entre ces deux équations on obtient (rappelons que $\mathbf{p}$ est le vecteur de coordonnées $(x\ y\ 1)$ correspondant au point p) :

$$(b_z\,\mathbf{r}_2\cdot\mathbf{p} - b_y\,\mathbf{r}_3\cdot\mathbf{p})\,x' + (b_x\,\mathbf{r}_3\cdot\mathbf{p} - b_z\,\mathbf{r}_1\cdot\mathbf{p})\,y' = b_x\,\mathbf{r}_2\cdot\mathbf{p} - b_y\,\mathbf{r}_1\cdot\mathbf{p} \quad [6.3]$$

qui est l'équation d'une droite de la forme :

$$a'x' + b'y' + c' = 0 \quad [6.4]$$

Les $'$ indiquent qu'il s'agit d'une droite exprimée dans le repère de la caméra droite. Cette droite n'est rien d'autre que la contrainte épipolaire étudiée également au chapitre 5. On peut remarquer par ailleurs que les paramètres de cette droite, a', b' et c', ont les expressions suivantes :

$$\begin{aligned}
a' &= (b_y r_{31} - b_z r_{21})x + (b_y r_{32} - b_z r_{22})y + (b_y r_{33} - b_z r_{23})\\
b' &= (b_z r_{11} - b_x r_{31})x + (b_z r_{12} - b_x r_{32})y + (b_z r_{13} - b_x r_{33})\\
c' &= (b_x r_{21} - b_y r_{11})x + (b_x r_{22} - b_y r_{12})y + (b_x r_{23} - b_y r_{13})
\end{aligned}$$

6.3.1. La matrice essentielle

Les expressions que nous venons d'établir pour les paramètres a', b' et c' peuvent se mettre sous forme matricielle :

$$\begin{pmatrix} a' \\ b' \\ c' \end{pmatrix} = \begin{pmatrix} 0 & -b_z & b_y \\ b_z & 0 & -b_x \\ -b_y & b_x & 0 \end{pmatrix} \begin{pmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{pmatrix} \begin{pmatrix} x \\ y \\ 1 \end{pmatrix} \quad [6.5]$$

C'est donc la transformation épipolaire qui, à un point de l'image de gauche $(x\ \ y\ \ 1)^t$, fait correspondre une droite de l'image de droite décrite par ses paramètres $(a'\ b'\ c')$:

$$\begin{pmatrix} a' \\ b' \\ c' \end{pmatrix} = \mathbf{E} \begin{pmatrix} x \\ y \\ 1 \end{pmatrix} \quad [6.6]$$

L'équation de la droite épipolaire s'écrit :

$$(x'y'1) \begin{pmatrix} a' \\ b' \\ c' \end{pmatrix} = 0$$

ou encore :

$$(x'y'1)\mathbf{E}\begin{pmatrix} x \\ y \\ 1 \end{pmatrix} = 0$$

soit

$$\mathbf{p}'^t\mathbf{E}\mathbf{p} = 0 \qquad [6.7]$$

La matrice $\mathbf{E}$ décrit la *transformation épipolaire* gauche-droite – elle permet de calculer l'équation d'une droite épipolaire passant par l'image de droite associée à un point de l'image de gauche. On remarque que la transformation épipolaire droite-gauche est donnée par la matrice transposée :

$$\mathbf{p}^t\mathbf{E}^t\mathbf{p}' = 0 \qquad [6.8]$$

La matrice $\mathbf{E}$ s'appelle la *matrice essentielle*. Elle est le produit de deux matrices : une matrice antisymétrique (de rang 2) et une matrice orthonormée (de rang 3). On peut remarquer que cette matrice peut être calculée à partir des paramètres b_x, b_y, b_z et $\mathbf{r}_1$, $\mathbf{r}_2$, $\mathbf{r}_3$. Ces paramètres sont fournis lors de la calibration du capteur stéréoscopique.

6.3.2. La matrice fondamentale

La relation entre les coordonnées caméra et les coordonnées image est donnée par (pour la caméra de gauche) :

$$\begin{pmatrix} u \\ v \\ 1 \end{pmatrix} = \begin{pmatrix} \alpha_u & 0 & u_0 \\ 0 & \alpha_v & v_0 \\ 0 & 0 & 1 \end{pmatrix}\begin{pmatrix} x \\ y \\ 1 \end{pmatrix}$$

ou :

$$\mathbf{m} = \mathbf{C}\mathbf{p}$$

On a la même expression pour la caméra de droite :

$$\mathbf{m}' = \mathbf{C}'\mathbf{p}'$$

Les éléments des matrices $\mathbf{C}$ et $\mathbf{C}'$ sont les paramètres intrinsèques des deux caméras. En remplaçant dans l'équation [6.7] on obtient :

$$\mathbf{m}'^t(\mathbf{C}'^{-1})^t\mathbf{E}\mathbf{C}^{-1}\mathbf{m} = 0$$

La matrice :

$$\mathbf{F} = (\mathbf{C}'^{-1})^t \mathbf{E} \mathbf{C}^{-1} \qquad [6.9]$$

est la matrice fondamentale qui décrit la géométrie épipolaire. L'équation :

$$\mathbf{m}'^t \mathbf{F} \mathbf{m} = 0 \qquad [6.10]$$

n'est rien d'autre que l'équation d'une droite épipolaire dans le repère image (et non pas dans le repère caméra comme dans l'équation [6.7]).

Comme la matrice essentielle, la matrice fondamentale peut se calculer pour un capteur stéréo préalablement calibré.

6.3.3. Estimation de la matrice fondamentale

La contrainte épipolaire, ou plus généralement la géométrie épipolaire, joue un rôle fondamental en vision stéréoscopique. Il serait intéressant d'analyser dans quelle mesure il est possible d'estimer la matrice fondamentale *sans calibrer* le capteur stéréo, à partir de correspondances gauche-droite. En effet la matrice fondamentale peut s'estimer grâce à l'équation [6.10].

La méthode la plus directe est une méthode linéaire [83]. Pour chaque correspondance (m $\leftrightarrow$ m$'$) on a une équation linéaire du type :

$$f_{11}uu' + f_{12}vu' + f_{13}u' + f_{21}uv' + f_{22}vv' + f_{23}v' + f_{31}u + f_{32}v + f_{33} = 0$$

Si l'on dispose de n correspondances gauche-droite on obtient l'equation matricielle suivante :

$$A\mathbf{f} = 0 \qquad [6.11]$$

Dans cette équation A est une matrice de taille $n \times 9$ et $\mathbf{f}$ est un vecteur dont les composantes sont les 9 éléments de la matrice fondamentale. Afin d'éviter la solution triviale on peut rajouter la contrainte suivante :

$$||\mathbf{f}|| = 1$$

Il est donc possible de résoudre l'équation [6.11] avec 8 correspondances. On obtient alors une solution exacte. Avec plus de 8 correspondances ($n > 8$) on obtient une solution en minimisant un critère quadratique.

Si les données dont on dispose sont parfaites alors la matrice A est de rang 8. Il a été, en effet, montré que sauf configurations exceptionnelles, la matrice

A a un rang égal à 8 [117]. En pratique il y a du bruit est le rang de A est égal à 9. On peut alors estimer $\mathbf{f}$ en minimisant le critère suivant :

$$\min_{\mathbf{f}} \left(||A\mathbf{f}||^2 + \lambda(||\mathbf{f}||^2 - 1) \right) \qquad [6.12]$$

La solution de ce problème de minimisation est bien connue (voir plus loin l'équation [7.24]) et $\mathbf{f}$ est donné par le vecteur propre correspondant à la plus petite valeur propre de la matrice $A^t A$.

Une propriété importante de la matrice fondamentale (et de la matrice essentielle) est qu'elle a un rang égal à 2. La façon la plus pratique d'introduire cette contrainte est de remplacer la matrice $\mathbf{F}$ estimée comme auparavant, par une matrice $\mathbf{F}'$ qui a un rang égal à 2. Pour faire ceci, considérons une décomposition en valeurs singulières de $\mathbf{F}$:

$$\mathbf{F} = UD(r, s, t)V^t$$

Dans cette expression D est une matrice diagonale 3×3 qui contient les valeurs singulières (r, s, et t) de $\mathbf{F}$, avec $r \geq s \geq t$. On peut alors forcer le rang de $\mathbf{F}'$ à être égal à 2 :

$$\mathbf{F}' = UD(r, s, 0)V^t$$

Dans un article paru récemment Hartley [83] montre comment on peut rendre la solution linéaire encore plus robuste numériquement tout simplement en opérant un changement de repère sur les données. Une solution linéaire qui tient explicitement compte de la géométrie du problème et qui donne également des résultats très satisfaisants a été recemment proposée par Boufama et Mohr [25].

La matrice fondamentale peut également être estimée par des méthodes non-linéaires. On peut, par exemple, minimiser la somme des carrés des distances d'un point à la droite épipolaire qui est censée passer par ce point. En effet, l'équation [6.10] exprime le fait que le point m$'$ se trouve sur la droite épipolaire dont les paramètres sont donnés par :

$$\begin{pmatrix} a \\ b \\ c \end{pmatrix} = \begin{pmatrix} f_{11} & f_{12} & f_{13} \\ f_{21} & f_{22} & f_{23} \\ f_{31} & f_{32} & f_{33} \end{pmatrix} \begin{pmatrix} u \\ v \\ 1 \end{pmatrix}$$

Plus généralement, la distance d'un point m$'$ à cette droite est donnée par la forme quadratique suivante :

$$d(\mathbf{m}', \mathbf{Fm}) = \mathbf{m}'^t \mathbf{Fm}$$

Les paramètres a et b de la droite étant :

$$a = f_{11}u + f_{12}v + f_{13}$$

$$b = f_{21}u + f_{22}v + f_{23}$$

En pratique, Faugeras, Luong & Maybank [56] proposent de considérer simultanément la distance de $\mathbf{m}'$ à la droite $\mathbf{Fm}$ et la distance de $\mathbf{m}$ à la droite $\mathbf{F}^t\mathbf{m}'$ afin de réduire les différences entre la géométrie épipolaire gauche-droite et droite-gauche. Le critère à minimiser est donc le suivant :

$$\min_{\mathbf{F}} \left(\sum_{i=1}^{i=n} \frac{(\mathbf{m}_i'^t \mathbf{F} \mathbf{m}_i)^2}{a_i^2 + b_i^2} + \sum_{i=1}^{i=n} \frac{(\mathbf{m}_i^t \mathbf{F}^t \mathbf{m}_i')^2}{a_i'^2 + b_i'^2} \right) \qquad [6.13]$$

pour n correspondances gauche-droite. Les paramètres a' et b' sont donnés par :

$$a' = f_{11}u' + f_{21}v' + f_{31}$$

$$b' = f_{12}u' + f_{22}v' + f_{32}$$

Comme la matrice essentielle, la matrice fondamentale est de rang 2. On peut alors écrire cette matrice, sans perte de généralité, de la façon suivante (les trois lignes de la matrice sont linéairement dépendantes) :

$$\mathbf{F} = \begin{pmatrix} x_1 & x_2 & x_3 \\ x_4 & x_5 & x_6 \\ x_7x_1 + x_8x_4 & x_7x_2 + x_8x_5 & x_7x_3 + x_8x_6 \end{pmatrix}$$

Cette fois-ci la minimisation de Q (équation [6.13]) fournira donc une estimation des paramètres x_1 à x_8 à partir desquels on déduit facilement les coefficients de la matrice fondamentale. Cette minimisation implique l'utilisation de techniques d'optimisation non-linéaires telles que Newton, Gauss-Newton ou Levenberg-Marquardt.

