



Institut National
Polytechnique de Lorraine
Inria Lorraine

Centre de Recherche en
Informatique de Nancy

~~SEN 89~~
~~443~~ A

Stratégie de localisation et identification d'objet à partir de quelques mesures tridimensionnelles

THÈSE

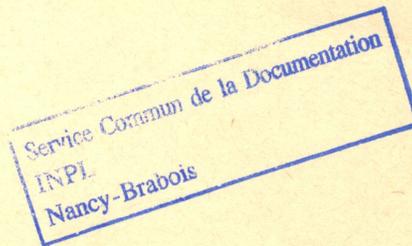
présentée et soutenue publiquement le 10 Juillet 1989

pour l'obtention du

Doctorat de l'Institut National Polytechnique de Lorraine
(Spécialité Informatique)

par

Shijia QIANG



Composition du jury :



D 136 037384 5

Pierre MARCHAND

Jim CROWLEY
Pierre MARCHAND

Examineurs :

Abdel BELAÏD
Roger MOHR

136037 3245

Institut National
Polytechnique de Lorraine
Inria Lorraine

Centre de Recherche en
Informatique de Nancy

[M] 1989 QIANG, S

Stratégie de localisation et identification d'objet à partir de quelques mesures tridimensionnelles

THÈSE

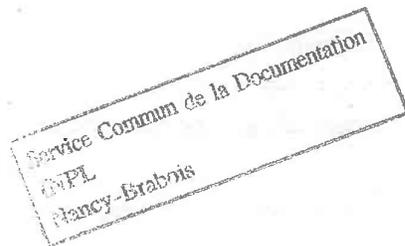
présentée et soutenue publiquement le 10 Juillet 1989

pour l'obtention du

Doctorat de l'Institut National Polytechnique de Lorraine
(Spécialité Informatique)

par

Shijia QIANG



Composition du jury :

Président : Pierre MARCHAND

Rapporteurs : Jim CROWLEY
Pierre MARCHAND

Examineurs : Abdel BELAÏD
Roger MOHR

Je tiens à exprimer ma profonde gratitude à Monsieur Roger MOHR, Professeur à l'Institut National Polytechnique de Lorraine, qui a bien voulu m'accueillir au sein de son équipe, et m'a suivie et guidée de manière spontanée, Je lui suis particulièrement reconnaissante pour ses conseils, ses encouragements et également pour sa patience. Ses conseils et son soutien m'ont été très précieux dans l'accomplissement de ce travail.

Je tiens à remercier profondément Monsieur le Professeur Pierre Marchand, Directeur du Département de l'informatique de l'Université de Nancy 1, d'avoir bien voulu consacrer une partie de son temps à la lecture et à la correction de ce mémoire et d'avoir accepté d'être rapporteur de cette thèse.

Je remercie très sincèrement Monsieur Jim Crowley, Professeur au LIFIA à Grenoble, du grand honneur qu'il me fait en acceptant de se déplacer à Nancy et d'être rapporteur du présent travail.

Je voudrais témoigner ma reconnaissance à Monsieur Abdel Belaïd, Chargé de Recherche au CNRS, d'avoir bien voulu accepter de participer à ce jury.

Je tiens également à exprimer ma reconnaissance à Monsieur Jean-Paul Haton, Professeur à l'Université de Nancy 1 et Directeur de Recherche INRIA au CRIN, pour m'avoir accueilli dans le laboratoire Reconnaissance des Formes et Intelligence Artificielle du CRIN, et également pour ses encouragements.

J'adresse mes plus vifs remerciements à toute l'équipe Vision du CRIN plus spécialement à Eric Thirion, Karl Tombre, Brigitte Wrobel-Dautcourt, Djamel Ziou pour avoir mis à ma disposition les outils logiciels qu'ils ont développés, et à Suzanne Collin, Claude Ingleber pour leur aide et leur sympathie.

Merci également à tous mes collègues et amis du CRIN pour l'atmosphère de travail sympathique dont j'ai bénéficié.

Résumé

L'objectif du présent travail est d'étudier la faisabilité d'un système d'identification et de localisation d'objet polyédrique à partir de quelques mesures tridimensionnelles.

Notre travail se compose en trois étapes : calibrage d'un système d'acquisition tridimensionnelle, modélisation d'objet à partir de plusieurs vues de l'objet, et identification et localisation d'objet polyédrique à partir de données tridimensionnelles éparses de la scène.

Le calibrage du système d'acquisition consiste à établir la transformation entre des coordonnées des points de l'espace tridimensionnel et celles de leurs images correspondantes. Les paramètres de cette transformation sont estimées par un processus d'optimisation.

Pour construire le modèle d'objet, plusieurs vues de cet objet sont nécessaires. Pour chaque vue de l'objet, une segmentation est effectuée pour obtenir une partition en région représentée par un graphe d'adjacence, le modèle complet de l'objet est obtenu par fusion des graphes correspondant aux différentes vues de l'objet.

L'identification et localisation s'effectuent par la mise en correspondance des données acquises et le modèle. La principale difficulté se révèle dans le processus de mise en correspondance par l'absence des primitives invariant de la position et de l'orientation de l'observateur. Avec des données éparses de la scène, il y a souvent plusieurs interprétations possibles qui sont toutes compatibles avec des données actuellement acquises, d'où une stratégie de prise de vue supplémentaire pour identifier l'interprétation correcte. Les fausses hypothèses sont rejetées par vérification avec des données acquises supplémentaires.

Mots clés :

- capteur 3D - calibrage - image de distance
- segmentation - appariement - modélisation
- reconnaissance - localisation -

Abstract

The aim of the present work is of studying the feasibility of a system for recognizing and locating polyhedral object from sparse light-stripe data.

Our work composes of three stages: calibration of the three dimensional scanning system, modelling of the object from multiple views, recognizing and locating polyhedral object from sparse range data.

The calibration of the 3D scanning system consists in establishing the transformation between 3-D world space coordinates of a point and their 2-D image pixel coordinates. The parameters of the transformation are estimated by a optimization procedure.

For constructing the model of a object, several views of the object are necessary. For each view of the object, a segmentation procedure is performed for obtaining a region partition represented by a adjacency graph, the complete model of the object is constructed by merging all the graphs corresponding to different views of the object.

The recognizing and locating of the object are accomplished by matching the scanned light-stripe data with the model. The principle difficulty involves in the matching procedure in the absence of the invariant features with respect to the position of the viewpoint. With sparse range data, it has usually more than one consistent interpretation generated by the matching procedure, the strategy for disambiguating feasible interpretations by additional scanning is made. Wrong interpretations are discarded through verification with further scanned data.

Key words :

- *range sensor - calibration - range image*
- *segmentation - matching - modelling*
- *recognizing - locating -*

Table des matières

1	Introduction	1
2	Systèmes d'acquisition tridimensionnelle	5
2.1	Systèmes passifs	5
2.2	Systèmes actifs	6
2.3	Choix des méthodes	9
3	Calibrage	11
3.1	Introduction	11
3.2	Principe du Système	13
3.3	Modèle de Caméra	15
3.3.1	Transformation du Repère Caméra au Repère Global	18
3.4	Paramètres du calibrage en cas général	19
3.5	Calcul des coordonnées	20
3.6	Discussion des hypothèses et corrections	24
3.6.1	Sources d'erreurs et corrections	24
3.7	Estimation des paramètres	29
3.7.1	Critère d'optimisation	29
3.7.2	Technique d'optimisation	30
3.8	Résultats expérimentaux	31
3.8.1	Considérations pratiques	31
3.8.2	Mesure des données — Etalonnage	36

Table des matières

3.8.3	Valeurs initiales des paramètres	39
3.8.4	Résultats	40
4	Techniques de prétraitement	43
4.1	Introduction	43
4.2	Binarisation	44
4.3	Squelettisation	46
4.4	Approximation Polygonale	51
4.5	Organisation des données 3D	57
5	Modélisation des objets	59
5.1	Généralité	59
5.1.1	Représentation d'objet tridimensionnel	59
5.1.2	Construction de modèle	63
5.2	Segmentation de l'Image de Distance	67
5.2.1	Segmentation par extraction de contour	68
5.2.2	Segmentation par extraction de région	70
5.2.3	Notre approche	71
5.3	Obtention de Modèle	88
5.3.1	Acquisition de plusieurs vues	88
5.3.2	Construction de graphe d'adjacence	89
5.3.3	Intégration des différentes vues — fusion des graphes	91
5.3.4	Obtention de modèle complet	96
5.3.5	Résultats	99
6	Reconnaissance et localisation des objets polyédriques	105
6.1	Introduction	105
6.1.1	Les différentes approches	106
6.1.2	Reconnaissance avec peu d'informations	110

Table des matières

6.1.3	Notre approche	112
6.2	Génération des hypothèses	114
6.2.1	Représentation de modèle	114
6.2.2	Mise en correspondance	114
6.2.3	Estimation des transformations	126
6.3	Vérification des hypothèses	127
6.3.1	Stratégie de prise des vues supplémentaires	128
6.3.2	Vérification	133
6.3.3	Résultats	135
7	CONCLUSION	147
A	Reconstitution d'une image de distance à partir des données 3D151	
B	Paramétrisation d'un segment de droite de longueur fixe dont les deux extrémités se situent sur deux arêtes connues	156
C	Courbes différentielles	160
	Bibliographie	163

1

Introduction

L'identification et localisation des objets dans l'espace sont des problèmes importants en vision par ordinateur. Ses applications industrielles sont multiples : inspection, la saisie et le tri automatique des pièces, robotique, etc. Les systèmes de vision bidimensionnelle ont connu ces dernières années un grand succès dans certaines de ces applications. Cependant la capacité de tels systèmes est limitée : ces systèmes ne peuvent distinguer que des formes bidimensionnelles, et en cas d'occultation, le problème devient difficile. En effet, l'interprétation d'une image d'intensité est une tâche difficile du fait que l'image est une projection de l'espace 3D dans l'espace 2D, les informations de profondeur sont perdues lors du processus de l'acquisition d'une part et les intensités dans une image sont fonctions de plusieurs facteurs tels que les positions et les orientations des surfaces, les réflectances des surfaces et la position de la source lumière d'autre part. Pour cette raison, au cours de ces dernières années s'est développée une grande tendance d'introduire directement des informations tridimensionnelles pour identifier des objets dans l'espace [Bolles 86], [Gordon 88], [Faugeras 86], [Nitzan 88], [Yang 85]. Les données tridimensionnelles peuvent être directement représentées par une image de distance.

Dans la littérature, de nombreuses techniques sont proposées pour identifier et localiser des objets en utilisant l'image de distance. Leurs méthodes sont fondées sur l'extraction des indices invariants par rapport à la position et l'orientation du capteur tels que l'angle entre des normales de facettes [Faugeras 86], [Dhome 87], la surface de facette [Oshima 83], ou bien la relation de voisinage des facettes [Bolles 86]. L'identification est effectuée par mise en correspondance des indices observés avec ceux du modèle. Ces méthodes peuvent traiter des objets complexes, éventuellement avec des recouvrements partiels. Mais elles demandent une acquisition complète de la scène qui est souvent très longue à effectuer. Par exemple,

dans le cas de la méthode par projection de plan lumineux, si la prise de vue s'effectue au rythme du cycle vidéo, plusieurs minutes seront nécessaires pour une acquisition complète de la scène. En plus, pour extraire des indices de l'image de distance, une segmentation est indispensable. Or, le processus de segmentation est souvent très coûteux en temps de calcul.

Pour des applications industrielles, l'identification d'objets doit s'effectuer le plus rapidement possible. L'idée principale de notre travail est qu'il serait intéressant si l'on puisse supprimer le processus de la segmentation en réduisant au maximum les données à acquérir. Le problème de base que nous nous posons est donc l'identification et la localisation d'objets à partir de données tridimensionnelles éparses de la scène afin d'accélérer le processus. Le même problème est également relevé par [Grimson 84], [Grimson 86], [Gordon 87]. Nous discuterons de leurs méthodes dans la suite de ce mémoire.

Notre travail se compose en trois étapes : calibrage d'un système d'acquisition tridimensionnelle, modélisation d'objet à partir de plusieurs vues de l'objet, et identification et localisation d'objet polyédrique à partir de données tridimensionnelles éparses de la scène.

Dans le chapitre suivant, nous présentons brièvement les techniques d'acquisition des données tridimensionnelles.

Le troisième chapitre de ce mémoire décrit la configuration expérimentale du système d'acquisition que nous avons adopté. Nous présenterons ensuite le processus du calibrage qui permet de déterminer la transformation entre des points dans l'image d'intensité et les points correspondants dans l'espace tridimensionnelle. Les coefficients de cette transformation sont des fonctions de paramètres géométriques de l'ensemble de système caméra/plan lumineux (la distance focale, les facteurs d'échelles, le coefficient de la distorsion de la lentille, la position et l'orientation de la caméra, etc). Les paramètres géométriques de cette transformation sont déterminées en minimisant la somme des carrés des distances entre des coordonnées d'un certain nombre de points actuellement observées et celles obtenues en appliquant la transformation. L'erreur résiduelle est approximativement d'un pixel par point. La précision relative obtenue est approximativement de 0.2 millimètres.

Puis, nous présentons dans le quatrième chapitre, des opérations de prétraitement de données bidimensionnelles et tridimensionnelles telles que la binarisation, la squelettisation d'image, l'élimination des reflets multiples, et la compression de

données.

Le chapitre suivant est consacré au problème de modélisation d'objet. Nous avons mis en œuvre une méthode de construction de modèle à partir de plusieurs vues de l'objet. Les différentes méthodes concernant la segmentation de l'image de distance sont présentées. Le modèle complet d'un objet est obtenu par fusion de modèles partiels.

Dans le sixième chapitre, nous abordons le problème de l'identification et la localisation d'objet. Notre méthode est fondée sur la mise en correspondance des données éparses de la scène avec le modèle. En fait, il est fort probable qu'il existe plusieurs interprétations compatibles avec les données éparses de la scène. Une stratégie de prises de vue supplémentaires est aussi évoquée. Les fausses interprétations sont ensuite éliminées par un processus de vérification.

Enfin, nous donnons la conclusion du présent travail.

2

Systèmes d'acquisition tridimensionnelle

La vision 3D a connu ces dernières années un développement considérable. De ce fait, une grande diversité de systèmes d'acquisition tridimensionnelle ont été développés, on peut en principe les classer en deux catégories :

- systèmes passifs
- systèmes actifs

2.1 Systèmes passifs

Comme l'image d'intensité est la projection bidimensionnelle de la scène tridimensionnelle, il est difficile d'en extraire les indications sur la troisième dimension. Plusieurs images prises en différentes positions (soit par plusieurs caméras soit par déplacement d'une caméra) sont nécessaires pour en déduire les informations tridimensionnelles de la scène. Le principe de la stéréovision est l'identification des primitives (pixels ou points de contour) correspondantes des deux images, l'information sur la profondeur en est déduite par la triangulation des primitives correspondantes. Pour résoudre le problème de correspondance, de nombreuses méthodes ont été proposées telles que la méthode par programmation dynamique [Ohta 85], la méthode de relaxation utilisant l'information contextuelle [Limozin 86], ainsi que la méthode "coarst to fine" [Grimson 81]. Remarquons que ces méthodes sont fondées sur des règles heuristiques et coûteuses en temps de calcul. On constate que la recherche de la correspondance est une tâche complexe, les principales difficultés sont les suivantes :

- Il existe des points qui ont des caractéristiques similaires, ce qui provoque des ambiguïtés pour l'appariement des points entre deux images.
- Il est difficile d'établir la correspondance pour des zones d'intensités uniformes.
- Il est impossible d'établir la correspondance pour des parties qui sont vues par une des deux caméras à cause d'effet d'occlusion ou de la limite de champ de vue d'une caméra.

De nombreuses recherches ont été effectuées au cours de ces dernières années en introduisant une troisième caméra pour résoudre ces problèmes [Ayache 87], [Gerhard 86], [Gurewitz 86], [Ohta 86], [Yachida 86]. La méthode proposée par [Ohta 86] a surtout pour objectif de résoudre la deuxième difficulté citée ci-dessus. Leur méthode consiste à effectuer respectivement la mise en correspondance pour deux couples de stéréo, la perception de profondeur est améliorée par fusion des résultats obtenus en utilisant la méthode de relaxation. Mais on constate qu'elle reste très coûteuse en temps de calcul.

2.2 Systèmes actifs

L'utilisation de capteur actif peut considérablement aplanir certaines difficultés rencontrées en vision passive. Dans cette catégorie on trouve la méthode par projection de lumières structurées, méthode par mesure du temps de vol.

Le système de mesure du temps de vol comprend généralement un couple d'émetteur/récepteur d'ultrason (ou lumière laser). L'émetteur envoie un signal sur l'objet, le récepteur reçoit le signal réfléchi par l'objet. C'est le temps de propagation de l'onde ou lumière qui permet de calculer la distance. Il est facile d'adapter ce type de système à un robot mobile dans le but de détecter des obstacles [Thorpe 88], et on peut éliminer le problème des parties cachées en installant coaxialement l'émetteur et le récepteur. Mais ce type de systèmes présente l'inconvénient d'être sensible aux propriétés de la surface réfléchissante. Par exemple, si l'émetteur envoie un signal sur une surface spéculaire ou semi-spéculaire, le récepteur ne reçoit aucun signal ou reçoit un faux signal à cause de réflexions multiples.

Les méthodes par projection de lumière structurée se fondent sur l'illumination de la scène à analyser par une source lumineuse structurée. La scène est observée

par une caméra, et les coordonnées tridimensionnelles des points illuminés sont directement obtenues par la triangulation. On trouve en générale trois variantes de cette méthode :

- Système par projection de point lumineux

C'est le système le plus simple (voir figure 2.1). On éclaire la scène par un point lumineux issu d'une source laser, un balayage bidimensionnel de la scène fournit des informations denses de l'objet à analyser [Pauchon 83], [Rioux 84].

Ce type de système est assez lent car l'acquisition et le calcul de données tridimensionnelles ne sont effectués que point par point.

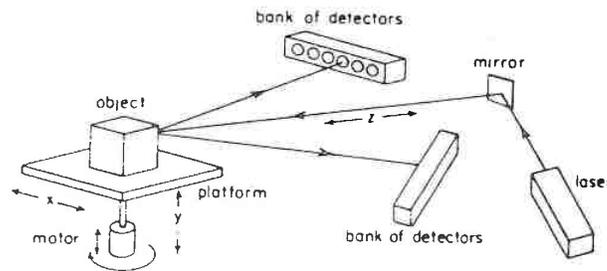


Figure 2.1. Méthode par projection de point lumineux [Faugeras86]

- Système par projection de plan lumineux

On éclaire l'objet à analyser par un plan lumineux produit par une source laser avec une lentille cylindrique, l'intersection du plan lumineux avec l'objet forme un fin rayon de lumière qui est observé par la caméra (voir figure 2.2). L'informations tridimensionnelles des points situés sur le rayon lumineux peuvent être calculées en connaissant la géométrie de la caméra et celle du plan lumineux. Pour obtenir toutes les informations tridimensionnelles de l'objet, un balayage unidimensionnel de la scène est nécessaire,

ce qui implique une image pour chaque prise de vue, et rend le temps de l'acquisition toujours important. Lelandais [Lelandais 84] a réalisé un système de ce type où l'objet à analyser est placé sur une table rotative. La caméra ainsi que la source laser étant fixées, différentes prises de vue sont obtenues par déplacement de l'objet d'un angle élémentaire. Les données de chaque prise de vue sont ensuite traitées par l'unité de mesure électronique. Supposons que l'on prend un pas de rotation de 1 degré, 360 prises de vue sont nécessaires pour balayer l'objet entier.

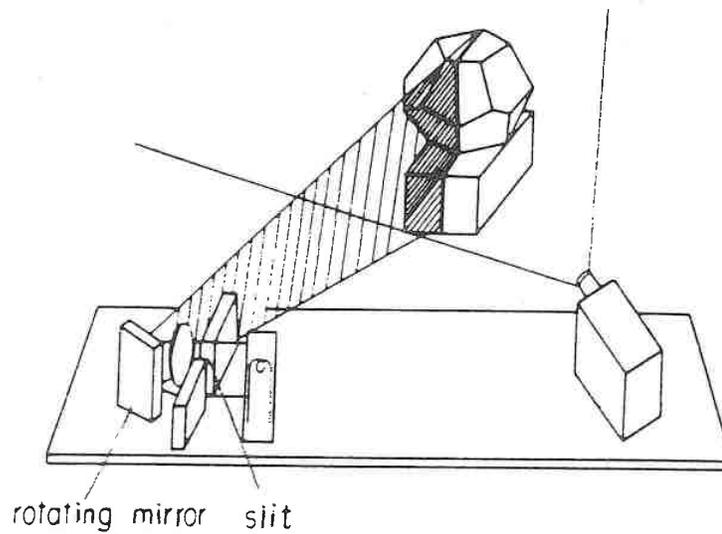


Figure 2.2. Méthode par projection de plan lumineux [Oshima83]

- Système par projection de grille lumineuse

On peut accélérer le processus d'acquisition par projection d'une grille lumineuse (voir figure 2.3). Dans ce cas une seule prise de vue est suffisante pour obtenir des informations tridimensionnelles de la scène. Mais le problème se révèle en correspondance entre des grilles observées par la caméra et des grilles originaux. La complexité du problème de correspondance conduit à de longs calculs [Guisser 87]. De nombreuses recherches ont été menées à bien sur ce problème. La méthode proposée par Boyer et al. [Boyer 87] consiste à

projeter une grille codée en couleur sur la scène. le problème de l'identification est simplifié ayant la connaissance a priori du codage de couleur.

Un compromis entre la méthode par projection de plan et la méthode par projection de grille est la méthode par projection des plans parallèles codés binaires [Yamamoto 86]. La scène est partitionnée en des zones éclairées (codées en 1) et des zones non éclairées (codées en 0) pour chaque vue. L'image résultante est obtenue en accumulant des codes pour chaque zone de différentes vues. Le problème de l'identification est résolu en décodant l'image résultante. Par rapport à la méthode de projection de plan, au lieu d'acquérir successivement 2^n vues, n vues sont suffisantes pour enregistrer toute la scène.

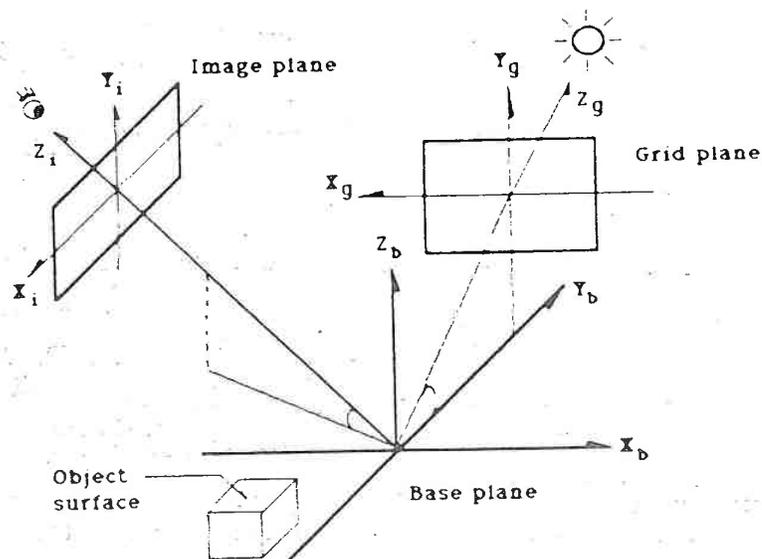


Figure 2.3. Méthode par projection de grille lumineuse [Wang87]

2.3 Choix des méthodes

La liste de différents systèmes d'acquisition tridimensionnelle que nous avons exposée est loin d'être exhaustive (voir [Jarvis 83],[Nitzan 88] pour une présentation plus complète). Toutefois on remarque que les systèmes passifs ont l'avantage

d'être simples et rapides au point de vue de l'acquisition proprement dite qui s'effectue dans les conditions d'éclairage naturel, permettant d'analyser des scènes naturelles . Mais le processus de perception de la profondeur est complexe. Comme les systèmes en stéréo, les systèmes par projection de lumière structurée posent aussi le problème des parties cachées (c.à.d. la partie soit non vue par la caméra, soit non éclairée par la source laser). Ce problème peut partiellement se résoudre en utilisant plusieurs caméras et/ou projecteurs de lumière. Mais le coût et la complexité augmentent dans ce type de systèmes qui ont également l'inconvénient d'être sensible aux propriétés réfléchissante des surfaces. L'avantage le plus important des systèmes actifs est que la reconstitution des données tridimensionnelles est simple et directe, mais ils ont l'inconvénient d'être relativement lents au niveau de l'acquisition des données. Nous pensons néanmoins que le développement des matériels spécialisés performants ouvre une voie optimiste pour ce type de systèmes.

Nous avons vu que chaque catégorie de méthodes a ses propres avantages et ses inconvénients. Le choix des méthodes dépendra essentiellement de l'application envisagée. La méthode d'acquisition fondée sur la projection de plan lumineux est adoptée dans notre travail pour sa simplicité et sa facilité de s'adapter au environnement industriel (par exemple, tapis roulant).

3

Calibrage

3.1 Introduction

Le calibrage joue un rôle fondamental pour la reconstruction des données tridimensionnelles : de ce processus dépend la qualité de données tridimensionnelles obtenues.

Le calibrage de la caméra consiste à établir les relations entre des coordonnées des points de l'espace tridimensionnel et celles de leurs images correspondantes. Dans la littérature, on trouve en général deux catégories de méthodes. La première catégorie de méthodes consiste à déterminer les paramètres intrinsèques et extrinsèques de la caméra en résolvant un système d'équations non linéaires [Sobel 74], [Haralick 80]. Les paramètres intrinsèques sont les caractéristiques de l'optique et de la géométrie interne de la caméra tels que la distance focale de la caméra, des facteurs d'échelles, le centre du plan rétinien (soit l'intersection de l'axe optique avec le plan rétinien), les coefficients de distorsion de la lentille, etc; et les paramètres extrinsèques sont la position et l'orientation de la caméra par rapport à un repère de l'espace. Un des principaux avantages de ces méthodes est l'obtention directe de tous les paramètres physiques de la caméra qui sont souvent indispensables pour certaines applications, mais ces méthodes nécessitent le choix judicieux des valeurs initiales de paramètres, ce qui pose des problèmes avec une caméra vidicon, car les paramètres intrinsèques ne sont pas disponibles pour ce type de caméra. De nombreuses techniques ont aussi été proposées pour estimer indépendamment du calibrage des autres paramètres, le facteur d'échelle horizontal, le centre du plan rétinien ainsi que la distance focale [Lenz 87] [Chou 86]. La deuxième catégorie consiste à estimer les éléments d'une matrice de transformation homogène étant données les coordonnées d'un certain nombre de points de l'espace ainsi que celles de leurs images [Yakimovsky 78] [Toscani 87]. Notons que ces éléments sont fonc-

tions des paramètres intrinsèques et extrinsèques de la caméra, en outre ils ne sont pas indépendants. L'avantage le plus important de ces méthodes est la simplicité, mais avec ces méthodes, il est difficile de prendre en compte des distorsions, et au cas où les paramètres intrinsèques et extrinsèques sont indispensables, la décomposition de ces éléments de matrice est nécessaire [Toscani 87].

Les méthodes citées ci-dessus ont été développées en vue d'une application en stéréo. En ce qui concerne les systèmes de projection de plan lumineux, le calibrage du système comprend en général deux étapes :

- le calibrage de la caméra: déterminer la géométrie de la caméra dans un repère de l'espace.
- le calibrage du plan lumineux: déterminer la géométrie du plan lumineux dans un repère de l'espace.

Bolles [Bolles] calibre d'abord la caméra en résolvant un système d'équations linéaires surdéterminés (la méthode similaire à celles de deuxième catégorie citée ci-dessus). Connaissant la matrice de transformation de caméra, étant donnée une des coordonnées (par exemple, x) d'un point, les deux autres coordonnées (ex , y et z) de ce point sont déterminées. Le calibrage du plan lumineux se fait donc en localisant un certain nombre de points se situant sur le plan lumineux et ayant une de coordonnées connues. L'obtention de la matrice de transformation qui détermine les coordonnées 3D d'un point à partir de son image nécessite l'inversion symbolique d'une matrice combinée des éléments de matrice de transformation de caméra et les paramètres de l'équation de plan lumineux. La précision qu'il a obtenu est de 2 millimètres pour la précision absolue, et de 0.2 millimètres pour la précision relative.

Agin et Borianne [Agin 87] [Borianne 84] calibrent *l'ensemble du système* en utilisant une méthode similaire, mais par localisation des points qui se situent sur le plan lumineux. La précision relative obtenue par [Agin 87] est de 1 millimètre suivant la direction horizontale et de 3,5 millimètres suivant la direction verticale pour une distance de 35,5 centimètres, et de 1,3 millimètres (direction horizontale) et de 7,6 millimètres (direction verticale) pour une distance de 50,8 centimètres. Les méthodes citées ci-dessus ne tiennent pas compte des distorsions.

Mansbach [Mansbach 86] a établi, en tenant compte de correction de nombreuses distorsions, les équations qui permettent de calculer les coordonnées d'un

point dans l'espace à partir des coordonnées de ce point dans l'image, les coefficients de ces équations sont des fonctions de paramètres géométriques et de paramètres de distortions du caméra/plan lumineux système.

Notre méthode est fondée sur le même principe que celle de [Mansbach 86], mais avec une configuration différente de l'ensemble du système et une stratégie d'optimisation différente.

3.2 Principe du Système

La figure 3.1 présente la configuration du système d'acquisition que nous avons adoptée. Le système se compose de trois parties:

- une source laser avec une lentille cylindrique
- une caméra CCD
- un plateau tournant et translatable

L'objet à analyser est placé sur le plateau tournant et translatable. La source laser émet un faisceau lumineux qui est transformé en un plan lumineux par une lentille cylindrique. La scène est éclairée par ce plan lumineux, l'intersection du plan lumineux avec l'objet forme une courbe lumineuse qui est observée par la caméra CCD. C'est la déformation de la courbe de l'image qui nous donne des indications de relief.

Comme le montre la figure 3.2, étant donné un point p_I dans l'image, le point de l'espace correspondant doit se situer sur la droite (dite droite de visée de p_I) définie par le point p_I et le centre optique de la caméra. Un point q_I de la courbe dans l'image correspond donc à un point Q de la courbe dans la scène qui est le point d'intersection de la droite de visée de q_I et le plan lumineux. Les coordonnées du point Q peuvent être uniquement déterminées en connaissant la géométrie de la caméra et celle du plan lumineux, ce qui n'est autre que l'objet du calibrage du système.

On transforme ainsi une image d'une courbe en une liste de coordonnées tridimensionnelles des points situés sur la courbe de la scène. La caméra ainsi que le plan lumineux étant fixes, en déplaçant le plateau d'un pas élémentaire, on obtient une nouvelle acquisition. Pour chaque position du plateau on aura une liste

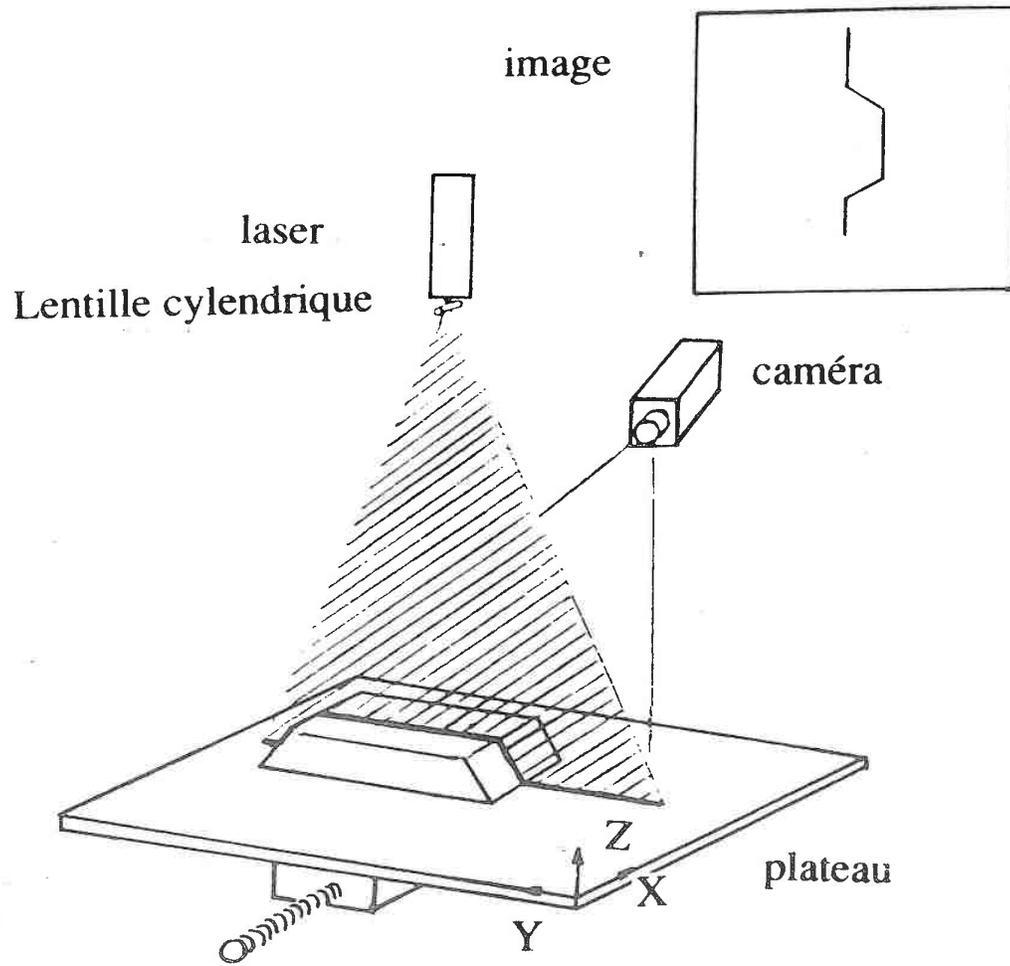


Figure 3.1. Configuration du système d'acquisition tridimensionnelle

de coordonnées des points de la courbe. En défilant le plateau devant la caméra, on obtient ainsi toutes les informations tridimensionnelles de la scène.

3.3 Modèle de Caméra

Le modèle classique de la caméra est le modèle à sténopé (soit la transformation perspective parfaite) décrit en figure 3.2. Un point P de l'espace tridimensionnel est observé par la caméra. La position P_I de ce point dans le plan rétinien est déterminée par l'intersection du plan rétinien avec la droite définie par le point P et le centre optique de la caméra.

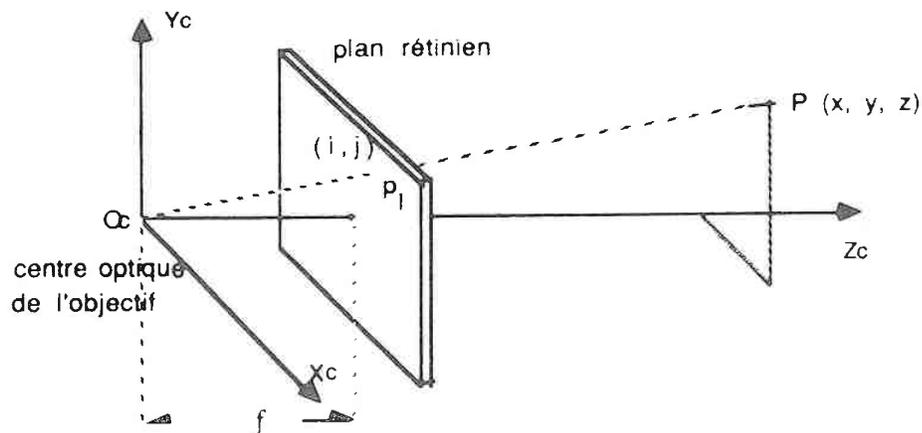


Figure 3.2. modèle classique de la caméra

Remarquons que ce modèle se repose sur l'hypothèse suivante:

la lentille de la caméra est une lentille idéale et fine, ce qui signifie que la lentille peut être parfaitement remplacée par un sténopé.

Nous discuterons, dans la suite de ce chapitre, des effets qui vont s'introduire en cas de non validité de cette hypothèse.