L'estimation non linéaire de la matrice fondamentale est moins sensible au bruit que la méthode linéaire. Les deux méthodes (linéaire et non linéaire) supposent cependant un appariement parfait de points entre les deux images. S'il y a des appariements incorrects, aucune des méthodes ne donne des résultats satisfaisants. Pour pallier au problème d'appariements incorrects, des méthodes statistiques robustes peuvent être utilisées, comme dans [183].

6.3.4. Reconstruction

La reconstruction tri-dimensionnelle à partir d'une paire d'images suppose que l'on dispose de correspondances entre les deux images. Le type de reconstruction qu'on peut alors effectuer dépend du type de calibrage dont on dispose. On peut distinguer les cas suivants :

1. le capteur stéréoscopique est calibré et on dispose alors de paramètres internes de chaque caméra ainsi que de la transformation rigide entre les deux caméras – dans ce cas on obtient une reconstruction euclidienne dans le repère de calibrage ;

2. les paramètres internes de chaque caméra sont connus mais la transformation rigide entre les deux caméras n'est pas connue. Il faut alors estimer la matrice essentielle [55] à partir de laquelle on peut extraire la transformation rigide entre les deux caméra *à un facteur d'echelle près* – dans ce cas on obtient une reconstruction euclidienne dans le repère de l'une ou l'autre des deux caméras ;

3. si aucun calibrage n'est disponible (ni les paramètres internes des caméras, ni la transformation rigide entre les deux caméras) il faut alors estimer la matrice fondamentale comme il a été expliqué auparavant. A partir de la matrice fondamentale on peut obtenir une reconstruction *projective* tridimensionnelle [54].

Si on se place dans le premier de ces cas et si le point p de l'image de gauche a été mis en correspondance avec le point p$'$ de l'image de droite, on a alors deux ensembles d'équations [5.16] et [5.17] :

$$\begin{aligned}
u &= \frac{m_{11}X + m_{12}Y + m_{13}Z + m_{14}}{m_{31}X + m_{32}Y + m_{33}Z + m_{34}} \\
v &= \frac{m_{21}X + m_{22}Y + m_{23}Z + m_{24}}{m_{31}X + m_{32}Y + m_{33}Z + m_{34}} \\
u' &= \frac{m'_{11}X + m'_{12}Y + m'_{13}Z + m'_{14}}{m'_{31}X + m'_{32}Y + m'_{33}Z + m'_{34}} \\
v' &= \frac{m'_{21}X + m'_{22}Y + m'_{23}Z + m'_{24}}{m'_{31}X + m'_{32}Y + m'_{33}Z + m'_{34}}
\end{aligned}$$

Les coordonnées X, Y et Z du point P reconstruit, dans le repère de calibrage, se calculent en résolvant ce système de 4 équations linéaires.

On peut également reconstruire le point P dans le repère de la caméra de gauche en utilisant une ou l'autre des équations [5.42] ou [5.43]. Les coordonnées

X, Y et Z du point P seront données dans ce cas par :

$$\begin{aligned} x' &= \frac{(r_{11}x + r_{12}y + r_{13})Z + b_x}{(r_{31}x + r_{32}y + r_{33})Z + b_z} \\ X &= xZ \\ Y &= yZ \end{aligned}$$

Ces équations nous permettent de constater que le déplacement entre la caméra gauche et la caméra droite doit compter une translation. Si la transformation gauche-droite est une rotation pure, la reconstruction n'est pas possible.

Finalement on peut remarquer que lorsque les images ont été rectifiées la matrice décrivant la transformation gauche-droite, équation [5.45], est réduite à une translation et les équations [5.42] et [5.43] deviennent :

$$\begin{aligned} x' &= x \\ y' &= \frac{Zy + b}{Z} \end{aligned}$$

Le point P peut alors être reconstruit dans le repère rectifié de la caméra gauche grâce aux équations suivantes :

$$\begin{aligned} Z &= \frac{b}{y' - y} \\ X &= xZ \\ Y &= yZ \end{aligned}$$

et on remarque que, d'un point de vue géométrique, la reconstruction n'est rien d'autre qu'une triangulation.

En notant :

$$D = y' - y \qquad [6.14]$$

on définit la notion de *disparité* (différence de position entre les deux projections d'un point de la scène) qui est abondamment utilisée en stéréoscopie. Dans la section suivante on développera une contrainte d'appariement en relation avec la variation de la disparité.

6.4. Contraintes pour la mise en correspondance

La question à laquelle nous essayons de répondre dans cette section est la suivante : quels sont les critères à mettre en œuvre pour décider qu'une paire de primitives gauche/droite est correcte ou pas ? En essayant de répondre à cette question, nous allons développer deux types de contraintes. Un premier type de contraintes nous permettra de valider un appariement primitive-gauche/primitive-droite. Il s'agira des contraintes épipolaire et d'orientation. Un deuxième type de contraintes nous permettra de valider la compatibilité entre deux appariements satisfaisant le premier type de contraintes. Il s'agit des contraintes d'ordre, d'unicité, de gradient de la disparité et figurale.

Supposons par exemple qu'on veuille apparier le point p_1 de l'image de gauche avec le point p'_1 de l'image de droite et le point p_2 de l'image de gauche avec le point p'_2 de l'image de droite. Nous devrons procéder de la façon suivante :

- on vérifie tout d'abord que les points $\mathrm{p}_1/\mathrm{p}'_1$ et $\mathrm{p}_2/\mathrm{p}'_2$ sont en correspondance épipolaire ;

- on vérifie ensuite grâce à l'équation [6.2] que les régions entourant ces paires de points se ressemblent ;

- finalement on vérifie que les deux appariements sont compatibles au sens des autres contraintes.

6.4.1. La contrainte épipolaire

La contrainte épipolaire est la seule contrainte géométrique intrinsèque au capteur stéréoscopique. En pratique nous avons étudié deux méthodes permettant de calculer la géométrie épipolaire d'un couple de caméras : soit en calibrant les deux caméras par rapport à une mire commune, soit en estimant la matrice fondamentale à partir d'appariements gauche-droite.

La contrainte épipolaire peut s'appliquer aux points et aux segments de droite. Soit P un point de la scène et soient p et p$'$ ses projections dans les deux images.

Comment se déplace p' le long de la ligne épipolaire lorsque P se trouve plus ou moins loin du capteur ? Puisque le point p est fixe, P se déplace le long de la droite de vue passant par F et p, figure 6.3. Lorsque le point P se trouve près du capteur tout en restant visible dans les deux images, il se projette en a'. Lorsque P se trouve très loin du capteur, il se projette en q' défini tel que la droite $F'q'$ soit parallèle à la droite Fp. Analytiquement, la position de q'

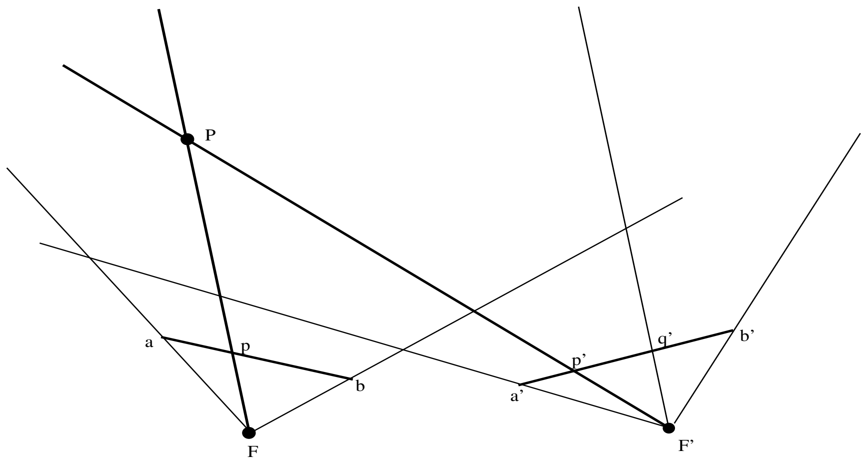

Figure 6.3. *Une vue d'un plan épipolaire : le point p' est contraint d'appartenir au segment $a'q'$.*

s'obtient des équations précédentes en faisant tendre Z vers l'infini :

$$\lim_{Z\to\infty} x' = \frac{\mathbf{r}_1 \cdot \mathbf{p}}{\mathbf{r}_3 \cdot \mathbf{p}}$$

$$\lim_{Z\to\infty} y' = \frac{\mathbf{r}_2 \cdot \mathbf{p}}{\mathbf{r}_3 \cdot \mathbf{p}}$$

p' et donc non seulement contraint le long de la ligne épipolaire mais il appartient également au segment $a'q'$.

Une analyse de la figure 6.3. révèle que les points de l'image de gauche voisins du point a ont intrinsèquement moins de correspondants que les points de l'image de gauche près du point b. Bien sûr, l'analyse gauche/droite que nous venons de faire est valide dans le cas droite/gauche.

La contrainte épipolaire s'applique facilement à des segments de droite. Il suffit pour cela de construire les droites épipolaires correspondant aux deux extrémités du segment. Remarquons toutefois la difficulté à définir une contrainte épipolaire pour des segments contenus dans le plan épipolaire, figure 6.4.

6.4.2. La contrainte d'orientation

Pour les primitives images auxquelles on s'intéresse plus particulièrement (des points de contours et des segments de droite), il y a, en plus de l'informa-

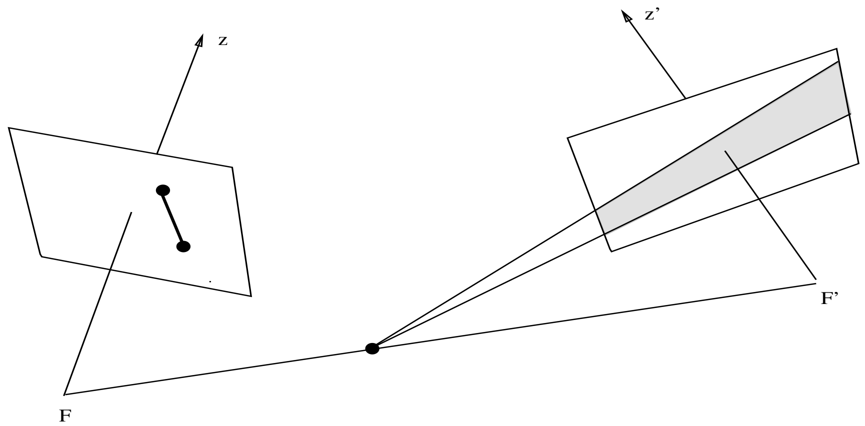

Figure 6.4. *Les deux lignes épipolaires associées aux deux extrémités du segment de l'image de gauche définissent une région dans l'image de droite.*

tion de position dans l'image, une information relative à son orientation par rapport au repère de l'image. Il s'agit de l'angle entre le vecteur directeur associé à cette primitive (dans le cas des points de contour il s'agit de la tangente au contour en ce point) et un des axes du repère de l'image. La valeur de cet angle peut être définie à 2π près en orientant le vecteur directeur par rapport aux valeurs moyennes des niveaux de gris de part et d'autre du contour, figure 6.5.