Nous distinguons trois repères liés à la caméra (voir figure 3.3) :

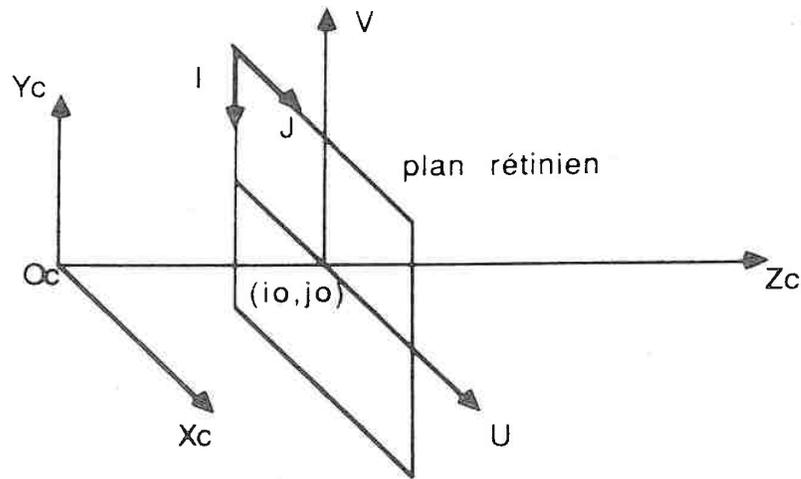


Figure 3.3. Définition de repères liés à la caméra

- Repère rétinien 2D (O, U, V)

Dans ce repère, le plan $U-V$ est confondu avec le plan rétinien, l'axe U est parallèle à la ligne de la matrice d'image, et l'axe V est perpendiculaire à l'axe U . L'origine de ce repère est au point d'intersection du plan rétinien avec l'axe optique de la caméra (soit le centre de rétine). Un point dans ce repère est identifié par ses distances (u, v) par rapport au centre du plan rétinien qui sont mesurées en millimètre dans notre cas.

- Repère caméra $R_C (O_c, X_c, Y_c, Z_c)$

L'origine de ce repère est au centre optique de la caméra, les axes X_c, Y_c sont respectivement parallèles aux axes U et V , l'axe Z_c s'aligne sur l'axe optique de la caméra. Remarquons que ce système de coordonnées ne repose pas sur un trièdre direct.

- Repère image 2D (O', I, J)

Dans une image digitalisée, un point est identifié par une paire d'entiers (i, j) indiquant respectivement les numéros de la ligne et de la colonne de ce point dans l'image. Dans notre cas, nous définissons le coin en haut à gauche de

l'écran comme l'origine de ce repère, et l'axe J est parallèle à la ligne de l'image, et l'axe I positionne de haut en bas.

Il pourrait s'avérer pratique de prendre le centre de la matrice image comme l'origine du repère rétinien. mais en réalité, le centre de matrice image ne coïncide pas exactement avec le point d'intersection du plan rétinien avec l'axe optique, Lenz et al. [Lenz 87] ressort l'influence importante de ce décalage sur la précision de la mesure 3D. Ses expérimentations montrent qu'un seul décalage de 20 pixels pourrait augmenter l'erreur relative de translation et de rotation par un facteur de dix.

En tenant compte de ce décalage, la transformation qui fait passer des coordonnées du repère image à celles du repère rétinien est la suivante (voir figure 3.3):

$$u = k_u(j - j_0) \quad (3.1)$$

$$v = -k_v(i - i_0) \quad (3.2)$$

où

$[k_u]$ facteur d'échelle horizontale (soit la distance suivant l'axe U du centre d'un pixel au centre d'un pixel adjacent mesurée en millimètre (mm/pixel)).

$[k_v]$ facteur d'échelle verticale (soit la distance suivant l'axe V du centre d'un pixel au centre d'un pixel adjacent mesurée en millimètre (mm/pixel)).

$[(i_0, j_0)]$ Les coordonnées du centre de plan rétinien exprimées dans le repère image.

Dans le repère caméra et le repère rétinien, on a les relations suivantes:

$$u = f \frac{x_c}{z_c} \quad (3.3)$$

$$v = f \frac{y_c}{z_c} \quad (3.4)$$

où f est la distance du centre optique au plan rétinien (soit la distance focale).

En combinant les équations 3.1 - 3.4, on a :

$$i = i_0 - \frac{fy_c}{k_v z_c} \quad (3.5)$$

$$j = j_0 + \frac{fx_c}{k_u z_c} \quad (3.6)$$

Récrivons les relations 3.5, 3.6 en coordonnées homogènes, on a :

$$\begin{pmatrix} wi \\ wj \\ w \end{pmatrix} = \begin{pmatrix} 0 & -\frac{f}{k_v} & i_0 & 0 \\ \frac{f}{k_u} & 0 & j_0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} X_c \\ Y_c \\ Z_c \\ 1 \end{pmatrix} \quad (3.7)$$

ou plus compactement:

$$\mathbf{I} = \mathbf{P}\mathbf{X}_c \quad (3.8)$$

où \mathbf{P} est souvent cité comme matrice de transformation perspective. Remarquons qu'étant donné un point de l'espace tridimensionnel visible par la caméra, il n'existe toujours qu'un seul point bidimensionnel dans l'image qui lui correspond; en revanche, un point p_I dans l'image peut être la projection perspective de nombreux points de l'espace tridimensionnel. En fait, tous les points de l'espace tridimensionnel qui se situent sur une droite passée par p_I et le centre optique de la caméra correspondent au même point p_I dans l'image.

3.3.1 Transformation du Repère Caméra au Repère Global

Il s'avère plus pratique d'exprimer les coordonnées des points de l'espace tridimensionnel dans un autre repère attaché à la scène (O, X, Y, Z) . Le repère caméra se déduit de ce repère par un déplacement rigide (c.à.d, une rotation \mathbf{R} et une translation \mathbf{T}).

$$\begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix} = \mathbf{R} \begin{pmatrix} X_c \\ Y_c \\ Z_c \\ 1 \end{pmatrix} + \mathbf{T} \quad (3.9)$$

En exprimant la relation 3.8 dans le repère global, on a :

$$\mathbf{I} = \mathbf{P}\mathbf{R}^{-1}(\mathbf{X} - \mathbf{T}) \quad (3.10)$$

3.4 Paramètres du calibrage en cas général

Admettons que R_o est le repère dans lequel on souhaite calculer des coordonnées de l'objet.

- Paramètres liés à la caméra

Remarquons que la distance focale apparaît toujours comme k_u/f ou k_v/f dans les équations, il y a donc dix paramètres indépendants à calibrer dont quatre sont les paramètres intrinsèques de la caméra (k_u/f , k_v/f , i_0 et j_0), et les six autres sont les paramètres extrinsèques (trois pour la rotation et trois pour la translation).

- Paramètres liés au plan lumineux

Supposons que le plan lumineux est un plan idéal. Le calibrage du plan lumineux consiste à estimer les coefficients de l'équation du plan lumineux dans le repère attaché à la scène.

L'équation du plan lumineux dans le repère attaché à la scène s'écrit:

$$a_1X + a_2Y + a_3Z + a_4 = 0 \quad (3.11)$$

Supposons que le plan n'est pas parallèle à l'axe X , on peut reformuler l'équation précédente en éliminant la solution triviale:

$$X + b_1Y + b_2Z + b_3 = 0 \quad (3.12)$$

les paramètres à calibrer sont donc b_1 , b_2 , b_3 .

- Paramètres liés à la translation de l'objet

La translation de l'objet peut avoir une direction quelconque dans le repère R_o . Elle est caractérisée par un vecteur $\vec{D}(x_d, y_d, 1)$, ce qui entraîne deux paramètres à estimer.

En cas général et sans tenir compte de corrections des distorsions potentielles, il faut calibrer donc en total quinze paramètres pour calibrer le système.

3.5 Calcul des coordonnées

Nous avons vu dans le paragraphe précédent qu'il faut estimer quinze paramètres pour le calibrage du système. Nous constatons que, en se fixant certaines contraintes sans nuire à la généralité du système, la géométrie de l'ensemble de système pourrait être considérablement simplifiées.

Choix de repères

La première contrainte consiste à choisir judicieusement des repères. Nous définissons deux repères de façon suivante (voir figure 3.4):

- Repère attaché à l'objet (repère absolu) $R_O(O_O, X_O, Y_O, Z_O)$

C'est le repère, qui lie l'objet de manière absolue, et dans lequel on souhaite calculer des coordonnées de l'objet. L'origine de ce repère est arbitrairement choisi sur la surface de plateau, le plan X_O-Y_O est confondu avec la surface de plateau, l'axe X_O positionne sur la direction inverse de la translation du plateau, l'axe Z_O est perpendiculaire à la surface de plateau, et l'axe Y_O est perpendiculaire aux deux autres.

- Repère associé au plan lumineux (repère relatif) $R_L(O_L, X_L, Y_L, Z_L)$

C'est le repère dans lequel on va calculer les coordonnées 3D des points d'une courbe acquise par la caméra à partir de coordonnées des points correspondants dans l'image. Ce repère ne prend pas en compte la translation de l'objet. Il se déduit de repère R_O par une translation suivant l'axe X_O , et l'axe Y_L s'aligne sur la ligne d'intersection du plan lumineux avec la surface du plateau, ce qui impose que le plan lumineux soit projeté de telle façon qu'il soit parallèle à l'axe Y_O .

Passage de R_L à R_O

Initialement, les deux repères R_O et R_L se sont confondus. A chaque position relative du plateau, on a :

$$\begin{pmatrix} x_O \\ y_O \\ z_O \end{pmatrix} = \begin{pmatrix} x_L \\ y_L \\ z_L \end{pmatrix} + \begin{pmatrix} n\Delta x \\ 0 \\ 0 \end{pmatrix} \quad (3.13)$$

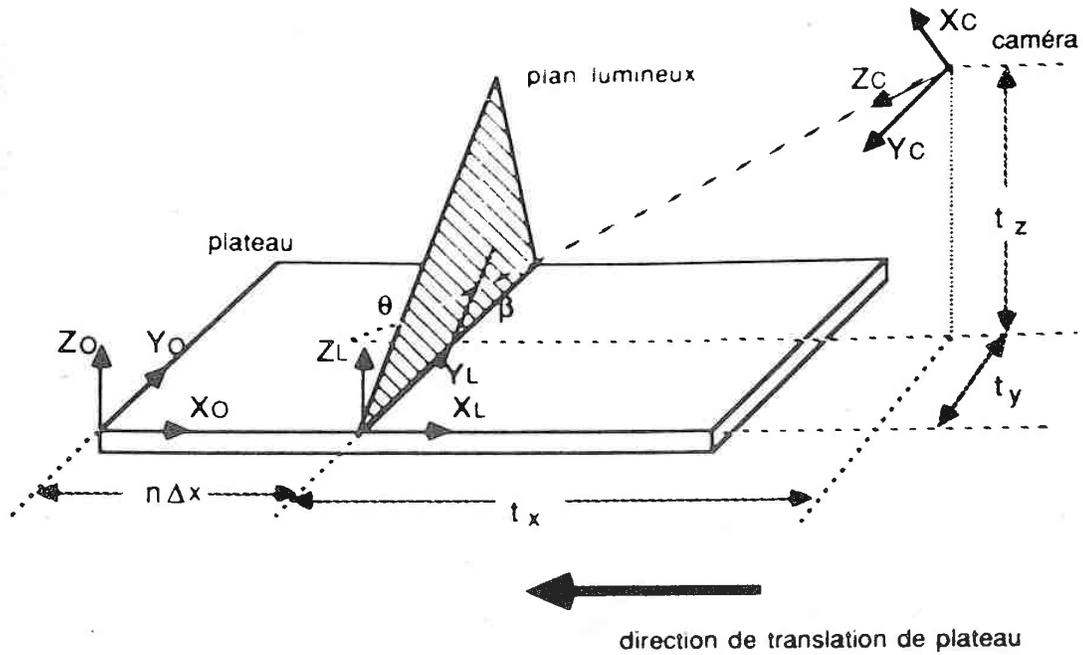


Figure 3.4. La géométrie de l'ensemble de système

où Δx est le pas de translation du plateau entre deux acquisitions consécutives, il peut être choisi de façon empirique en tenant compte de la complexité de la scène ; et n est un entier indiquant le nombre de pas effectués.

Contraintes sur la géométrie externe de la caméra

En respectant les règles suivantes (figure 3.4):

- le centre optique de la caméra se situe en un point (t_x, t_y, t_z) dans le repère attaché au plan lumineux R_L ;
- l'axe Y_c du repère caméra positionne sur la direction inverse de l'axe Y_O ;
- l'axe optique de la caméra Z_c s'oriente vers la scène (le plateau), et forme un angle θ ($\leq \pi/2$) avec la surface de plateau.

l'ensemble de transformations qui fait passer les coordonnées des points dans le repère R_C à celles dans le repère R_L se composent donc (fig. 3.5):

- des translations respectivement de t_x, t_y, t_z suivant les axes X_L, Y_L, Z_L .
- d'une rotation de l'angle π autour de Z_L
- d'une rotation de l'angle $(\theta - \frac{\pi}{2})$ autour du nouveau axe Y_L
- d'une transformation qui fait passer un système de coordonnées gaucher au système de coordonnées droitier (c.à.d, $X = X', Y = Y', Z = -Z'$).

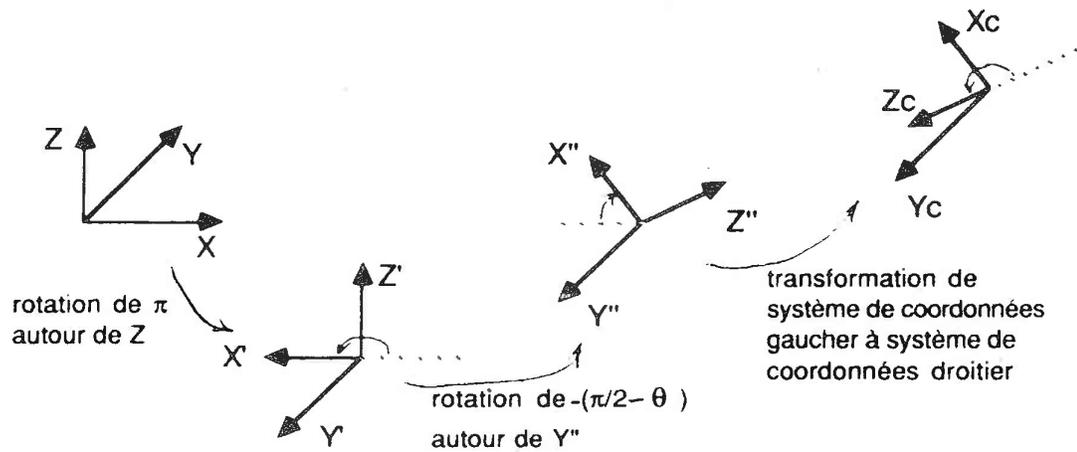


Figure 3.5. Décomposition de rotations

En coordonnées homogènes on a :

$$\mathbf{R} = \begin{pmatrix} -\sin \theta & 0 & -\cos \theta & 0 \\ 0 & -1 & 0 & 0 \\ \cos \theta & 0 & -\sin \theta & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

et le vecteur $\mathbf{T} = (t_x \ t_y \ t_z \ 0)^t$.

En posant $\alpha_u = \frac{f}{k_u}$ et $\alpha_v = \frac{f}{k_v}$ la relation 3.10 devient suivante:

$$i = i_0 - \frac{\alpha_v (y_L - t_y)}{\cos \theta (x_L - t_x) + \sin \theta (z_L - t_z)} \quad (3.14)$$

$$j = j_0 + \frac{\alpha_u \sin \theta (x_L - t_x) - \alpha_u \cos \theta (z_L - t_z)}{\cos \theta (x_L - t_x) + \sin \theta (z_L - t_z)} \quad (3.15)$$

Remarquons que les équations 3.14, 3.15 représentent la droite passant par le centre optique de la caméra et un point (i, j) de l'image sur laquelle se situe le point correspondant de l'espace. Sachant aussi que le point correspondant de l'espace se situe sur le plan lumineux, en connaissant la géométrie du plan lumineux, les coordonnées de ce point dans l'espace sont uniquement déterminées.

Sous les contraintes citées ci-dessus, et en exprimant l'équation 3.12 dans le repère attaché au plan lumineux, on a :

$$x_L - ctan(\beta)z_L = 0 \quad (3.16)$$

où β est l'angle formé par le plan lumineux et la surface de plateau ($\theta < \beta$).

En résolvant les équations 3.14, 3.15, 3.16, les coordonnées du point dans l'espace correspondant à un point (i, j) dans l'image sont les suivantes:

$$y_L = t_y + \frac{\frac{\alpha_u}{\alpha_v}(t_x \sin \beta - t_z \cos \beta)(i - i_o)}{\cos(\beta - \theta)(j - j_o) + \alpha_u \sin(\beta - \theta)} \quad (3.17)$$

$$z_L = \frac{[(t_x \cos \theta + t_z \sin \theta)(j - j_o) - \alpha_u(t_x \sin \theta - t_z \cos \theta)] \sin \beta}{\cos(\beta - \theta)(j - j_o) + \alpha_u \sin(\beta - \theta)} \quad (3.18)$$

$$x_L = ctan(\beta)z_L \quad (3.19)$$

Dans ce cas, les paramètres à estimer sont $i_o, j_o, \alpha_u, \alpha_v, t_x, t_y, t_z, \theta$ et β . Nous rappelons que les significations physiques de ces paramètres sont les suivantes:

$[(i_o, j_o)]$ les coordonnées du centre du plan rétinien exprimées dans le repère image.

$[\alpha_u]$ la distance focale exprimée en fonction de dimensionnement horizontal d'un pixel.

$[\alpha_v]$ la distance focale exprimée en fonction de dimensionnement vertical d'un pixel.

$[(t_x, t_y, t_z)]$ les coordonnées du centre optique de la caméra exprimées dans le repère associé au plan lumineux.

[θ] l'angle formé par l'axe optique de la caméra avec la surface du plateau.

[β] l'angle entre la normale au plan lumineux et celle à la surface de plateau.

3.6 Discussion des hypothèses et corrections

Nous avons vu que la déduction des équations ci-dessus s'est effectuée sous certaines hypothèses simplificatrices. Dans la pratique, la violation de ces hypothèses entraînera des erreurs sur des données obtenues. Nous discutons ici des sources d'erreur possibles ainsi que des corrections que l'on peut développer au cas où ces hypothèses ne sont pas rigoureusement valides.

3.6.1 Sources d'erreurs et corrections

Distorsion

En réalité, il est peu réaliste de supposer que la caméra puisse être modélisée parfaitement par un sténopé, en négligeant la distorsion tangentielle, cela se traduit par le fait que la droite de visé passant par un point proche du bord de l'image converge en un point légèrement plus proche (ou plus loin) du plan rétinien que celle passant par le centre de l'image (c.à.d, le centre optique de la caméra n'est pas unique), ce qui est l'effet de la distorsion (voir figure 3.6). Admettons que (u, v) sont les coordonnées d'un point actuellement observées, et (u_p, v_p) les coordonnées de ce point qu'on devrait obtenir en utilisant une caméra parfaitement modélisée par un sténopé, avec une approximation de premier ordre, on a des relations suivantes:

$$u = (1 + \kappa r_p^2)u_p \quad (3.20)$$

$$v = (1 + \kappa r_p^2)v_p \quad (3.21)$$

où $r_p^2 = u_p^2 + v_p^2$, et κ est le seul coefficient de distorsion à déterminer.

Ce coefficient peut être déterminé indépendamment du calibrage des autres paramètres, avec le modèle mathématique de distorsion ci-dessus (eq. 3.20, 3.21) en utilisant la contrainte : l'image d'une droite est une droite [Toscani 87] et en considérant que le centre d'image ainsi que les facteurs d'échelle soient connus.

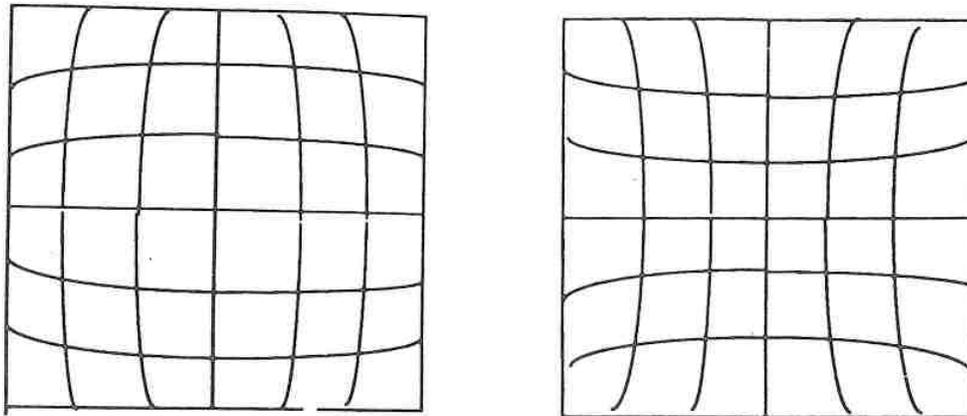


Figure 3.6. distorsion

Séparation de centres optiques

Si la lentille n'est pas fine, il y aura deux centres optiques au lieu d'un seul. Comme l'illustre la figure 3.7, les rayons incidents au premier centre optique émergent du deuxième centre optique, l'angle d'incidence étant égal à l'angle d'émergence.

Nous n'avons pas exploré cette correction pour les raisons suivantes:

- elle peut être partiellement compensée par les autres paramètres déjà définis dans le processus de l'estimation des paramètres.
- En général, la distance entre ces deux centres optiques D_s est considérablement plus petite que l'épaisseur de la lentille, et l'effet de cette séparation sur l'image est proportionnelle à D_s/z_c , ce qui est tout à fait négligeable pour des points se situant loin de la caméra.

Non alignement de caméra/plan lumineux

Nous avons contraint l'axe Y_c à être parallèle à la droite (soit l'axe Y_L) formée par l'intersection du plan laser avec la surface de plateau. Dans la pratique, il est probable que les deux axes ne soient pas parfaitement parallèles à cause d'un

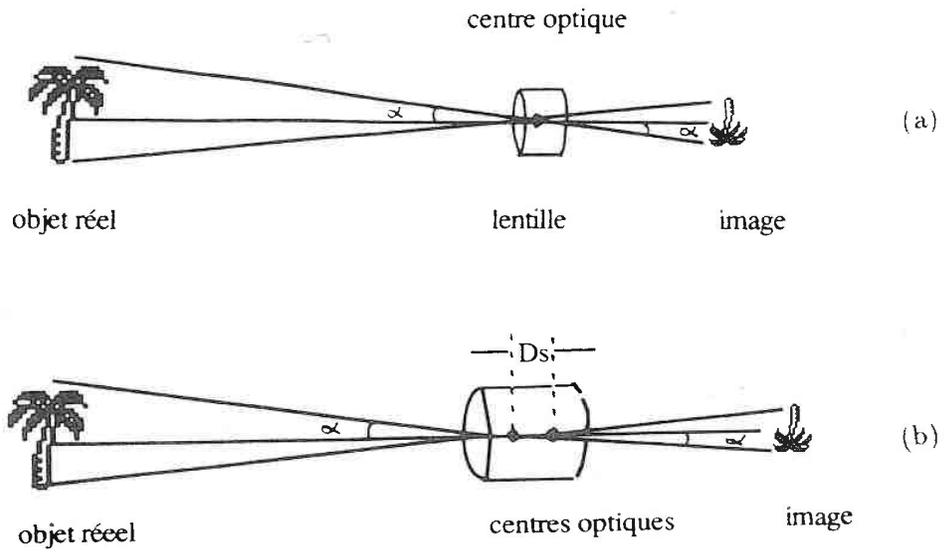


Figure 3.7. (a). Formation d'image avec une lentille idéale (b). Formation d'image avec une lentille non fine

assemblage imparfait du système (voir figure 3.8). Dans ce cas, un paramètre supplémentaire γ est ajouté pour rendre compte de ce non parallélisme.

Supposant que l'axe Y'_L du repère (O_L, X'_L, Y'_L, Z'_L) soit parallèle à l'axe Y_c (voir figure 3.8), on a :

$$x'_L = x_L \cos \gamma - y_L \sin \gamma \quad (3.22)$$

$$y'_L = x_L \sin \gamma + y_L \cos \gamma \quad (3.23)$$

où γ est l'angle entre les axes Y_L et Y'_L .

Pour un petit angle γ (rad.), l'approximation de premier ordre est:

$$x'_L = x_L - \gamma y_L \quad (3.24)$$

$$y'_L = \gamma x_L + y_L \quad (3.25)$$

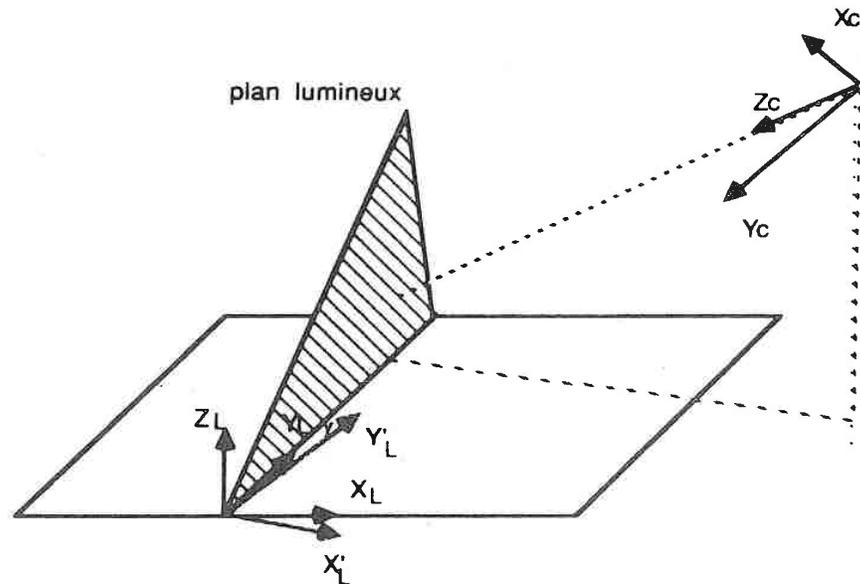


Figure 3.8. La géométrie du système en cas de non alignement de caméra/plan lumineux

Plan lumineux non parallèle à l'axe Y_0

Nous avons supposé que la ligne d'intersection du plan lumineux avec la surface de plateau est orthogonale à la direction de translation de plateau (soit s'aligne sur l'axe Y_L). Ceci peut aisément assuré en plaçant une mire au dessus de plateau. Nous ne prenons donc pas en compte de ce problème.

Épaisseur du plan lumineux

Nous avons supposé que le plan lumineux était un plan idéal. En réalité, ce plan lumineux a toujours une certaine épaisseur, ce qui entraîne des erreurs sur la mesure de coordonnées des points dans l'image. Cela est partiellement corrigé par l'étape de prétraitement de l'image que nous présenterons dans le chapitre suivant.

Aspect digitalisation d'image

Pour qu'une image puisse être traitée par un ordinateur, il faut convertir l'image en une forme discrète (ex: une matrice). Le signal vidéo a donc subi deux processus:

- **échantillonnage**, qui consiste à représenter l'image par une grille discrète, seules les valeurs du signal dans les nœuds de la grille sont conservées;

Deux sortes d'erreur s'introduisent lors du processus de l'échantillonnage:

- Erreur alias¹ due à un sous-échantillonnage.
- Si l'on suppose que la fréquence d'échantillonnage satisfait la critère de Shannon, une autre erreur due à la taille limitée de la matrice d'image s'introduit, qui affecte les bords de l'image.

- **quantification**, qui consiste à représenter les niveaux de gris de l'image par des entiers, ce qui introduit une erreur de quantification.

Calcul des coordonnées après correction de γ

En prenant en compte de la correction de γ , on a donc:

$$i = i_o + \frac{\alpha_u[-y_L \sin \beta + \gamma \cos \beta z_L + (t_y - \gamma t_x) \sin \beta]}{\gamma \sin \beta \cos \theta y_L + \cos(\beta - \theta) z_L - (t_x \cos \theta + t_z \sin \theta + \gamma t_y \cos \theta) \sin \beta} \quad (3.26)$$

$$j = j_o + \frac{\alpha_u[\gamma \sin \beta \sin \theta y_L - \sin(\beta - \theta) z_L - (t_x \sin \theta - t_z \cos \theta + \gamma t_y \sin \theta) \sin \beta]}{\gamma \sin \beta \cos \theta y_L + \cos(\beta - \theta) z_L - (t_x \cos \theta + t_z \sin \theta + \gamma t_y \cos \theta) \sin \beta} \quad (3.27)$$

en posant $\alpha'_u = 1/\alpha_u$, $\alpha'_v = 1/\alpha_v$ et en prenant $\gamma^2 = 0$, la transformation inverse des équations 3.26, 3.27 qui détermine des coordonnées 3D des points se situant sur le plan lumineux à partir de leurs images devient:

$$y_L = \frac{A(i - i_o) + B(j - j_o) + C}{G(i - i_o) + H(j - j_o) + I} \quad (3.28)$$

1. en Anglais: aliasing error

$$z_L = \frac{D(i - i_0) + E(j - j_0) + F}{G(i - i_0) + H(j - j_0) + I} \quad (3.29)$$

$$x_L = \operatorname{ctan}(\beta) z_L \quad (3.30)$$

dont

$$A = \alpha'_v(t_x \sin \beta + \gamma t_y \sin \beta - t_z \cos \beta) \quad (3.31)$$

$$B = \alpha'_u[\gamma(t_x \cos \theta + t_z \sin \theta) \cos \beta + (t_y - \gamma t_x) \cos(\beta - \theta)] \quad (3.32)$$

$$C = [(t_y - \gamma t_x) - \gamma \cos \beta(t_x \sin \theta - t_z \cos \theta)] \sin(\beta - \theta) \quad (3.33)$$

$$D = \alpha'_v t_z \gamma \sin \beta \quad (3.34)$$

$$E = \alpha'_u \sin \beta(t_x \cos \theta + t_z \sin \theta) \quad (3.35)$$

$$F = -(t_x \sin \theta - t_z \cos \theta) \sin \beta \quad (3.36)$$

$$G = \alpha'_v \gamma \sin \beta \quad (3.37)$$

$$H = \alpha'_u \cos(\beta - \theta) \quad (3.38)$$

$$I = \sin(\beta - \theta) \quad (3.39)$$

3.7 Estimation des paramètres

L'idée principale est de choisir des paramètres de telle façon que les données réelles soient bien représentées par la transformation, c'est à dire la déviation entre des points actuellement observés et ceux calculés soit minimale.

3.7.1 Critère d'optimisation

Admettons que les coordonnées $(x_{L_k}, y_{L_k}, z_{L_k})$ ($k = 1, \dots, N$) des N points (dits points de calibrage) dans le repère attaché au plan lumineux et celles de leurs images correspondantes (I_k, J_k) , ($k = 1, \dots, N$) soient connues, les paramètres sont obtenus en minimisant la somme des carrés des distances entre les coordonnées calculées en appliquant les équations 3.26, 3.27 et celles actuellement observées :

$$Q = \sum_{k=1}^N [(i_{cal_k} - I_k)^2 + (j_{cal_k} - J_k)^2] \quad (3.40)$$

3.7.2 Technique d'optimisation

La technique d'optimisation que nous avons adoptée est une technique de moindres carrés non linéaires fondée sur la correction différentielle [McCalla].

Récrivons l'équations 3.26, 3.27 de façon suivante:

$$i = f(y_L, z_L, \mathbf{C}) \quad (3.41)$$

$$j = g(y_L, z_L, \mathbf{C}) \quad (3.42)$$

où \mathbf{C} est un vecteur, $\mathbf{C} = (i_o, j_o, \alpha_u, \alpha_v, t_x, t_y, t_z, \theta, \beta, \gamma)^T$.

Supposons que les N paires de coordonnées $(y_{L_k}, z_{L_k}) \iff (I_k, J_k)$ des points de calibrage soient connues, admettons également que les dix paramètres ont chacun une valeur initiale et $N > 10$, la fonction f peut se développer autour des valeurs initiales des paramètres $\mathbf{C}^o = (c_1^o, c_2^o, \dots, c_{10}^o)^T$ en utilisant l'approximation de premier ordre de série Taylor:

$$f(y_L, z_L, \mathbf{C}) \approx f(y_L, z_L, \mathbf{C}^o) + \sum_{i=1}^{10} \frac{\partial f}{\partial c_i} (c_i - c_i^o) \quad (3.43)$$

En posant $\delta c_i = c_i - c_i^o$, et $\frac{\partial f_k}{\partial c_i} = \frac{\partial f}{\partial c_i} |_{y_L=y_{L_k}, z_L=z_{L_k}, c_i=c_i^o}$, et de la même façon pour la fonction g , le critère à minimiser (eq. 3.40) devient:

$$Q = \sum_{k=1}^N \{ [f(y_{L_k}, z_{L_k}, \mathbf{C}^o) - I_k + \sum_{i=1}^{10} \frac{\partial f_k}{\partial c_i} \delta c_i]^2 + [g(y_{L_k}, z_{L_k}, \mathbf{C}^o) - J_k + \sum_{i=1}^{10} \frac{\partial g_k}{\partial c_i} \delta c_i]^2 \} \quad (3.44)$$

L'équation ci-dessus indique que Q est fonction de δc_i ($i = 1, \dots, 10$). Q aura sa valeur minimale lorsque toutes ses dérivées par rapport à δc_i sont simultanément zéros, ce qui nous ramène à un système d'équations linéaires:

$$\mathbf{M}_{10 \times 10} \begin{pmatrix} \delta c_1 \\ \delta c_2 \\ \vdots \\ \delta c_{10} \end{pmatrix} = \begin{pmatrix} -\sum_{k=1}^N [R_{I_k} \frac{\partial f_k}{\partial c_1} + R_{J_k} \frac{\partial g_k}{\partial c_1}] \\ -\sum_{k=1}^N [R_{I_k} \frac{\partial f_k}{\partial c_2} + R_{J_k} \frac{\partial g_k}{\partial c_2}] \\ \vdots \\ -\sum_{k=1}^N [R_{I_k} \frac{\partial f_k}{\partial c_{10}} + R_{J_k} \frac{\partial g_k}{\partial c_{10}}] \end{pmatrix} \quad (3.45)$$

où

$$\mathbf{M}_{10 \times 10} = \begin{pmatrix} \sum_{k=1}^N [(\frac{\partial f_k}{\partial c_1})^2 + (\frac{\partial g_k}{\partial c_1})^2] & \sum_{k=1}^N [\frac{\partial f_k}{\partial c_2} \frac{\partial f_k}{\partial c_1} + \frac{\partial g_k}{\partial c_2} \frac{\partial g_k}{\partial c_1}] & \dots & \sum_{k=1}^N [\frac{\partial f_k}{\partial c_{10}} \frac{\partial f_k}{\partial c_1} + \frac{\partial g_k}{\partial c_{10}} \frac{\partial g_k}{\partial c_1}] \\ \sum_{k=1}^N [\frac{\partial f_k}{\partial c_1} \frac{\partial f_k}{\partial c_2} + \frac{\partial g_k}{\partial c_1} \frac{\partial g_k}{\partial c_2}] & \sum_{k=1}^N [(\frac{\partial f_k}{\partial c_2})^2 + (\frac{\partial g_k}{\partial c_2})^2] & \dots & \sum_{k=1}^N [\frac{\partial f_k}{\partial c_{10}} \frac{\partial f_k}{\partial c_2} + \frac{\partial g_k}{\partial c_{10}} \frac{\partial g_k}{\partial c_2}] \\ \vdots & \vdots & \ddots & \vdots \\ \sum_{k=1}^N [\frac{\partial f_k}{\partial c_1} \frac{\partial f_k}{\partial c_{10}} + \frac{\partial g_k}{\partial c_1} \frac{\partial g_k}{\partial c_{10}}] & \sum_{k=1}^N [\frac{\partial f_k}{\partial c_2} \frac{\partial f_k}{\partial c_{10}} + \frac{\partial g_k}{\partial c_2} \frac{\partial g_k}{\partial c_{10}}] & \dots & \sum_{k=1}^N [(\frac{\partial f_k}{\partial c_{10}})^2 + (\frac{\partial g_k}{\partial c_{10}})^2] \end{pmatrix}$$

$$R_{I_k} = f(y_{L_k}, z_{L_k}, \mathbf{C}^o) - I_k,$$

$$R_{J_k} = g(y_{L_k}, z_{L_k}, \mathbf{C}^o) - J_k,$$

Ce système d'équations peut être tout simplement résolu en utilisant la méthode de Gauss. Les résolutions $\delta \mathbf{C}$ de ce système d'équations sont des approximations de premier ordre de changements nécessaires en \mathbf{C}^o pour assurer les meilleures estimations des paramètres \mathbf{C} . Si l'un de $|\delta c_i| \geq \varepsilon$, on remplace c_i^o par $c_i^o + \delta c_i$ ($i = 1, \dots, 10$), et la procédure se répète en utilisant les nouvelles estimations.

Nos expérimentations montrent que ce processus converge assez rapidement (en général moins de 7 itérations), et l'erreur résiduelle est approximativement d'un pixel par point.

3.8 Résultats expérimentaux

3.8.1 Considérations pratiques

Au cours de nos expérimentations, la caméra que nous avons utilisée est une caméra Hitachi KP-141 de 500×582 photoéléments avec un objectif ANGENIEUX de 25 millimètres. La source lumière que nous avons utilisée est un laser à l'hélium-néon de faible puissance (5 mw). Le plateau tournant et translatable a une précision de l'ordre du 10ème de millimètre pour la translation et du 10ème de degré pour la rotation.