Soit maintenant une primitive spatiale définie dans le repère de gauche par sa position (X, Y et Z) et son orientation donnée par son vecteur directeur $\mathbf{V} = (V_x, V_y, V_z)^t$. La projection de cette primitive (point et vecteur) dans une image donne naissance à un point p et un vecteur $\mathbf{v} = (v_x, v_y, 0)$. On peut remarquer que le centre de projection F forme avec le point P et le vecteur $\mathbf{V}$ un plan, qu'on appellera *plan d'interprétation*. Le vecteur $\mathbf{v}$ est l'intersection de ce plan avec le plan image, figure 6.6. :

$$\mathbf{v} = \mathbf{k} \wedge (\mathbf{V} \wedge \overrightarrow{FP})$$

Le vecteur $\mathbf{k}$ est le vecteur normal au plan image. Avec le choix de coordonnées qui a été fait ce vecteur n'est autre que la direction de l'axe optique. En développant ce double produit vectoriel et tenant compte du fait que $X = xZ$ et $Y = yZ$ on obtient :

$$v_x = Z(V_x - xV_z)$$

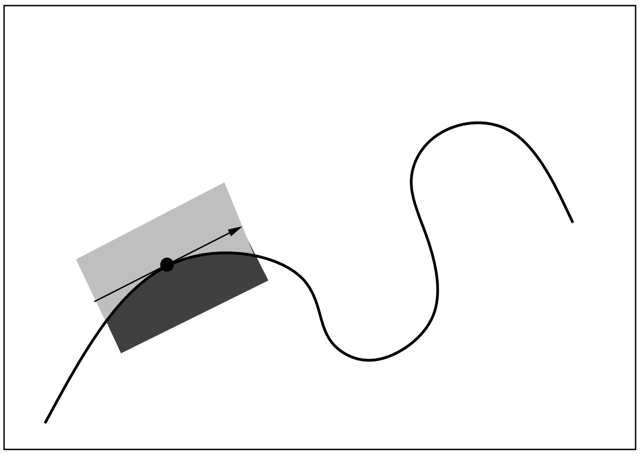

Figure 6.5. *Un point de contour est défini par sa position dans l'image, et la direction de la tangente au contour en ce point. Cette direction peut être définie à 2π près en tenant compte de la moyenne des niveaux de gris de part et d'autre du contour aux environs du point.*

$$\begin{aligned} v_y &= Z(V_y - yV_z) \\ v_z &= 0 \end{aligned}$$

La pente de cette primitive (v) est mesurable dans l'image et est donnée par :

$$\tan\theta = \frac{v_x}{v_y}$$

En exprimant la même chose dans le repère de l'image de droite on obtient une formule semblable pour la pente d'une primitive dans l'image de droite :

$$\tan\theta' = \frac{v'_x}{v'_y}$$

La relation entre les coordonnées du vecteur $\mathbf{V}$ exprimées dans les deux repères s'écrit ($\mathbf{A}$ est la matrice de transformation entre les deux caméras) :

$$\begin{pmatrix} V'_x \\ V'_y \\ V'_z \\ 0 \end{pmatrix} = \mathbf{A} \begin{pmatrix} V_x \\ V_y \\ V_z \\ 0 \end{pmatrix}$$

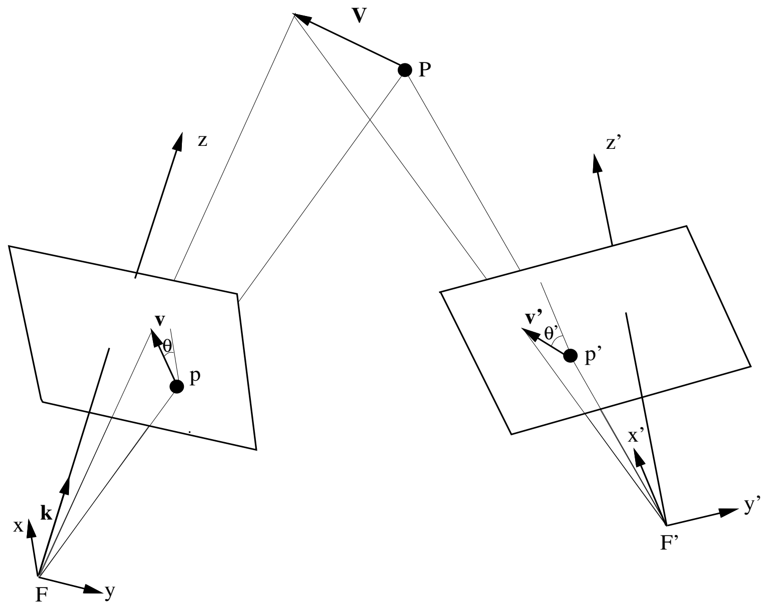

Figure 6.6. *La géométrie de la projection d'un vecteur dans les images.*

De plus on peut exprimer x' et y' en fonction de x et y grâce aux equations [5.42] et [5.43]. On peut ainsi obtenir une relation entre tan θ' et tan θ. Dans le cas général cette relation est inexploitable. Si les plans images et axes optiques sont parallèles (cas des images rectifiées) on a :

$$\begin{aligned} x' &= x \\ y' &= y + b/Z \\ V'_x &= V_x \\ V'_y &= V_y \\ V'_z &= V_z \end{aligned}$$

On obtient :

$$\tan\theta' = \frac{\tan\theta}{1 - \frac{b}{Z}\frac{V_z}{V_y - yV_z}}$$

ou plus simplement :

$$\tan\theta' = \frac{v_x}{v_y - bV_z}$$

On peut remarquer que si le vecteur **V** est parallèle au plan image, soit pour $V_z = 0$ on obtient :

$$\theta' = \theta$$

Dans ce cas la recherche d'un appariement gauche/droite est triviale. En effet, seules les primitives ayant la même inclinaison dans les deux images peuvent être appariées.

On peut également remarquer que plus le vecteur **V** sera incliné par rapport au plan des images (plus V_z est grand) et plus il y aura de différence entre θ et θ'.

Dans le cas géneral, on veut trouver une constante ε qui sera la borne de la différence des deux orientations :

$$| \theta - \theta' | < \varepsilon$$

Imposer une limite à la différence d'inclinaison entre les vecteurs se correspondant dans les deux images revient à rejeter les appariements qui correspondent à des vecteurs **V** fortement inclinés par rapport au plan des images. Cette limitation de la différence d'orientation dans les deux images est cohérente avec la limite du gradient de la disparité qui interdit également des surfaces trop inclinées par rapport au plan des images.

La valeur de la constante ε peut se calculer en fonction de la distribution des inclinaisons spatiales de **V** et pour une valeur de b (distance entre les centres de projection) donnée [4]. Choisissons une primitive de l'image de gauche. Plus ε sera grand, plus il y aura des appariements possibles entre cette primitive et des primitives de l'image de droite. Plus il y aura d'appariements gauche/droite ainsi formés plus la recherche d'une mise en correspondance globale sera complexe.

6.4.3. La contrainte d'ordre

Les contraintes épipolaires et d'orientation nous ont permis de réduire le nombre d'appariements entre des primitives de l'image de gauche et des primitives de l'image de droite. Soient maintenant deux appariements établis sur la base de ces contraintes : (g_1, d_1) et (g_2, d_2). Dans quelle mesure ces appariements sont compatibles entre eux ? La première contrainte qu'on va étudier est la contrainte d'ordre.

Nous allons nous intéresser à des primitives points. Soit P_1 un point et soient p_1 et p'_1 ses projections dans les deux images. Les directions de projection de ce point partitionnent le plan épipolaire en deux régions, une région (hachurée) intérieure et une région extérieure, figure 6.7. Soient P_2 et P_3 deux points de l'espace et soient p_2, p_3 et p'_2, p'_3 leurs projections respectives dans l'image de gauche et de droite. Le point P_2 se trouve dans la région intérieure alors que le point P_3 se trouve dans la région extérieure. Regardons l'ordre de projection de ces points dans les deux images.

- dans l'image de gauche l'ordre est p_3, p_1, p_2.
- dans l'image de droite l'ordre est p'_3, p'_2, p'_1.

Si on prend P_1 comme référence, les projections de P_3 apparaissent dans le même ordre dans les deux images alors que l'ordre de projection de P_2 est inversé.

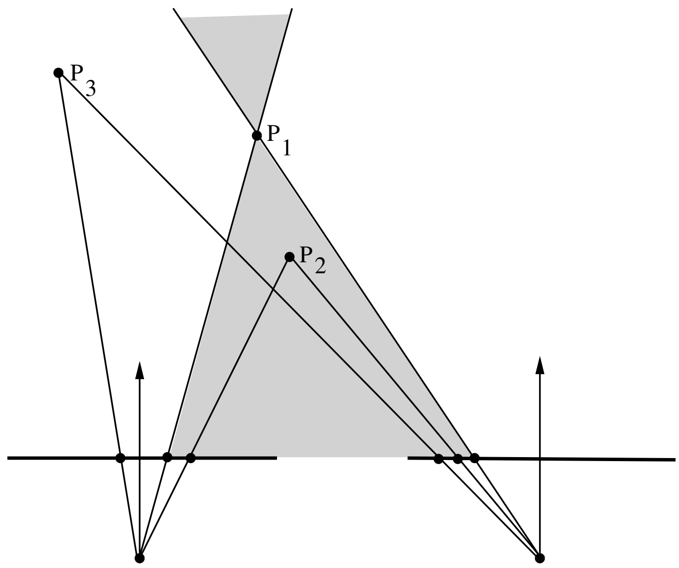

Figure 6.7. *P_1 et P_3 se projettent dans le même ordre dans les deux images alors que l'ordre des projections de P_1 et P_2 est inversé.*

Supposons maintenant que P_1 et P_3 se trouvent sur la surface d'un objet. Le fait que P_3 se trouve dans la région extérieure de P_1 correspond au fait que l'objet est relativement peu incliné par rapport au plan des images (il s'agit de l'objet 1 de la figure 6.8.). Soit maintenant l'objet 2 de la même figure sur lequel se trouvent les points P_1 et P_2. Le fait que le point P_2 se trouve dans la région intérieure de P_1 correspond au fait que l'objet 2 est très incliné par rapport au plan des images. Si cet objet est opaque les projections des deux points ne peuvent pas être vues à la fois dans les deux images. Si l'objet est transparent, les projections seront vues dans les deux images mais dans un ordre inversé.

En conclusion, la seule configuration pour laquelle l'ordre des projections est inversé est le cas de *surfaces transparentes fortement inclinées* par rapport au plan des images. Si on suppose donc que la scène observée contient peu (ou pas) de surfaces transparentes fortement inclinées, l'ordre des projections est le même dans les deux images.