Dans les paragraphes précédents, nous avons présenté les équations pour calculer des coordonnées. Les coefficients de ces équations sont fonctions de paramètres géométriques (internes et externes) de système caméra/plan lumineux. Remarquons que les valeurs de paramètres géométriques internes sont fixées par le constructeur de caméra; en revanche, il nous revient de déterminer les valeurs de paramètres externes. Le problème qu'on se pose maintenant est de savoir quelles

sont les valeurs de ces paramètres qui conduiront à une erreur minimale. Notons que les angles θ et β sont particulièrement concernés.

Paramètres t_x, t_y, t_z

Dans nos expérimentations, nous avons voulu que la courbe acquise par la caméra soit dans une zone proche du centre de l'image. Dans ce cas, il faut que les paramètres t_x, t_z satisfassent la contrainte suivante:

La coordonnée z_L de point d'intersection de l'axe optique de la caméra avec le plan lumineux soit supérieure à zéro, soit:

$$t_z - t_x \tan \theta \geq 1 - \tan \theta \tan(\beta)$$

Paramètre θ

Nous allons voir quelle est l'erreur qui va s'introduire sur la variation de l'angle θ .

Sans tenir compte de la correction de γ , considérons les équations 3.17, 3.18, 3.19:

$$dx_L = \frac{\cos \beta (-t_x \sin \beta + t_z \cos \beta) [\alpha_u^2 + (j - j_o)^2]}{[\cos(\beta - \theta)(j - j_o) + \alpha_u \sin(\beta - \theta)]^2} d\theta \quad (3.46)$$

$$dy_L = \frac{\alpha_u (t_x \sin \beta - t_z \cos \beta) [\sin(\beta - \theta)(j - j_o) - \alpha_u \cos(\beta - \theta)] (i - i_o)}{\alpha_v [\cos(\beta - \theta)(j - j_o) + \alpha_u \sin(\beta - \theta)]^2} d\theta \quad (3.47)$$

$$dz_L = \frac{\sin \beta (-t_x \sin \beta + t_z \cos \beta) [\alpha_u^2 + (j - j_o)^2]}{[\cos(\beta - \theta)(j - j_o) + \alpha_u \sin(\beta - \theta)]^2} d\theta \quad (3.48)$$

L'erreur sur un point P est donc:

$$dP = \frac{|t_x \sin \beta - t_z \cos \beta| \sqrt{[\alpha_u^2 + (j - j_o)^2]^2 \alpha_v^2 + \alpha_u^2 (i - i_o)^2 [(j - j_o) \sin(\beta - \theta) - \alpha_u \cos(\beta - \theta)]^2}}{\alpha_v [\cos(\beta - \theta)(j - j_o) + \alpha_u \sin(\beta - \theta)]^2} d\theta \quad (3.49)$$

Les courbes sont tracées pour les quatre valeurs différentes de θ en fonction de j (voir figures 3.9, 3.10, 3.11) ou en fonction de i (figure 3.12). Remarquons

qu'en fixant β , la variation de θ introduit un effet important pour la coordonnée Z . Les coordonnées seront plus précises lorsque l'angle formé par le plan lumineux et l'axe optique de la caméra (soit l'angle $(\beta - \theta)$) est plus grand. Néanmoins cet angle est limité par le problème de parties cachées: plus l'angle $(\beta - \theta)$ est grand, plus la partie soit non vue par la caméra, soit non éclairée par le plan laser est importante. Dans nos expérimentations, nous avons pris 40 degrés pour l'angle $(\beta - \theta)$ qui nous paraît un compromis raisonnable.

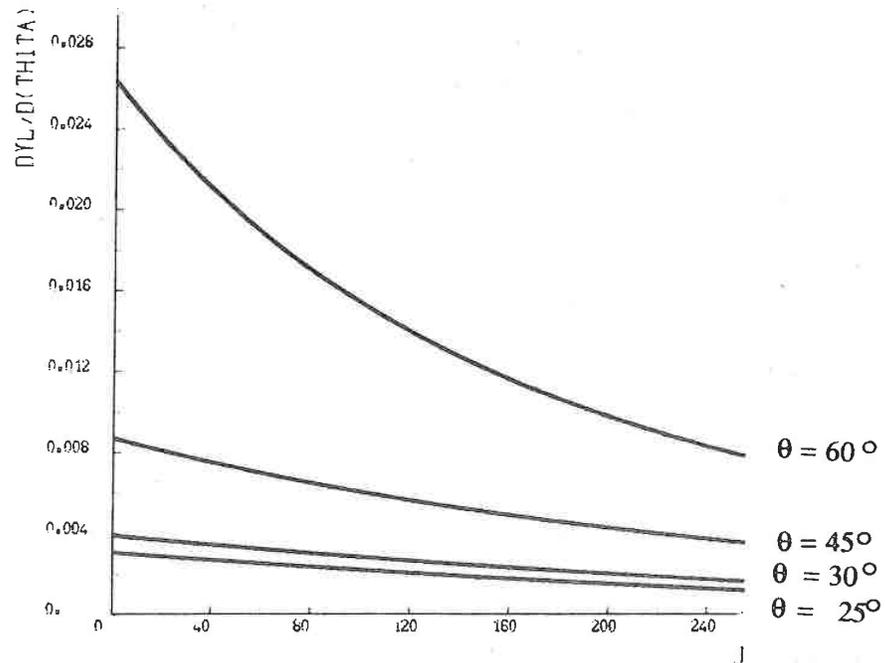


Figure 3.9. la courbe $|dy_L/d\theta|$ en fonction de j .

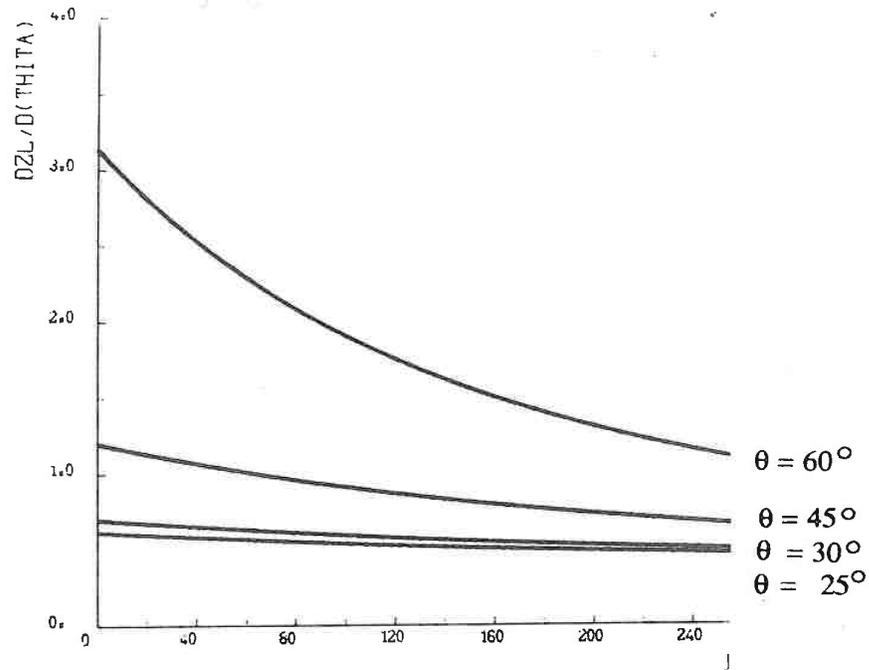


Figure 3.10. la courbe $|dz_L/d\theta|$ en fonction de j

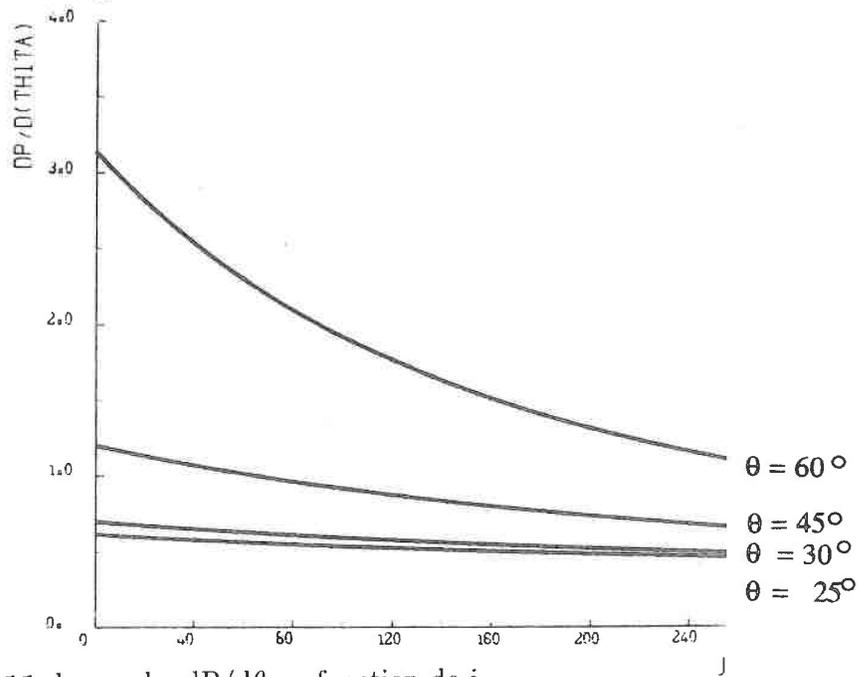


Figure 3.11. la courbe $dP/d\theta$ en fonction de j
 avec $i_0 = 162$, $j_0 = 127$, $\alpha_u = 863.8$, $\alpha_v = 1310.5$, $t_x = 429.1$, $t_y = 97.2$, $t_z = 543.6$,
 $i = 150$ et $\beta = \pi/2$

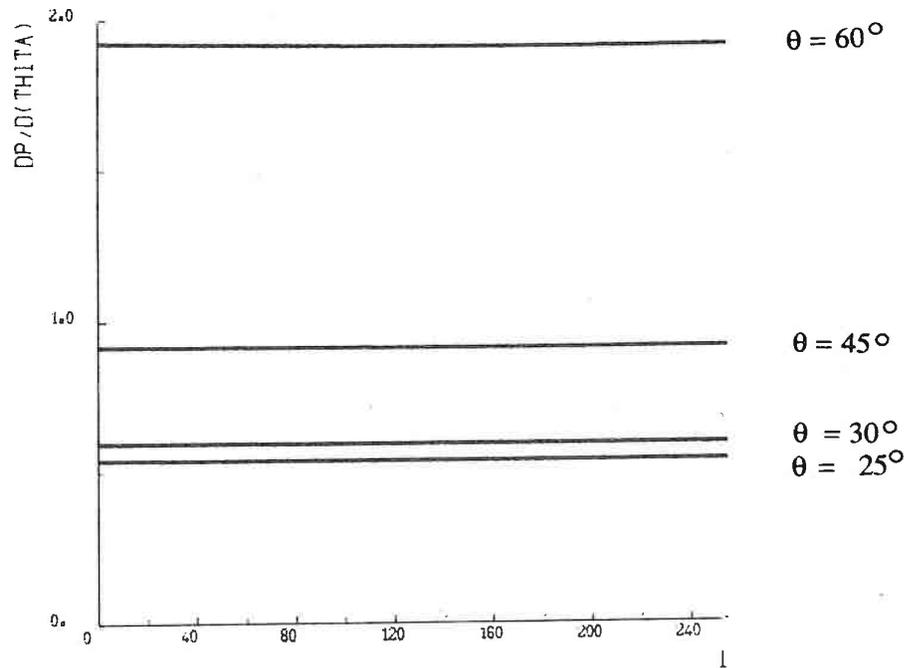


Figure 3.12. la courbe $dP/d\theta$ en fonction de i
avec $i_0 = 162$, $j_0 = 127$, $\alpha_u = 863.8$, $\alpha_v = 1310.5$, $t_x = 429.1$, $t_y = 97.2$, $t_z = 543.6$,
 $j = 150$ et $\beta = \pi/2$

Paramètre β

$$dx_L = \frac{-[(t_x \cos \theta + t_z \sin \theta)(j - j_o) - \alpha_u(t_x \sin \theta - t_z \cos \theta)][(j - j_o) \sin \theta + \alpha_u \cos \theta]}{[\cos(\beta - \theta)(j - j_o) + \alpha_u \sin(\beta - \theta)]^2} d\beta \quad (3.50)$$

$$dy_L = \frac{\alpha_u[(t_x \cos \theta + t_z \sin \theta)(j - j_o) - \alpha_u(t_x \sin \theta - t_z \cos \theta)](i - i_o)}{\alpha_v[\cos(\beta - \theta)(j - j_o) + \alpha_u \sin(\beta - \theta)]^2} d\beta \quad (3.51)$$

$$dz_L = \frac{[(t_x \cos \theta + t_z \sin \theta)(j - j_o) - \alpha_u(t_x \sin \theta - t_z \cos \theta)][(j - j_o) \cos \theta - \alpha_u \sin \theta]}{[\cos(\beta - \theta)(j - j_o) + \alpha_u \sin(\beta - \theta)]^2} d\beta \quad (3.52)$$

Pour un point P:

$$dP = \frac{|(t_x \cos \theta + t_z \sin \theta)(j - j_o) - \alpha_u(t_x \sin \theta - t_z \cos \theta)| \sqrt{\alpha_u^2(i - i_o)^2 + \alpha_v^2(j - j_o)^2 + \alpha_u^2 \alpha_v^2}}{\alpha_v[\cos(\beta - \theta)(j - j_o) + \alpha_u \sin(\beta - \theta)]^2} d\beta \quad (3.53)$$

En fixant θ , plus l'angle $(\beta - \theta)$ est grand, plus $dP/d\beta$ sera petit (car la valeur de α_u est beaucoup plus grande que $|j - j_o|$). Les courbes montrées en annexe C confirment donc nos remarques pour la variation de θ . Nous avons pris 90 degrés pour l'angle β dans nos expérimentations.

3.8.2 Mesure des données — Etalonnage

Le processus de calibrage comprend la mesure précise des coordonnées 3D de points de calibrage et celles de leurs images. Pour effectuer cette mesure, nous avons construit et utilisé un certain nombre d'objets étalons (voir figure 3.13). L'objet étalon est placé sur le plateau, et est illuminé par le plan lumineux, la courbe observée par la caméra est constituée des segments de droite, ce sont des points angulaires que nous prendrons comme des points de calibrage.

Coordonnées 3D des points de calibrage

Les coordonnées 3D (seulement (y_L, z_L) sont nécessaires) de ces points sont mesurées à la main. Pour faciliter cette mesure, l'objet étalon est toujours posé sur

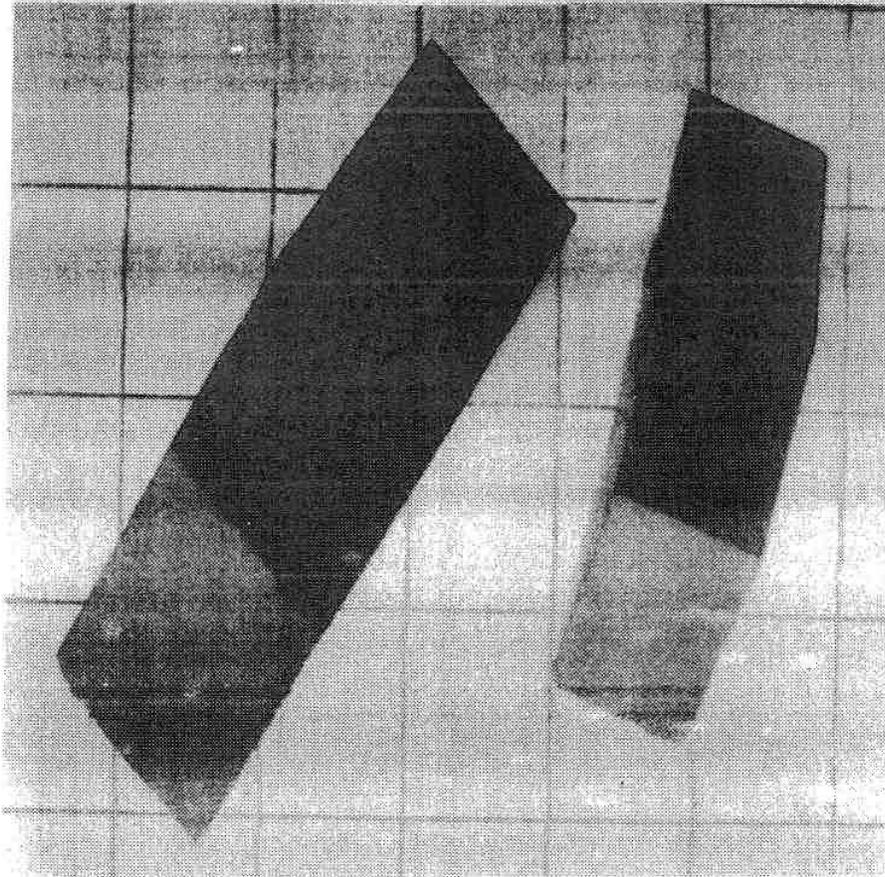


Figure 3.13. Objets étalons

le plateau de telle façon que ses faces latérales soient parallèles à l'axe Y_L (soit la ligne d'intersection du plan lumineux avec la surface du plateau)(voir figure 3.14).

Notons qu'il y a toujours des erreurs associées à la mesure des positions de points de calibrage dans l'espace. On ne peut donc pas totalement garantir que les points de calibrage soient coplanaires. Pour corriger cette erreur, il est préférable que le nombre de points de calibrage utilisés dans le processus de minimisation soit supérieur à vingt.

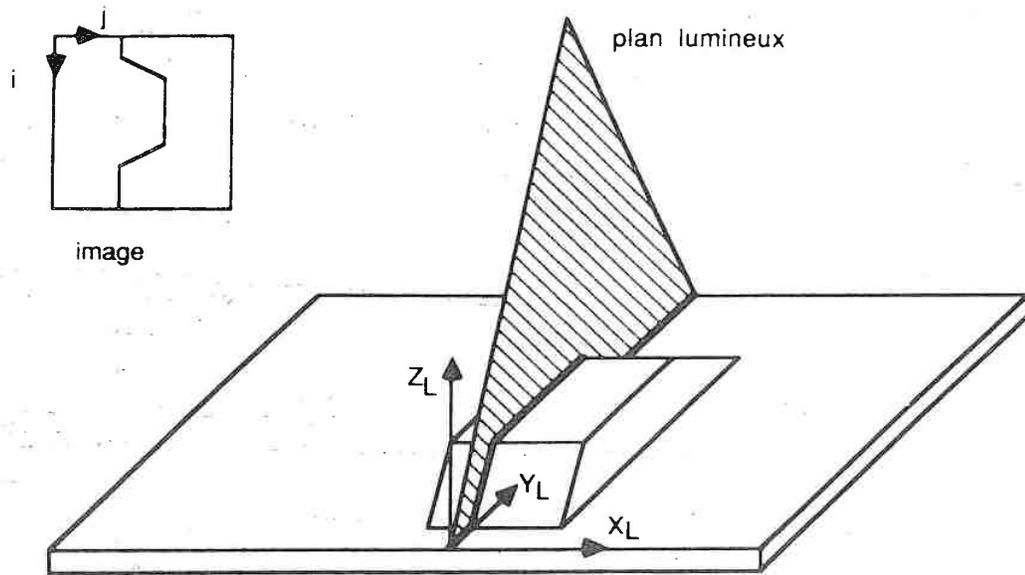


Figure 3.14. Position de l'objets étalons

Coordonnées 2D des points de calibrage

Les coordonnées 2D de ces points sont calculées de façon suivante (cf chapitre 4):

- squelettiser la courbe à une épaisseur d'un pixel.
- approximation de l'ensemble de points par les meilleures droites au sens de moindres carrés.
- l'intersections de ces droites entre elles nous donnent les coordonnées de ces points angulaires.

Remarquons que les erreurs associées à la position de point image dues à l'épaisseur du plan lumineux et à l'aspect digitalisation de l'image sont partiellement corrigées, les coordonnées 2D ainsi calculées sont donc assez précises et fiables.

Notons aussi que, en général les choix exacts des paramètres par le processus de minimisation ne sont pas tout à fait indépendants des positions de points de

calibrage, ce qui signifie que la précision de calibrage n'est pas uniforme pour des points situés sur des positions différentes dans l'image. Les points qui sont proches des points de calibrage auraient une meilleure précision, ou inversement. Pour remédier cela, un multi-étalonnage est effectué. Les objets étalons ont des hauteurs différentes, et ils sont posés sur le plateau dans des positions différentes (suivant l'axe Y_L), ce qui nous donne des points de calibrage bien répartis dans la zone proche du centre d'image où nous voulons une meilleure précision de calibrage.

3.8.3 Valeurs initiales des paramètres

Les paramètres intrinsèques peuvent être estimés à partir des informations fournies par le constructeur de la caméra.

- f est la distance focale nominale engravée sur la monture de l'objectif.
- $k_u = s_u d_u$, où

$$s_u = \frac{\text{nombre de éléments dans une ligne horizontale}}{\text{nombre de pixels dans la ligne de la matrice image}}$$

et d_u est la distance entre deux éléments photosensibles adjacents dans la direction horizontale.

- k_v peut se calculer de la même façon que k_u .
- Les valeurs initiales pour (i_0, j_0) peuvent être soit $(128, 128)$ (pour une image de 256×256) soit des valeurs obtenues par des expériences antérieures en utilisant la même caméra et l'objectif.

Les paramètres extrinsèques tels que t_x, t_y, t_z, β , et θ peuvent tous être approximativement mesurés. Le paramètre de correction γ est plus difficile à mesurer, comme il est très faible, on peut prendre zéro pour valeur initiale.

3.8.4 Résultats

Nous présentons dans le tableau ci-dessous des résultats obtenus:

PARAMETRES	ESTIMATION DES PARAMETRES			
	valeurs initiales	valeurs estimées $N_p = 23$	valeurs initiales	valeurs estimées $N_p = 18$
io (pixel)	128	162	128	152
jo (pixel)	128	127	128	128
α_u (pixel)	735.3	863.8	735.3	753.3
α_v (pixel)	961.5	1310.5	961.5	1017.5
tx (mm)	443	429.14	500	482.2
ty (mm)	115.5	97.18	150	144.5
tz (mm)	502	543.61	600	571.3
θ (deg.)	53	52.7	50	48.7
β (deg.)	90	92.2	90	89.44
γ (deg.)	0	1.3		
erreur résiduelle (pixel)	67.8	1.2	44.3	1.6

Figure 3.15. les estimations des paramètres

où N_p est le nombre de points de calibrage utilisés.

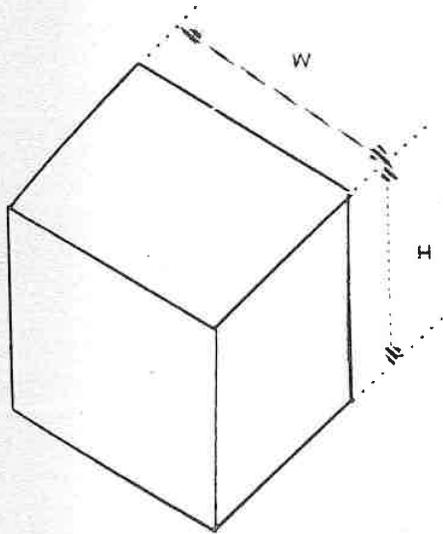
Précision

Nous distinguons deux sortes de précision:

- **Précision absolue:** C'est la précision avec laquelle le système localise un point par rapport à l'origine de repère attaché à l'objet (repère absolu).
- **Précision relative:** C'est la précision avec laquelle le système effectue une mesure dimensionnelle d'un objet.

Nous avons testé la précision relative du calibrage en calculant les dimensions W et H d'un cube (voir figure 3.16). le résultat est présenté dans le tableau ci-dessous (figure 3.16).

(avec $Dw = | W_{réel} - W_{calculé} |$. $DH = | H_{réel} - H_{calculé} |$):



$W = 6 \text{ cm}$, $H = 6 \text{ cm}$

Distance de la caméra à l'origine de repère RL: 67.9 cm	
Dw (mm)	0.22
DH (mm)	0.27

Figure 3.16. Cube de test et précision relative obtenue

Remarquons que la précision est meilleure dans la direction horizontale que celle dans la direction verticale.

Les précisions absolues obtenues sont montrées en tableau ci-dessous (le nombre de points testés étant de 40):

ERREUR	Y (mm)	Z (mm)
moyenne	0.402	0.586
maximale	0.9	1.9

4

Techniques de prétraitement

4.1 Introduction

Dans le chapitre précédent, nous avons présenté comment on peut extraire des informations tridimensionnelles d'un objet en utilisant une caméra et une source lumineuse plane. Notons que les résultats de l'acquisition ne sont pas directement des coordonnées tridimensionnelles, les prétraitements de l'image sont nécessaires pour convertir une image d'un objet intersecté par un plan lumineux en une liste de coordonnées tridimensionnelles.

Dans le cas d'image bidimensionnelle traditionnelle, les trois principaux motifs de prétraitement d'image sont les suivants:

- améliorer la qualité de l'image, par exemple, filtrer l'image pour éliminer des bruits (filtre médian, lissage, etc), et modifier l'histogramme de l'image pour que l'image ait une distribution spécifiée de niveaux de gris.
- convertir l'image en une forme plus convenable pour des traitements ultérieurs, par exemple, segmenter l'image en des régions de caractéristiques similaires, ou binariser l'image pour séparer la forme et le fond, etc.
- extraire des primitives, par exemple, extraction des contours, des régions, etc.

Dans le cas de système fondé sur la projection du plan lumineux, étant donnée une image d'un objet intersecté par le plan lumineux, le prétraitement d'image consiste à extraire les informations les plus pertinentes et filtrer certains bruits pour que le système ne détecte que les parties de l'objet qui se situent sur le plan lumineux.

Etant donnée une image de la courbe lumineuse, une technique de seuillage fondée sur l'analyse de histogramme est effectuée pour extraire la courbe lumineuse du fond, un processus de squelettisation utilisant le graphe de ligne d'adjacence est ensuite effectué en ne conservant que des centres des intervalles des nœuds. Les bruits souvent rencontrés dans le cas d'image bidimensionnelle tels que "salt-and-pepper" ou faux contours dus au processus de quantification ne nous posent pas de problème, car l'acquisition de l'image s'effectue dans une "chambre noire". Dans des conditions d'expérimentations plus délicates, on peut aussi ajouter devant l'objectif de la caméra un filtre passe-bande centré sur la longueur d'onde de lumière laser [Borianne 84]. Les bruits dont nous nous préoccupons dans notre cas sont ceux dus aux réflexions multiples de lumière, ces bruits sont éliminés du fait que le plan lumineux n'intersecte l'objet qu'une fois.

4.2 Binarisation

La segmentation d'image consiste à diviser l'image en des régions de caractéristiques similaires. La binarisation est la technique de segmentation d'image la plus simple.

Dans notre cas, l'extraction de la courbe s'effectue en comparant le niveau de gris de chaque point d'image avec un seuil prédéfini, ceux qui ont une valeur supérieure au seuil sont considérés comme des points qui sont effectivement sur la courbe, en revanche, ceux qui ont une valeur inférieure au seuil appartiennent au fond. Le seuil est déterminé en analysant l'histogramme de l'image. Si l'on regarde l'image d'une coupe (Figure 4.1), tout porte à croire qu'elle ait un histogramme bimodal représentant deux pics bien séparables: l'un correspondant au fond, un autre correspondant à la courbe. Dans ce cas, la recherche automatique de seuil est facile à réaliser: il suffit de choisir un seuil optimal entre ces deux pics. Or, comme la figure 4.2 montre, l'histogramme n'a qu'un pic majeur correspondant au fond, le pic correspondant à la courbe étant trop faible. De ce fait, au lieu de calculer l'histogramme pour tous les points de l'image, nous ne l'évaluons que pour les points qui sont proches des bords de régions, c'est à dire, pour chaque point p de l'image, on compare la valeur de p avec la moyenne des points dans un voisinage $n \times n$ de p , on n'évalue l'histogramme pour le point p que si la déviation est supérieure à un certain seuil. Ainsi on obtient un histogramme parfaitement bimodal. Dans ce cas, le choix automatique du seuil devient très simple:

Admettons que $H(n)$ soit l'histogramme de l'image, on trouve deux maximums locaux de l'histogramme $H(n_1)$, $H(n_2)$, avec $|n_1 - n_2|$ supérieur à un seuil. Supposons que $H(n_3)$ soit un minimum local, et n_3 est entre n_1 et n_2 , si $H(n_3)$ est très petit devant $\min\{H(n_1), H(n_2)\}$, n_3 est alors le seuil choisi.

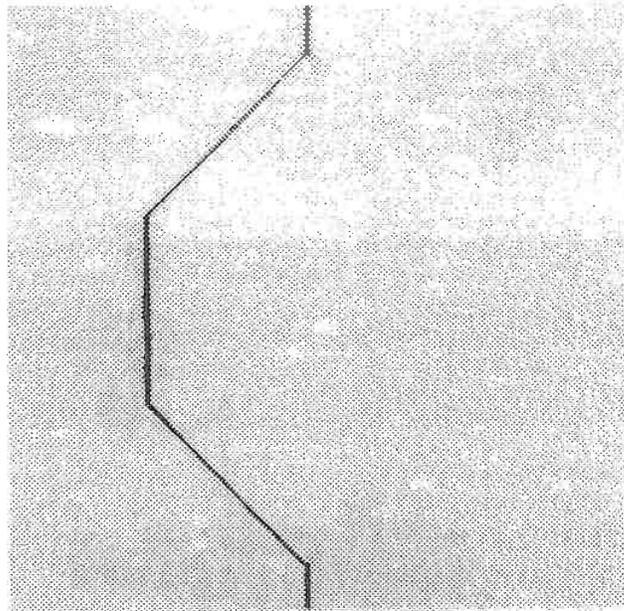


Figure 4.1. l'image d'une coupe

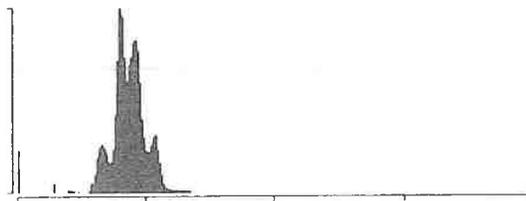


Figure 4.2. l'histogramme de l'image fig.4.1

4.3 Squelettisation

Nous l'avons dit dans le chapitre précédent, la courbe acquise a toujours une certaine épaisseur, souvent de 4 à 8 pixels. Le problème est de déterminer quels sont les points qui se situent effectivement sur le plan lumineux. Borianne [Borianne 84] détermine ces points en calculant le gradient pour chaque point de l'image, et en conservant l'un des deux bords correspondants aux maxima locaux du module de gradient. Nous pensons que ce n'est pas un choix optimal, car l'épaisseur de segment varie en fonction de l'orientation de la surface de l'objet par rapport au plan lumineux incident. On observe une épaisseur minimale de segment lorsque la normale du plan lumineux est orthogonale à celle de la surface. De plus le faisceau laser peut présenter des irrégularités. Nous avons choisi d'utiliser la technique de squelettisation pour déterminer les points qui se situent dans la partie médiane du plan lumineux.

De nombreux algorithmes de squelettisation ont été proposés. Les contraintes que tous les algorithmes de squelettisation doivent vérifier sont les suivantes:

- *préservation de la connexité des régions* : élimination d'un point ne doit pas altérer la connexité de son voisinage
- *conservation des segments d'épaisseur 1* : les segments fins ne doivent pas être éliminés

Une des méthodes les plus connues procède par élimination (érosion) des points du bords qui ne font pas partie du squelette. Le problème essentiel est de définir quels sont les points du bord qui doivent être définitivement éliminés, et quels sont les points du squelette qui doivent être conservés. Etant donnée une région S , les points appartenant au bord sont les points de S qui ont au moins un voisin appartenant à la région \bar{S} (complément de S). Les points du squelette sont les points qui ont un voisinage vérifiant certaines configurations [Pavlidis 82], [Tombre 87].

Notons que le squelette (c.à.d, l'ensemble de points du squelette) est très sensible aux bruits, la moindre irrégularité de contour altère la forme de squelette, de plus, après une telle squelettisation, une droite de certaine épaisseur donne un squelette qui a deux branches aux extrémités, ce qui n'est évidemment pas ce que nous voulions dans notre cas.

Remarquons aussi que cette méthode n'est pas très efficace car plusieurs parcours de l'image sont nécessaires. Toutefois, ce processus peut être implémenté par

des opérations locales qui peuvent facilement être câblées en utilisant des circuits spécialisés.

Dans notre cas, grâce à une caractéristique spécifique de l'image (l'image ne contient que des formes minces), une simple méthode modifiée [Pavlidis 82] utilisant la structure de graphe de ligne d'adjacence¹ peut être adoptée.

Nous rappelons d'abord quelques définitions que nous utiliserons ultérieurement:

- un intervalle

Dans une image binaire, un intervalle correspond à un ensemble de points noirs consécutifs d'une même ligne, la largeur de cet intervalle est le nombre de ces points consécutifs.

- graphe de ligne d'adjacence

C'est un graphe dont les nœuds correspondent aux intervalles, et les arêtes lient les nœuds s'ils sont voisins.

- voisinages d'un nœud

Par une définition stricte, un nœud A est voisin d'un nœud B si les conditions suivantes sont vérifiées :

les intervalles correspondants se situent sur les lignes adjacentes et leurs projections suivant la direction de balayage se superposent (fig. 4.3(b)).

Admettons que J_{k-1} et J_k représentent les coordonnées des extrémités d'un intervalle, alors les deux intervalles $[J_{m-1}, J_m)$ et $[J_{n-1}, J_n)$ ont des projections superposées si les contraintes ci-dessous sont satisfaites :

$$J_{m-1} \leq J_n$$

et

$$J_{n-1} \leq J_m$$

dans notre cas, pour intégrer le processus d'élimination des reflets multiples dans le programme de squelettisation, nous avons étendu la définition de voisinage comme suite :

si les intervalles correspondants se situent sur les lignes adjacentes, et les projections suivant la direction de balayage de ces deux intervalles se sont

1. en Anglais: line adjacency graph

superposées ou sont assez proches de l'un de l'autre, alors ces deux nœuds sont considérés comme des voisins (fig. 4.3(c)).

- degrés d'un nœud

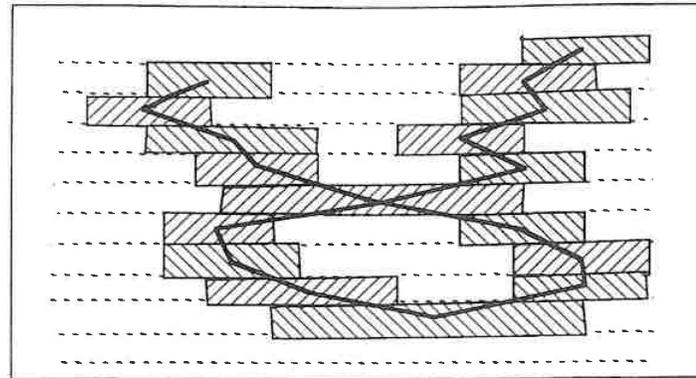
- *degré supérieur* : nombre de nœuds voisins qui se situent sur la ligne précédente de ce nœud.
- *degré inférieur* : nombre de nœuds voisins qui se situent sur la ligne suivante de ce nœud.

On désigne (*sup*, *inf*) les degrés d'un nœud dont *sup* correspond au degré supérieur et *inf* le degré inférieur.

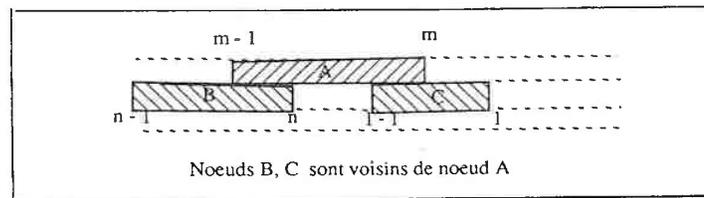
- chemin

un chemin est une liste des nœuds voisins ayant les degrés (1, 1) sauf pour les deux nœuds extrêmes.

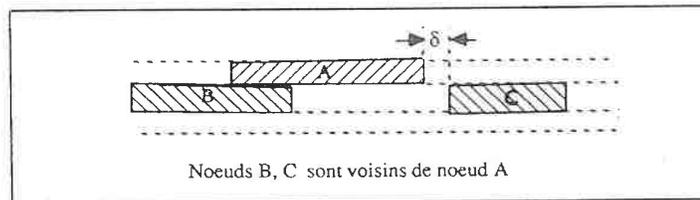
le processus de squelettisation consiste donc simplement à parcourir l'ensemble des nœuds du graphe, le centre d'un nœud ayant degrés (m, n) ($m \leq 1$ et $n \leq 1$) est considéré comme un point du squelette. lorsque l'on rencontre un nœud *H* de degrés (m, n) avec $n \geq 1$, un test de priorité est effectué: le chemin le plus long parmi tous les chemins débutant de nœud *H* a la priorité, mais dans le cas où les chemins ont approximativement les mêmes longueurs, la priorité est d'abord conservée au chemin qui continue dans le même sens que le chemin précédent, sinon, la priorité est donnée au chemin dont les nœuds ayant une largeur plus importante. Le chemin prioritaire est donc conservé, les autres chemins sont alors considérés comme des bruits dûs aux réflexions multiples et sont ensuite éliminés (figure 4.4).



(a)



(b)



(c)

Figure 4.3. Définition du graphe de ligne d'adjacence: (a) Exemple d'un graphe LAG; (b) définition stricte de voisinage d'un nœud; (c) définition étendue de voisinage d'un nœud

L'algorithme :

Soit $Adjacent(P, nbr\text{-}voisin, Voisin)$ une fonction qui retourne le nombre de voisins se situant sur la ligne suivante de P et leurs positions qui sont sauvegardées dans le vecteur $Voisin[]$.

$Thinning()$:

mise a jour le graphe LAG

pour chaque noeud P du graphe

si P n'est pas encore visite alors

$stop = faux$

tant que $stop \neq vrai$

$marquer$ le noeud P déjà visite

$appeler$ la fonction $Adjacent(P, nbr\text{-}voisin, Voisin)$

si $nbr\text{-}voisin = 0$ alors $stop = vrai$

sinon

si $nbr\text{-}voisin = 1$

$mise a jour$ le point du squelette (le centre du noeud)

$P = Voisin[0]$

sinon

$test$ de la priorite

le chemin prioritaire est conserve

$mise a jour$ les points du squelette (les centres

des noeuds du chemin prioritaire)

(les autres chemins sont enleves)

fsi

fsi

ftant que

fsi

fpour

Remarquons que cet algorithme est très rapide car un seul parcours du graphe est nécessaire. Cependant, cet algorithme donne des faux squelettes lorsque l'image contient des chemins relativement courts dont tous les nœuds ont une largeur supérieure à un seuil (voir figure 4.5). Dans notre cas, nous ne travaillons que sur des objets polyédriques, l'intersection du plan lumineux avec l'objet forme des segments droites et cet algorithme nous donne des résultats satisfaisants. Si l'on

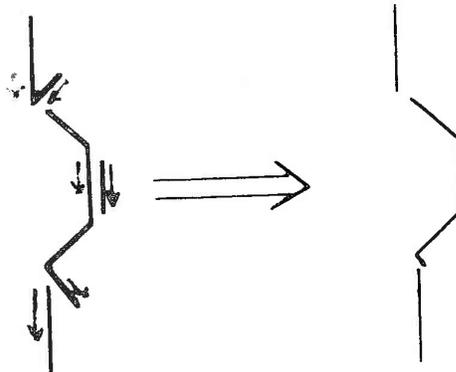


Figure 4.4. Elimination des reflets multiples

travaillait sur les objets quelconques, cela pourrait poser de sérieux problèmes, car l'image est une projection perspective, et pour une surface non plane, il pourrait y avoir plusieurs intersections valides par ligne. Dans ce cas on peut envisager d'utiliser une autre méthode [Tombre 87] qui consiste, après une approximation polygonale, à appairer des segments situés les uns en face des autres et parallèles ou quasi-parallèles, la ligne encadrée par ces deux segments est alors conservée. La figure 4.6 schématise une telle opération.

4.4 Approximation Polygonale

Une approximation polygonale consiste à diviser un ensemble de points en plusieurs groupes, les points de chaque groupe sont ensuite approximés par une droite.

Dans la littérature on trouve de nombreux algorithmes. Le problème essentiel de l'approximation polygonale est de choisir un critère de test qui décide si un groupe de points est susceptible d'être approximé par une droite ou non. La méthode la plus simple est celle du suivi séquentiel en utilisant le critère de distance :

On cherche séquentiellement le plus long segment acceptable en respectant la contrainte : toutes les distances des points délimités par les deux points extrêmes

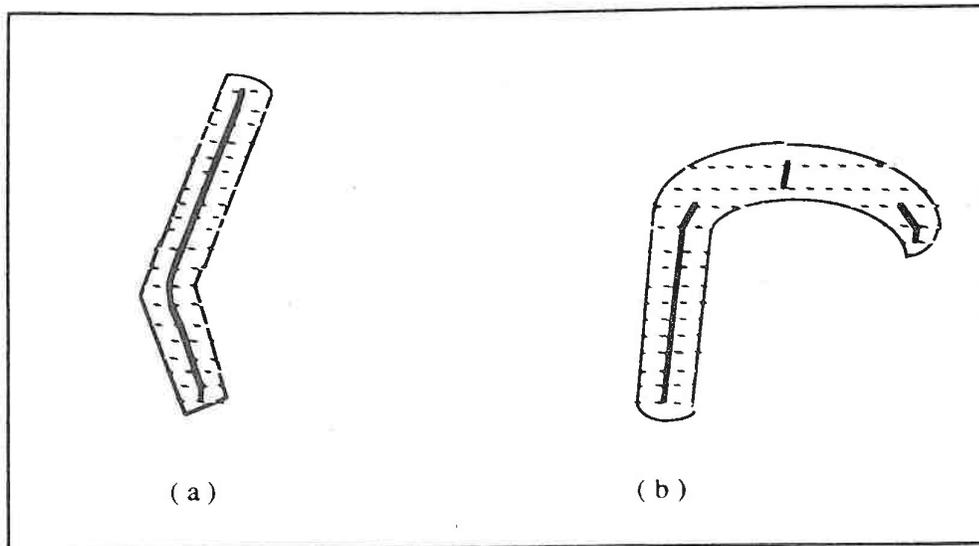


Figure 4.5. Exemple de squelette : (a) cas favorable; (b) cas d'échec

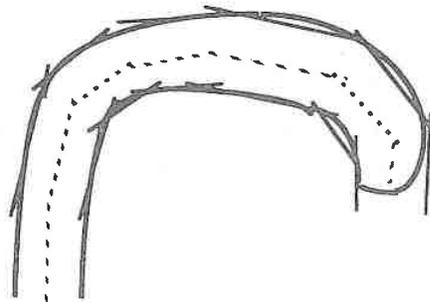


Figure 4.6. Méthode par appariement de contours opposés

à ce segment sont inférieures à un seuil prédéfini. Lorsque l'on rencontre un point ayant une distance supérieure au seuil, on découpe la courbe en ce point, et on recommence le processus à partir de ce point. Cette méthode est inefficace car à chaque étape, il faut parcourir tous les points délimités par deux points extrêmes. Depuis, de nombreuses améliorations sont abordées en utilisant d'autres critères. Toujours avec la méthode de suivi séquentiel, Wall et al. [Wall 84] utilisent un critère de déviation de l'aire, ce qui permet d'effectuer l'approximation polygonale en un seul parcours. Cette méthode est très rapide, toutefois, puisque elle est séquentielle, l'approximation n'est pas optimale surtout aux endroits où il y a un changement brutal de l'orientation des segments.

Nous avons adopté une méthode de découpage récursive [Duda 73] dans notre travail. Etant donné un ensemble de points, la méthode de découpage récursive s'effectue comme la suite (figure 4.7) :

- trouver les deux points A, B les plus éloignés
- pour tous les points restant, calculer les distances de ces points à la droite AB
- si toutes les distances sont inférieures à un seuil, le processus s'arrête.
- sinon trouver le point le plus éloigné de la droite AB, dit point C.
- couper la droite initiale AB en deux nouvelles droites AC et CB.
- le processus se répète respectivement sur les deux nouvelles droites.

Cette méthode a l'inconvénient d'être très sensible aux bruits, en effet, s'il y a un point (bruit) se situant en dehors de zone de l'ensemble de points, le résultat de cette méthode sera complètement faux. Nous n'avons pas attaché trop d'importance à ce problème, car nous avons constaté que les bruits sont éliminés par les traitements précédents. L'avantage important de cette méthode est qu'elle préserve bien les points anguleux, ce qui est moins évident par les autres méthodes [Tombre 87]. Rappelons que c'est exactement ce que nous cherchons lorsque nous repérons les coordonnées 2D des points de calibrage (chapitre 3).

Après l'approximation polygonale, une fusion des segments consécutifs est effectuée:

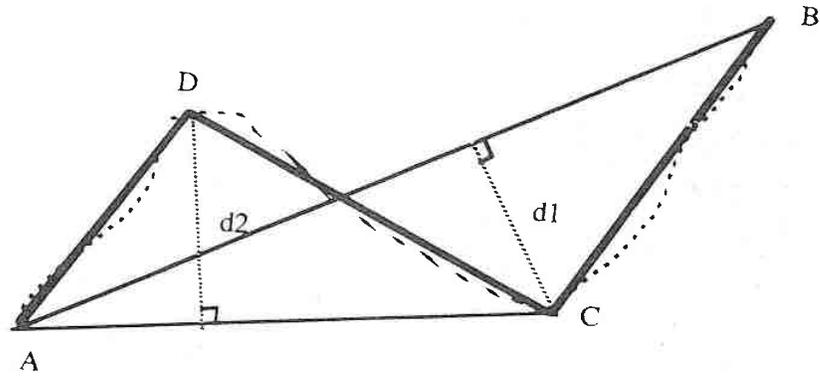
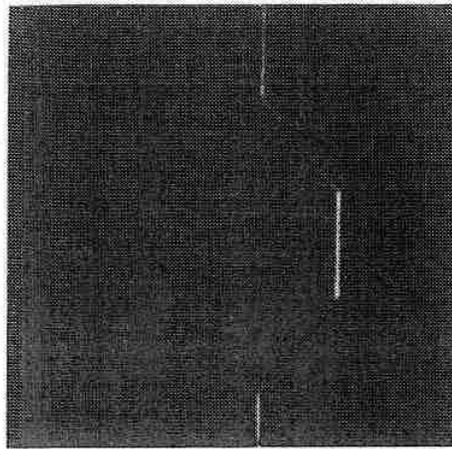


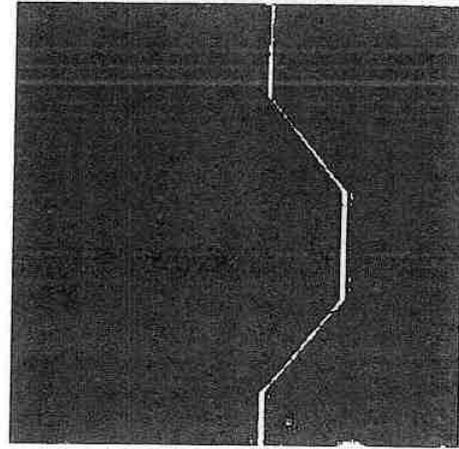
Figure 4.7. Découpage récursive

pour deux segments consécutifs non connexes, si leurs extrémités sont très proches l'un de l'autre, alors les deux segments sont étendus jusqu'à ce qu'ils rencontrent le point d'intersection des deux droites support de ces segments.

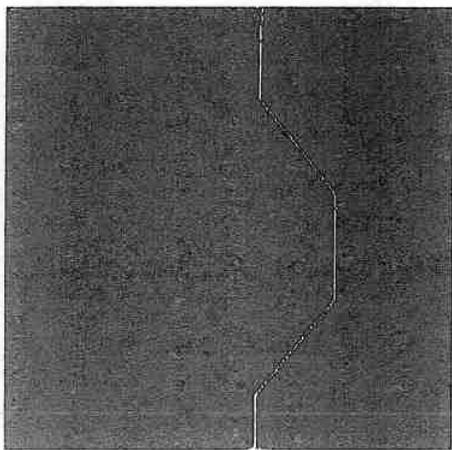
Les résultats de prétraitement sont montrés sur les figures 4.8, 4.9.



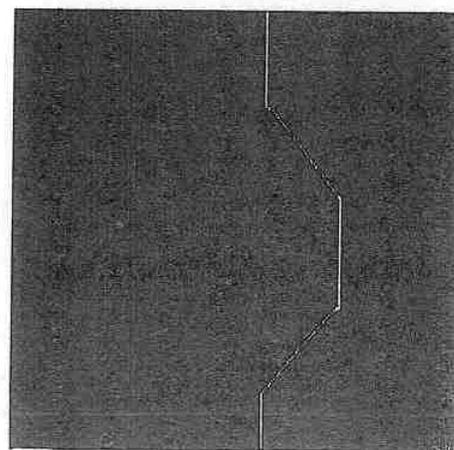
(a) image originale



(b) image binaire

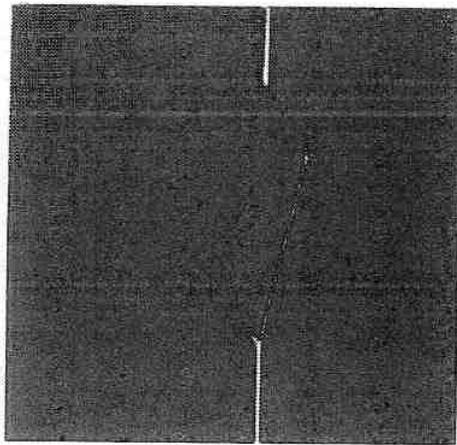


(c) image squelettisée

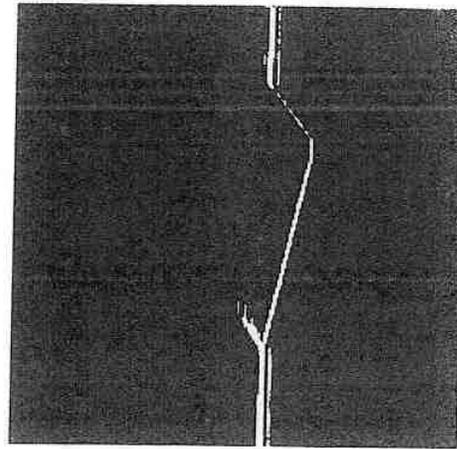


(d) image après l'approximation polygonale

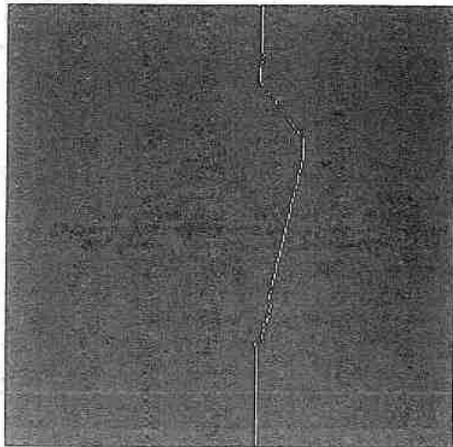
Figure 4.8. Résultat de prétraitement : (a) image originale; (b) image binaire; (c) image squelettisée (et après l'élimination de bruit); (d) image après le processus d'approximation polygonale



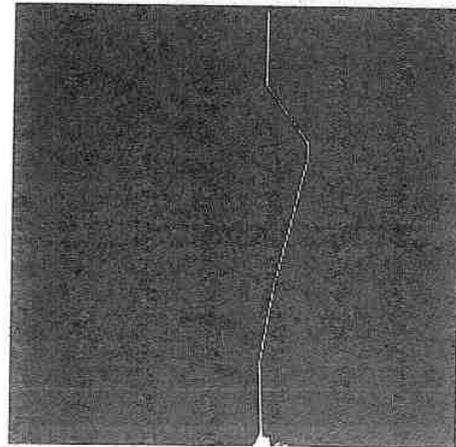
(a) image originale



(b) image binaire



(c) image squelettisée



(d) image après l'approximation polygonale

Figure 4.9. Exemple d'élimination de bruits dus aux réflexions multiples: (a) image originale; (b) image binaire; (c) image après squelettisation et élimination des reflets multiples; (d) image après approximation polygonale

4.5 Organisation des données 3D

Les données tridimensionnelles peuvent être stockées sous la forme d'une carte de profondeur (une image de distance) $z = f(x, y)$ ou sous la forme des points éparpillés x, y, z . La première forme est plus compacte, car au lieu de trois matrices, une seule matrice est suffisante pour représenter les données. Pour obtenir une telle représentation, un échantillonnage bidimensionnel et une interpolation bilinéaire sont nécessaires (voir Annexe A). Nous avons ensuite quantifié les valeurs de profondeur à 8 bits (soit 256 niveaux de gris). Une image de distance a donc un format identique que celui d'une image d'intensité, seul l'interprétation de la valeur d'un pixel est différente.

Dans la phase de construction de modèle (chap. 5), les informations denses de la scène est nécessaire, les données tridimensionnelles de la scène sont stockées sous forme d'image de distance (figure 4.10). En revanche, durant la phase de reconnaissance (chap. 6), nous ne disposons que de quelques segments tridimensionnelles, les données seront donc organisées en listes correspondant aux différents segments (figure 4.11).

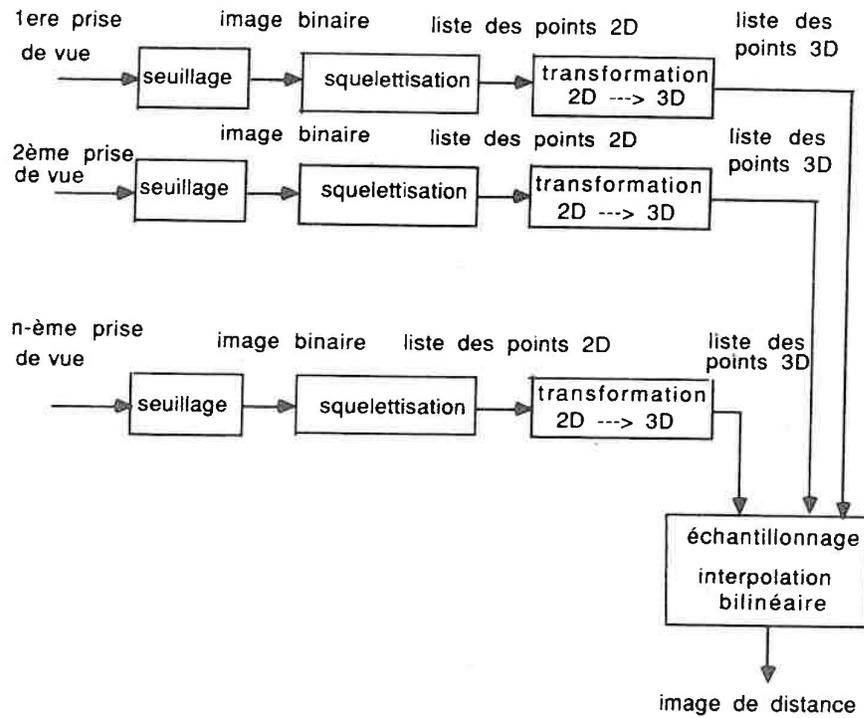


Figure 4.10. Les prétraitements effectués pour obtention des données durant la phase de modélisation

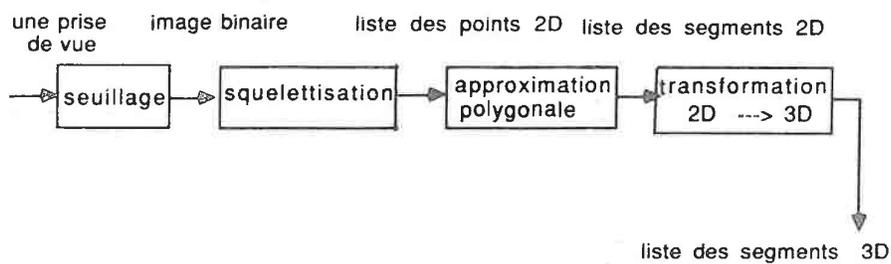


Figure 4.11. Les prétraitements effectués pour obtention des données durant la phase de reconnaissance

5

Modélisation des objets

5.1 Généralité

La modélisation d'objet consiste à recueillir des informations décrivant l'objet et à sauvegarder ces informations [Besl 85]. Pour de nombreuses applications telles que la vision par ordinateur, la robotique, le synthèse d'image, etc, il est nécessaire d'affronter le problème de la construction automatique de modèles d'objets tridimensionnels. Les différentes conditions préalables ainsi que les buts différents de ces applications requièrent une représentation différente. Cependant, une propriété commune que tous les types de représentations doivent posséder est la suivante :

le modèle (ou la représentation) d'un objet doit être indépendant de la position et l'orientation de l'observateur.

Dans ce paragraphe, nous exposons d'abord deux grandes familles de représentations, et nous présentons ensuite les problèmes que nous devons résoudre lors de la phase de modélisation.

5.1.1 Représentation d'objet tridimensionnel

Il existe de nombreuses représentations pour décrire un objet tridimensionnel, mais on peut les classer dans l'un de ces deux grandes familles :

1. Représentation volumétrique

La représentation volumétrique d'objet fournit explicitement des informations sur les espaces occupés par l'objet et les points de l'espace se situant à l'intérieur de l'objet. Dans cette catégorie de représentations, les trois représentations volumétriques les plus utilisées sont les suivantes:

- Constructif Solide Géométrie (CSG) :

Dans cette représentation, l'objet est représenté par un arbre binaire dont chaque feuille représente une simple primitive volumétrique qui pourrait subir un déplacement rigide, et chaque nœud interne spécifie le type d'opération Booléan (union, intersection ou soustraction) sur ses descendants. Les instances des simples primitives volumétriques sont typiquement des cubes, des cylindres, des sphères et des cônes. Un exemple de l'arbre CSG est illustré dans la figure 5.1(b).

Notons qu'une représentation CSG n'est pas unique, car un objet pourrait être construit de différentes façon.

Cette représentation est très compacte, cependant elle n'est pas efficace pour l'application de la reconnaissance et localisation d'objets du fait que le temps de calcul pour évaluer des informations surfaciques est souvent prohibitif, toutefois, elle est souvent adopté par les applications dans le domaine CAO ¹/FAO ². Elle offre en particulier de grande facilité de conception d'objets.

- Représentation par arbre octal :

Un arbre octal représente un objet par un ensemble de régions cubiques dans un espace tridimensionnel.

La structure utilisée pour sauvegarder cette représentation est un arbre dont chaque nœud interne représente un cube et pointe sur les huit autres nœuds qui décrivent respectivement l'occupation volumétrique de l'objet dans les huit sous-cubes correspondants de ce cube.

Dans certaines applications comme celle de la navigation d'un robot mobile dans un univers inconnu, cette représentation est souvent adoptée pour représenter les espaces occupés par les obstacles ainsi que les espaces libres [Arimoto 87].

- Cône généralisé

Un cône généralisé se compose d'une forme planaire arbitraire (dite une coupe) qui balaye le long d'une courbe 3D arbitraire (dite l'axe). La taille ainsi que la forme de la coupe pourrait varier le long de l'axe. Rao et al. [Rao 88] ont utilisé une telle représentation pour construire une description volumétrique des objets à partir de contours 3D obtenus par la stéréo.

1. Conception Assistée par Ordinateur

2. Fabrication Assistée par Ordinateur

2. Représentation surfacique

La représentation la plus familière dans cette catégorie de représentations est la représentation par **frontière**³ (voir figure 5.1(c)).

Dans cette représentation, un objet est représenté par les frontières (surfaces enfermées) qui le composent, mais la description de surface est différente selon le type de surface.

La plus simple représentation par frontière est la **représentation polyédrique**, qui représente un objet par une liste des facettes polygonales. Les surfaces de type arbitraire peuvent être approximées à tous les degrés de précision désirés en utilisant un nombre plus ou moins important de polygones. Hebert [Hebert 83] a utilisé une telle représentation dans son système de reconnaissance d'objets.

Pour décrire plus précisément les morceaux de surfaces de forme générale (non planaire), les différentes représentations de surface ont été proposées. Nous en citons ci-dessous les deux les plus utilisées :

- représentation par quadrique [Han 87],[Hebert 83],[Besl 86].
- représentation par B-spline[Yang 85]

La représentations surfacique est bien adoptée aux applications comme la reconnaissance et localisation d'objets et la synthèse d'image grâce à la disponibilité des représentations explicites des surfaces, des contours et des sommets [Faugeras 86].

On constate que le choix du paradigme de représentation dépend de l'application visée. Le problème auquel nous nous intéressons est la reconnaissance et la localisation d'objets, dans la suite de ce chapitre, nous ne discutons donc que de cette dernière représentation.

3. en Anglais : boundry représentation

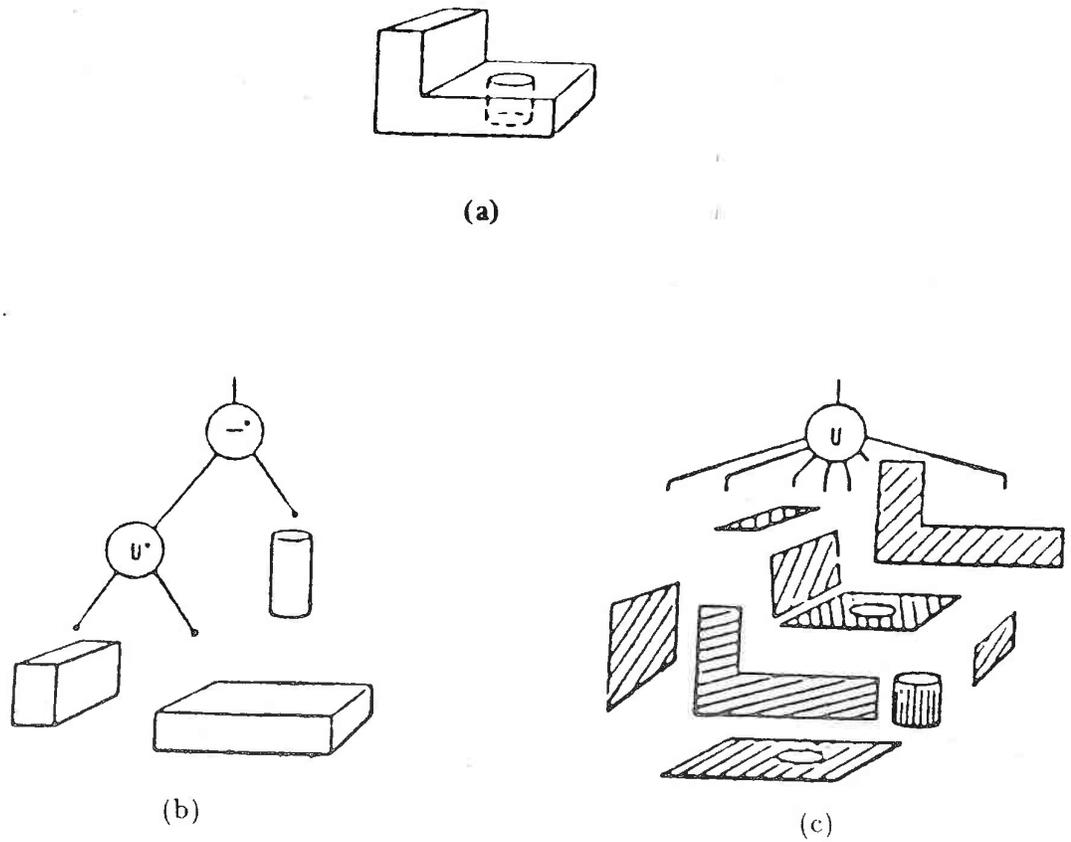


Figure 5.1. Représentation d'objet tridimensionnel [Ballard82]:
(a) Un objet à représenter; (b) L'arbre CSG; (c) Représentation de frontière

5.1.2 Construction de modèle

La construction de modèle d'un objet 3D comporte en général deux étapes :

1. Extraction des primitives caractéristiques décrivant les surfaces visibles pour chaque vue de points différents.

Le problème est de choisir le type de primitives à extraire et la méthode d'extraction.

- **le choix des primitives surfaciques**

Le choix de primitives pour représenter l'objet dépend d'une part de la facilité d'extraire ces primitives de l'image de distance, et de l'efficacité d'apparier les primitives de la scène avec ceux du modèle d'autre part. D'une manière générale, les primitives doivent satisfaire des propriétés suivantes [Faugeras 88] :

- n'être pas sensibles au bruit de mesure
- être invariants de la position et l'orientation de l'observateur, c'est-à-dire, les primitives exprimées dans deux repères différents doivent être reliées par la même transformation que celle qui relie ces deux repères.

Faugeras et al. [Faugeras 86] et Muller [Muller 84] ont choisi les plans et les quadriques pour approximer des morceaux de surface dans leur représentation.

Besl et Jain [Besl 86] ont indiqué que la géométrie différentielle fournit une description locale complète de surfaces. La courbure moyenne et la courbure gaussienne sont des caractéristiques surfaciques de second ordre qui possèdent des propriétés invariantes par rapport au changement de la paramétrisation de surface. On peut classer une région en l'un de huit types élémentaires selon les signes de ces deux courbures. Les huit types élémentaires de surfaces sont *plat*, *pic*, *fosse*, *minimal*, *crête*, *selle-crête*, *vallée* et *selle-vallée* qui sont illustrés sur la figure 5.2.

Hoffman et al. [Hoffman 87] ont distingué les régions *convexes*, *concaves*, et *plates* comme des primitives dans leur représentation.

Dans notre cas, les objets sont des polyèdres, les primitives sont donc des surfaces planes qui sont caractérisées par les normales de surfaces,

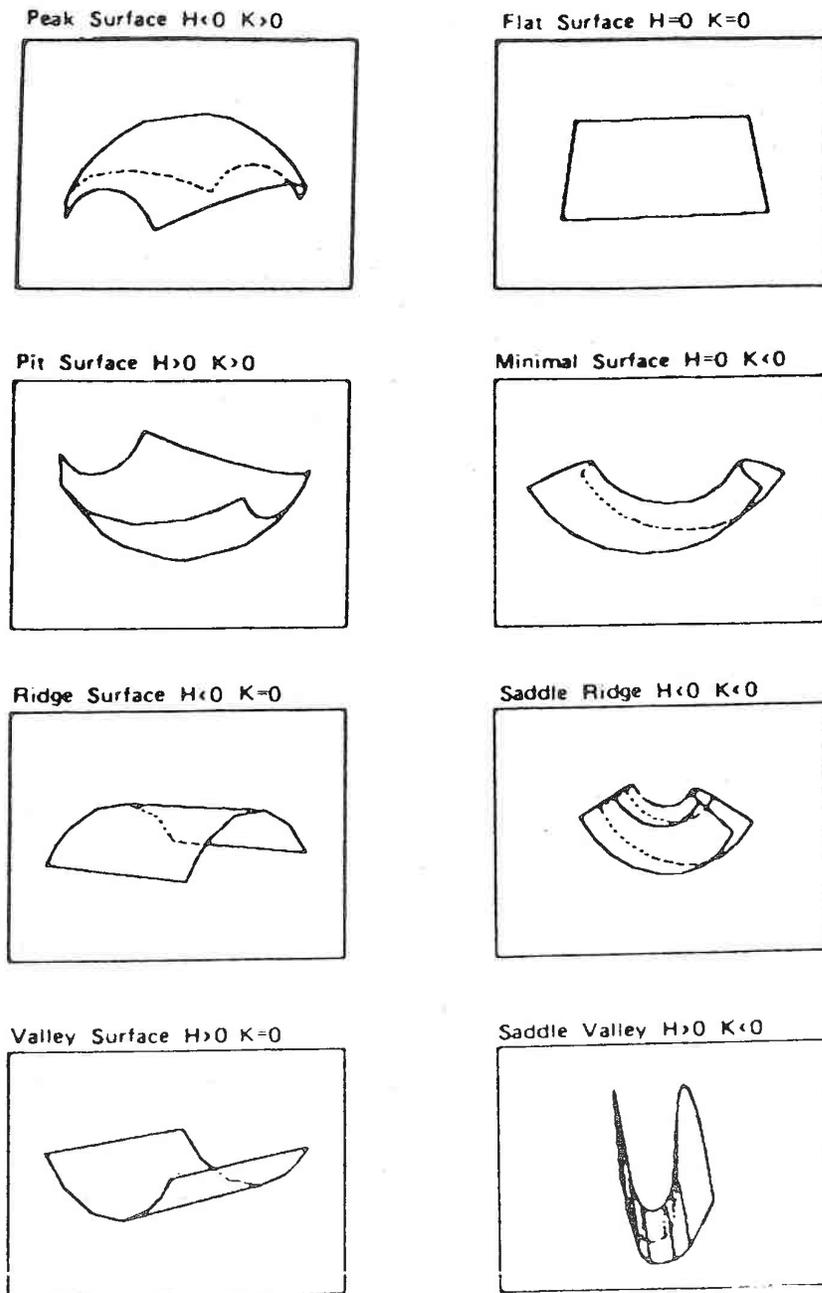


Figure 5.2. Les huit types élémentaires de surfaces [Besl86]:
 H est la courbure moyenne, et K est la courbure gaussienne

les distances des surfaces par rapport à l'origine de repère, les contours entourant la surface, et les sommets.

- **extraction de ces primitives**

Ce problème fait l'objet de la segmentation que nous présentons en détail dans le paragraphe suivant.

2. Intégration des descriptions partielles inférées de différentes vues pour une description complète de l'objet.

Le problème qu'on va résoudre dans cette étape est d'établir des correspondances entre des primitives extraites de vue de points différents, c'est-à-dire, repérer des vues multiples dans un système de coordonnées commun.

La principale difficulté se révèle dans la deuxième étape, c'est-à-dire, le repérage de différentes vues. Dans la méthode de Potmesil [Potmesil 83], les segments de surfaces de différentes vues sont acquis de telle sorte qu'ils se chevauchent partiellement. Le repérage de différentes vues est donc accompli par la mise en correspondance des parties communes des surfaces de différentes vues en minimisant les différences sur la position, l'orientation et la courbure entre ces surfaces. L'inconvénient majeur de cette méthode est qu'un nombre relativement important de vues (plus de 12) est nécessaire pour assurer les recouvrements partiels entre les différentes vues. Une façon pour simplifier le problème de repérage de différentes vues est d'acquérir les différentes vues en connaissant les transformations entre elles. Wang et al. [Wang 87] ont proposé une méthode de construction de modèle d'objet en exploitant conjointement les données acquises par la technique d'acquisition passive et celle d'acquisition active. Le principe en est le suivant :

- Une description de volume par frontières est d'abord construite en utilisant les contours extérieurs ⁴ extraits de différentes vues par l'acquisition passive en supposant que:

- les transformations entre ces différentes vues soient connues.
- le modèle de la caméra soit la projection orthogonale.

4. en Anglais : occluding contours. les contours extérieurs délimitent l'étendue de l'objet, et ils se produisent à l'endroit où il y a un changement brutal en profondeur

- les structures surfaciques partielles de différentes vues obtenues par l'acquisition active sont ensuite incorporées pour raffiner la description de volume par frontières initiales.

Cette méthode a l'avantages de permettre de construire le modèle avec un nombre variable de vues (éventuellement peu de vues) du fait que :

- Il ne nécessite pas de révéler tous les surfaces tridimensionnelles de l'objet.
- Il n'est pas obligatoire que les différentes vues se chevauchent partiellement.

Toutefois, il nous paraît difficile de modéliser un objet complexe avec une précision suffisante avec cette méthode, et le modèle de la projection parallèle devient peu raisonnable dans le cas où la distance entre l'objet et la caméra n'est pas suffisamment grande devant la distance focale de la caméra.

Connolly et al. [Connolly 86] présentent une méthode de construction de modèle polyédrique avec plusieurs vues de l'objet. Pour chaque vue, les points de contours sont extraits en analysant les courbures de surfaces. Les contours sont ensuite squelettisés et étendus pour obtenir des sommets. Un plan est approximé par un certain nombre de sommets qui lui appartiennent, les positions des sommets trièdres sont plus précisément obtenues en calculant l'intersections de trois plans concernés. Pour les sommets d'ordre supérieur, ils calculent l'intersection de tous les combinaisons des trois plans, le centre de gravité de ces points d'intersection est considéré comme la position du sommet. Une fois les positions des sommets modifiées, les plans concernés sont reestimés, et ce processus se répète jusqu'à ce que l'erreur maximale entre des sommets et les plans approximés soit inférieure à un seuil. Ils utilisent ensuite un algorithme de Wesley et Markowsky [Markowsky 80] pour construire le polyèdre pour chaque vue. Cet algorithme assure l'intégralité géométrique et la fermeture du modèle créé. Le modèle final est obtenu par l'intersection de tous les polyèdres créés par chaque vue.

Au lieu de traiter chaque vue séparément, et de les fusionner ensuite, Bhanu [Bhanu 84] propose une autre méthode qui traite ensemble toutes les vues. pour créer un modèle 3D d'un objet, l'acquisition d'une vue de l'objet s'effectue tous les trente degrés en tournant le plateau où est placé l'objet. Enfin une vue de dessus et une vue de dessous de l'objet sont acquises. Ces deux dernières vues sont mises en correspondance avec les autres vues à l'aide de trois points de contrôle qui sont visibles dans la première vue et également dans chacune de ces deux vues. Après

l'élimination du fond, tous les points 3D de ces 14 vues sont transformés dans un repère commun. Les régions planaires sont détectées par la méthode de "three points seed". Les points appartenant aux plusieurs surfaces sont détectés. Enfin, les régions sont approximées par des polygones.