Si on revient à l'exemple précédent et si on respecte la contrainte d'ordre on aura : l'appariement (p_1, p'_1) est compatible avec (p_3, p'_3) et *incompatible* avec (p_2, p'_2).

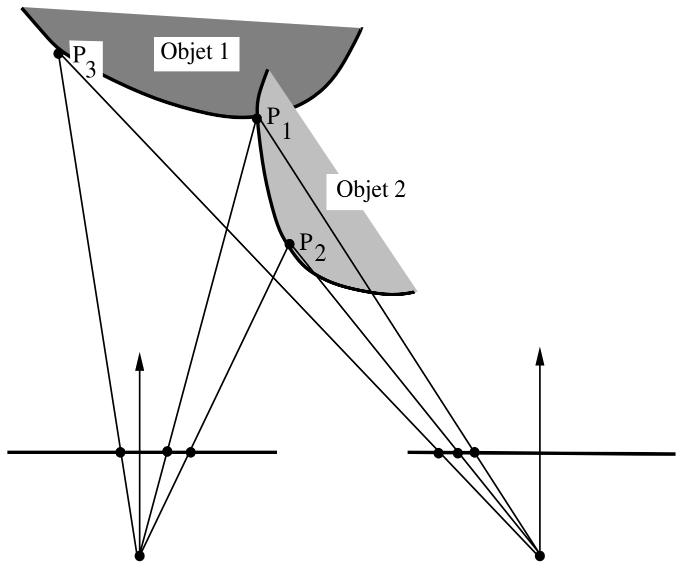

Figure 6.8. *Lorsque deux points se trouvent sur la surface d'un objet peu incliné (1) l'ordre de leurs projections est le même dans les deux images. Si l'objet est fortement incliné (2) les projections ne sont pas vues simultanément dans les deux images si l'objet est opaque ou l'ordre est inversé si l'objet est transparent.*

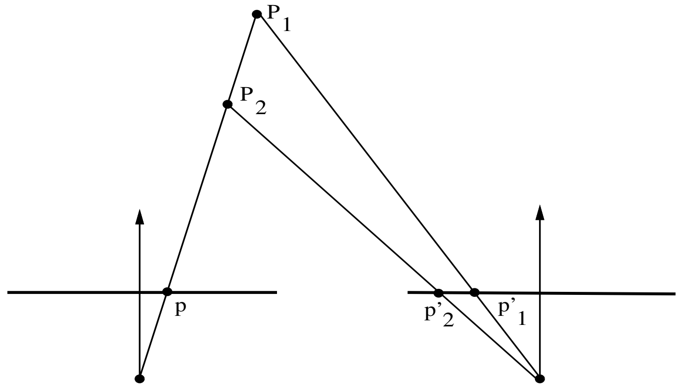

Figure 6.9. *La contrainte d'unicité n'est pas respectée lorsque deux points se projettent en un même point dans une image et en deux points différents dans l'autre image.*

La contrainte d'ordre a été utilisée pour la première fois par Baker et Binford [13] pour simplifier la combinatoire de la mise en correspondance. Une caractérisation analytique de cette contrainte a été proposée par Yuille et Poggio [182]. Cette caractérisation consiste essentiellement à interdire la compatibilité entre deux appariements qui placerait un point dans la région intérieure de l'autre point. Nous verrons plus loin que la contrainte d'ordre peut être exprimée dans le contexte plus général de la limite du gradient de disparité.

6.4.4. La contrainte d'unicité

Considérons maintenant un cas limite de la contrainte d'ordre. Ce cas est illustré par la figure 6.9. Le point P_2 se trouve sur la frontière des régions extérieure et intérieure du point P_1. Les deux points se projettent en un même point dans l'image de gauche et en deux points différents dans l'image de droite. Si on désigne par p la projection commune de P_1 et de P_2 dans l'image de gauche et par p'_1 et p'_2 leurs projections dans l'image de droite on doit avoir en même temps les appariements (p, p'_1) ET (p, p'_2). Accepter que ces deux appariements soient compatibles revient à augmenter la difficulté de la mise en correspondance car, à l'exception de l'alignement décrit plus haut, un point d'une image devrait avoir un et un seul correspondant dans l'autre image.

En pratique nous verrons au paragraphe suivant que la contrainte d'unicité peut, comme la contrainte d'ordre, être exprimée analytiquement dans le cadre de la limite du gradient de disparité.

6.4.5. La limite du gradient de disparité

Nous avons vu que la contrainte d'ordre est directement associée à l'inclinaison d'une surface observée par rapport au plan des deux images. On peut exprimer cette contrainte d'une autre manière. Reprenons pour cela la formule qui permet de calculer la profondeur d'un point en fonction de la disparité :

$$Z = \frac{b}{D} = \frac{b}{y' - y}$$

Soient deux points sur la surface d'un objet, P et Q et soient Z_P et Z_Q leurs profondeurs. Une autre façon d'exprimer le fait que la surface est peu inclinée est d'imposer que la différence entre ces deux profondeurs est faible. La relation avec la contrainte d'ordre est triviale. Les projections des points se trouvant sur un objet peut incliné (objet 1 de la figure 6.8.) ont le même ordre dans les deux images alors que les projections des points se trouvant sur un objet fortement incliné sont inversées dans les deux images.

La profondeur d'un point est associée à la disparité (la différence entre les positions des deux projections dans les deux images). La variation de la profondeur est donc associée à la variation de la disparité. Cependant la relation entre la variation de la disparité et la variation de la profondeur n'est pas facile à exploiter en pratique car la profondeur est inversement proportionnelle à la disparité. Burt et Julesz [31] ont défini le *gradient de disparité* et Pollard et al. [152], [153] ont établi une caractérisation analytique intéressante grâce à la limite du gradient de disparité.

Soient p et p' les projections du point P et q et q' les projections du point Q, P et Q ne se situant pas nécessairement dans le même plan épipolaire, figure 6.10. Le gradient de disparité est défini comme la différence des disparités de deux couples de points image divisée par la *séparation cyclopéenne*. Soit un repère commun aux deux images et soient les couples de points (p, p') et (q, q'). $\mathbf{p}$, $\mathbf{p}'$, $\mathbf{q}$ et $\mathbf{q}'$ sont les vecteurs associés à ces points dans le repère commun. Le gradient de disparité est défini par l'expression suivante :

$$G_d = \frac{||(\mathbf{p}' - \mathbf{p}) - (\mathbf{q}' - \mathbf{q})||}{||\frac{1}{2}(\mathbf{p}' + \mathbf{p}) - \frac{1}{2}(\mathbf{q}' + \mathbf{q})||} \qquad [6.15]$$

Le gradient de disparité peut s'écrire sous la forme suivante :

$$\frac{G_d}{2} = \frac{||(\mathbf{p}' - \mathbf{q}') - (\mathbf{p} - \mathbf{q})||}{||(\mathbf{p}' - \mathbf{q}') + (\mathbf{p} - \mathbf{q})||}$$

Le gradient de disparité est relié à la pente d'une droite passant par P et Q. La pente de cette droite rend finalement compte de l'inclinaison de la surface autour de ces points. Interdire des appariements gauche/droite qui correspondent à des surfaces fortement inclinées revient à imposer une limite supérieure au gradient de disparité. Si on choisit :

$$0 < \frac{G_d}{2} \leq k \text{ avec } 0 < k < 1 \qquad [6.16]$$

on obtient l'inégalité suivante :

$$||(\mathbf{p}' - \mathbf{q}') - (\mathbf{p} - \mathbf{q})|| < k||(\mathbf{p}' - \mathbf{q}') + (\mathbf{p} - \mathbf{q})||$$

Utilisant les propriétés classiques des inégalités :

$$\begin{aligned} ||a|| - ||b|| &\leq ||a - b|| \\ ||a + b|| &\leq ||a|| + ||b|| \end{aligned}$$

on obtient $(0 < k < 1)$:

$$\frac{||\mathbf{p}' - \mathbf{q}'||}{||\mathbf{p} - \mathbf{q}||} < \frac{1 + k}{1 - k} = \frac{1}{\lambda} \qquad [6.17]$$

Pour un point P fixé et pour une valeur de k on va chercher le lieu des points Q satisfaisant la contrainte exprimée par l'équation [6.17]. On cherche d'abord le lieu des points Q satisfaisant l'égalité :

$$\frac{||\mathbf{p}' - \mathbf{q}'||}{||\mathbf{p} - \mathbf{q}||} = \frac{1}{\lambda} \qquad [6.18]$$

Soit le repère commun aux deux images, le repère de l'image de gauche. Les points p et p' ont comme coordonnées : $\mathbf{p} = (x_1 \; y_1)^t$ et $\mathbf{p}' = (x_1 \; y_1')^t$. De même on a : $\mathbf{q} = (x_2\ y_2)^t$ et $\mathbf{p}' = (x_2\ y_2')^t$ Les coordonnées de q et q' sont reliées aux coordonnées de $Q = (X\ Y\ Z)^t$ par les formules suivantes :

$$\begin{aligned} x_2 &= X/Z \\ y_2 &= Y/Z \\ y_2' &= y_2 + b/Z \end{aligned}$$

L'égalité précédente, soit l'équation [6.18] devient alors :

$$\frac{(x_1 Z - X)^2 + (y_1 Z - Y)^2}{(x_1 Z - X)^2 + (y_1' Z - (b+Y))^2} = \lambda^2$$

Le lieu des points Q est finalement un cône dont le sommet se trouve en P, figure 6.10. :

$$\begin{aligned} &(1-\lambda^2)(X^2+Y^2) + (x_1^2 + y_1^2 - \lambda^2(x_1^2 + y_1'^2))Z^2 + 2(\lambda^2 - 1)x_1 XZ + \\ &2(\lambda^2 y_1' - y_1)YZ - 2\lambda^2 bY + 2\lambda^2 y_1' bZ - \lambda^2 b^2 = 0 \qquad [6.19] \end{aligned}$$

Revenant maintenant au lieu des points Q satisfaisant l'inégalité, soit l'équation [6.17], on peut voir facilement que ces points doivent se trouver à l'extérieur du cône. L'intérieur du cône est la zone *interdite* soit la région qui correspond à des surfaces trop inclinées par rapport au plan des images.

Pour conclure, on dira que deux appariements gauche/droite (p, p') et (q, q') sont compatibles si la valeur du gradient de disparité qui leur est associée est inférieure à 2. On peut vérifier que cette limite garantit également la contrainte d'ordre et la contrainte d'unicité.

6.4.6. La contrainte de continuité figurale

La contrainte de continuité figurale est une contrainte simple et efficace lorsque les points à mettre en correspondance se trouvent le long de contours.

La figure 6.11. montre deux images et des points le long de contours appartenant à ces images. Les lignes épipolaires sont également représentées sur cette figure. Les points appartenant à un contour de l'image de gauche ont été mis en correspondance avec des points de l'image de droite n'appartenant pas nécessairement au même contour.