Nous avons choisi de mettre en œuvre une méthode de construction de modèles des objets polyédriques à partir de plusieurs vues de l'objet. Les différentes vues sont acquises par une rotation connue de l'objet. Pour chaque vue, au lieu d'extraire des arêtes de données 3D, et d'approximer ensuite des plans à partir de ces arêtes extraites comme ce que font Connolly et al. [Connolly 86], nous extrayons directement des régions planes, et approximations ensuite le plan par tous les points appartenant à cette région. Nous constatons que le plan ainsi estimé est plus précis, car toutes les faces entourées par plus de trois arêtes ne sont parfaitement planaires que par chance à cause de :

- l'erreur entachée aux données 3D
- l'erreur introduite par localisation des arêtes (des contours)

Ayant obtenue la partition en régions, chaque vue est représentée par un graphe d'adjacence. Le modèle complet est finalement obtenu par fusion de tous les graphes correspondant aux différentes vues.

On constate que le modèle construit à partir des données acquises est inévitablement entaché de bruit, il est donc plus difficile d'apparier des primitives de la scène avec ceux du modèle ainsi construit que ceux du modèle idéal lors de la phase de reconnaissance.

Dans la suite de ce chapitre, nous présentons d'abord des techniques de segmentation pour extraire des primitives, et nous décrirons ensuite la reconstruction de modèle d'objet.

5.2 Segmentation de l'Image de Distance

La segmentation d'une image de distance est un processus fondamental tant pour décrire la scène que pour reconnaître des objets de la scène. De fait, la segmentation consiste à partitionner l'image en des parties qui ont des caractéristiques similaires (plus précisément, des régions du même type de surface) ou à en extraire des primitives importantes dans le but de reconnaître des objets par la méthode de

mise en correspondance. On trouve dans la littérature de nombreuses techniques de segmentation, en général, on peut les classer en deux catégories :

- méthode fondée sur les régions :

Cette méthode consiste à grouper des pixels en des régions en respectant la règle d'homogénéité ou de similarité (ex: des caractéristiques de géométrie différentielle).

- méthode fondée sur le contour :

Cette méthode consiste à extraire des discontinuités sur les distances (profondeurs) ou sur les normales des surfaces, ou les lignes de courbure, et à former ensuite des contours fermés.

Il existe une autre catégorie (dite méthode hybride) qui consiste à combiner les deux méthodes citées ci-dessus.

Nous exposons dans ce paragraphe d'abord les différentes catégories des méthodes citées ci-dessus, et nous présentons ensuite la méthode que nous avons choisie.

5.2.1 Segmentation par extraction de contour

L'extraction de contours de l'image de distance est une tâche difficile. Les principales causes sont les suivantes :

- Les images de distance sont souvent plus bruitées que les images d'intensité dû à l'aspect inhérent d'acquisition des données 3D. en fait, une image de distance subit des erreurs provenant de la calibration du système proprement dite, et de la limite de la précision mécanique du système (précision de translation ou rotation), et enfin de la quantification des valeurs de distance.
- La deuxième cause est liée directement aux types de contours existants dans l'image de distance.

Dans le cas d'image d'intensité, les contours sont souvent formés par la discontinuité des valeurs de l'intensité, c'est-à-dire, de type "marche d'escalier". Or, dans le cas de l'image de distance, on distingue en général trois types importants des contours (figure 5.3) :

- contour de type "marche d'escalier" :
Ce sont l'ensemble des points où les valeurs de distances (profondeurs) sont discontinues. Ce type de contour se produit lorsqu'un objet occulte un autre objet.
- contour de type "toit" :
Ce sont l'ensemble des points sur lesquels les normales des surfaces sont discontinues (par exemple, la ligne d'intersection de deux plans).
- contour "lisse" :
Ce type de contour est caractérisé par la continuité sur les normales des surfaces, mais la discontinuité sur les courbures.

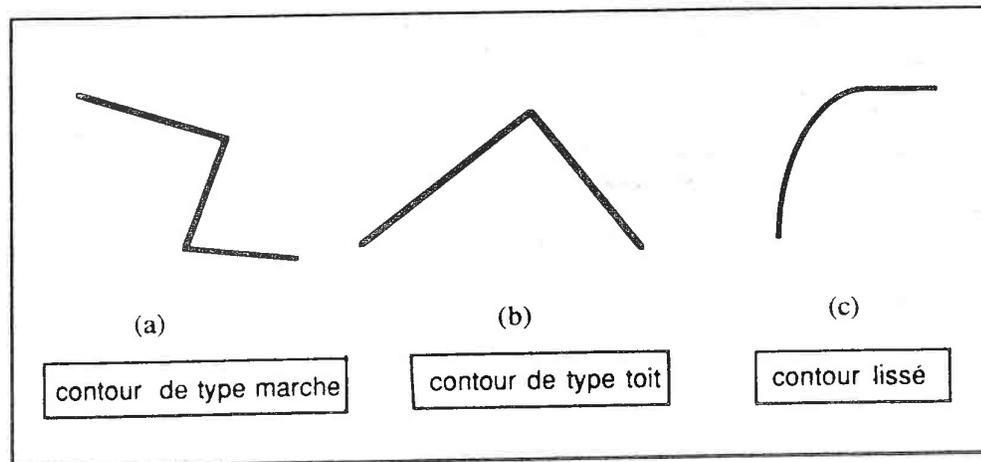


Figure 5.3. Les différents types de contours

Ces différents types de contours ont des caractéristiques différentes, il est donc difficile de détecter tous les types de contours avec un seul opérateur.

La détection de contours de type "marche" est la plus simple. Il suffit d'évaluer les différences de premier ordre des valeurs de profondeur en des points adjacents. Les contours correspondent aux maxima de ces différences [Mitiche 83].

Ponce et al. [Ponce 85] détectent des différents types de contours en analysant des caractéristiques liées à la géométrie différentielle de surfaces. En effet,

les contours du type "marche" correspondent aux passages par zéro de la courbure gaussienne; et ceux du type "toit" correspondent aux extrêma de la courbure principale.

En revanche, à notre connaissance, aucun travail n'a été mené sur la détection des contours lisses.

On constate que les méthodes fondées sur contour sont souvent locales, elles sont plus sensibles au bruit.

5.2.2 Segmentation par extraction de région

Transformée de Hough

La transformée de Hough a été d'abord utilisée pour détecter des lignes droites dans une image. Elle est ensuite généralisée pour détecter des courbes (ou des surfaces) paramétrées dans un ensemble de points [Ballard 82].

Etant donnée une courbe paramétrée $f(x, \mathbf{v}) = 0$, où \mathbf{v} est un vecteur dont les composants sont des paramètres à chercher. La transformée de Hough consiste à discrétiser d'abord l'espace des paramètres, et à former une matrice d'accumulateur. Pour chaque point x de l'image, tous les points de la matrice d'accumulateur dont les paramètres \mathbf{v} satisfont $f(x, \mathbf{v}) = 0$ sont incrémentés. Les maxima locaux dans la matrice correspondent donc aux points se situant sur les courbes recherchées dans l'image.

Hebert [Hebert 83] a utilisé cette technique pour détecter des morceaux de plans, de cylindres et de cône sur la surface des objets.

Leur méthode procède en quatre étapes :

1. On présente les normales aux points par leurs extrémités sur la sphère unité :
 - Une régions plane sur l'image est caractérisée par un petit nuage de points (idéalement, un seul point) dans cette sphère unité.
 - l'ensemble des normales sur un cylindre est représenté par un cercle sur cette sphère unité.
 - l'ensemble des normales sur un cône est représenté par un petit cercle sur cette sphère unité.

2. la technique de la transformée de Hough est utilisée pour détecter ce nuage de points et ces cercles.
3. Le troisième étape consiste à récupérer les points correspondants à chaque région détectée dans l'image.
4. Le dernier étape consiste à séparer, en utilisant les relations de voisinage de point ainsi que les coordonnées (x, y, z) , des régions qui ne peuvent pas être distinguées par la seule utilisation des normales (par exemple, deux régions planes distinctes mais parallèles ou deux cylindres avec des axes parallèles).

Ce type d'algorithme fournit souvent une segmentation assez stable, en revanche, il ne permet que de détecter des primitives surfaciques simples. Lorsque le nombre de paramètres dépasse deux, le temps de calcul ainsi que son coût en mémoire deviennent importants.

Croissance de région

Le principe de cette méthode est la suivante :

Supposons que l'on dispose d'une segmentation initiale de l'image, on définit un critère d'erreur E , à chaque étape on fusionne deux régions voisines R_i, R_j qui produisent une erreur $E(R_i \cup R_j)$ minimale. Le processus s'arrête lorsqu'aucun couple de régions ne peut plus être fusionné.

Cette méthode est plus générale car elle permet de segmenter des surfaces de forme arbitraire.

Nous avons choisi de mettre en œuvre une segmentation de ce type [Monga 86], [WrobelDautcourt 88].

5.2.3 Notre approche

I. Définition de problème

La segmentation fondée sur l'approximation par des fonctions est couramment adoptée dans le cadre de l'analyse de l'image de distance.

Le problème de segmentation fondée sur l'approximation par des fonctions peut être défini comme suite [Besl 88] :

Etant donnée une surface digitalisée $g(i, j)$ spécifiée par les valeurs z_{ij} , une *segmentation/reconstruction* consiste à trouver N fonctions approximatives $\hat{f}_i(x, y)$ et N régions \hat{R}_i sur lesquelles ces fonctions sont évaluées de telle sorte que :

- L'erreur total de l'approximation:

$$\varepsilon_{total} = \|g(i, j) - \hat{g}(x(i), y(j))\|_I$$

soit petite ;

où $\hat{g}(x, y)$ est la fonction d'image reconstruite :

$$\hat{g}(x, y) = \sum_{l=1}^N \hat{f}_l(x, y) \chi(x, y, \hat{R}_l)$$

avec

$$\chi(x, y, \hat{R}_l) = \begin{cases} 1 & \text{si } (x, y) \in \hat{R}_l \\ 0 & \text{autres} \end{cases}$$

et I est l'ensemble des points de l'image.

- Le nombre total de régions N soit petit.

Remarquons que la première condition indique que les fonctions surfaciques définies sur les régions de l'image doivent représenter mathématiquement les faces naturelles (visibles) des objets de la scène ; et la vérification de la deuxième condition évite une sur-segmentation.

Dans notre cas, les objets sont polyédriques. La qualité de segmentation est maximale lorsque les régions segmentées correspondent aux facettes naturelles visibles des objets de la scène.

II. Choix de prédicat d'homogénéité

Dans l'analyse de l'image d'intensité, il est très fréquent de segmenter l'image par la méthode de croissance de régions. Dans une image d'intensité, la seule caractéristique dont on dispose est le niveau de gris de chaque point de l'image. Or, le niveau de gris d'un pixel représente le nombre de photons visibles incident dans la position (i, j) du plan image, qui est une fonction complexe faisant intervenir l'éclairage, la réflectance de la surface de l'objet, la géométrie de l'objet ainsi que la

position et l'orientation de l'observateur. Il est donc difficile de définir un prédicat d'homogénéité caractérisant tous ces différents facteurs [WrobelDautcourt 88]. Il se peut que, dans une image d'intensité, deux zones de caractéristiques différentes appartiennent pourtant à la même surface d'un objet à cause de l'ombre. Une image de distance a le même format qu'une image d'intensité, mais elle s'en distingue par l'interprétation de la valeur d'un pixel. En fait, la valeur d'un pixel dans une image de distance représente la distance d'un point d'une surface physique à une surface de référence. On dispose donc directement des caractéristiques géométriques des objets de la scène. La définition d'un prédicat d'homogénéité est plus facile et claire au sens physique du fait qu'une image de distance représente une vraie surface digitalisée.

Dans le cas de segmentation fondée sur l'approximation par des fonctions, le prédicat d'homogénéité requiert que les points appartenant à une même région puissent être bien représentés par une même fonction surfacique. Les fonctions utilisées dans la littérature sont multiples : les surfaces planaires ou quadriques [Oshima 83], les sphères, les cylindres, les cônes [Hebert 83], des polynômes d'ordre variable [Besl 88], etc.

Dans notre cas, les régions sont planaires, le prédicat d'homogénéité est donc un critère d'erreur :

$$E(R_i \cup R_j) = \text{Min}_{a,b,d} \sum_{i=1}^n (z_i - ax_i - by_i - d)^2$$

avec $(x_i, y_i, z_i) \in (R_i \cup R_j)$

les paramètres a, b, d sont estimées par la méthode de moindre carré :

en dérivant respectivement E par rapport à a, b , et d , et en les mettant ensuite à zéro, les paramètres a, b et d sont :

$$a = \frac{(\sum_i x_i z_i - \bar{x} \sum_i z_i)(\sum_i y_i^2 - \bar{y} \sum_i y_i) - (\sum_i x_i y_i - \bar{x} \sum_i y_i)(\sum_i y_i z_i - \bar{y} \sum_i z_i)}{(\sum_i x_i^2 - \bar{x} \sum_i x_i)(\sum_i y_i^2 - \bar{y} \sum_i y_i) - (\sum_i x_i y_i - \bar{x} \sum_i y_i)^2}$$

$$b = \frac{(\sum_i x_i^2 - \bar{x} \sum_i x_i)(\sum_i y_i z_i - \bar{y} \sum_i z_i) - (\sum_i x_i y_i - \bar{x} \sum_i y_i)(\sum_i x_i z_i - \bar{x} \sum_i z_i)}{(\sum_i x_i^2 - \bar{x} \sum_i x_i)(\sum_i y_i^2 - \bar{y} \sum_i y_i) - (\sum_i x_i y_i - \bar{x} \sum_i y_i)^2}$$

$$d = \bar{z} - a\bar{x} - b\bar{y}$$

où \bar{x} , \bar{y} et \bar{z} sont respectivement les valeurs moyennes des coordonnées x , y , z des points.

III. Algorithme

L'algorithme que nous présentons ici est analogue à ceux développés par [Monga 86], [WrobelDautcourt 88] et est une extension en segmentation de l'image 3D.

L'algorithme de segmentation comporte deux étapes :

- **partition initiale**

L'image originale est divisée en des petites régions très homogènes qui ne contiennent pas de discontinuité.

- **reconstruction des surfaces**

La partition initiale est itérativement raffinée par fusion des régions adjacentes vérifiant certains prédicats.

(A). Partition initiale

Nous avons utilisé le processus de la création de la partition initiale développé par B. Wrobel-Dautcourt [WrobelDautcourt 88] en modifiant tout simplement les prédicats d'homogénéité. Nous en décrivons ici brièvement le principe :

La partition se fait par une division récursive de l'image originale en utilisant la structure de l'arbre quaternaire ⁵.

Pour une image de taille $2^N \times 2^N$, son arbre quaternaire a N niveaux :

le niveau 0 correspond à l'ensemble de l'image ; et le niveau N correspond au pixel de l'image.

Partant de niveau 0, un test d'homogénéité est effectué. Si le test échoue, on divise la zone en quatre sous-zones de même taille, et le processus se répète jusqu'à ce que soit tous les zones sont homogènes, soit la division atteint le niveau final, c'est-à-dire, le niveau pixel.

Dans cette étape, le problème qui nous préoccupe est que les régions segmentées par cette partition initiale ne doivent pas contenir des discontinuités

5. en Anglais : quadtree

(des points de contour de type "marche" et de type "toit"). Pour assurer ceci, nous avons utilisé à la fois le prédicat sur la distance et le prédicat sur la normale.

Les prédicats utilisés dans cette étape :

Pour la division en niveau n ($n = 0$ à $N-2$) :

$$z_{imax} - z_{imin} < seuil_1$$

où z_{imax} , z_{imin} sont respectivement la valeur maximale et la valeur minimale d'une zone.

Pour la division en niveau $N-1$:

- parmi ses quatre fils, il n'existe pas de points de contour de type "marche".
- l'angle maximal entre les normale de ses quatre fils $< seuil_2$

Remarquons que, pour le niveau $N-1$, nous n'avons pas utilisé le même prédicat que les niveaux inférieurs. Car avec le premier prédicat, une zone planaire qui a une grande pente risque d'être divisée. Notons aussi que les points de contours de type "marche" sont pré-calculés.

(B) Croissance de région

Dans notre implantation, une partition et les attributs de ses éléments ainsi que les relations des éléments adjacents sont représentés par un graphe.

- Chaque nœud N_i du graphe est attaché un élément (une région) de la partition courante et ses attributs.

Les attributs associés à un nœud sont les suivants :

- * Les caractéristiques qui permettent d'évaluer les paramètres de l'équation de cette région par la méthode de moindre carré: $\sum x_i$, $\sum y_i$, $\sum z_i$, $\sum x_i y_i$, $\sum x_i z_i$, $\sum y_i z_i$, $\sum x_i^2$, $\sum y_i^2$, $\sum z_i^2$.
- * Les paramètres \vec{n}_i , d_i de cette région: $\vec{n}_i \cdot \vec{X} + d_i = 0$
- * le nombre de points appartenant à cette région : Np
- * un pointeur pointant sur une liste des nœuds adjacents de ce nœud : ADJ

- A chaque arc sont attachés les attributs du couple de nœuds correspondants.

pour un couple de régions (R_i, R_j) , les attributs associés à un arc sont :

- * le produit scalaire des normals unitaires de deux régions : $\vec{n}_i \cdot \vec{n}_j$
- * le nombre des couples de points de la frontière entre les deux régions R_i, R_j : $Np(F(R_i, R_j))$
 $F(R_i, R_j)$ est l'ensemble des couple de points situés le long de la frontière entre les deux régions R_i, R_j .
- * l'écart moyen des distances entre des couples de points situés le long de la frontière entre les deux régions R_i , et R_j :

$$Ecart_{moy}(F(R_i, R_j)) = \sum_{k \in R_i, l \in R_j} \frac{|z_k - z_l|}{Np(F(R_i, R_j))}$$

A chaque itération, la procédure de croissance de régions consiste à chercher, parmi tous les arcs, l'arc qui vérifie le prédicat de distance et le prédicat de la normale, et qui produit une erreur $E(R_i \cup R_j)$ minimale, et à fusionner ensuite les deux nœuds pointés par cet arc. Le nouveau nœud $(R_i \cup R_j)$ remplace le nœud R_i , et le nœud R_j est enlevé.

Les prédicats utilisés :

- le prédicat de distance :

$$Ecart_{moy}(F(R_i, R_j)) < seuil_3$$

- le prédicat de normale : $\vec{n}_i \cdot \vec{n}_j > seuil_4$

- critère d'erreur : $E(R_i \cup R_j) < seuil_5$

$$\text{avec } E(R_i \cup R_j) = \text{Min}_{(a,b,d)} \sum_{i=1}^n (z_i - ax_i - by_i - d)^2,$$

et $(x_i, y_i, z_i)_{i=1,n} \in (R_i \cup R_j)$.

L'accès rapide au meilleur arc est assuré par un tri des arcs en fonction du critère d'erreur de fusion $E(R_i \cup R_j)$ à l'aide de la structure de l'arbre binaire. Le couple (R_i, R_j) qui produit une erreur $E(R_i, R_j)$ minimale est placé à la feuille gauche de l'arbre.

La procédure de croissance-de-région :

Etant donnée une partition initiale : $S^0 = R_1^0, \dots, R_i^0, \dots, R_n^0$:

arrêt = faux

tant que non arrêt faire

construire l'arbre binaire (trier des couples de régions)

si l'arbre binaire est vide

arrêt = vrai

fsi

tant que l'arbre binaire n'est pas vide faire

 * *fusionner le couple de régions (R_i^{k-1}, R_j^{k-1}) qui se trouve à la feuille gauche de l'arbre binaire*

 * *mise à jour la partition courante :*

$$S^k = S^{k-1} - R_i^{k-1} - R_j^{k-1} \cup R_i^k, \text{ avec } R_i^k = R_i^{k-1} \cup R_j^{k-1}$$

 * *enlever le couple de régions fusionnées de l'arbre binaire*

 * *remise à jour des attributs associés à R_i^k*

 * *remise à jour des attributs associés aux arcs reliant R_i^k*

 * *enlever les arcs reliant R_i^{k-1} ou R_j^{k-1} de l'arbre binaire*

 * pour tous les arcs reliant R_i^k

si le couple de régions satisfait les prédicats d'homogénéité

 · *insérer l'arc dans l'arbre binaire*

fsi

fpour

ftantque

affecter des régions qui ne comportent qu'un seul point à leurs plus proches régions voisines.

ftantque

Remise à jour des attributs associés à R_i^k

- Les caractéristiques permettant d'évaluer l'équation de la région telles que $\sum x_i, \sum y_i \dots$ sont respectivement les sommes des caractéristiques correspondantes associées à R_i^{k-1} et celles associées à R_j^{k-1} :

$$\text{par exemple, } \sum x_i = \sum x_{i'} + \sum x_j$$

$$\text{avec } x_i \in R_i^k, x_{i'} \in R_i^{k-1} \text{ et } x_j \in R_j^{k-1}.$$

- réestimer les paramètres de l'équation de la nouvelle région.
- le nombre de points appartenant à la nouvelle région :

$$Np(R_i^k) = Np(R_i^{k-1}) + Np(R_j^{k-1})$$

- la liste de régions adjacentes avec la nouvelle région :

$$ADJ(R_i^k) = (ADJ(R_i^{k-1}) - R_j^{k-1}) \cup (ADJ(R_j^{k-1}) - R_i^{k-1}) - \sum R_i^{k-1}$$

$$\text{avec } R_i^{k-1} \in (ADJ(R_i^{k-1}) \cap ADJ(R_j^{k-1}))$$

Remise à jour des attributs associés aux arcs reliant à R_i^k

Pour tous les arcs (R_i^{k-1}, R_i^{k-1})

- si R_i^{k-1} est aussi adjacent avec R_j^{k-1}

– mise à jour le nouvel arc (R_i^k, R_i^k)

– les attributs associés au nouvel arc :

$$* \text{Ecart}_{\text{moy}}(F(R_i^k, R_i^k)) = \frac{Np(F(R_i^{k-1}, R_i^{k-1})) \cdot \text{Ecart}_{\text{moy}}(F(R_i^{k-1}, R_i^{k-1})) + Np(F(R_i^{k-1}, R_j^{k-1})) \cdot \text{Ecart}_{\text{moy}}(F(R_i^{k-1}, R_j^{k-1}))}{Np(F(R_i^{k-1}, R_i^{k-1})) + Np(F(R_i^{k-1}, R_j^{k-1}))}$$

$$* Np(F(R_i^k, R_i^k)) = Np(F(R_i^{k-1}, R_i^{k-1})) + Np(F(R_i^{k-1}, R_j^{k-1}))$$

* réévaluer l'erreur $E(R_i^k, R_i^k)$

* recalculer le produit scalaire des normales des régions

- sinon

– mise à jour le nouvel arc (R_i^k, R_i^k)

– les attributs associés au nouvel arc :

$$* \text{Ecart}_{\text{moy}}(F(R_i^k, R_i^k)) = \text{Ecart}_{\text{moy}}(F(R_i^{k-1}, R_i^{k-1}))$$

- * $Np(F(R_i^k, R_i^k)) = Np(F(R_i^{k-1}, R_i^{k-1}))$
- * réévaluer l'erreur $E(R_i^k, R_i^k)$
- * recalculer le produit scalaire des normales des régions

- fsi

fpour

pour tous les arcs (R_i^{k-1}, R_j^{k-1})

- si R_i^{k-1} est aussi voisin de R_i^{k-1}

- enlever cet arc

- sinon

- l'arc (R_i^k, R_i^k) remplace l'arc (R_i^{k-1}, R_j^{k-1})

- $Ecart_{moy}(F(R_i^k, R_i^k)) = Ecart_{moy}(F(R_i^{k-1}, R_j^{k-1}))$

- $Np(F(R_i^k, R_i^k)) = Np(F(R_i^{k-1}, R_j^{k-1}))$

- réévaluer l'erreur $E(R_i^k, R_i^k)$

- recalculer le produit scalaire des normales des régions

- fsi

fpour

Remarquons que, après une fusion, les évaluations des attributs de nouveau nœud $(R_i \cup R_j)$ ainsi que celles des attributs associés aux arcs reliant le nouveau nœud ne nécessitent que d'exploiter les valeurs des attributs précédents associés respectivement à R_i , R_j et les arcs reliant R_i ou R_j , sans revenir aux données initiales.

IV. Estimation des normales

L'orientation d'une surface de l'objet est une caractéristique importante. L'orientation d'un point p sur la surface est représentée par un vecteur qui est normale au plan tangentiel en ce point, et est désigné par \vec{N}_p .

Les valeurs de profondeur dans une image de distance peuvent être représentées comme une fonction des coordonnées x et y , $z = f(x, y)$, où f est une fonction

différentielle au premier ordre sauf en des points de contours. Dans un petit voisinage d'un point (x, y) , comme $f(x, y)$ est continue, la normale unité $\vec{N}(x, y)$ est définie par l'équation suivante :

$$\vec{N}(x, y) = \frac{\nabla f(x, y)}{\|\nabla f(x, y)\|}$$

où ∇ est un opérateur de gradient.

Si l'on approxime la fonction $f(x, y)$ par l'approximation Taylor de premier ordre, dans ce cas, $f(x, y)$ décrit un plan :

$$ax + by + cz + d = 0$$

Puisque dans notre système de coordonnées, l'axe z s'oriente vers le haut, et que la surface est visible, on a donc toujours $c > 0$. Sans nuire à la généralité, en normalisant c à la unité, on a donc :

$$z = f(x, y) = ax + by + d$$

Les constantes a et b peuvent être calculées de la façon suivante :

$$a = \frac{\partial f(x, y)}{\partial x} = \lim_{\Delta x \rightarrow 0} \frac{f(x + \Delta x, y) - f(x, y)}{\Delta x}$$

$$b = \frac{\partial f(x, y)}{\partial y} = \lim_{\Delta y \rightarrow 0} \frac{f(x, y + \Delta y) - f(x, y)}{\Delta y}$$

Par approximation, on a :

$$a \approx f(x + 1, y) - f(x, y)$$

$$b \approx f(x, y + 1) - f(x, y)$$

Or, ce calcul est trop sensible au bruit, pour remédier cela, nous avons donc estimé ces constantes en approxinant au mieux la fonction f par un plan sur un voisinage de taille $M \times M$ du point.

Nous définissons un critère d'erreur à minimiser dans le voisinage de ce point avec $N = M \times M$:

$$E = \sum_{i=0}^{N-1} [z_i - f(x_i, y_i)]^2 = \sum_{i=0}^{N-1} [z_i - ax_i - by_i - d]^2$$

Z_{16}	Z_{15}	Z_{14}	Z_{13}	Z_{12}
Z_{17}	Z_4	Z_3	Z_2	Z_{11}
Z_{18}	Z_5	Z_0	Z_1	Z_{10}
Z_{19}	Z_6	Z_7	Z_8	Z_9
Z_{20}	Z_{21}	Z_{22}	Z_{23}	Z_{24}

5 × 5 voisinage autour de point Z_0

En dérivant E respectivement par rapport à a , b , et d , et en les mettant à zéro, les paramètres à estimer satisfont donc le système d'équation suivant:

$$\begin{pmatrix} \sum_{i=0}^{N-1} x_i^2 & \sum_{i=0}^{N-1} x_i y_i & \sum_{i=0}^{N-1} x_i \\ \sum_{i=0}^{N-1} x_i y_i & \sum_{i=0}^{N-1} y_i^2 & \sum_{i=0}^{N-1} y_i \\ \sum_{i=0}^{N-1} x_i & \sum_{i=0}^{N-1} y_i & N \end{pmatrix} \begin{pmatrix} a \\ b \\ d \end{pmatrix} = \begin{pmatrix} \sum_{i=0}^{N-1} x_i z_i \\ \sum_{i=0}^{N-1} y_i z_i \\ \sum_{i=0}^{N-1} z_i \end{pmatrix}$$

A noter que le choix du voisinage pourrait poser certain problème. Si l'on choisissait un voisinage trop petit, l'estimation de la normale serait très sensible au bruit et à l'erreur de la quantification des valeurs de distance; un grand voisinage pourrait donner des résultats plus fiables, mais il risque, dans un tel voisinage, de contenir des discontinuités (des contours) à proximité desquels l'estimation de la normale serait altérée. Dans notre expérimentation, nous avons choisi un voisinage de taille 5×5 qui est, il nous semble, un bon compromis.

Pour un voisinage de 5×5 , il est facile de vérifier que les estimations peuvent être implantées par les convolutions suivantes:

a: $\frac{1}{50 \times \Delta x}$

- 2	- 2	- 2	- 2	- 2
- 1	- 2	- 1	- 1	- 1
0	0	0	0	0
1	1	1	1	1
2	2	2	2	2

$$b: \frac{1}{50 \times \Delta y} \begin{array}{|c|c|c|c|c|} \hline -2 & -1 & 0 & 1 & 2 \\ \hline -2 & -1 & 0 & 1 & 2 \\ \hline -2 & -1 & 0 & 1 & 2 \\ \hline -2 & -1 & 0 & 1 & 2 \\ \hline -2 & -1 & 0 & 1 & 2 \\ \hline \end{array}$$

et

$$d = \frac{1}{N} \left(\sum_{i=0}^{N-1} z_i - a \sum_{i=0}^{N-1} x_i - b \sum_{i=0}^{N-1} y_i \right)$$

où Δx et Δy sont respectivement les pas d'échantillonnage suivant les axes X et Y .

Le vecteur de la normale estimée est donc $(-a, -b, 1)$.

V. Résultats

Les résultats de segmentation sont illustrés par les figures 5.4 – 5.7. les différentes régions segmentées sont codées en différents niveaux de gris. Les résolutions de ces deux images de distance originales sont respectivement ⁶ : $\Delta x = 0.952$, $\Delta y = 0.866$, $\Delta z = 0.213$ pour l'image de figure 4.4 ; et $\Delta x = 0.952$, $\Delta y = 0.866$, $\Delta z = 0.399$ pour l'image de figure 4.6. Les seuils choisis pour l'image en figure 4.4 sont les suivants :

$$\text{seuil}_1 = 15 \text{ pixels}$$

$$\text{seuil}_2 = 6 \text{ degrés}$$

$$\text{seuil}_3 = 2 \text{ mm}$$

$$\text{seuil}_4 = 0.95 \text{ (correspond à 18 degrés)}$$

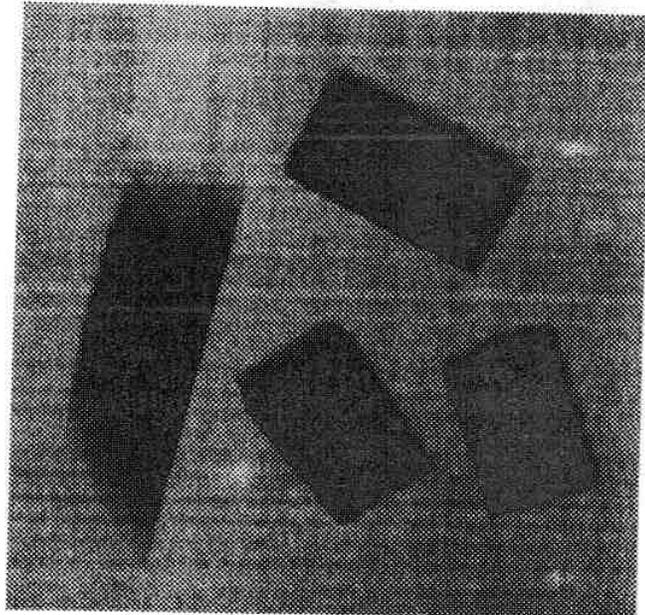
$$\text{seuil}_5 = 0.64 \text{ (correspond à une distance moyenne de 0.8 mm des points au plan estimé)}$$

Les mêmes seuils sont choisis pour l'image en figure 4.6, à l'exception de $\text{seuil}_1 = 12 \text{ pixels}$.

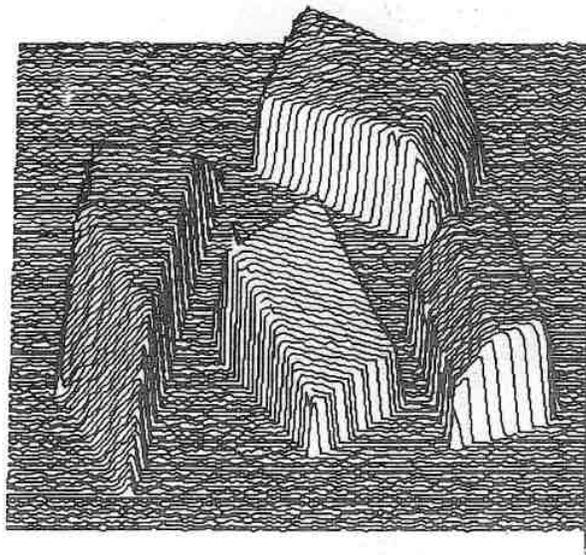
Remarquons que les régions non verticales sont assez bien segmentées, et l'algorithme est assez stable en raison de la signification physique des prédicats, il est facile de choisir des seuils ; remarquons aussi que les contours ne sont pas

6. Ces deux images de distance sont fournies par l'Université de Clermont-ferrand

bien localisés, ce phénomène est dû essentiellement à l'aspect d'échantillonnage de l'espace et à la distortion des normales à la proximité des points de contours. les contours peuvent être localisés plus précisément par l'intersection de deux régions correspondantes.

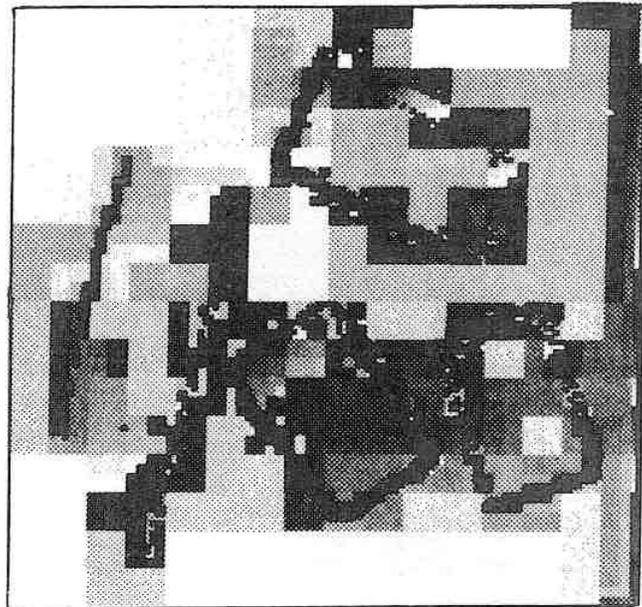


(a)

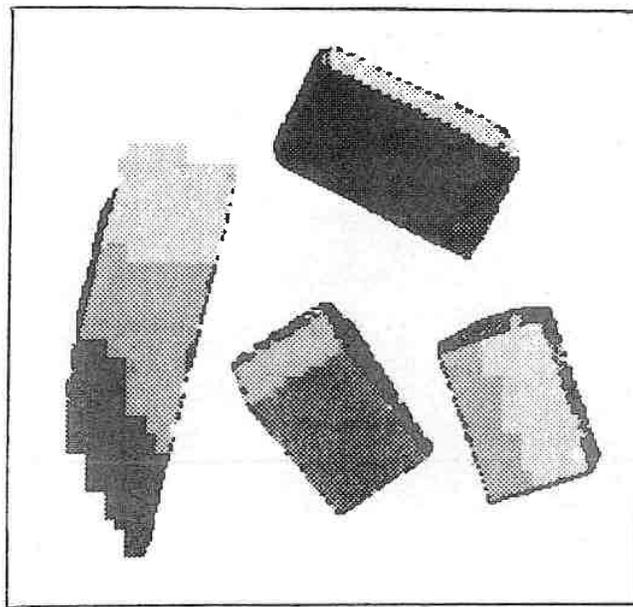


(b)

Figure 5.4. Résultat de segmentation en région : (a) image de distance originale
(b) une vue perspective

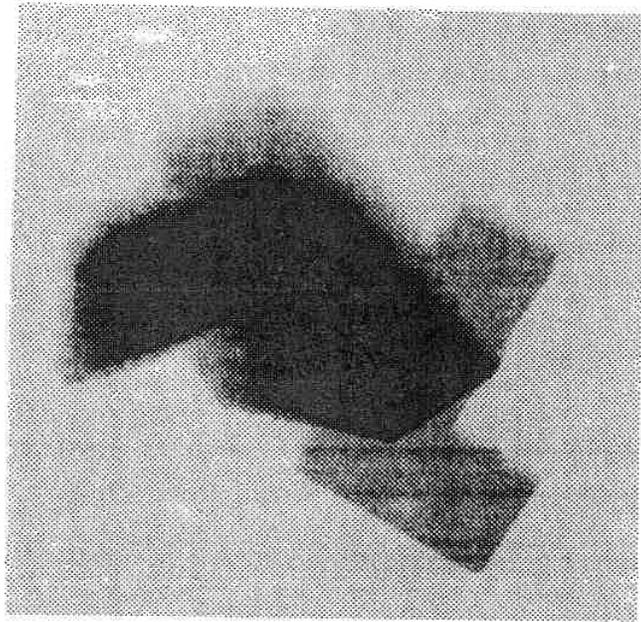


(c)

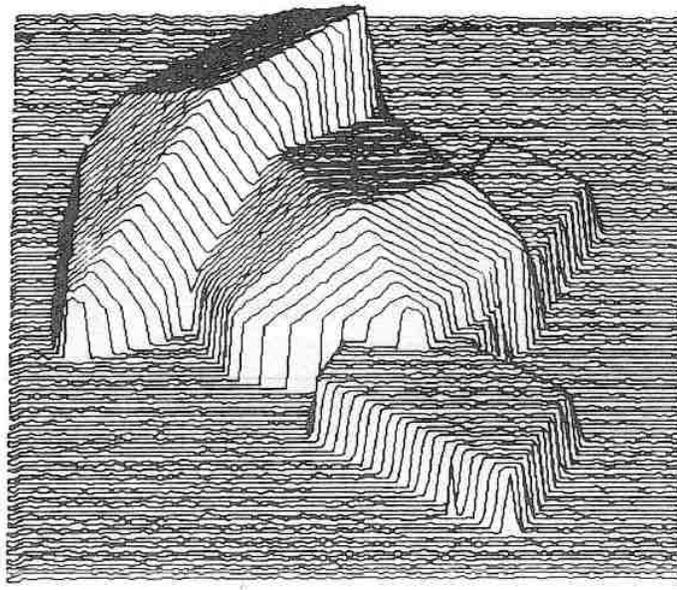


(d)

Figure 5.5. Résultat de segmentation en région : (c) partition des régions initiales
(d) des régions segmentées
le nombre de régions initiales est de 7312, le nombre d'arcs initiaux est de 15991
(sur l'image (c), les points noirs sont des points isolés)

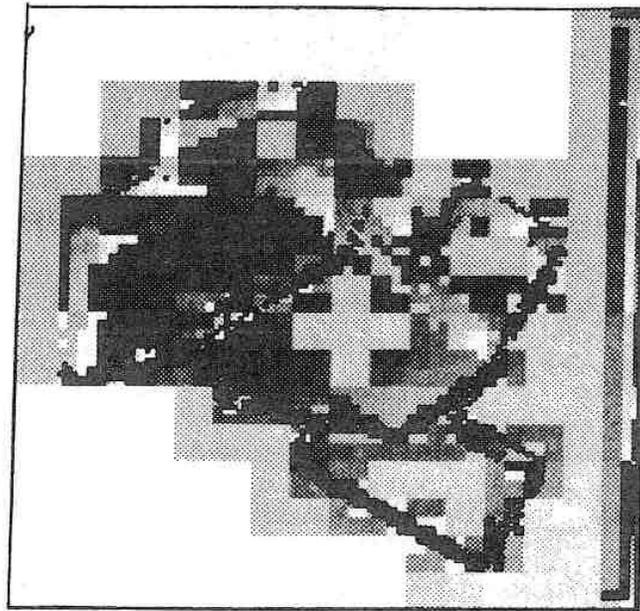


(a)

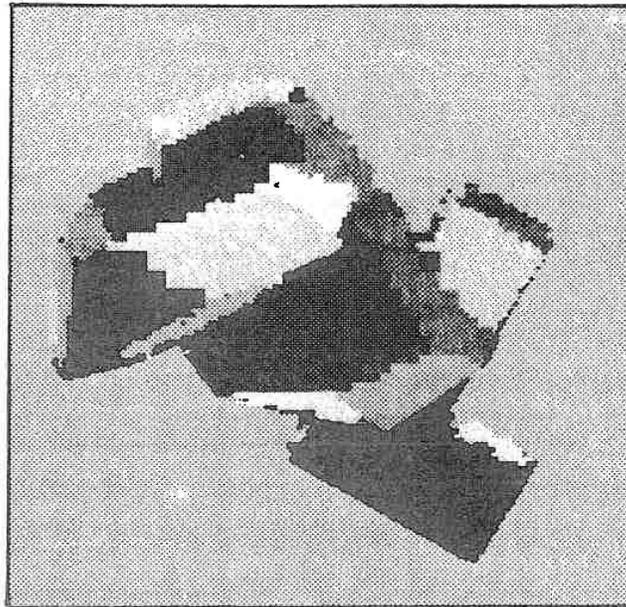


(b)

Figure 5.6. Résultat de segmentation en région : (a) image de distance originale
(b) une vue perspective



(c)



(d)

Figure 5.7. Résultat de segmentation en région : (c) partition des régions initiales
(d) des régions segmentées
le nombre de régions initiales est de 6022, le nombre d'arcs initiaux est de 13103
(sur l'image (c), les points noirs sont des points isolés)

5.3 Obtention de Modèle

Etant données plusieurs images de distance correspondant aux différentes vues de l'objet à modéliser, pour obtenir le modèle complet de l'objet, les procédures suivantes doivent être effectuées :

- construction d'un graphe d'adjacence pour chaque vue

Pour chaque image de distance, la segmentation est effectuée pour obtenir une partition en régions, la partition est représentée par un graphe dont chaque nœud représente une région, et les arcs représentent des relations d'adjacence entre des régions (voir paragraphe précédent).

- transformation des primitives de chaque vue dans un repère commun lié à l'objet, dans notre cas, tous les autres vues sont transformées dans le repère de la première vue.
- Ayant transformé les différentes vues dans un repère commun, l'intégration des différentes vues est effectuée par fusion des graphes correspondant aux différentes vues en un graphe complet. Les coordonnées des sommets du modèle sont obtenues par l'intersection des facettes voisines.

Le modèle obtenu est représenté par une liste des facettes qui le compose, et chaque facette est représentée par une liste des arêtes qui l'entourent, enfin, chaque arête est représentée par ses deux extrémités.

5.3.1 Acquisition de plusieurs vues

L'objet à modéliser est placé sur un plateau tournant. Les différentes vues sont acquises par une rotation dans le plan $X-Y$ (la surface) du plateau d'un angle connu. En connaissant l'origine de la rotation du plateau, la transformation entre la i -ème vue et la première vue est la suivante :

$$\vec{x} = \mathbf{R}^{-1}(\theta_i)(\vec{x}' - \vec{t}') + \vec{t}$$

\vec{x} est un point dans le repère de la première vue, et \vec{x}' est un point dans le repère de la i -ème vue. $\mathbf{R}(\theta_i)$ est la matrice de la rotation :

$$\mathbf{R} = \begin{pmatrix} \cos(\theta_i) & -\sin(\theta_i) & 0 \\ \sin(\theta_i) & \cos(\theta_i) & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

$\vec{t} = (t_x, t_y, 0)$ est le vecteur de translation, $(t_x, t_y, 0)$ sont les coordonnées de l'origine de la rotation dans le repère de la première vue.

5.3.2 Construction de graphe d'adjacence

Après la segmentation, un test de consistance spatiale est nécessaire :

dans le graphe obtenu par la segmentation, deux régions séparées par une discontinuité du type "marche" (par exemple, deux régions séparées par une facette verticale sont considérées comme des régions adjacentes, car les relations d'adjacence sont établies par la projection des régions sur le plan XOY . Il est facile de détecter ce type de fausse adjacence. En fait, il suffit de tester si l'écart moyen des distances entre des couples de points situés le long de la frontière entre deux régions dépasse un certain seuil ou non.

Après la segmentation de l'image de distance, étant donnés des points appartenant à une même région, on construit une image dont les points appartenant à une même région sont colorés par une même couleur, et les régions différentes sont colorées d'une couleur différente. Pour chaque région, à exclusion de la région du fond correspondant à la surface sur laquelle est placé l'objet, on trace les contours d'une région en utilisant un algorithme de suivi de contour [Pavlidis 82]. On utilise une notion de N -voisin ($N = 0, 7$) qui est illustrée sur la figure 5.8. Partant du point se trouvant au plus gauche de la région, le programme cherche à longer le contour de la région, le processus s'arrête lorsque l'on revient au point de départ. En connaissant l'équation du plan auquel la région appartient, les points de contours sont reprojétés sur le plan, On obtient ainsi une liste des points 3D de contours.

3	2	1
4	P	0
5	6	7

Figure 5.8. Notion de N -voisin d'un point P : $N = 0, 7$

(i_0, j_0) étant le point appartenant à la région, et se situant sur le plus gauche de la région :

Procédure suivi-de-contour :

point-courant = (i_0, j_0)

premier = vrai

direction de recherche D = 6

Liste-des-points-de-contour = vide

arrêt = faux

tant que non arrêt

si *premier* = faux et *point-courant* = (i_0, j_0)

arrêt = vrai

sinon

– *trouvé* = faux

– tant que non trouvé

si le $(D-1)$ -voisin appartenant à la région

ajouter $(D-1)$ -voisin dans *Liste-des-points-de-contour*

point-courant = $(D-1)$ -voisin

trouvé = vrai

$D = D - 2$ (modulo 8)

sinon

si le D -voisin appartenant à la région

ajouter D -voisin dans *Liste-des-points-de-contour*

point-courant = D -voisin

trouvé = vrai

sinon

si le $(D+1)$ -voisin appartenant à la région

ajouter $(D+1)$ -voisin dans *Liste-des-points-de-contour*

point-courant = $(D+1)$ -voisin

trouvé = vrai

sinon

$D = D + 2$ (modulo 8)

fsi
fsi
fsi
 — premier = faux

fsi

ftantque

Après la segmentation et le suivi de contours de chaque région, chaque vue est décrite par un graphe où chaque nœud correspond à une région visible de l'objet et où chaque arc représente une relation d'adjacence entre deux régions. Les attributs associés à un nœud sont les suivants :

- Les paramètres (\vec{n}_i, d_i) du plan :
 $\vec{n}_i \cdot \vec{x} + d_i = 0$ étant l'équation du plan.
- centre de gravité de la région : \vec{x}_c
- une liste des points du contour 3D délimitant la région.
- une liste des régions adjacentes.

5.3.3 Intégration des différentes vues — fusion des graphes

Ayant construit un graphe d'adjacence pour chaque vue, à cette étape, pour construire le modèle complet, on peut construire d'abord le modèle partiel correspondant à chaque vue, et fusionner ensuite les modèles partiels pour obtenir le modèle complet. On constate que, étant donnée une vue de l'objet, les arrêtes (du type "toit") peuvent être précisément déterminées en calculant l'intersection des couples de plans auxquels elles appartiennent, cependant, ce n'est pas le cas pour les arrêtes occultantes qui ne délimitent qu'une facette visible. Nous avons donc utilisé une autre méthode qui consiste à fusionner le graphe (graphe 1) correspondant à la première vue avec ceux correspondant aux autres vues. Chaque fois une fusion est effectuée, le graphe 1 est remplacé par la réunion de ces deux graphes.

Etant donné le graphe 1 correspondant à la première vue et le graphe i correspondant à la i -ème vue, l'angle de rotation entre ces deux vues étant θ_i , la fusion de ces deux vues fait appel à deux opérations principales :

- détection de sous-graphe commun, c-à-d, identifier les nœuds correspondent à la même facette visible de l'objet (appariement de deux nœuds) ;
- pour les nœuds du graphe i qui n'ont pas été appariés avec ceux du graphe 1 (les facettes correspondantes sont cachées dans la première vue), créer les nœuds correspondants dans le graphe 1 .

Appariement des nœuds de deux graphes

Supposons que $G_1(V_1, E_1)$ et $G_i(V_i, E_i)$ soient respectivement le graphe 1 et le graphe i correspondant aux deux vues de l'objet (la première et la i -ème), où V_1 et V_i désignent l'ensemble des nœuds, et E_1 et E_i désignent l'ensemble des arcs. soit $V_1 = \{v_{1k}\}$, $V_i = \{v_{ik}\}$, et (\vec{n}_{1k}, d_{1k}) , (\vec{n}_{ik}, d_{ik}) les paramètres des équations des plans correspondants exprimés respectivement dans le repère de la première vue et celui de la i -ème vue. Les conditions pour qu'un nœud du graphe 1 et un nœud du graphe i correspondent à une même facette visible de l'objet sont les suivantes :

- l'angle entre les normales des plans soit petit :

$$|\vec{n}_{1k} - \vec{n}_{ik} \mathbf{R}(\theta_i)| < \epsilon 1 \quad (5.1)$$

- l'écart des distances de l'origine aux plans soit petit:

$$|d_{1k} - d_{ik} - \vec{n}_{ik} \cdot \vec{t} + \vec{n}_{ik} \mathbf{R} \vec{t}| < \epsilon 2 \quad (5.2)$$

- la distance entre les centres de gravité des régions soit inférieure à un seuil de distance :

$$|\vec{x}_{c_{1k}} - (\mathbf{R}^{-1}(\vec{x}_{c_{ik}} - \vec{t}) + \vec{t})| < \text{seuil} \quad (5.3)$$

Remarquons que les deux premières conditions sont des conditions nécessaires pour que deux plans se superposent, la vérification de la troisième condition évite l'appariement de deux régions non superposées ayant une même équation.

En connaissant la transformation entre ces deux vues, le problème est de les identifier en réduisant l'espace de recherche. Si l'on dispose d'un couple de nœuds appariés, la recherche des appariements pour les autres nœuds peut être effectuée en n'appariant que des nœuds voisins des nœuds déjà appariés. Ceci peut être accompli par une exploration en largeur d'abord :

Procédure Exploration :

nœuds du graphe 1 : $\{v_{1j}\}$ ($j = 1, m_1$)

nœuds du graphe i : $\{v_{ik}\}$ ($k = 1, m_2$)

La procédure *test-identique*() teste si les deux nœuds correspondent à une même facette de l'objet en utilisant les conditions 5.1, 5.2, 5.3. La procédure retourne "vrai" si ces deux nœuds correspondent effectivement à une même facette de l'objet; sinon, elle retourne "faux".

Etant donné le couple d'appariements $v_{11}, v_{i1} : IDENTIQUE[v_{i1}] = v_{11}$

Noeud-courant = v_{i1}

N -identique = $IDENTIQUE[Noeud-courant]$

arrêt = faux

tant que non arrêt faire

pour tous les noeuds voisins N -voisin- i de Noeud-courant

si N -voisin- i n'a pas été marqué et non apparié

pour tous les noeuds voisins N -voisin-1 de N -identique

test-identique(N -voisin-1, N -voisin- i)

si identique = vrai

marquer deux noeuds appariés

$IDENTIQUE[N$ -voisin- $i]$ = N -voisin-1

ajouter N -voisin- i dans Liste-des-noeuds-appariés

sinon

marquer N -voisin- i

fsi

fpour

fsi

fpour

si Liste-des-noeuds-appariés non vide

Noeud-courant = un noeud de Liste-des-noeuds-appariés

Liste-des-noeuds-appariés = enlever ce noeud de la liste

sinon

arrêt = vrai

fsi

ftantque

La procédure complète pour la recherche de sous-graphe commun est la suivante :

Procédure recherche-de-sous-graphe-commun() :

cherche-premier-appariement : $v_{ik} \leftrightarrow v_{1j}$

IDENTIQUE[v_{ik}] = v_{1j}

marquer le noeud v_{ik}

exploration(v_{ik})

En connaissant la transformation entre les deux vues, la recherche du premier appariement est facile à effectuer en testant les conditions 5.1, 5.2, 5.3. La procédure de la recherche du sous-graphe commun est donc très efficace.

Correction de l'estimation de la transformation

En tenant compte de petits déplacements qui pourraient avoir lieu lors de la rotation du plateau, la transformation peut être corrigée par la méthode des moindres carrés.

Supposons que m paires des nœuds entre deux graphes sont appariés :

$$v_{1l}(\vec{n}_{1l}, d_{1l}) \leftrightarrow v_{il}(\vec{n}_{il}, d_{il}) \quad (l = 1, m)$$

$$\text{avec } \vec{n}_{1l} = (n_{1lx}, n_{1ly}, n_{1lz}), \text{ et } \vec{n}_{il} = (n_{ilx}, n_{ily}, n_{ilz})$$

Le critère pour minimiser l'erreur sur l'angle de la rotation étant :

$$\epsilon = \sum_{l=1}^m \|\vec{n}_{1l} - \vec{n}_{il} \mathbf{R}(\theta_i)\|$$

En dérivant ϵ par rapport à θ_i , et le mettant ensuite en zéro, on a :

$$\text{si} \quad \sum_{l=1}^m (n_{1lx} n_{ilx} + n_{1ly} n_{ily}) = 0$$

$$\text{alors} \quad \theta_i = \frac{\pi}{2}$$

$$\text{sinon} \quad \theta_i = \arctan\left(\frac{\sum_{l=1}^m (n_{1lx} n_{ily} - n_{ilx} n_{1ly})}{\sum_{l=1}^m (n_{1lx} n_{ilx} + n_{1ly} n_{ily})}\right)$$

Ayant réestimé l'angle de la rotation, le critère pour minimiser l'erreur sur les translations est la suivant :

$$E = \sum_{l=1}^m (d_{1l} - d_{il} - \vec{n}_{il}\vec{t} + \vec{n}_{il}\mathbf{R}\vec{t})^2$$

avec $\vec{t} = (t_x, t_y, 0)$.

En dérivant E respectivement par rapport à t_x , t_y , et les mettant à zéro, on obtient :

$$t_x = \frac{\sum_{l=1}^m b_l^2 \sum_{l=1}^m (d_{il} - d_{1l})a_l - \sum_{l=1}^m a_l b_l \sum_{l=1}^m (d_{il} - d_{1l})b_l}{\sum_{l=1}^m a_l^2 \sum_{l=1}^m b_l^2 - (\sum_{l=1}^m a_l b_l)^2}$$

$$t_y = \frac{\sum_{l=1}^m a_l^2 \sum_{l=1}^m (d_{il} - d_{1l})b_l - \sum_{l=1}^m a_l b_l \sum_{l=1}^m (d_{il} - d_{1l})a_l}{\sum_{l=1}^m a_l^2 \sum_{l=1}^m b_l^2 - (\sum_{l=1}^m a_l b_l)^2}$$

avec

$$a_l = n_{il_x}(\cos(\theta) - 1) + n_{il_y} \sin(\theta)$$

$$b_l = n_{il_y}(\cos(\theta) - 1) - n_{il_x} \sin(\theta)$$

Création d'un nœud correspondant dans le graphe 1

Etant donné un nœud non apparié du graphe i , la création de son homologue consiste à mettre à jour les attributs associés au nouveau nœud créé ainsi que les relations d'adjacence entre le nœud créé et les autres nœuds déjà existants dans le graphe 1.

Supposons que v_{il} soit un nœud du graphe i qui n'a pas été apparié avec celui du graphe 1, v_{1k} soit le nœud correspondant à créer :

1. Mise à jour des attributs associés à v_{1k}

Les attributs associés au nœud créé v_{1k} sont les transformations (corrigées) de ceux associés au nœud v_{il} .

2. Mise à jour des relations d'adjacence

Comme la figure 5.9 illustre, pour tous les nœuds appariés du graphe 1, si son homologue dans le graphe i est voisin de nœud v_{il} , alors ce nœud est voisin du nœud créé v_{1k} . Sinon, le nœud n'est pas voisin du nœud créé v_{1k} . Pour les nœuds du graphe 1 qui n'ont pas été appariés avec ceux du graphe i , un test de relation d'adjacence entre ces nœuds avec le nœud créé est effectué :

Etant donné v_{1j} le nœud non apparié du graphe 1,

- si les deux plans associés respectivement à v_{1k} et v_{1j} sont parallèles, alors ces deux nœuds ne sont pas voisins.
- si parmi un échantillon des points de contours associés à nœud v_{1k} il y a un certain nombre de points qui sont proches du plan associé à nœud v_{1j} et si parmi les points de contours associés à nœud v_{1j} il y a aussi un certain nombre de points qui sont proches du plan associés à nœud v_{1k} , alors ces deux nœuds sont adjacents (figure 5.10); sinon, ils ne sont pas adjacents.

La distance entre un point \vec{p} et un plan (\vec{n}_i, d_i) est la suivante :

$$d = |\vec{n}_i \cdot \vec{p} + d_i| \text{ avec } \|\vec{n}_i\| = 1$$

Le programme complet de la fusion des graphes est le suivant :

Pour toutes les vues i ($i = 2, n$) faire
recherche-de-la-sous-graphe-commun(1, i)
corriger-estimation-de-la-transformation(1, i)
pour tous les nœud non appariés du graphe i faire
creer-un-noeud-correspondant-dans-graphe-1
fpour
fpour

5.3.4 Obtention de modèle complet

Le processus de fusion des graphes s'arrête lorsque la formule d'Euler est vérifiée :

$$\text{le nombre de facettes} + \text{le nombre de sommets} - \text{le nombre d'arêtes} = 2$$

Nous prenons la convention suivante:

deux facettes ne sont adjacentes que si elles partagent une frontière entre elles.

Le nombre de facettes est égal au nombre des nœuds du graphe, le nombre d'arêtes est égal au nombre d'arcs du graphe, or le nombre de sommets n'est pas

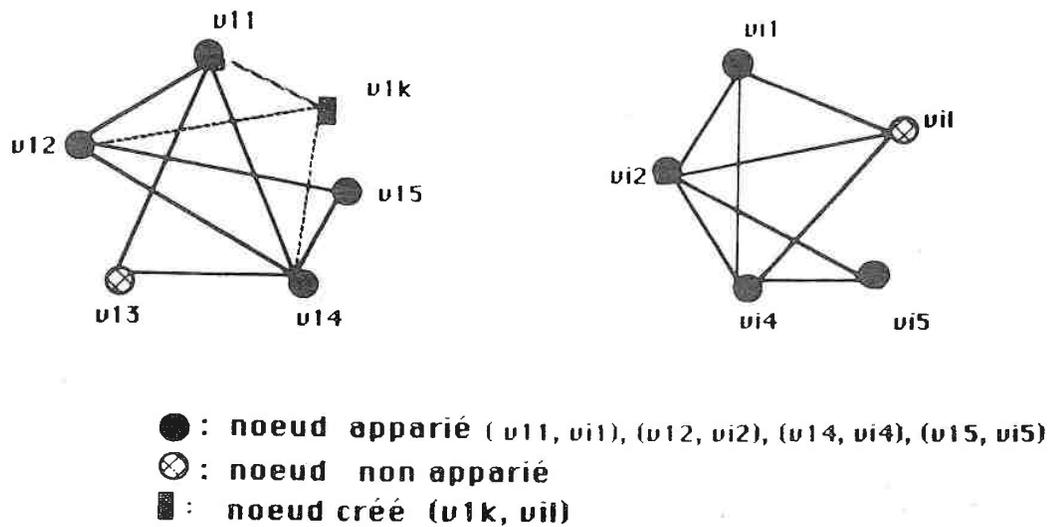
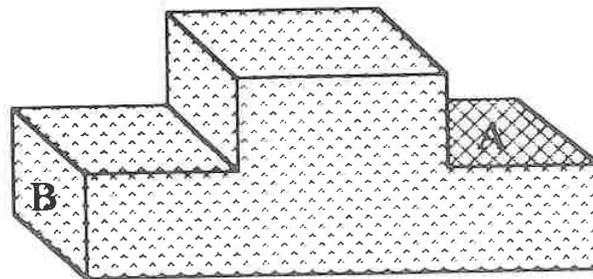


Figure 5.9. Détermination des relations d'adjacence entre le noeud créé et les noeuds déjà existants : les relations d'adjacence avec les noeuds appariés sont faciles à déterminer : mais un test d'adjacence plus compliqué est nécessaire pour le noeud créé v_{1k} et le noeud non apparié v_{13}



A : un noeud créé
 B : un noeud non apparié

Figure 5.10. test de la relation d'adjacence entre un noeud non apparié et un noeud créé : un certain nombre de points de contours de la région B sont proches du plan auquel la région A appartient, mais aucun point de contours de la région A n'est proche du plan auquel appartient la région B . les régions A , B ne sont donc pas voisins

aussi évident, car un sommet est un point d'intersection de trois plans mutuellement adjacents, mais les points d'intersection de trois plans mutuellement adjacents ne forment pas tous effectivement un sommet de l'objet (figure 5.11).

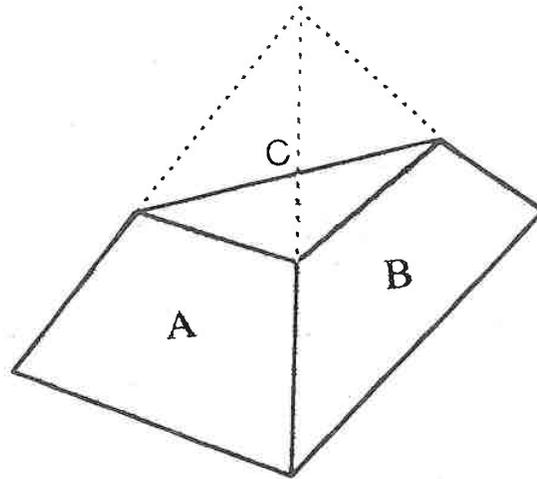


Figure 5.11. Le point d'intersection de trois plans mutuellement adjacents n'est pas nécessairement un sommet de l'objet : le point d'intersection des plans A, B, C

Pour déterminer le nombre de sommets, on cherche tous les cliques possibles d'un graphe en utilisant un algorithme de recherche de clique (voir chap. 6.2). si tous les cliques trouvés ne comportent que trois composants, alors le nombre de sommets est égale au nombre de cliques trouvées.

Ayant obtenu un graphe décrivant l'objet complet, en prenant la règle citée ci-dessous, les coordonnées des sommets sont déterminées en calculant le point d'intersection des facettes mutuellement adjacentes.

Règle :

Pour un objet trièdre, étant donné un graphe d'adjacence de l'objet, le point d'intersection de trois plans auxquels appartiennent trois nœuds mutuellement adjacents n'est effectivement un sommet de l'objet que s'il y a au plus un nœud qui est adjacent avec chacun de ces trois nœuds.

5.3.5 Résultats

Comme l'acquisition d'image se fait manuellement (le déplacement de plateau est manuellement contrôlé), Nous n'avons pris que deux vues de l'objet pour illustrer la méthode de modélisation. Le pas de translation du plateau est de 2 mm, la deuxième vue est prise en tournant le plateau d'un angle de 105 degrés. La dimension du plateau étant 15 cm × 15 cm, le centre de rotation est supposé au centre du plateau. Les données 3D acquises pour chaque vue sont ensuite représenté par une image de distance par un échantillonnage bidimensionnel et une interpolation bilinéaire (Annexe A). Les figures 5.12, 5.13 montrent les deux images de distance ainsi obtenues dont les résolutions sont $\Delta x = 1mm$, $\Delta y = 0.588mm$, $\Delta z = 0.157mm$. Les parties cachées (non vue par la caméra) sont codées en zéro (noir). L'observateur est supposé au dessus de l'objet. Sur l'image de distance, plus un point est loin de la surface du plateau, plus la valeur du niveau de gris sera grande (plus claire).

Les figures 5.14, 5.15 sont les images après la segmentation en régions. Les seuils utilisés sont les suivants :

pour la partition initiale :

$$\text{seuil}_1 = 15 \text{ pixels} \approx 2.355 \text{ mm}$$

$$\text{seuil}_2 \approx 5 \text{ degrés}$$

pour la fusion des régions :

$$\text{seuil}_3 = 3 \text{ mm}$$

$$\text{seuil}_4 \approx 18 \text{ degrés}$$

$\text{seuil}_5 = 0.72$ (correspond à une distance moyenne de 0.85 mm des points appartenant à la région au plan estimé)

Le nombre de régions initiales pour la première vue est de 4111 pour la première vue et de 5872 pour la deuxième vue.

Les images de deux vues après le processus de suivi de contour sont montrées en figure 5.16, 5.17. Les graphes d'adjacence correspondent respectivement la partition en région de deux vues sont illustrés en figure 5.18.

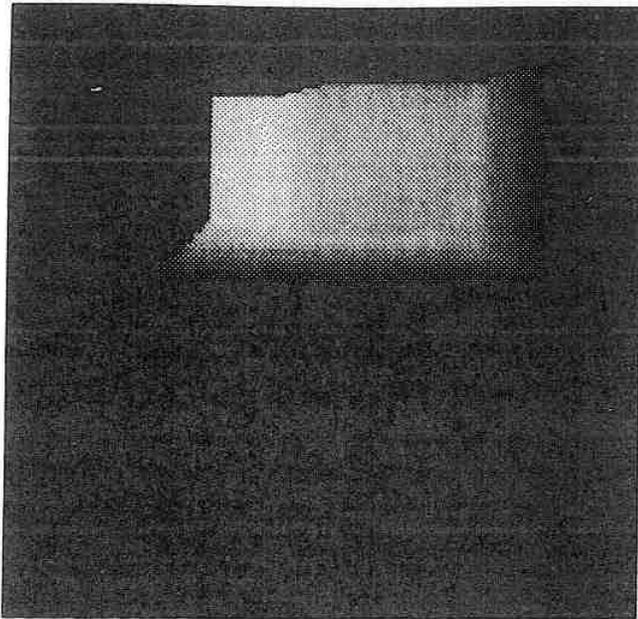


Figure 5.12. image de distance de la première vue de l'objet

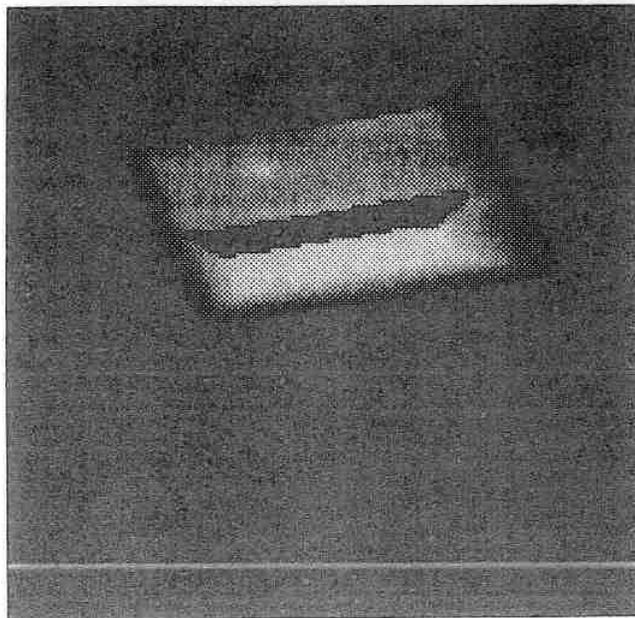


Figure 5.13. image de distance de la deuxième vue de l'objet

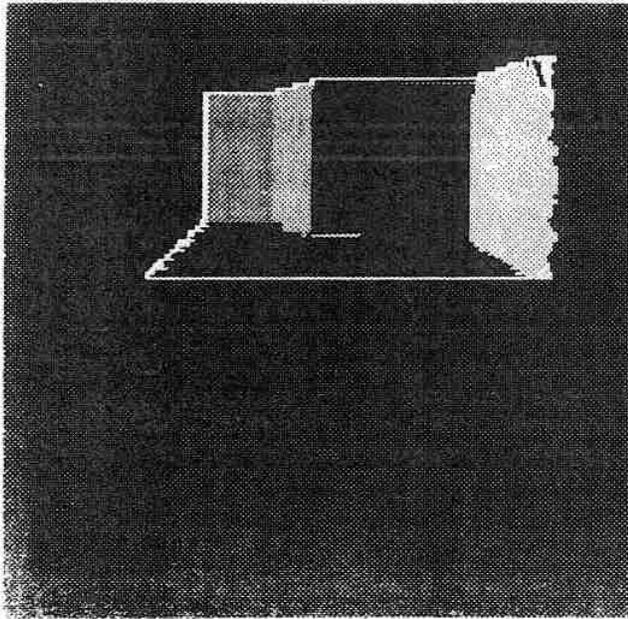


Figure 5.14. partition en régions de la première vue

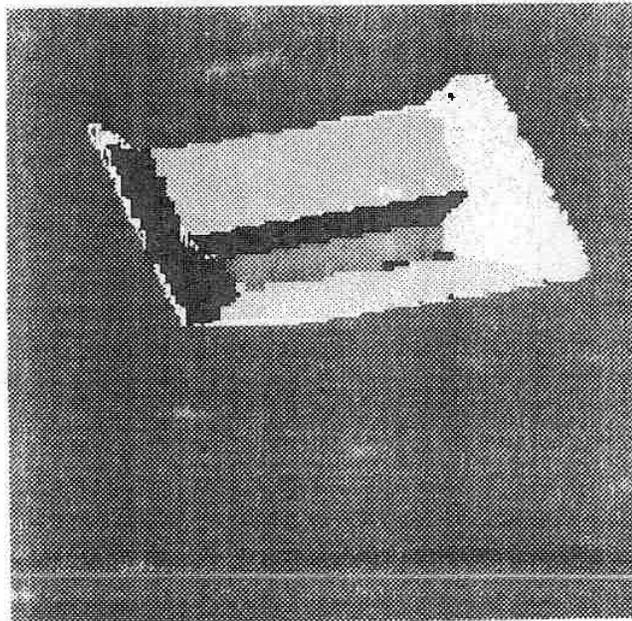


Figure 5.15. partition en régions de la deuxième vue

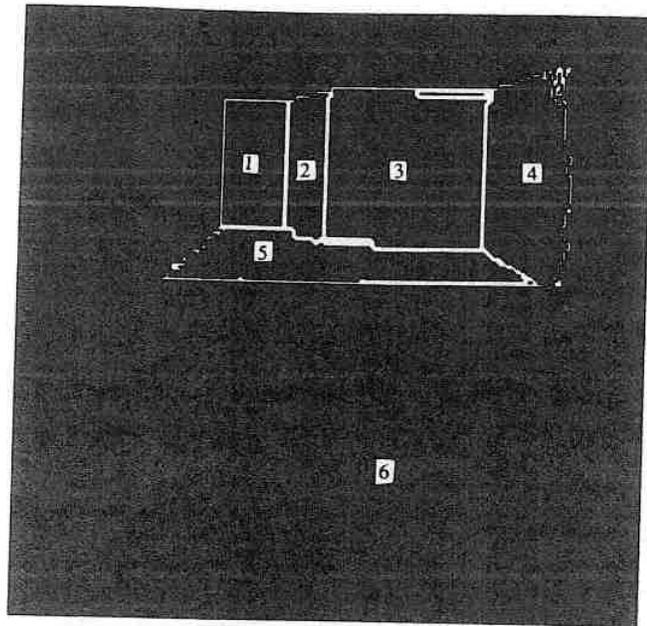


Figure 5.16. image de la première vue après suivi de contour

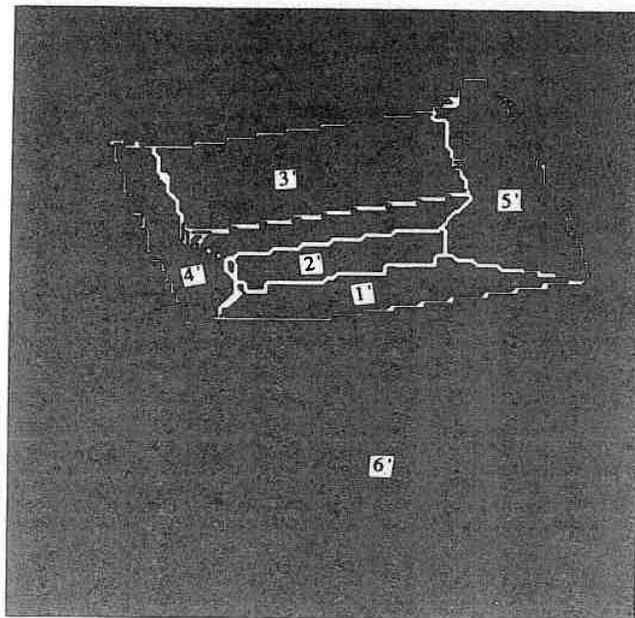


Figure 5.17. image de la deuxième vue après suivi de contour

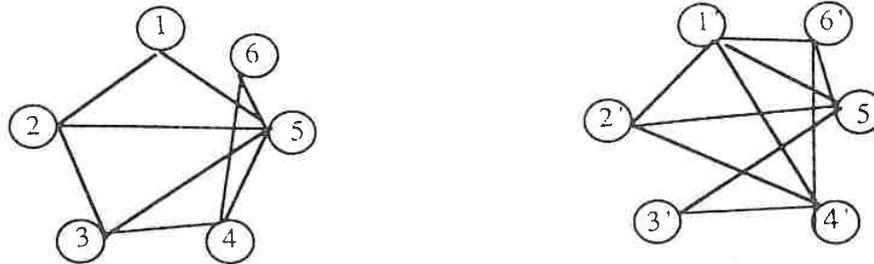


Figure 5.18. graphes d'adjacence correspondent à la partition en régions de deux vues

Ayant construit les graphes d'adjacence pour chaque vue, la fusion des graphes est effectuée. Les appariements trouvés sont :

$$n\text{œud } 1 \leftrightarrow n\text{œud } 2'$$

$$n\text{œud } 3 \leftrightarrow n\text{œud } 3'$$

$$n\text{œud } 5 \leftrightarrow n\text{œud } 5'$$

$$n\text{œud } 6 \leftrightarrow n\text{œud } 6'$$

Les nœuds (nœuds 7, 8) correspondent aux nœuds non appariés du *graphe 2* (nœuds 1', 4') sont créés dans le *graphe 1*. Lorsque le *graphe 1* est complet, les coordonnées des sommets du modèle sont estimées en calculant le point d'intersection des facettes qui le composent.