Ce type d'erreur peut être facilement corrigé en imposant aux points de l'image de droite d'appartenir au contour possédant déjà une majorité de points

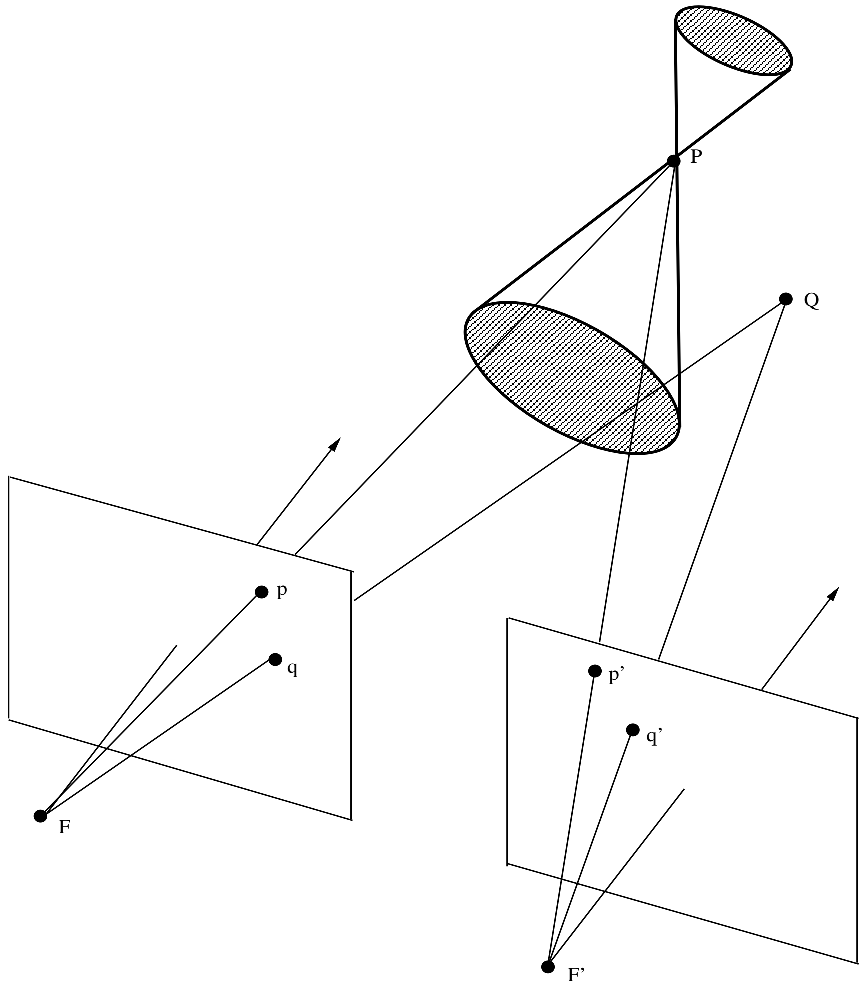

Figure 6.10. *Le lieu des points Q, tels que le gradient de disparité des appariements (p, p') et (q, q') est inférieur à 2, est la région extérieure à un cône dont le sommet se trouve en P.*

qui ont été mis en correspondance avec les points d'un contour de l'image de gauche.

Cependant Mohan, Medioni et Nevatia [125] remarquent à juste titre que ce genre de raisonnement *local* peut être source d'erreur. En effet, considérons la figure 6.12. Dans cette figure il y a une majorité de points appartenant à un contour de l'image de droite qui n'est pas le bon contour. La contrainte de continuité figurale doit donc être utilisée dans un contexte global et non pas local.

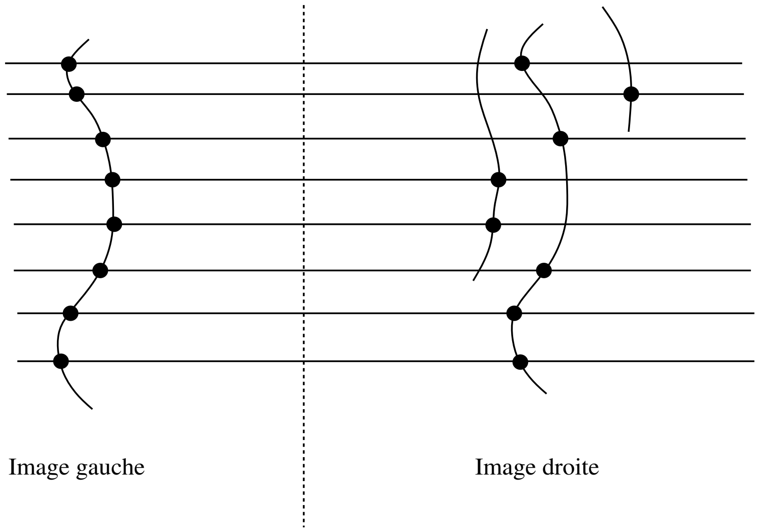

Figure 6.11. *L'erreur de mise en correspondance illustrée dans cette figure peut être corrigée en tenant compte de la continuité figurale.*

6.5. Mise en correspondance stéréo

Comme on peut s'en douter il n'existe pas actuellement une méthode de mise en correspondance suffisamment générale pour qu'elle puisse s'appliquer à une paire d'images stéréo indépendamment du contenu et du type de ces images. Des domaines d'application aussi variés que :

- reconstruction de terrain à partir d'images aériennes ;
- modélisation de scènes d'intérieur ;

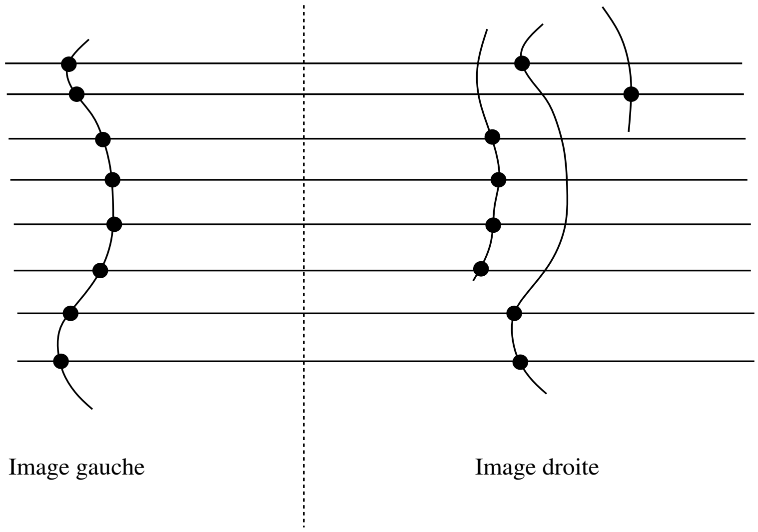

Figure 6.12. *La correction des erreurs de mise en correspondance selon un raisonnement local conduirait dans ce cas à un mauvais résultat.*

- modélisation de scènes d'extérieur ;
- images médicales et biomédicales ;

manipulent des images très différentes du point de vue de leur contenu.

Parmi le large spectre de méthodes proposées nous en avons sélectionné quatre que nous allons présenter en détail. Ces quatre méthodes sont les suivantes :

- *mise en correspondance hiérarchique.* Les images sont représentées à plusieurs niveaux de résolution, du niveau le plus "fin" au niveau le plus "grossier". Le résultat de la mise en correspondance au niveau le plus grossier est utilisé pour initialiser la mise en correspondance au niveau le plus fin.

- *mise en correspondance par programmation dynamique.* Cette méthode exploite trois contraintes parmi celles que nous avons développées : la contrainte épipolaire, la contrainte d'unicité et la contrainte d'ordre. Les points se trouvant le long de deux droites en correspondance épipolaire (qu'on appelle également deux droites épipolaires conjuguées) sont appariés d'une façon unique par programmation dynamique.

• *mise en correspondance par relaxation.* La contrainte épipolaire combinée avec une mesure de ressemblance permettent d'établir un certain nombre d'appariements. Un algorithme de relaxation permet ensuite de sélectionner, parmi cet ensemble d'appariements initiaux, un sous-ensemble d'appariements compatibles au sens des contraintes d'unicité et de limite du gradient de disparité.

• *mise en correspondance par isomorphisme de graphes.* Chaque image est vue comme un graphe. Le problème d'appariement se ramène alors à un problème d'appariement de deux graphes.

6.5.1. Mise en correspondance hiérarchique

Nous allons décrire une méthode qui a été développée à SRI International pour la reconstruction de terrain à partir de 2 vues aériennes [78], [79]. Les images utilisées ont une taille de 2048×2048 pixels et chaque image représente une surface au sol d'environ 2 kilomètres carrés. La hauteur du terrain varie entre 360 et 540 mètres. Il s'agit donc d'un terrain relativement plat.

A partir d'une paire d'images on produit une *hiérarchie* d'images. Une partie de cette hiérarchie est montrée sur la figure 6.13..

Cette hiérarchie consiste en deux *pyramides d'images*, une pyramide pour chaque image initiale [39]. Chaque pyramide contient l'image initiale ainsi que des images de taille 1024×1024, 512×512, 256×256, 128×128, 64×64, 32×32 et 16×16. Chaque pyramide a donc 8 niveaux, du niveau le plus fin (2048×2048) ou haute résolution au niveau le plus grossier (16×16) ou basse résolution. La construction de cette pyramide est simple : on réduit une image en moyennant la valeur des pixels dans un carré $N \times N$. Ainsi pour $N = 2$, un carré 2×2 d'une image est remplacé par un seul pixel dans l'image de résolution inférieure, et ainsi de suite. On peut remplacer ce moyennage par des techniques plus sophistiquées qui convoluent l'image avec des filtres gaussiens et en sous-échantillonnant pour réduire en même temps la taille de l'image (voir à ce sujet le chapitre 2).

La première étape de toute méthode de mise en correspondance est d'extraire de chaque image (gauche et droite) les points qui seront les candidats à l'appariement. La méthode, développée à SRI, extrait des points d'intérêt de l'image de gauche uniquement. Pour chacun de ces points on cherche un

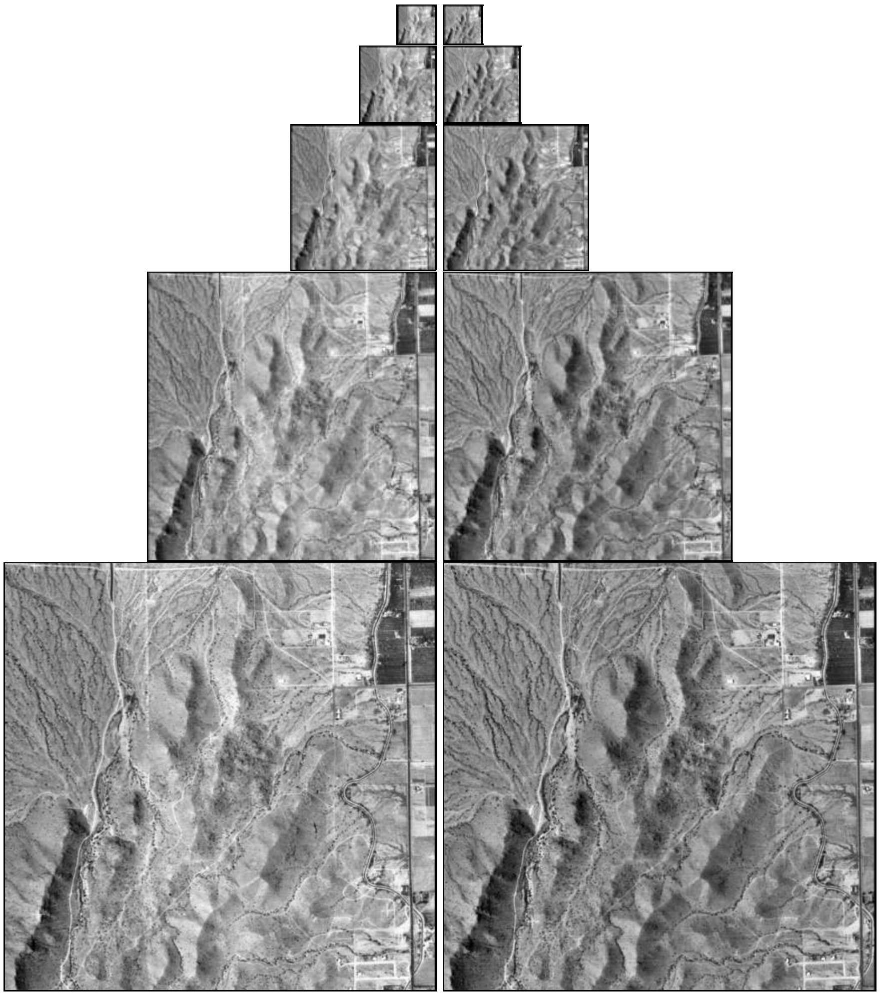

Figure 6.13. *Deux pyramides formées à partir d'une paire d'images à mettre en correspondance. Les images de cette figure sont de taille 512×512, 256×256, 128×128, 64×64 et 32×32.*

correspondant dans l'image de droite de la façon suivante (voir également la figure 6.14.)

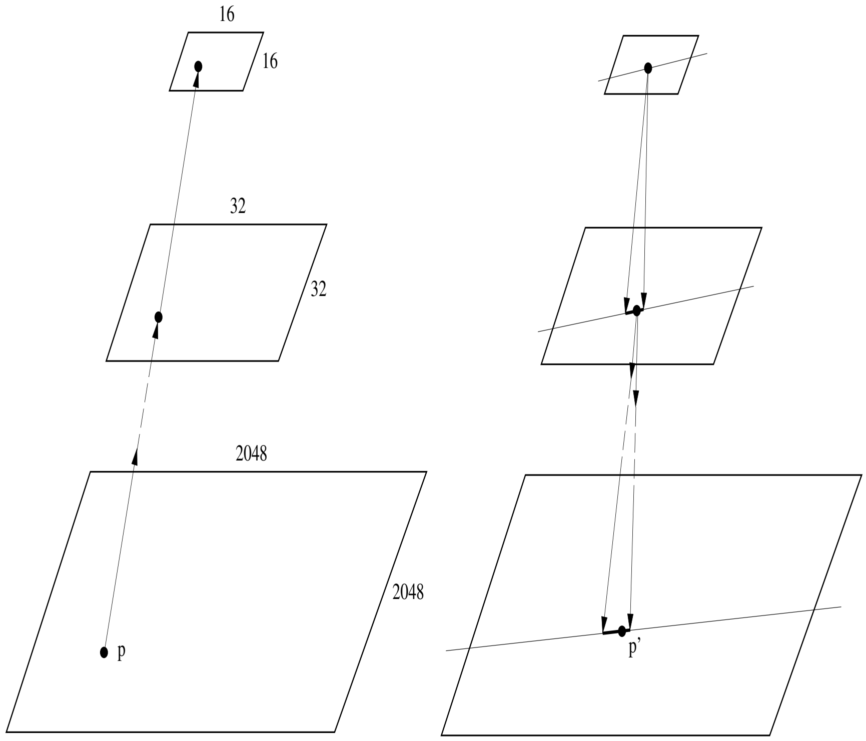

Figure 6.14. *L'algorithme de mise en correspondance hiérarchique développé à SRI utilise une représentation multi-résolution (pyramidale) des images, ce qui facilite la recherche d'appariements gauche/droite même lorsque la contrainte épipolaire est peu ou mal connue.*

Soit un point p de l'image de gauche haute résolution (2048×2048). Pour chaque image de la pyramide de gauche on calcule successivement les coordonnées de ce point, jusqu'à l'image de basse résolution (16×16). En pratique cela revient à diviser successivement par 2 les coordonnées du point p. On s'intéresse ensuite à la recherche d'une correspondance au niveau de résolution le plus faible. Pour cela on utilise la contrainte épipolaire et on cherche le long de la droite épipolaire le point de l'image de droite le plus ressemblant au point de l'image de gauche. On utilise le produit de corrélation comme mesure de ressemblance, équation [6.2]. Les images étant de taille réduite, une recherche du point de corrélation maximum le long de la ligne épipolaire n'est pas très coûteuse.

La prochaine étape consiste à parcourir la pyramide droite de l'image 16×16 vers l'image 2048×2048 de façon à combiner la contrainte épipolaire avec la prediction fournie par l'image basse résolution. Le point selectionné dans l'image 16×16 correspond à une région (de taille 2×2 dans ce cas-ci) dans l'image 32×32. L'intersection de cette région avec la contrainte épipolaire constitue la zone d'initialisation pour la recherche du point le plus ressemblant (au sens du produit de corrélation). Cette stratégie est poursuivie jusqu'à atteindre l'image 2048×2048.

On peut remarquer que cette méthode peut rester efficace lorsque la contrainte épipolaire est peu ou mal connue. En effet cela entraîne une recherche quasi exhaustive du meilleur correspondant dans l'image 16×16 ce qui n'est pas très coûteux. On peut donc utiliser cette méthode pour trouver des appariements afin de déterminer la géométrie épipolaire conformément à la section 6.3.3. de ce chapitre.

6.5.2. Appariement par programmation dynamique

La programmation dynamique est une technique utile pour apparier deux séquences tout en respectant l'ordre des éléments à l'intérieur de chaque séquence. Soient par exemple les séquences $\{a_1\ a_2\ a_3\ a_4\}$ et $\{b_1\ b_2\ b_3\ b_4\}$. Si l'élément a_2 est apparié avec b_2, les éléments à droite de a_2 ne pourront être appariés qu'avec les éléments à droite de b_2.

L'algorithme de Viterbi est une méthode de programmation dynamique qui trouve le meilleur appariement entre deux séquences parmi tous les appariements possibles. Les éléments des séquences à apparier définissent les deux dimensions d'une matrice. Chaque élément de cette matrice représente donc le coût d'appariement de deux éléments, soit une paire d'éléments. Un *chemin* consistera en une séquence de telles paires, chaque paire correspondant à un élément de la matrice. L'objectif sera de trouver un chemin optimal dans cette matrice. Le coût d'un chemin est la somme des coûts de ses éléments. Afin de trouver le chemin optimal (de coût maximal) deux contraintes sont utilisées : (i) l'ordre doit être respecté et (ii) chaque élément d'une séquence n'intervient qu'une seule fois.

Dans le cas de la stéréo, les deux séquences correspondent à des points le long de deux droites épipolaires conjuguées. Le coût de mise en correspondance de deux points peut être calculé à partir d'une mesure de ressemblance telle que celle donnée par l'équation [6.2]. Il est possible qu'un point d'une image n'ait pas de correspondant dans l'autre image. Dans ce cas le coût associé à une paire contenant un tel point est très faible voire nul. Le chemin optimal est une séquence d'éléments de la matrice. Un chemin est formé de transitions verticales (un point sans correspondant), horizontales (idem) ou diagonales (un appariement).

L'algorithme de Viterbi n'est pas nouveau. Il a été proposé pour la première fois pour comparer deux signaux en reconnaissance de la parole. Pour la mise en correspondance stéréo il a été utilisé par Baker et Binford [13] et par Ohta et Kanade [143] pour ne citer que les travaux les plus connus. Outre la contrainte épipolaire, la contrainte d'ordre et la contrainte d'unicité, on peut prendre également en compte la contrainte de continuité figurale [143].

6.5.3. Appariement par relaxation

Le problème de mise en correspondance stéréo peut être vu comme un problème d'étiquetage : aux points d'une image on associe les points d'une autre image. Le problème étant combinatoire, on est amené à faire "coopérer" les points d'une image entre eux en mettant en œuvre des relations découlant des contraintes géométriques et/ou physiques. Cette coopération peut être réalisée dans le cadre d'un processus de relaxation qui peut être décrit de la façon suivante (voir [40] pour une introduction à la relaxation) :

- un ensemble d'étiquettes (appariements) est sélectionné pour chaque point. Une mesure de confiance est ensuite associée à chaque étiquette ;

- les différentes étiquettes (avec leurs confiances) possibles pour chaque point sont comparées avec celles des points voisins. Cette comparaison est faite sur la base des relations entre les points voisins. Les étiquettes ainsi que leurs confiances sont ensuite modifiées itérativement pour réduire les inconsistances d'étiquetage entre points voisins.

De nombreux travaux ont proposé la relaxation comme méthodologie de mise en correspondance stéréo. Un algorithme "générique" est proposé dans l'ouvrage de Ballard et Brown [15]. On peut également se référer à la thèse de

Pascale Long-Limozin [112] qui utilise la relaxation pour mettre en correspondance des segments de droite.

Nous décrivons ici l'algorithme PMF (Pollard-Mayhew-Frisby) mis au point à l'Université de Sheffield et développé dans le cadre de la thèse de Stephen Pollard [151]. Cette méthode utilise la limite du gradient de disparité et la contrainte d'unicité. PMF procède en deux étapes : (i) une initialisation et (ii) une phase itérative. Pendant l'initialisation on calcule la confiance associée à chaque appariement. Cette confiance est calculée sur la base du “support” que cet appariement reçoit plusieurs autres appariements voisins. Le processus itératif de la deuxième phase utilise la contrainte d'unicité pour finalement classer les appariements initiaux en *corrects* et *incorrects*.

6.5.3.1. *La confiance d'un appariement*

Soit p un point de l'image de gauche et p′ son correspondant dans l'image de droite. Bien évidemment le couple p/p′ est compatible avec la contrainte épipolaire. La confiance d'un appariement peut être définie par l'expression suivante (voir également la figure 6.15.) :

$$S(p/p') = \sum_{i \in N(p)} \left(\max_{j'} \left(\frac{C(i,j')G(p/p',i/j')}{d(p,i)} \right) \right) \qquad [6.20]$$

Dans cette expression :

- $N(p)$ est un voisinage du point p dans l'image de gauche et i est un point de ce voisinage qui a au moins un correspondant j' dans l'image de droite ;
- $C(i, j')$ est le produit de corrélation défini par l'équation [6.2] et détermine la “qualité” de l'appariement i/j' ou la ressemblance entre ces deux points ;
- $d(p, i)$ est une fonction proportionnelle à la distance entre les points p et i ;
- G est le support dû au gradient de disparité associé aux appariements p/p' et i/j' et se calcule grâce à l'expression suivante (voir la section 6.4.5.) :

$$G(p/p',i/j') = \begin{cases} 1 & \text{si} \quad G_d \leq 2k \\ 0 & \text{si} \quad G_d > 2k \end{cases}$$

où G_d et k sont définis par les équations [6.15] et [6.16].