Les résultats sont montrés ci-dessous :

Les paramètres des équations des facettes :

```

facette 1: (0.001558 -0.028695 0.999587 -28.063982)
facette 2: (0.002152 0.656693 0.754155 -63.693249)
facette 3: (0.003143 -0.016131 0.999865 -19.408211)
facette 4: (0.001482 0.700942 0.713217 -92.031395)
facette 5: (0.709373 -0.002995 0.704827 -84.798553)
facette 6: (0.000243 -0.865100 0.501700 22.055000)
facette 7: (-0.866000 -0.001687 0.500000 17.457254)
facette 8: (0.000000 0.000000 1.000000 0.000000)

```

Les coordonnées des sommets calculées :

```

point d'intersection: (1, 2, 5) x=90.264214, y=62.552135, z=29.730562
point d'intersection: (1, 2, 7) x=37.250916, y=62.628456, z=29.815382
point d'intersection: (2, 3, 5) x=99.698730, y=73.373024, z=20.281178
point d'intersection: (2, 3, 7) x=31.848225, y=73.350838, z=20.494101
point d'intersection: (3, 4, 5) x=99.266663, y=109.850327, z=20.871027
point d'intersection: (3, 4, 7) x=32.116089, y=109.778702, z=21.080956
point d'intersection: (4, 5, 8) x=120.093414, y=131.042816, z=0.000000
point d'intersection: (5, 1, 6) x=90.754059, y=42.425880, z=29.152037
point d'intersection: (6, 1, 7) x=36.956188, y=42.459965, z=29.236868
point d'intersection: (8, 4, 7) x=19.902803, y=131.254669, z=0.000000
point d'intersection: (5, 6, 8) x=119.647926, y=25.527771, z=0.000000
point d'intersection: (6, 7, 8) x=20.108818, y=25.499809, z=0.000000

```

L'angle moyen entre les normales des facettes réelles du modèle et celles des facettes estimées est de 2.1 degrés (l'angle maximal est de 4.5 degrés. et l'angle minimal est de 0.52 degrés). la distance moyenne entre les sommets réels du modèle et les sommets calculés est de 2.25 mm (la distance minimale est de 0.35 mm. et la distance maximale est de 4.51 mm).

6

Reconnaissance et localisation des objets polyédriques

6.1 Introduction

Nous avons constaté que la caractéristique la plus attrayante de l'image de distance est la disponibilité explicite des informations sur les surfaces visibles des objets. Grâce à ces informations tridimensionnelles d'objets, le problème de la reconnaissance et de la localisation d'objets devient plus facile en utilisant l'image de distance qu'en utilisant l'image d'intensité.

Un ensemble de modèles étant constitué, le problème est d'utiliser les données tridimensionnelles pour opérer l'identification des formes en présence. Les problèmes à résoudre sont :

- quels sont les modèles compatibles avec des données
- pour chacun d'entre eux, déterminer la position dans l'univers observé.

Dans la littérature, de nombreuses méthodes existent pour résoudre ce problème [Besl 85], [Brady 88]. Ces méthodes sont fondées sur le même principe : elle consiste en l'extraction de l'image de distance de primitives invariantes par rapport à la position et l'orientation de l'observateur telles que l'angle entre les normales des surfaces [Faugeras 86], [Dhome 87], l'aire de morceau de surface [Oshima 83], les contours circulaires, ou des relations entre des surfaces adjacentes [Bolles 86] ; la reconnaissance des objets est donc accomplie par mise en correspondance des primitives extraites de l'image avec celles du modèle. Toutefois ces méthodes diffèrent par le choix de la représentation (des primitives) et la stratégie de contrôle pour

la mise en correspondance. Notons que la représentation et la technique de mise en correspondance sont étroitement liées.

6.1.1 Les différentes approches

Parmi les méthodes proposées dans la littérature, nous en présentons quelques unes sans être exhaustive :

- La méthode de reconnaissance des objets 3D proposée par Oshima et Shirai [Oshima 83] procède comme suit :

- Décrire la scène en termes des propriétés des régions ainsi que leurs relations entre elles.
 - * les points 3D sont d'abord groupés en petits éléments en supposant que chaque élément soit planaire. Les éléments sont ensuite fusionnés en des régions ;
 - * ces régions sont classifiées en régions planaires ou régions non planaires ;
 - * extension des régions non planaires par fusion des régions non planaires adjacentes, et approximation de ces régions par des surfaces quadriques.

les attributs associés à une région sont le type de celle-ci (planaire, cône, cylindre, ellipsoïde, etc), le nombre de régions adjacentes, son aire, périmètre, rayon de courbure moyen, minimum, et maximum.

les relations entre deux régions sont caractérisées par le type de l'intersection (concave, convexe), l'angle entre deux régions, et la position relative des barycentres de ces deux régions.

- Mise en correspondance

Le système choisit d'abord une région "noyau" qui est plus fiable et utile pour l'appariement. La priorité est donnée aux grandes régions planaires ayant plus de régions adjacentes. La région noyau ainsi que les régions adjacentes sont appariées avec celles des modèle. Les meilleurs appariements avec des modèles sont retenus et l'appariement continue pour les régions adjacentes de régions déjà appariées jusqu'à ce que une description compatible de la scène soit engendrée.

- Faugeras et al. [Faugeras 86] utilisent une méthode de recherche en arbre pour engendrer des interprétations possibles de la scène. Les primitives qu'ils utilisent sont des surfaces planes. Le programme de mise en correspondance traite des primitives par l'ordre décroissant de surface. Le programme choisit d'abord deux appariements. En connaissant deux appariements, la rotation associée à ces deux appariements peut être calculée. On peut alors prévoir efficacement les appariements compatibles avec ces deux premiers appariements en classant les normales des surfaces planes de la scène sur la sphère de Gauss :

Pour chaque surface plane (avec la normale \vec{n}_m) du modèle, la cellule de la sphère discrétisée correspondant à $\mathbf{R}\vec{n}_m$ (\mathbf{R} est la matrice de rotation) contient la liste des facettes candidates de la scène. Pour chaque appariement d'une facette du modèle avec une facette candidate de la scène, deux tests sont effectués :

soit $((\vec{n}_s, d_s), (\vec{n}_m, d_m))$ l'appariement à tester, et $((\vec{n}_{so}, d_{so}), (\vec{n}_{mo}, d_{mo}))$ l'appariement initial :

- la distance entre les barycentres de deux facettes de la scène doit être approximativement égale à la distance entre les barycentres de deux facettes correspondantes du modèle.
- $\vec{n}_s \cdot \vec{n}_{so} \approx \vec{n}_m \cdot \vec{n}_{mo}$

Les appariements satisfaisant ces deux tests sont considérés compatibles avec l'hypothèse initiale.

L'estimation de la transformation associée est améliorée au fur à mesure que les appariements sont trouvés.

- Bolles et al. présentent un système 3DPO¹ [Bolles 86] dont le but est d'identifier et localiser des pièces industrielles identiques dans une image de distance. Le système 3DPO utilise un modèle CAO pour représenter les objets. Le modèle utilisé lors du processus d'appariement a plusieurs niveaux : volume-surface-contour-sommet, comportant ainsi des informations très redondantes. Contrairement aux autres méthodes fondées sur les régions, ils choisissent les contours comme des primitives : contour droit, contour circulaire, contour tangentiel. Ces primitives sont trouvées en détectant les

1. Three-Dimensional Part Orientation

discontinuités de la distance et du gradient. L'algorithme cherche ensuite des surfaces adjacentes qui forment ces contours, l'estimation de la position de ces contours est améliorée en calculant l'intersection de ces surfaces adjacentes.

La génération des hypothèses sur les positions et les orientations des pièces s'effectue par mise en correspondance des contours extraits avec ceux du modèle. Les primitives les plus distinctives et les plus contraignantes (par exemple, le contour circulaire détermine cinq degrés de liberté parmi six) sont considérés comme des primitives clé ("focus feature"). La mise en correspondance se fait par appariement initial des primitives clé, et appariement séquentiel des autres primitives adjacentes avec chaque primitive clé et compatible avec le modèle. Les positions et les orientations hypothétiques des pièces sont ainsi déterminées.

Etant donnée une hypothèse sur la position et l'orientation de la pièce, une image de distance est prédite en connaissant le modèle de l'objet ainsi que la position et l'orientation du capteur. La vérification s'effectue en comparant l'image de distance prédite et celle observée.

- Dhome et Kasvand [Dhome 87] présentent une méthode fondée sur la transformée de Hough pour reconnaître et localiser les polyèdres. Leur méthode procède comme suit :

Les régions planaires sont extraites par la méthode de Hough en analysant un histogramme bidimensionnel sur l'orientation (θ, ϕ) de la normale de surface. La primitive choisie pour l'appariement est un couple de facettes adjacentes dont les attributs associés sont l'angle entre les normales de ces deux facettes adjacentes, un nombre booléen indiquant si l'intersection de ces deux facettes est concave ou non, les normales de deux facettes et les centres de gravité de ces deux facettes. Le couple des facettes de la scène et celui du modèle ayant le même type d'intersection et possédant un angle égal entre les normales des facettes sont appariés. Chaque appariement permet d'engendrer une hypothèse sur la transformation qui permet de superposer le couple de facettes du modèle avec celui de la scène. Pour éviter de travailler avec une matrice de six dimensions, ils déterminent trois paramètres équivalents qui caractérisent la position correspondante du modèle. Ces paramètres sont l'axe de vue du modèle, l'orientation du modèle autour de l'axe de vue, et la position du centre du modèle. Pour tous les appariements possibles,

l'algorithme cherche séquentiellement ces trois paramètres. Les hypothèses les plus probables (celles possédant les plus grands nombres de hypothèses compatibles) sont retenues. Parmi les hypothèses retenues, ils définissent un critère de qualité qui mesure la cohérence entre la scène et le modèle transformé. Pour ce faire, ils utilisent une méthode similaire à celle de Bolles [Bolles 86] : pour chaque hypothèse possible, une image de distance prédite est calculée en appliquant la transformation associée à cette hypothèse au modèle CAO. La meilleure solution est sélectionnée de façon à minimiser l'erreur moyenne globale entre l'image de distance originale et celle prédite.

- Fan et al. [Fan 88] présentent un système de reconnaissance de formes qui prend des données 3D denses comme données de départ et fournit automatiquement une description symbolique des objets dans la scène en termes de morceaux de surfaces visibles. Une telle description est représentée par un graphe dont les nœuds représentent les morceaux de surfaces, les arcs entre deux nœuds sont valués par des attributs représentant les relations entre ces deux nœuds. Les attributs associés à un nœud sont l'aire du morceau de surface reconstruite, l'orientation de la surface, la courbure principale moyenne, un poids estimé d'occultation et les coordonnées 3D du centre de cette région. Les attributs associés aux arcs sont le type de contour entre ces deux régions, un nombre mesurant la confiance d'appartenance de ces deux nœuds à un même objet. Dans leur système de reconnaissance, les modèles utilisés sont dépendant de la position et de l'orientation de l'observateur. C'est à dire que chaque modèle comporte plusieurs vues (de deux à six) de l'objet. Chaque vue dans le modèle est également représentée par un graphe. Le but du processus de mise en correspondance consiste à trouver, pour chaque objet de la scène, la vue la plus semblable du modèle. Le processus de mise en correspondance procède en trois étapes :

- Pour réduire l'espace de recherche, ils utilisent une méthode heuristique pour ordonner les vues des modèles pour chaque objet de la scène suivant la différence entre la description d'objet et celle des vues des modèles (nombre de nœuds, nombre d'arcs, etc.)
- Parmi une liste comportant un petit nombre de vues candidates de modèles issues de l'étape précédente, le processus cherche le plus grand sous-graphe dans la vue du modèle dont chaque nœud correspond à un nœud de graphe d'objet.

- La dernière étape consiste à détecter et à corriger les erreurs commises lors du processus de la segmentation du graphe de la scène en graphes disjoints correspondant aux objets individuels et les faux appariements résultant des arguments heuristiques invoqués lors du processus de mise en correspondance.

6.1.2 Reconnaissance avec peu d'informations

Dans le paragraphe précédent, nous avons exposé quelques méthodes principales de reconnaissance des objets 3D. Remarquons que ces méthodes peuvent traiter des objets de types très variés même avec des occultations partielles. Toutefois, deux processus indispensables en général très coûteux en temps de calcul sont impliqués dans ces méthodes :

- Acquisition

Ces méthodes utilisent des données 3D denses, ce qui exige que toute la scène soit balayée.

- Segmentation

Pour décrire la scène, il faut extraire des primitives de l'image de distance.

Il pourrait donc être intéressant que l'on puisse supprimer le processus de la segmentation en réduisant au maximum les données à acquérir. Dans certaines applications, a priori, on a une vague idée sur la position de l'objet recherché, dans ce cas, au lieu de balayer toute la scène, il suffit d'acquérir des données 3D éparées sur l'objet (quelques points ou quelques coupes selon le système d'acquisition utilisé) pour reconnaître l'objet et déterminer précisément la position et l'orientation de l'objet. Nous avons donc envisagé d'utiliser une telle approche qui consiste à reconnaître et à localiser l'objet en utilisant peu de données. Le même problème a été abordé par différents auteurs [Grimson 84], [Gaston 84], [Gordon 87], [Gordon 88], [LozanoPerez 87].

Grimson et al. [Grimson 84] proposent une méthode qui se repose sur les données éparées acquises en des points se situant sur l'objet polyédrique. Les normales des surfaces en ces points étant supposées connues, les interprétations possibles sont alors engendrées par appariement des points acquis avec les facettes des modèles en utilisant des contraintes locales telles que les angles entre des normales

des surfaces en points, les distances entre des points, etc. Ayant obtenu plusieurs interprétations des points acquis, la transformation est estimée pour chaque interprétation, puis un test de modèle est effectué : si les points transformés se situent effectivement sur les facettes du modèle délimitées par les arêtes, l'interprétation est considérée comme valide. sinon, l'interprétation est rejetée. Remarquons que la contrainte sur l'angle entre des normale des surfaces est très efficace lors du processus de mise en correspondance.

Gordon et al. [Gordon 87], [Gordon 88] travaillent aussi avec les objets polyédriques. Ils supposent que l'intersection du plan lumineux avec un objet forme au moins trois segments de droite. En général, trois segments de droite se situant sur trois facettes différentes d'un objet comportent des informations suffisantes pour localiser l'objet avec six degrés de liberté pourvue que les conditions suivantes soient satisfaites :

- aucun couple de segments de droite ne doit être parallèle.
- aucun couple de facettes sur lesquelles les segments de droite se situent ne soit parallèle.
- la normale commune de n'importe quel couple des segments de droite ne doit pas être en parallèle avec la ligne d'intersection des deux facettes sur lesquelles se situent ce couple des segments acquis.

Ils supposent que les appariements de ces trois segments de droite avec les facettes du modèle sont connus. En connaissant la géométrie de ces trois segments dans l'espace, la rotation qui fait passer le modèle dans l'espace est donc déterminée du fait que les vecteurs directionnels de ces trois segments doivent être respectivement perpendiculaires aux normales transformées des facettes correspondantes du modèle.

Silverman et al. [Silverman] présentent également une méthode pour localiser un polyèdre. Ils supposent que les coordonnées tridimensionnelles de trois points (dits points de contour) se situant sur trois arêtes d'un polyèdre sont connues, ce qui est équivalent à une autre représentation : trois segments de droite connexes de longueurs fixes tombent entre trois arêtes d'un polyèdre. Les extrémités d'un segment de droite de longueur fixe peuvent se déplacer dans une direction uniforme le long de deux arêtes dans l'espace. Chaque point de contour est localisé à l'intérieur d'un intervalle sur l'arête correspondante par une recherche en arbre

en utilisant la propriété monotone de ce déplacement. Le problème de localisation d'objet est finalement résolu par localisation de ces trois points de contour. Notons que le problème de l'appariement des points de contour avec les arêtes du modèle n'est pas résolu, les auteurs ont supposé que ces appariements étaient connus.

6.1.3 Notre approche

Nous avons donc décidé de mettre en œuvre une méthode de reconnaissance et de localisation d'objet avec peu de données [Qiang 88].

L'organigramme de notre méthode est illustré sur la figure 6.1. On suppose qu'un objet polyédrique est le seul à être placé sur un support plan (un plateau plus précisément) dans une position stable. L'objet a donc trois degrés de liberté par rapport au capteur. Les différentes positions stables de cet objet sur le support sont traitées comme les objets indépendants, c'est-à-dire, nous avons un modèle correspondant à chaque position stable de l'objet. Nous prenons d'abord deux prises de vue (deux coupes) de la scène qui sont effectuées l'une après l'autre par une translation du plateau avec une distance connue suivant l'axe X . Avec une scène qui ne contient qu'un polyèdre, l'image de chaque prise de vue ne contient que des segments de droite. Les prétraitements décrits au chapitre 4 sont effectués pour chaque coupe :

- extraire des points suréclairés par le plan lumineux du fond ;
- squelettisation ;
- effectuer l'approximation polygonale pour obtenir une liste des segments de droite.

Une image d'une coupe est donc transformée en une liste des segments de droite. Les différentes configurations des segments de droite sont possibles :

- segment oblique
- segment horizontal
- segments connexes

Remarquons que, comme nous l'avons mentionné au paragraphe 6.1.1, les méthodes proposées dans la littérature pour reconnaître et localiser les objets

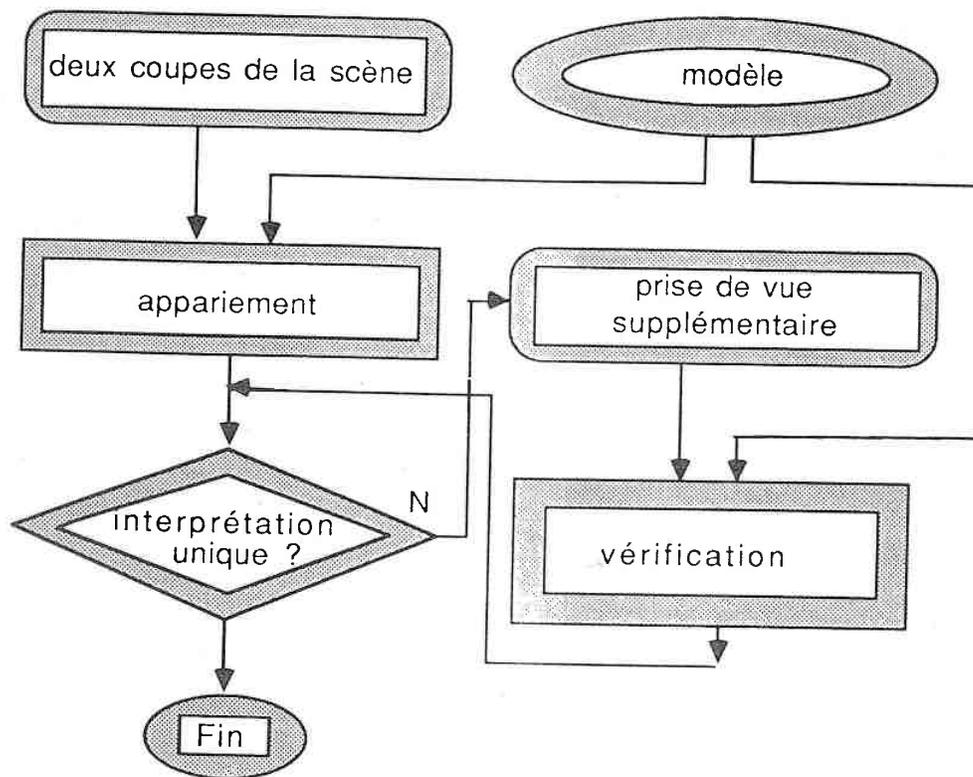


Figure 6.1. l'organigramme de la méthode de reconnaissance

à partir des données denses sont efficaces du fait que les primitives invariantes de la position et de l'orientation de l'observateur sont utilisées: le type de région (plane, cône, ellipsoïde, cylindre) [Oshima 83], l'angle des normales des régions [Faugeras 86], [Dhome 87], l'aire de la région, le type d'intersection (concave, convexe) entre deux régions adjacentes [Oshima 83], [Fan 88], et les relations d'adjacence entre des régions [Bolles 86]. or, avec des données éparses de la scène, les primitives invariantes par rapport à la position et l'orientation de l'observateur sont souvent absentes. Nous ne pouvons donc pas reconnaître et localiser l'objet avec ces primitives. Cela rend le problème de mise en correspondance plus difficile.

Ayant deux listes des segments de droite acquis, le programme de reconnaissance cherche donc les interprétations compatibles avec ces segments de droite acquis. Avec des données éparées de la scène, il y aurait souvent plusieurs interprétations possibles qui sont toutes compatibles avec des données actuellement acquises, une stratégie de prise de vue supplémentaire est prise pour identifier l'interprétation correcte. Les fausses hypothèses sont rejetées par vérification avec des données acquises supplémentaires.

Dans la suite de ce chapitre, nous discutons d'abord du processus de la mise en correspondance des données éparées avec le modèle. Nous présentons ensuite la stratégie de prise de vue supplémentaire, et nous donnons enfin les résultats expérimentaux.

6.2 Génération des hypothèses

6.2.1 Représentation de modèle

Nous rappelons brièvement la représentation du modèle d'un objet (voir chap. 5). Nous supposons qu'une base de données des modèles des objets 3D est disponible. Le modèle 3D d'un objet est représenté par une liste des facettes et de leur relations d'adjacence.

Une facette est décrite par des attributs suivants :

- le vecteur unitaire de la normale de la surface et la distance de l'origine au plan.
- une liste des arêtes ordonnées entourant la facette.
- la plus longue et la plus courte des distances entre des sommets de la facette.

La direction d'une arête est imposée de telle façon que la facette soit toujours à la gauche de cette arête.

6.2.2 Mise en correspondance

Dans une scène ne contenant que des polyèdres, nous observons par conséquent des segments de droite isolés ou des segments de droite connexes. Avec deux listes

des segments de droite acquis, le problème est de déterminer les appariements de deux extrémités de chaque segment de droite avec les arêtes du modèle. Etant donné un segment de droite acquis S_i et un modèle d'objet comportant P facettes et chaque facette a m_k arêtes. Le nombre des appariements possibles pour des extrémités du segment de droite avec les arêtes du modèle est de $\sum_{k=1}^P \frac{m_k!}{2(m_k-2)!}$. Le problème de recherche des correspondances est donc un problème combinatoire. Dans le cas où l'objet a six degrés de liberté, le problème pour exploiter efficacement l'espace de recherche est très difficile par manque des primitives pertinentes. Dans notre cas, nous supposons que l'objet n'a que trois degrés de liberté (déplacement dans un plan qui correspond par exemple au cas d'un objet en position d'équilibre sur un convoyeur), l'espace de recherche est largement réduit en tenant compte des configurations des segments de droite acquis (horizontale, oblique, ou connexes) et en exploitant les contraintes locales telles que les longueurs des segments de droite, la distance entre les segments de droite, et les critères sur les hauteurs des extrémités d'un segment de droite.

Appariement des extrémités des segments de droite avec les arêtes du modèle

Un segment de droite de longueur L ne peut joindre que deux arêtes du modèle qui vérifient que leur distance minimale d_{min} et maximale d_{max} satisfont la relation $d_{min} \leq L \leq d_{max}$ (figure 6.2). Cette contrainte peut être efficacement implantée par un tableau précalculé [Grimson 84]:

Par exemple, étant donné un modèle avec m arêtes, on construit un tableau de taille $m \times m$, la distance minimale entre le couple d'arêtes i, j est sauvegardée dans le triangle supérieur du tableau $\text{Tab}[\min\{i, j\}, \max\{i, j\}]$, et la distance maximale entre le couple d'arêtes est sauvegardée dans le triangle inférieur du tableau $\text{Tab}[\max\{i, j\}, \min\{i, j\}]$. Pour les couples d'arêtes qui n'appartiennent pas à une même facette, les distances minimales ainsi que les distances maximales sont mises à une valeur négative.

Pour tous les couples des arêtes appartenant à une même facette, on calcule la distance minimale et la distance maximale entre elles:

La distance minimale entre deux arêtes est déterminée comme le minimum parmi toutes les distances entre les extrémités de deux arêtes ainsi que toutes les distances perpendiculaires entre les extrémités de l'un de deux arêtes à l'autre

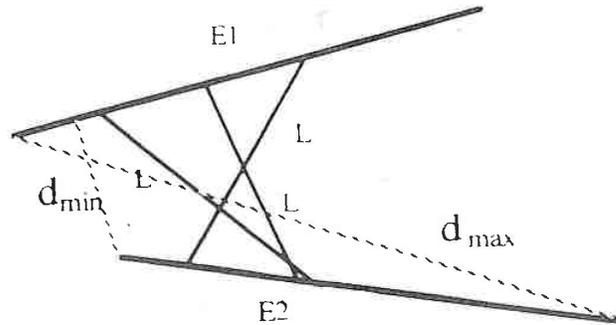


Figure 6.2. l'appariement des extrémités d'un segment de droite avec les arêtes du modèle

arête si la projection de l'extrémité est sur l'arête. si les deux arêtes se joignent à un sommet, la distance minimale est égale à zéro.

La distance maximale entre deux arêtes est déterminée comme le maximum parmi toutes les distances entre les combinaisons des couples d'extrémités de deux arêtes.

Un segment de droite de longueur L ne peut joindre que deux arêtes i, j du modèle si la condition suivante est satisfaite :

$$Tab[\min\{i, j\}, \max\{i, j\}] < L < Tab[\max\{i, j\}, \min\{i, j\}]$$

avec $i \neq j$.

Localisation des extrémités des segments de droite sur les arêtes connues du modèle

Connaissant les appariements des segments de droite acquis avec des arêtes du modèle, le problème de localiser l'objet n'est toujours pas résolu. Le processus de mise en correspondance cherche donc à localiser les extrémités de chaque segment

de droite acquis sur une arête connu. Nous considérons séparément les différentes configurations des segments de droite acquis :

1. cas d'un segment de droite oblique

Etant donné un segment de droite oblique AB avec longueur l et deux arêtes $e1$, $e2$ du modèle entre lesquelles les extrémités du segment AB devraient se situer, nous désignons le segment de droite correspondant de AB dans le modèle par $A'B'$, le segment recherché $A'B'$ doit satisfaire les conditions suivantes :

$$(|\text{Longueur}(A'B') - l| < \varepsilon) \cap (Z(A') = Z(A)) \cap (Z(B') = Z(B))$$

où ε est un seuil d'erreur et $Z(p)$ désigne la valeur z d'un point p .

Les différentes configurations sont illustrées dans la figure 6.3. Notons que dans le cas où les deux arêtes du modèle sont parallèles, seules la rotation correspondante (au maximum, il y a deux rotations possibles), la direction ainsi que la marge de translation correspondante sont définies.

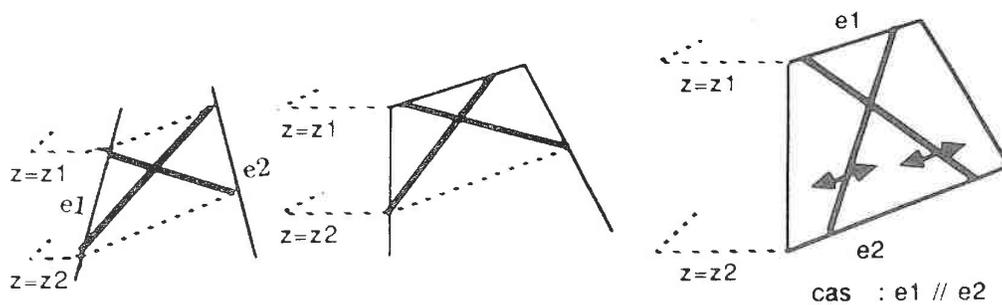


Figure 6.3. les différentes configurations en cas de segment oblique

2. cas d'un segment horizontal

Considérant un segment de droite horizontal acquis avec longueur l_1 dont les extrémités se situent sur deux arêtes $e1$, $e2$ du modèle. Avec une seule

contrainte sur la longueur du segment, il y aurait un grand nombre de segments avec la même longueur qui tombent entre ces deux arêtes. Cependant peu parmi eux sont compatibles avec les autres segments de droite observés. Pour chercher efficacement ces segments compatibles, nous considérons un second segment de droite horizontal acquis dans une autre liste des segments en utilisant la contrainte sur la distance entre ces deux segments pour réduire le nombre des hypothèses possibles pour ces deux segments.

Il nous faut alors distinguer deux cas :

- Le couple d'arêtes du modèle n'est pas parallèle

Nous associons un repère 2D à ce couple des arêtes non parallèle avec les conventions suivantes (figure 6.4) :

- l'origine de ce repère est au sommet formé par ces deux arêtes (ou leurs extensions) :
- l'axe X coïncide avec une arête, et position sur la même direction que cette arête (la direction de l'arête est orientée de telle sorte que la facette se trouve toujours à gauche de cette arête).

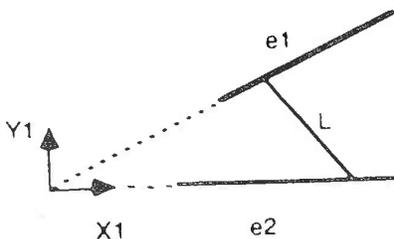


Figure 6.4. le repère associé à un couple d'arêtes non parallèles

Nous pouvons obtenir une représentation paramétrique pour les segments de droite recherchés (voir annexe B) :

Les extrémités des segments de droite recherchés sont les points d'intersection de chaque arête avec un cercle ayant un rayon fixe qui passe par l'origine du repère associé, et le lieu du centre de ce cercle est aussi

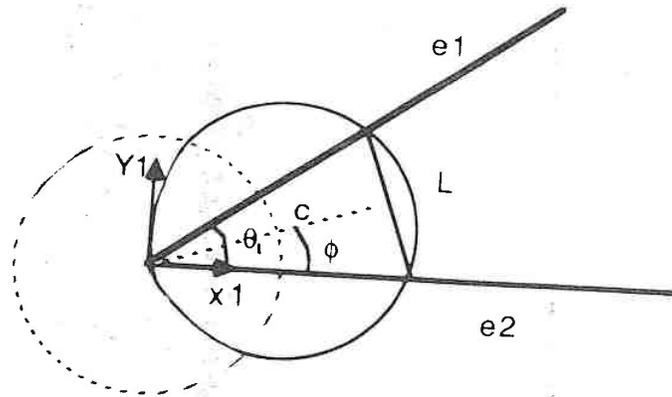


Figure 6.5. segment recherché avec longueur L déterminé par le cercle

un cercle de même rayon qui est centré sur l'origine du repère associé (figure 6.5).

Les segments de droite recherchés peuvent donc être représentés par un paramètre ϕ qui est l'angle polaire du centre de cercle.

$$y_1 = \frac{1}{\tan(\phi - \theta_1)} \left(x_1 - \frac{l_1 \cos(\phi)}{\sin(\theta_1)} \right) \quad (6.1)$$

avec $\theta_1 - \frac{\pi}{2} < \phi < \frac{\pi}{2}$

La même représentation peut être obtenue pour le second segment horizontal CD avec longueur l_2 et deux arêtes connues e_3, e_4 du modèle :

$$y_2 = \frac{1}{\tan(\phi' - \theta_2)} \left(x_2 - \frac{l_2 \cos(\phi')}{\sin(\theta_2)} \right) \quad (6.2)$$

Avec $\theta_2 - \frac{\pi}{2} < \phi' < \frac{\pi}{2}$

(les équations (6.1), (6.2) sont respectivement exprimées dans le repère associé au couple des arêtes e_1, e_2 et celui associé au couple des arêtes e_3, e_4).

où θ_1 et θ_2 sont les angles formés respectivement par les couples des arêtes (e_1, e_2) et (e_3, e_4) .

En exprimant l'équation 6.1 dans le repère $x_2o_2y_2$, la pente de la droite $A'B'$ devient $k = 1/\tan(\phi - \theta_1 + \theta_{r21})$. le parallélisme entre les deux segments horizontaux AB et CD impose le parallélisme entre les segments correspondants $A'B'$ et $C'D'$, nous avons donc (figure 6.6) :

$$\phi' = k\pi + \phi - \theta_1 + \theta_2 + \theta_{r21} \quad (6.3)$$

avec $k = 0, 1$

où θ_{r21} est l'angle connu de la rotation du repère associé au couple des arêtes $e1, e2$ à celui associé au couple des arêtes $e3, e4$.

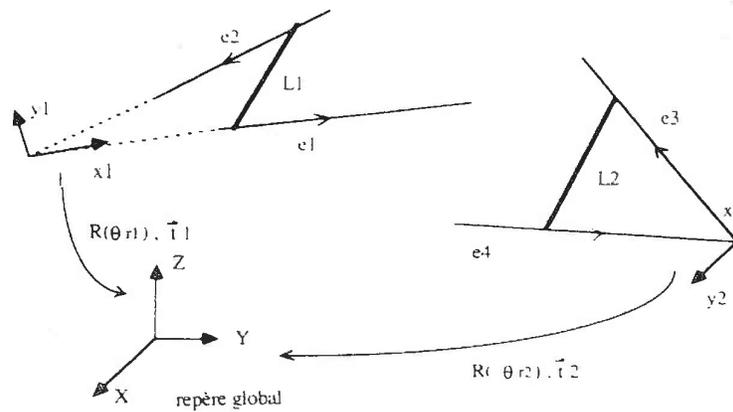


Figure 6.6. les transformations entre les deux repères associés à deux couples des arêtes

En projetant les segments correspondants $A'B'$ et $C'D'$ sur un même plan horizontal, la distance entre ces deux segments projetés est connue, qui est juste la distance de translation d suivant l'axe X prise pour acquérir deux coupes de la scène :

$$d = D(A'B', C'D') = \mathcal{F}(\phi, \phi') \quad (6.4)$$

En combinant les équations (6.3), (6.4), nous avons

$$\mathcal{G}(\phi) = d \quad (6.5)$$

où \mathcal{G} est une fonction trigonométrique en fonction de ϕ qui représente la distance entre les segments projetés des $A'B'$ et $C'D'$.

Remarquons qu'il est difficile de trouver une solution analytique pour l'équation (6.5), nous résolvons donc cette équation par une recherche dichotomique qui suit la démarche suivante :

Etant donné un intervalle initial d'angle ($\phi_1 = \theta_1 - \pi/2$, $\phi_n = \pi/2$) :

recherche-angle(ϕ_1, ϕ_n) :

$$\phi_m = (\phi_1 + \phi_n)/2$$

calcul $\mathcal{G}(\phi_1)$, $\mathcal{G}(\phi_m)$, $\mathcal{G}(\phi_n)$

si $|\mathcal{G}(\phi_m) - d| \leq \text{seuil1}$

enregistrer le résultat ϕ_m .

sinon

si $d \in [\min\{\mathcal{G}(\phi_1), \mathcal{G}(\phi_m)\}, \max\{\mathcal{G}(\phi_1), \mathcal{G}(\phi_m)\}]$

recherche-angle(ϕ_1, ϕ_m)

fsi

si $d \in [\min\{\mathcal{G}(\phi_m), \mathcal{G}(\phi_n)\}, \max\{\mathcal{G}(\phi_m), \mathcal{G}(\phi_n)\}]$

recherche-angle(ϕ_m, ϕ_n)

fsi

si $d \notin [\min\{\mathcal{G}(\phi_1), \mathcal{G}(\phi_m)\}, \max\{\mathcal{G}(\phi_1), \mathcal{G}(\phi_m)\}]$

et $d \notin [\min\{\mathcal{G}(\phi_m), \mathcal{G}(\phi_n)\}, \max\{\mathcal{G}(\phi_m), \mathcal{G}(\phi_n)\}]$

calcul la tangente à ϕ_m

$d1 = \text{distances-de-}\phi_1\text{-à-la-tangente}$

$dn = \text{distances-de-}\phi_n\text{-à-la-tangente}$

$dmax = \max\{d1, dn\}$

si $dmax \geq \text{seuil2}$

recherche-angle(ϕ_1, ϕ_m)

recherche-angle(ϕ_m, ϕ_n)

fsi

fsi

fsi

Les segments recherchés sont trouvés lorsque les angles ϕ et ϕ' sont connus.