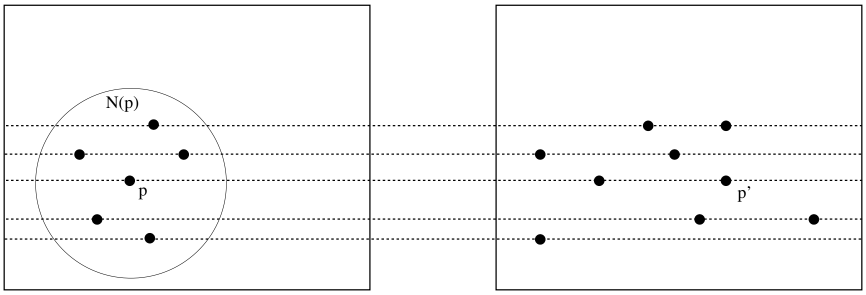

Figure 6.15. *La confiance d'un appariement p/p' peut se calculer à partir d'un voisinage $N(p)$ du point p et des appariements des points appartenant à ce voisinage. Les confiances respectives de p/p' (gauche-droite) et p'/p (droite-gauche) sont différentes dans le cas général car la structure locale des deux images n'est pas la même.*

On peut remarquer que l'expression définissant la confiance d'un appariement n'est pas symétrique, autrement dit dans le cas général on a :

$$S(p/p') \neq S(p'/p)$$

En effet, la configuration du voisinage de p dans l'image de gauche n'est pas la même que la configuration du voisinage de p' dans l'image de droite – surtout lorsque l'appariement p/p' est incorrect.

6.5.3.2. Appariement itératif

Les résultats expérimentaux obtenus par Stephen Pollard montrent que les confiances associées aux appariements corrects sont en général beaucoup plus élevées que celles associées aux appariements incorrects. Par conséquent, l'hypothèse faite dans PMF est que les appariements ayant une confiance maximale sont ceux qui sont physiquement corrects.

A chaque itération de l'algorithme PMF les appariements ayant la plus grande confiance pour les deux points les constituant – $S(p/p')$ et $S(p'/p)$ – sont choisis comme corrects. Ensuite, en accord avec la contrainte d'unicité, tous les autres appariements associés avec l'un ou l'autre des points des appariements choisis sont éliminés. Ceci permet à d'autres appariements qui on été ni acceptés ni été éliminés d'être choisis comme corrects parce qu'ils ont maintenant la plus grande confiance pour leurs deux points les constituant. La convergence – lorsqu'il n'y a plus que des appariements corrects ou éliminés – a lieu au bout de 4 à 5 itérations. En pratique Stephen Pollard a observé que 73% des appariements sont choisis lors de la première itération et 11% lors des 3 itérations suivantes. Les points non appariés n'ont pas, en général, de correspondant.

L'algorithme PMF peut être vu comme une forme de relaxation car à l'initialisation chaque point de l'image de gauche a un certain nombre d'étiquettes (appariements) possibles et à chaque étiquette est associée une confiance. Au cours des itérations de l'algorithme les confiances évoluent et à la convergence chaque point n'a plus que deux étiquettes possibles : apparié ou éliminé.

Plus formellement, chaque pas d'itération de cette méthode de mise en correspondance stéréo se décompose en les étapes suivantes :

Etape 1 : Pour chaque point de l'image de gauche on choisit un appariement dans l'image de droite qui maximise son support :

$$\max_{p'} S(p/p')$$

On obtient ainsi un certain nombre d'appariements gauche-droite ;

Etape 2 : Pour chaque point de l'image de droite on choisit un appariement dans l'image de gauche qui maximise son support :

$$\max_{p} S(p'/p)$$

On obtient ainsi un certain nombre d'appariements droite-gauche ;

Etape 3 : On compare les appariements gauche-droite avec les appariements droite-gauche et on conserve ceux qui sont identiques.

Etape 4 : Les points qui viennent d'être appariés ne peuvent plus participer à d'autres appariements (en vertu de la contrainte d'unicité) et il faut donc éliminer de la liste initiale d'appariements ceux qui contiennent les points qui viennent d'être appariés. Ceci aura pour effet de modifier le support des appariements restants.

Il y a donc, au cours de l'algorithme trois types d'appariements : retenus, éliminés et non traités. L'algorithme s'arrête dès qu'il n'y a plus d'appariements non traités.

6.5.3.3. Prise en compte de la continuité figurale

Il est possible d'incorporer la contrainte de continuité figurale dans l'algorithme PMF. En effet, après chaque itération on considère les points appariés de l'image de gauche appartenant à un même contour. Leurs correspondants doivent également en principe appartenir au même contour de l'image de droite. Si ce n'est pas le cas, on sélectionne le contour qui contient la majorité de ces points et on élimine les quelques appariements qui ne satisfont pas cette contrainte – voir figure 6.11.

On peut remarquer que ce processus a un caractère moins local que le processus de relaxation puisqu'un contour peut traverser toute l'image.

6.5.4. Appariement par isomorphisme de graphes

Le problème de mise en correspondance stéréo peut être vu comme un problème d'isomorphisme de graphes. En effet l'ensemble de primitives extraites d'une image peut constituerl'ensemble des nœuds d'un graphe. Des relations entre ces primitives constitueront alors les arcs (ou arêtes) de ce graphe. C'est l'approche qui a été suivie dans la thèse de Thomas Skordas [166], [96]. Dans le cadre de cette méthode les primitives à mettre en correspondance sont des segments de droite et les relations entre ces segments sont : (i) à gauche de, (ii) à droite de, (iii) colinéaire avec et (iv) a une jonction commune avec. Chaque image est donc décrite par un réseau de segments de droite et de relations entre ces segments, soit un graphe relationnel. Le problème de mise en correspondance est donc le problème de trouver le meilleur appariement entre les nœuds et les arcs de ces deux graphes.

Cependant, pour un certain nombre de raisons ces deux graphes ne sont pas identiques :

- un segment vu dans une image peut être caché ou partiellement caché dans l'autre image ;

- un segment "entier" dans une image peut être coupé en plusieurs petits segments dans l'autre image ;

- deux segments peuvent former une jonction (ou être colinéaires) dans une image et non pas dans l'autre image ;

• la relation d'ordre peut parfois être violée : un segment se trouvant à gauche d'un autre segment dans une image peut se trouver à droite de ce même segment dans l'autre image.

En conclusion, le problème d'appariement des deux graphes n'est pas simplement un problème d'isomorphisme de graphes mais plutôt un problème d'isomorphisme de sous-graphes maximal. Ceci dit, le problème est de chercher tous les isomorphismes entre sous-graphes d'un graphe et sous-graphes de l'autre graphe et de sélectionner ensuite le plus grand isomorphisme – celui qui met en correspondance le plus grand nombre de nœuds.

L'approche suivie dans la méthode de Thomas Skordas pour chercher le meilleur appariement entre deux graphes consiste à construire un *graphe d'association* ou *graphe de correspondance*. Le graphe d'association est construit de la manière suivante. Les contraintes épipolaires, de position et d'orientation (voir sections 6.4.1. et 6.4.2.) sont tout d'abord utilisées pour permettre la construction d'un certain nombre d'appariements potentiels. Ces appariements segment/segment sont les nœuds du graphe d'association. Ensuite on construit un arc entre deux nœuds dès lors que les deux nœuds (ou les deux appariements sont compatibles). La notion de compatibilité entre deux appariements découle de l'utilisation des contraintes d'ordre, d'unicité et figurale (voir les sections 6.4.3., 6.4.4., 6.4.6.) ainsi que d'un certain nombre de propriétés projectives.

Dans ce contexte, la meilleure correspondance entre les graphes des deux images devient équivalente au plus grand ensemble de nœuds mutuellement compatibles dans le graphe d'association. Dans un graphe, un ensemble de nœuds mutuellement compatibles forme un sous-graphe complétement connecté, ou une *clique* – dans une clique chaque nœud est connecté aux autres nœuds de la clique. Une clique maximale est une clique qui ne peut être incluse dans une autre clique. Une clique maximale dans le graphe d'association correspond donc à un isomorphisme de sous-graphes. La plus grande clique maximale correspond au plus grand nombre d'appariements possible.

Les figures 6.16. et 6.17. montrent le résultat de mise en correspondance sur quelques segments de droite des images de la figure 6.1.. Pour plus de détails voir [166] et [96].

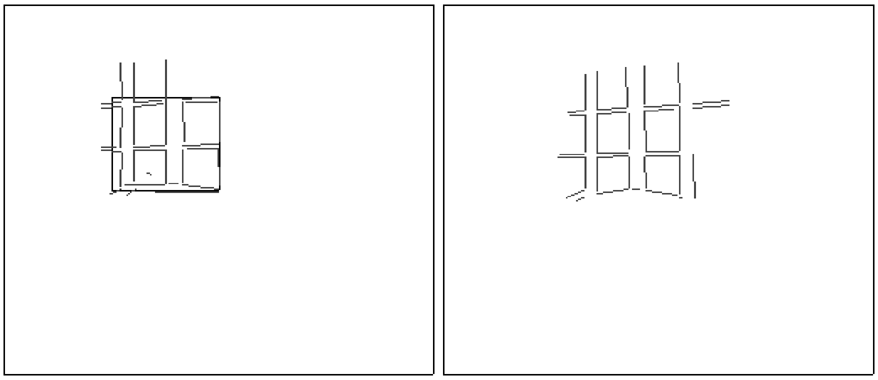

Figure 6.16. *Cette figure montre un ensemble de segments de l'image de gauche (segments se trouvant à l'intérieur d'un rectangle) ainsi que tous leurs correspondants potentiels dans l'image de droite.*

6.6. Systèmes stéréoscopiques particuliers

Les méthodes de mise en correspondance stéréoscopique que nous avons étudié comportent toutes une forme ou une autre de recherche des meilleurs appariements dans un ensemble fini. Malgré les contraintes mises en œuvre pour réduire au maximum l'espace de recherche, la complexité de ces méthodes reste prohibitive pour certaines applications où l'on ne dispose que d'une puissance de calcul limitée. Dans ce qui suit nous allons étudier deux cas particuliers de mise en correspondance stéréoscopique.

Dans le premier cas on utilise 3 caméras (et donc 3 images) au lieu de 2, ces 3 caméras étant "judicieusement" placées de façon que la recherche d'appariements devienne une simple vérification. La vision *trinoculaire* est couramment utilisée dans de nombreux projets de recherche et la présentation qui suit s'inspire des travaux de Nicholas Ayache et Francis Lustman [8], [118], [12].

Dans le deuxième cas on suppose que la scène observée est planaire et on peut alors décrire simplement la transformation qui permet de passer d'un point projeté dans une image à un point projeté dans l'autre image. Lorsqu'un objet est présent dans la scène (un obstacle posé sur un sol plat, par exemple) les projections des points de cet objet dans les deux images ne sont plus reliées par la transformation précédente. Encore une fois, une simple vérification permet de détecter des points de la scène qui n'appartiennent pas au plan de référence.

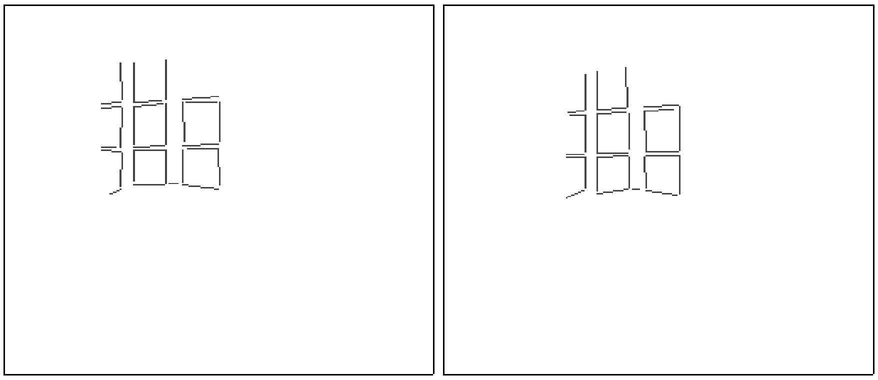

Figure 6.17. *Le résultat d'appariement obtenu sur les images de la figure précédente.*

6.6.1. Vision trinoculaire

La figure 6.18. illustre la géométrie épipolaire d'un capteur stéréoscopique avec 3 caméras. Un point P de la scène se projette en trois points p_1, p_2 et p_3 selon des droites passant par les centres de projection des caméras. Il y a 6 épipolaires associées à trois caméras et aux trois projections de P : l'épipolaire e_{ij} correspond à la projection de la droite Pp_i sur l'image j. Ces 6 épipolaires forment 3 paires d'épipolaires conjuguées : l'épipolaire e_{ij} est conjuguée avec l'épipolaire e_{ji}.

Lorsque p_1, p_2 et p_3 forment un appariement correct, le point p_i $(i \in \{1, 2, 3\})$ se trouve nécessairement à l'intersection des épipolaires e_{ki} et e_{ji}. Ceci nous conduit naturellement au processus d'appariement suivant : Soit un point q_1 de l'image 1 et un point q_2 de l'image 2. Si l'appariement q_1/q_2 est correct, alors q_3 (leur homologue dans l'image 3) se trouve précisément à l'intersection de deux épipolaires. On peut remarquer qu'un tel système est parfaitement symétrique et qu'on peut donc vérifier dans l'image k un appariement i/j.

6.6.2. Détection d'obstacles

Reprenons les équations [5.42] et [5.43]:

$$x' = \frac{Z\,\mathbf{r}_1 \cdot \mathbf{p} + b_x}{Z\,\mathbf{r}_3 \cdot \mathbf{p} + b_z} \qquad [6.21]$$

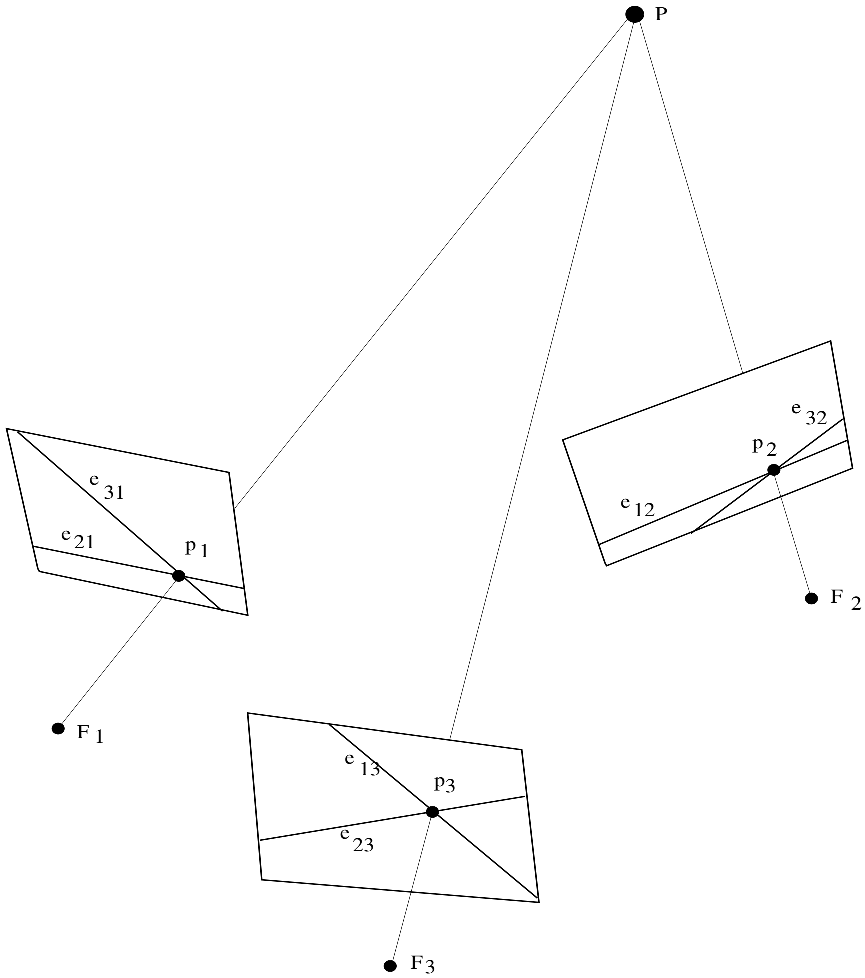

Figure 6.18. *Dans le cas d'un capteur stéréoscopique comportant 3 caméras on a 3 paires de droites épipolaires conjuguées et on peut ainsi réduire le problème de mise en correspondance de points à une vérification.*

$$y' = \frac{Z\,\mathbf{r}_2 \cdot \mathbf{p} + b_y}{Z\,\mathbf{r}_3 \cdot \mathbf{p} + b_z} \tag{6.22}$$

qui expriment la relation entre les coordonnées x' et y' d'un point p' d'une image et un point p de l'autre image de coordonnées x et y et $\mathbf{p} = (x\,y\,1)^t$.

Supposons de plus que la scène est une surface plane. Les coordonnées du point P vérifient alors l'équation :

$$\alpha X + \beta Y + \gamma Z = 1$$

et de plus on a $x = X/Z$ et $y = Y/Z$ en substituant on obtient :

$$\frac{1}{Z} = \alpha x + \beta y + \gamma = \mathbf{n} \cdot \mathbf{p}$$

avec $\mathbf{n} = (\alpha\;\beta\;\gamma)^t$.

Les équations [6.21] et [6.22] peuvent maintenant s'écrire :

$$x' = \frac{\mathbf{r}_1 \cdot \mathbf{p} + b_x(\mathbf{n} \cdot \mathbf{p}}{\mathbf{r}_3 \cdot \mathbf{p} + b_z(\mathbf{n} \cdot \mathbf{p})}$$

$$y' = \frac{\mathbf{r}_2 \cdot \mathbf{p} + b_y(\mathbf{n} \cdot \mathbf{p})}{\mathbf{r}_3 \cdot \mathbf{p} + b_z(\mathbf{n} \cdot \mathbf{p})}$$

Ceci peut s'écrire sous forme matricielle de la façon suivante ($\mathbf{n}^t = (n_x\,n_y\,n_z)$) :

$$\begin{pmatrix} sx' \\ sy' \\ s \end{pmatrix} = \left(\mathbf{R} \;+ \mathbf{b}\mathbf{n}^t\right) \begin{pmatrix} x \\ y \\ 1 \end{pmatrix}$$

soit plus simplement :

$$\begin{pmatrix} sx' \\ sy' \\ s \end{pmatrix} = \mathbf{N} \begin{pmatrix} x \\ y \\ 1 \end{pmatrix}$$

La matrice $\mathbf{N}$ est une matrice 3×3 exprimant la transformation projective entre un point d'une image et son correspondant dans l'autre image dans l'hypothèse où le point P de la scène appartient à un plan. On a donc :

$$\mathbf{N} = \left(\mathbf{R} \;+ \mathbf{b}\mathbf{n}^t\right) = \begin{pmatrix} n_{11} & n_{12} & n_{13} \\ n_{21} & n_{22} & n_{23} \\ n_{31} & n_{32} & n_{33} \end{pmatrix} \tag{6.23}$$

Les coefficients de la matrice $\mathbf{N}$ peuvent se calculer, à un facteur multiplicatif près, avec seulement 4 points de correspondance. En effet, chaque correspondance fournit 2 contraintes linéaires quant aux coefficients de la matrice $\mathbf{N}$ et pour n correspondances ($n \geq 4$) on a à résoudre un système linéaire surcontraint.

Supposons, par exemple, qu'on ait déterminé la matrice $\mathbf{N}$ pour un système stéréo et par rapport au plan du sol, figure 6.19. Deux points p et p' correspondant aux projections d'un point P appartenant au sol vérifient donc la relation :

$$p' = \mathbf{N}p$$

alors que les points q et q' qui ne correspondent pas à un point du sol ne la vérifient pas :

$$q' \neq \mathbf{N}q$$

Ceci fournit un moyen simple de détection d'obstacles pour un robot mobile se déplaçant sur un sol plat. Néanmoins cette méthode de détection d'obstacles suppose que la mise en correspondance ait été obtenue au préalable.

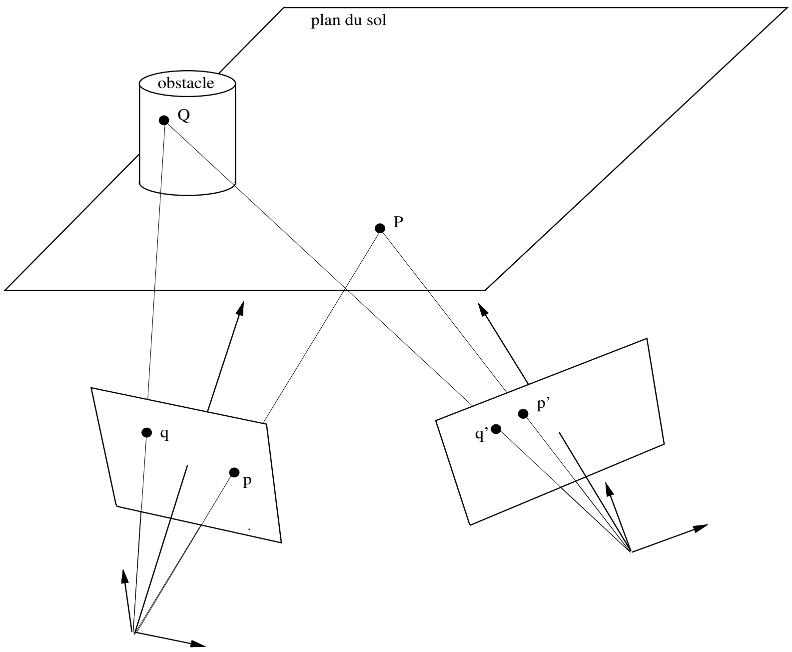

Figure 6.19. *En calibrant un système stéréo par rapport au plan du sol on peut détecter des obstacles à condition que l'appariement gauche/droite ait été préalablement établi.*