Notons que plusieurs solutions pourraient être trouvées, un test supplémentaire est nécessaire pour tester si les extrémités du segment trouvé

se situent effectivement sur les arêtes correspondantes, et non pas sur leurs extensions (figure 6.7).

Notons aussi que les appariements incompatibles des $AB \leftrightarrow A'B'$ et $CD \leftrightarrow C'D'$ peuvent être rejetés rapidement par cette procédure.

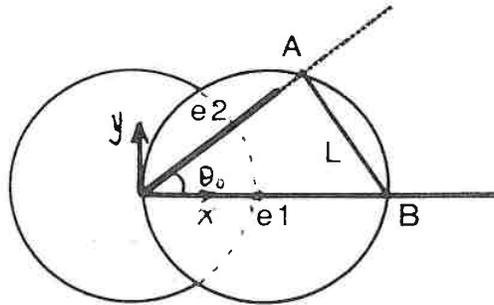


Figure 6.7. test de la validité de la solution ϕ

- au moins un couple des arêtes consiste en des arêtes parallèles
 Dans ce cas le problème peut être plus facilement résolu car les rotations possibles peuvent être aisément déterminées en recherchant le segment correspondant qui tombe sur ce couple des arêtes parallèles. Notons que si les deux couples d'arêtes sont parallèles, la transformation associée au appariement n'est définie que partiellement comme l'illustre la figure 6.8.

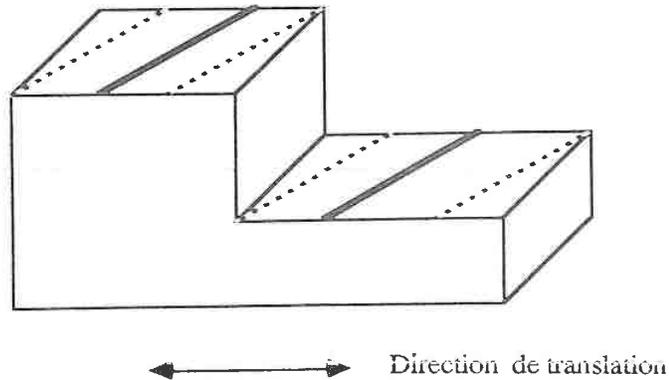


Figure 6.8. cas de deux couples d'arêtes parallèles: la transformation n'est définie que partiellement

3. Cas de segments connexes

Il est également facile de trouver des segments correspondants dans le modèle au cas de segments connexes en utilisant les contraintes suivantes [Qiang 88] (fig. 6.9) :

Soit AB , CD les deux segments connexes de la scène, et $A'B'$, $C'D'$ les deux segments correspondants dans le modèle :

- les deux segments de droite $A'B'$ et $C'D'$ doivent se situer sur deux facettes adjacentes du modèle et ils se rejoignent à un point commun :

- conservation des longueurs :

$$\text{Longueur}(A'B') \approx \text{Longueur}(AB)$$

$$\text{Longueur}(C'D') \approx \text{Longueur}(CD)$$

- conservation de l'angle :

l'angle formé par les deux segments $A'B'$, $C'D'$ \approx l'angle formé par les deux segments AB , CD

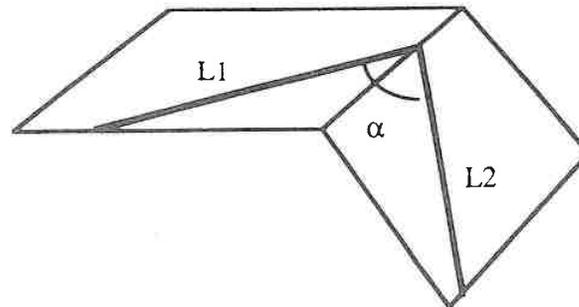


Figure 6.9. cas de segments connexes

Ayant trouvé tous les appariements possibles pour les segments de droite acquis, on transforme cet ensemble des appariements possibles en un graphe (figure 6.10). Chaque nœud de ce graphe représente une hypothèse pour l'appariement d'un segment de droite acquis (ou un couple des segments de droite dans le cas de segment horizontal et dans le cas de segments connexes) avec un segment de droite (ou un couple de segments de droite) se situant sur la (ou les) facette(s) du

modèle, et chaque arc entre deux nœuds indique que les deux hypothèses correspondantes sont mutuellement compatibles. Deux hypothèses ne sont considérées mutuellement compatibles que si les critères suivants sont satisfaits :

Supposons que l'on a deux hypothèses : $S_i \longleftrightarrow M_{il}$, $S_j \longleftrightarrow M_{jk}$

où S représente les segments de droite dans la scène, et M représente les segments de droite correspondants se situant sur des facettes du modèle.

- la distance entre deux segments de droite dans le modèle doit être approximativement égale à la distance entre deux segments de droite acquis.
- en projetant les segments M_{il} et M_{jk} sur le plan horizontal $X-Y$, les segments projetés doivent être approximativement parallèles.

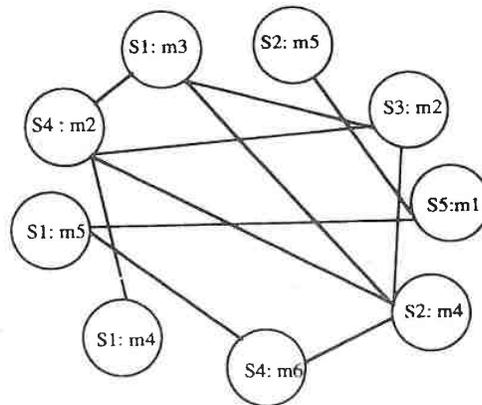


Figure 6.10. graphe représentant les appariements mutuellement compatibles

Ayant un graphe des appariements possibles, le problème est de localiser les plus larges groupes d'appariements mutuellement compatibles. Chaque groupe correspondant à un sous-graphe dont les nœuds sont complètement connexes représente donc une interprétation possible. Nous utilisons un algorithme de recherche de clique maximale [Bolles 82] pour résoudre ce problème.

Les notations utilisées dans l'algorithme :

- un ensemble P est représenté par un vecteur dont le nombre de composants est égal au nombre de nœuds du graphe, si un nœud i existe dans P , alors $P[i] = 1$, sinon, $P[i] = 0$;
- $C = Adj(x)$ — un ensemble des nœuds qui sont adjacents avec le nœud x : $C[i] = 1$ si nœud i est adjacent avec le nœud x , sinon, $C[i] = 0$;
- $R = complément(Q : un ensemble)$ — un ensemble de nœuds dont $R[v] = 1$ si $\{v | v \notin Q\}$, sinon, $R[v] = 0$;
- $S = intersection(P : un ensemble, Q : un ensemble)$ — un ensemble de nœuds dont $S[v] = 1$ si $\{v | v \in (P \cap Q)\}$, sinon, $S[v] = 0$;

L'algorithme est décrit suivant : (Q : un ensemble)

Maximal-clique(*Clique*, P , PS) :

Clique : une liste des nœuds P , PS : un ensemble

Si PS est vide

sauvegarder *Clique*

Sinon

choisir un nœud x qui a le moins de voisins dans PS

$Q = complément(Adj(x))$

pour chaque nœud y dans $intersection(P, Q)$

$P = P - y$

$Clique = Clique \cup y$

Maximal-clique(*Clique*, $intersection(P, Adj(y))$, $intersection(PS, Adj(y))$)

fpour

fsi

Pour détecter tous les cliques, initialement, les trois paramètres sont :

$Clique = vide$, $P = \{\text{tous les nœuds du graphe}\}$, $PS = \{\text{tous les nœuds du graphe}\}$

A partir de ces appariements mutuellement compatibles, on peut donc estimer la transformation.

6.2.3 Estimation des transformations

Pour chaque interprétation, nous estimons la transformation qui fait passer les coordonnées des points du modèle à celle des points de la scène. Supposons qu'un vecteur \vec{V}_m dans le repère du modèle soit transformé dans le repère de la scène par la transformation suivante :

$$\vec{V}_s = \mathbf{R}\vec{V}_m + \vec{t}$$

où \mathbf{R} est une matrice de rotation :

$$\mathbf{R} = \begin{pmatrix} \cos(\theta) & -\sin(\theta) & 0 \\ \sin(\theta) & \cos(\theta) & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

et \vec{t} est un vecteur de translation :

$$\vec{t} = (t_x, t_y, 0)$$

Pour chaque interprétation, nous allons donc déterminer ces trois paramètres.

• Estimation de la rotation

Etant donné un couple des appariements : $S_i \leftrightarrow M_k$

Supposons que \vec{s}_{s_i} et \vec{s}_{m_k} sont respectivement les vecteurs directionnels unitaires de deux segments de droite S_i et M_k , la direction de la rotation étant $\vec{r} = (0, 0, 1)$ (l'axe de Z), nous allons déterminer l'angle θ de la rotation autour de \vec{r} .

Il est facile de vérifier que l'angle de la rotation θ autour de \vec{r} satisfait les conditions suivantes :

$$\cos(\theta) = 1 - \frac{1 - (\vec{s}_{s_i} \cdot \vec{s}_{m_k})}{1 - (\vec{r} \cdot \vec{s}_{s_i})(\vec{r} \cdot \vec{s}_{m_k})}$$

$$\sin(\theta) = \frac{(\vec{r} \wedge \vec{s}_{s_i}) \cdot \vec{s}_{m_k}}{1 - (\vec{r} \cdot \vec{s}_{s_i})(\vec{r} \cdot \vec{s}_{m_k})}$$

l'angle de θ est donc :

$$\theta = \arcsin\left(\frac{(\vec{r} \wedge \vec{s}_{s_i}) \cdot \vec{s}_{m_k}}{1 - (\vec{r} \cdot \vec{s}_{s_i})(\vec{r} \cdot \vec{s}_{m_k})}\right)$$

Notons que, si le couple de segments appariés sont obliques, les directions de ces segments sont orientées vers le haut, il y aura donc un seul angle de rotation possible, tandis qu'au cas où deux segments appariés sont horizontaux, il y aura deux angles de rotation possible: θ et $\theta + \pi$.

- **Estimation de la translation**

Après avoir déterminé l'angle de la rotation, l'estimation de la translation est très simple :

Supposons que P_s et P_m soient respectivement les centres des segments appariés S_i , M_k :

$$\vec{t} = \vec{P}_s - \mathbf{R}\vec{P}_m$$

De fait, un couple d'appariements contenant un couple de directions et un couple de points est suffisant pour estimer respectivement la rotation et la translation. En tenant compte des erreurs attachées aux données acquises, nous estimons l'angle de rotation associée à une interprétation par la valeur moyenne de tous les angles de rotations correspondant à chaque appariement de cette interprétation ; et la translation est estimée comme le centre de gravité de toutes les translations associées aux tous les appariements de cette interprétation. La variance de transformation associée à chaque interprétation est aussi estimée. Les interprétations dont la variance de transformation est supérieure à un seuil sont rejetées. Pour les interprétations dont la transformation associée n'est définie que partiellement, l'angle de rotation ainsi que la marge de transformation sont estimées.

6.3 Vérification des hypothèses

Étant données plusieurs interprétations compatibles avec des données acquises, l'étape suivante est de déterminer la prise de vue supplémentaire pour identifier efficacement des interprétations. La vérification se fait par la comparaison de données prédites par chaque hypothèse et celles issues de la prise de vue supplémentaire. Une interprétation ne sera considérée comme valide que si les

données prédites par cette hypothèse sont cohérentes avec des nouvelles données acquises.

6.3.1 Stratégie de prise des vues supplémentaires

Après le processus de mise en correspondance, nous disposons donc d'une liste des interprétations compatibles avec des données éparses acquises. Par interprétation, cela signifie d'une part les appariements des segments de droite acquis avec ceux se situant sur les facettes du modèle et la transformation qui fait superposer les segments de droite correspondants dans le modèle avec ceux acquis. Pour déterminer l'interprétation correcte, une prise de vue supplémentaire est donc nécessaire. Une façon simple de le faire, c'est de prendre aléatoirement des mesures sur l'objet, et de comparer les nouveaux points mesurés avec les interprétations, jusqu'à ce qu'une seule interprétation reste valide. Mais Grimson [Grimson 86] a révélé qu'une telle stratégie de prise de vue supplémentaire avait souvent une convergence lente. Il faut donc prendre une stratégie de prise de vue fondée sur les caractéristiques des interprétations. Notons qu'une stratégie de prise de vue supplémentaire sera meilleure si :

- elle permet d'identifier rapidement l'interprétation correcte ;
- elle n'est pas sensible aux erreurs associées aux interprétations.

Etant donnée une liste des interprétations compatibles avec des données éparses acquises, Grimson propose une méthode pour déterminer la position et l'orientation optimales de nouvelles mesures. Nous illustrons sa méthode dans le cas bidimensionnel :

Supposons qu'on dispose d'une méthode pour prendre la mesure sur un point de l'objet : la normale de surface en ce point et la position de ce point sont mesurées. Supposons qu'on ait deux interprétations (ou plus) compatibles avec des données acquises, étant donnée une direction sur laquelle qu'on va prendre une nouvelle mesure, on définit une ligne qui est perpendiculaire à la direction de la prise de mesure comme ligne d'acquisition (scanning line) (dans le cas 3D, cela devrait être un plan d'acquisition). Pour chaque interprétation, et pour chaque facette du modèle, si la facette est visible, on projete les contours de cette facette sur la ligne d'acquisition. Dans le cas 2D, cette projection est un segment sur la ligne d'acquisition (figure 6.11). Chaque segment de la ligne d'acquisition est étiqueté

par la normale de la facette correspondante et la marge de distance entre la facette et la ligne d'acquisition. Lorsque toutes les facettes visibles sont projetées sur la ligne d'acquisition, on obtient une partition sur la ligne d'acquisition. On effectue cette opération pour toutes les interprétations, on obtient ainsi une partition différente de la ligne d'acquisition. La nouvelle partition est obtenue par intersection de l'ensemble de toutes les partitions. Dans la nouvelle partition, l'étiquette associée à chaque segment est l'union des étiquettes des segments correspondants des partitions individuelles. La nouvelle mesure de point sera donc prise à la position où la direction de prise de mesure intersecte le segment portant plusieurs étiquettes de la nouvelle partition. Si la normale de surface en ce nouveau point mesuré ne correspond pas à une étiquette associée à une interprétation, cette interprétation sera rejetée.

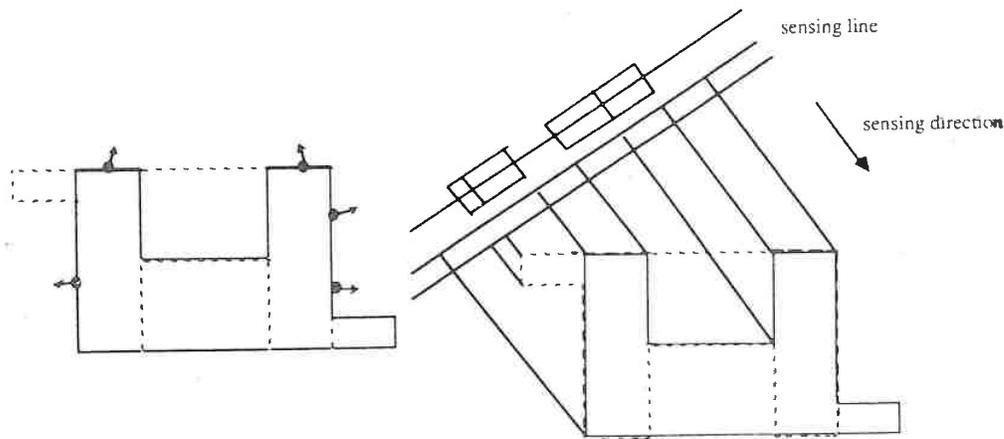


Figure 6.11. Stratégie de prise de vue supplémentaire [Grimson86a]

- (a) deux interprétations compatibles avec cinq points mesurés
- (b) partition de la ligne d'acquisition

Dans notre cas, une prise de vue supplémentaire est une nouvelle coupe par le plan lumineux avec l'objet ; et par la position de prise de vue, nous entendons la position de la ligne d'intersection du plan lumineux avec la surface du plateau. Rappelons que notre système d'acquisition pose des contraintes suivantes :

- la caméra ainsi que le plan lumineux sont fixes.
- les différentes prises de vue peuvent être effectuées par une translation ou une rotation du plateau où se pose l'objet.

Pour déterminer la position de prise de vue supplémentaire, tous les modèles possibles sont transformés de l'espace du modèle dans l'espace de la scène et chaque modèle est projeté sur un plan horizontal. Chaque projection horizontale est associée à son polygone. La position de la prise de vue supplémentaire peut être déterminée par la technique de découpage des polygones.

Dans le cas où la transformation associée à une interprétation est définie partiellement, pour déterminer complètement la transformation, il suffit de prendre une nouvelle prise de vue en tournant le plateau de telle sorte que la ligne formée par l'intersection du plan lumineux avec la surface du plateau soit parallèle à la direction de la translation associée à cette hypothèse. Nous ne discuterons donc dans la suite que du problème d'identifier les interprétations en cas où la transformation associée à chaque interprétation est complètement définie.

Etant données deux interprétations, il est naturel de penser à prendre une nouvelle prise de vue dans une position où le plan lumineux passe par la zone de différence de chaque polygone associé à chaque interprétation. Etant donnée la position de la nouvelle prise de vue, pour chaque interprétation, une zone séparative est estimée. Par exemple, dans la figure 6.12, la zone séparative pour l'interprétation *A* est $((a_x, a_y), (b_x, b_y))$ et $((c_x, c_y), (d_x, d_y))$ pour l'interprétation *B*. S'il y a certaines données parmi les données provenant de la nouvelle prise de vue tombent dans la zone séparative de l'une de ces interprétations, cette interprétation est donc considérée comme valide.

Cependant cette méthode a les inconvénients suivants :

- Dans le cas où il y a plus de deux interprétations compatibles avec des données acquises, plusieurs prises de vue supplémentaires (dépendant du nombre des interprétations compatibles) sont nécessaires pour identifier uniquement l'objet.
- Cette méthode échouera lorsque la partie de l'objet dans la zone séparative est cachée (la partie de l'objet ne sera visible par la caméra que si l'angle entre l'axe optique de la caméra et la normale de la surface correspondante est inférieur à un certain seuil).

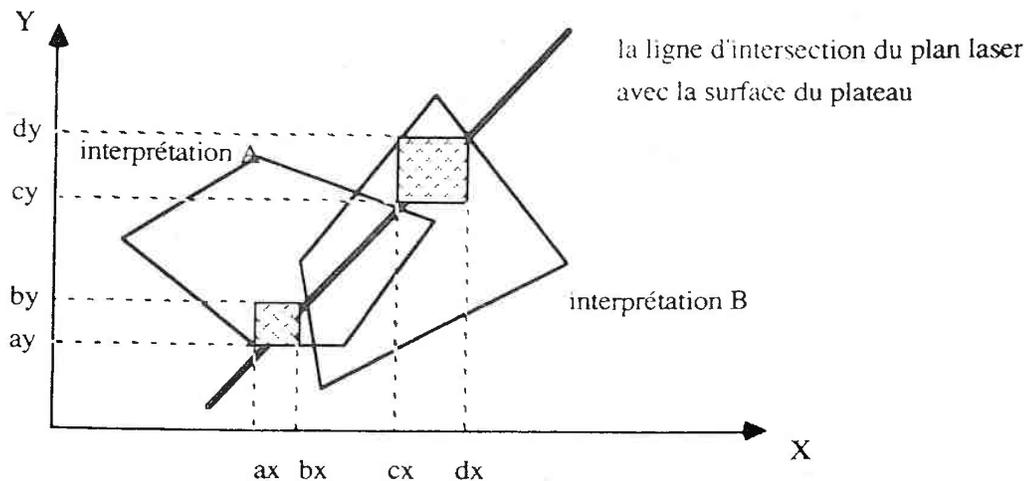


Figure 6.12. Les zones séparatives de deux interprétations

Nous avons donc adopté une autre méthode fondée sur l'intersection de tous les polygones associés aux interprétations possibles. Notons que l'intersection de tous les polygones ne peut être vide du fait que les interprétations ont au moins en commun les mesures précédentes. En considérant que la position de la nouvelle prise de vue ne doit pas être sensible à l'erreur associée à chaque interprétation, la nouvelle coupe est prise lorsque le plateau a été translaté d'une certaine distance de telle sorte que la ligne formée par l'intersection du plan lumineux avec la surface du plateau passe par le centre de l'intersection de tous les polygones :

Etant donné n polygones associés à n modèles compatibles avec des données acquises, les notations utilisées pour l'algorithme d'intersection des polygones sont les suivantes :

polygone- i — une liste des sommets ordonnées de polygone i ($i = 1, n$).

PetQ — une liste des sommets ordonnées de polygone résultant de l'intersection

PetQ = *intersection-de-deux-polygones*(*polygone 1*, *polygone 2*)

pour tous les polygone i ($i = 3, n$) faire

PetQ = *intersection-de-deux-polygones*(*polygone- i* , *PetQ*)

fpour

Nous avons adopté le programme de O'Rourke [Hegron] pour obtenir l'intersection de deux polygones. le principe en est le suivant :

étant donnés deux polygones P et Q (figure 6.13) dont les directions des arêtes sont orientées de telle façon que le polygone soit toujours à la gauche de l'arête. partant de deux arêtes pc (p_k étant le point courant, et p_{k-1} le point précédent) et qc (q_j étant le point courant, et q_{j-1} le point précédent) choisies arbitrairement de deux polygones, on progresse alternativement sur l'une ou l'autre des arêtes de telle sorte que l'une des deux arêtes pointe vers l'autre. cela est assuré par les règles suivantes :

- $\underline{si} \ \vec{q_c} \wedge \vec{p_c} \geq 0 \ \underline{et} \ pc \in \text{demi-plan}(q_{j-1}q_j)$ alors $qc = qc\text{-suivant}$
- $\underline{si} \ \vec{q_c} \wedge \vec{p_c} \geq 0 \ \underline{et} \ pc \notin \text{demi-plan}(q_{j-1}q_j)$ alors $pc = pc\text{-suivant}$
- $\underline{si} \ \vec{q_c} \wedge \vec{p_c} \leq 0 \ \underline{et} \ qc \in \text{demi-plan}(p_{k-1}p_k)$ alors $pc = pc\text{-suivant}$
- $\underline{si} \ \vec{q_c} \wedge \vec{p_c} \leq 0 \ \underline{et} \ qc \notin \text{demi-plan}(p_{k-1}p_k)$ alors $qc = qc\text{-suivant}$

où $\text{demi-plan}(p_{k-1}p_k)$ est le demi-plan qui contient le polygone P défini par la droite passant par p_{k-1} , p_k , de même pour $\text{demi-plan}(q_{j-1}q_j)$.

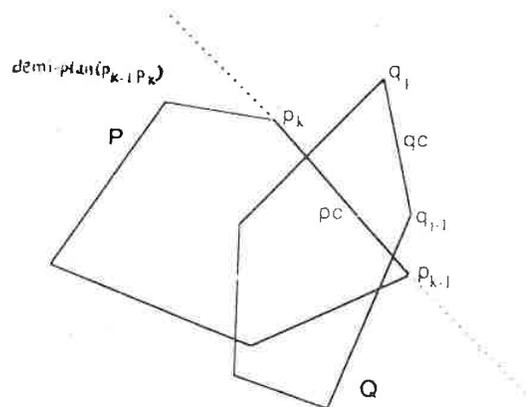


Figure 6.13. intersection des polygones

A chaque pas, on teste si deux arêtes se coupent. si c'est le cas, $PetQ = PetQ \cup$ le point d'intersection. à cette étape, si l'on va progresser sur $pc\text{-suivant}$ et si $pc \in \text{demi-plan}(q_{j-1}q_j)$, alors $PetQ = PetQ \cup q_j$; sinon, si l'on va progresser sur $qc\text{-suivant}$ et $qc \in \text{demi-plan}(p_{k-1}p_k)$, alors $PetQ = PetQ \cup q_j$. Le processus se répète jusqu'à ce que soit le point d'intersection revient à la première

intersection soit le nombre d'itérations est supérieur à $2(M+N)$, dont M et N sont respectivement le nombre des sommets de deux polygones.

6.3.2 Vérification

Pour chaque interprétation, la liste des segments de droite 3D prédits qui devaient être formés par l'intersection du plan lumineux avec le modèle est estimée. Le processus de vérification cherche alors la meilleure cohérence entre des données acquises supplémentaires et la liste des segments de droite prédits pour chaque interprétation. Les interprétations dont les listes des segments de droite 3D ne sont pas cohérentes avec les données acquises supplémentaires sont rejetées.

la liste de segments prédits associée à une interprétation

Etant données une interprétation possible et la position de la nouvelle prise de vue ($x = x_c$ est l'équation du plan laser), le modèle associé à cette interprétation étant transformé dans la scène, on teste pour chaque facette de modèle transformé si elle se coupe avec le plan laser $x = x_c$. Pour ce faire, il suffit de tester, dans le plan $X - Y$, si la droite $x = x_c$ intersecte les polygones formés par la projection verticale de ces facettes (fig. 6.14). cela est facilement accompli par un programme de découpage d'une droite par un polygone [Hegron]. On obtient ainsi des coordonnées 2D (x, y) des points d'intersection, les coordonnées z de ces points sont obtenues en reprojétant les points d'intersection sur les facettes correspondantes. On obtient ainsi une liste de segments 3D prédits pour chaque interprétation possible.

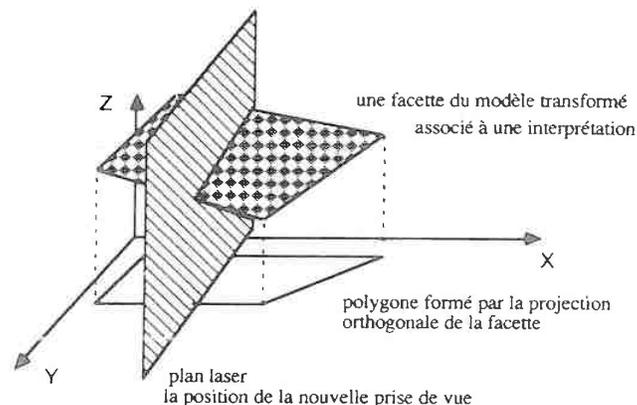


Figure 6.14. Estimation des segments prédits pour une interprétation

Critère de cohérence

Nous définissons un critère de qualité pour juger la cohérence entre les données acquises supplémentaires et la liste de segments de droite 3D prédits associée à une interprétation (figure 6.15) :

Pour chaque nouveau point acquis appartenant à l'intervalle limité par deux extrémités d'un segment de droite prédit, si la distance verticale entre le point acquis et le segment de droite prédit est inférieure à un seuil, ce point est alors considéré comme point cohérent avec cette liste des segments prédits; dans le cas contraire, le point sera considéré incohérent avec la liste des segments prédits.

Pour chaque interprétation possible, nous avons donc le critère suivant :

$$q_i = \frac{\text{nombre de nouveaux points acquis cohérents}}{\text{nombre total de nouveaux points acquis}}$$

L'interprétation correspondant au critère de qualité d'une valeur la plus grande est considérée comme interprétation valide.

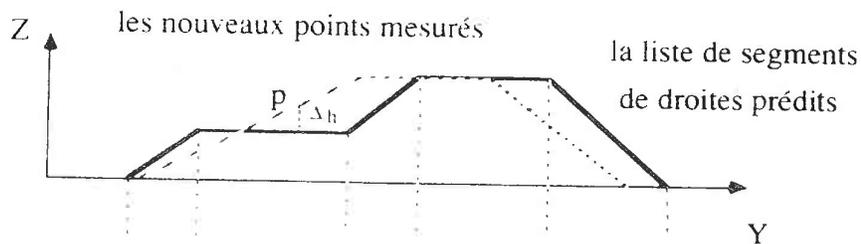


Figure 6.15. vérification

6.3.3 Résultats

Nous avons testé cette approche avec plusieurs objets polyédriques. L'identification des objets différents peut être effectuée en répétant ce processus avec des modèles des objets différents. Nous ne montrons ici que les résultats avec deux modèles d'objets (fig. 6.16).

Pour le test 1, avec le modèle 1, les appariements pour les segments 1, 2, 3 de la scène (fig. 6.17) sont trouvés, comme les figures 6.18, 6.19 le montrent, deux interprétations sont possibles. On trouve aussi une interprétation possible avec le modèle 2 (fig. 6.20). Les résultats d'estimation de la transformation pour chaque interprétation sont montrés au tableau 6.1.

Pour les interprétations 1 et 2 (fig. 6.18, 6.19) les segments 4' et 4'' reliant respectivement les deux extrémités des segments appariés 2', 3' et 2'', 3'' devraient apparier respectivement avec le segment 4 de la scène (fig. 6.17), or les longueurs de deux segments 4' et 4'' ainsi que les angles entre les segments 2', 4' et les segments 2'', 4'' ne sont pas compatibles avec la longueur du segment 4 de la scène et l'angle entre deux segments 2, 4, ces deux interprétations sont donc rejetées. L'interprétation valide pour test 1 est donc la troisième interprétation. La figure 6.20 montre la superposition des segments acquis de la scène avec le modèle ayant été transformé dans la scène.

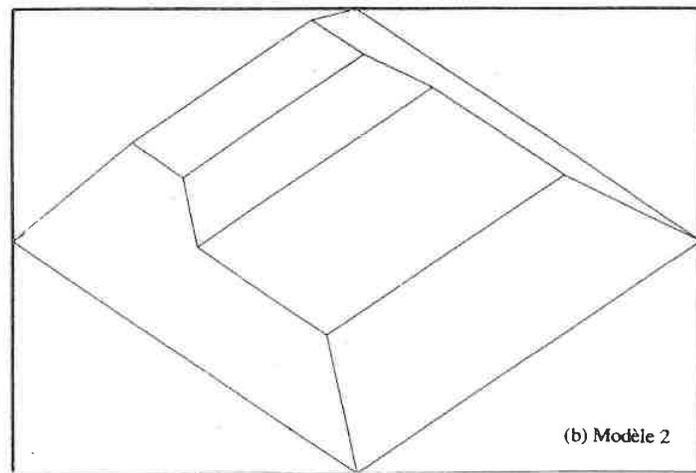
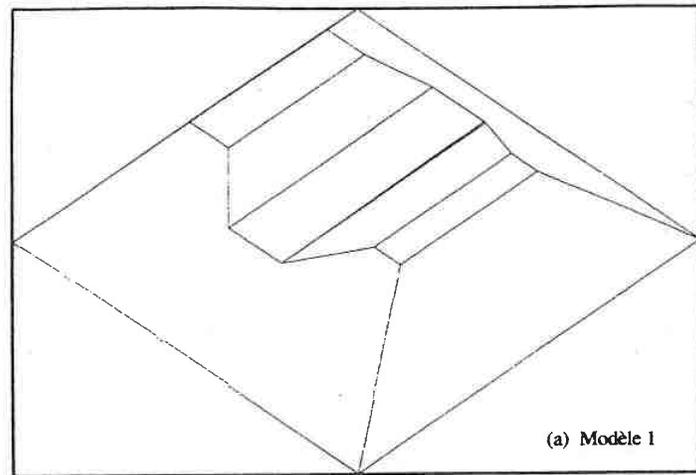


Figure 6.16. deux modèles utilisés pour les tests : (a) modèle 1 (b) modèle 2

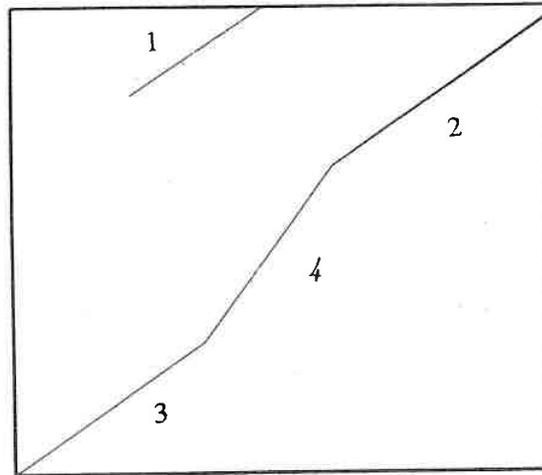


Figure 6.17. test 1 : les segments acquis de la scène

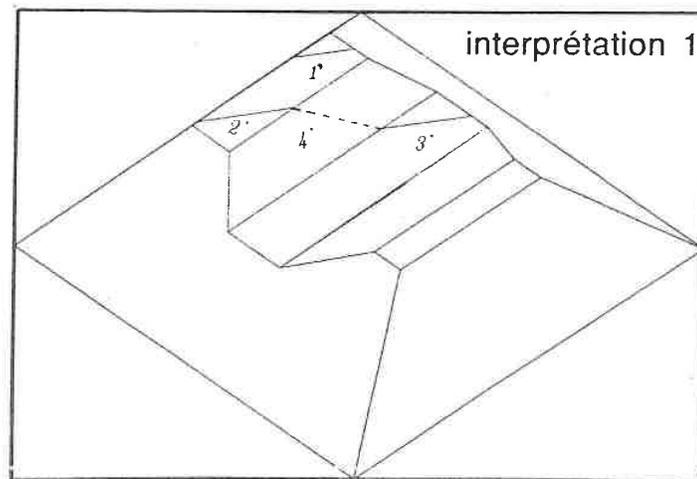


Figure 6.18. test 1 (interprétation 1) : les segments hypothétiques correspondant aux segments 1, 2, 3 de la scène sont trouvés dans le modèle 1

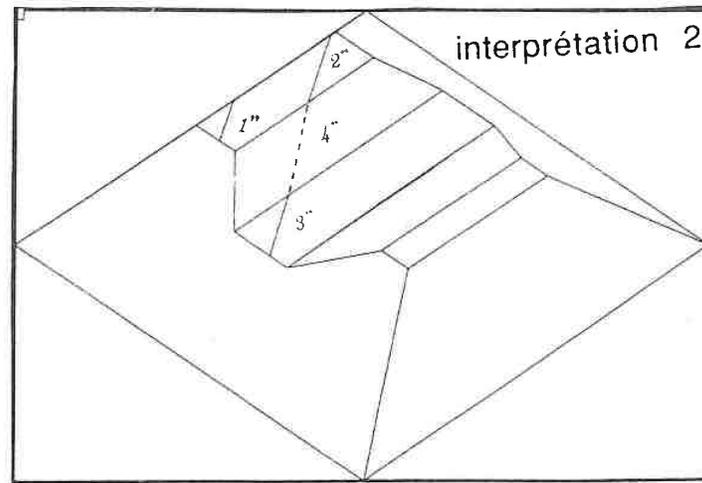


Figure 6.19. *test 1 (interprétation 2)* : les segments hypothétiques correspondant aux segments 1, 2, 3 de la scène sont trouvés dans le *modèle 1*

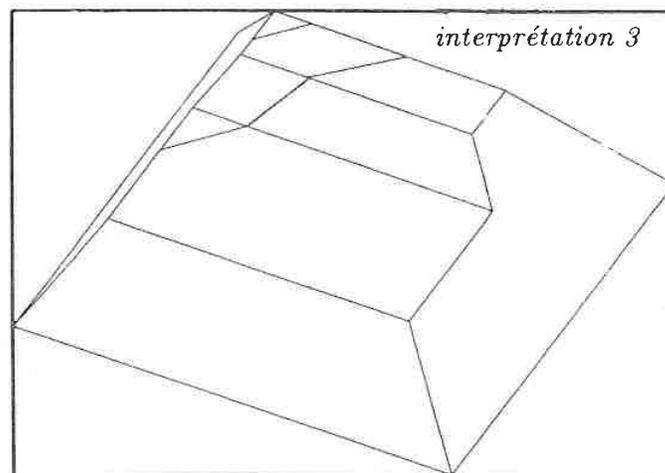


Figure 6.20. *test 1 (interprétation 3)* : la superposition des segments acquis de la scène avec le *modèle 2* ayant été transformé dans la scène

TEST 1		ESTIMATION DE LA TRANSFORMATION			VARIANCES SUR LES ESTIMATIONS		
		Ang. (deg.)	Tx (mm)	Ty (mm)	V(ang) (deg.)	V(Tx) (mm)	V(ty) (mm)
modèle 1	interprétation 1	- 32.78	15.3	104	0.27	0.123	17.9
	interprétation 2	32.8	69.36	53.6	0.224	0.227	1.33
modèle 2	interprétation 3	47.8	97.5	48.5	1.59	1.7	1.02

Figure 6.21. Tab 6.1

Pour le test 2 (fig. 6.22), aucun appariement n'est trouvé pour le modèle 1, mais une seule interprétation est trouvée pour modèle 2. La figure 6.23 montre la superposition des segments acquis avec le modèle transformé. Le résultat d'estimation de la transformation pour cette interprétation est montré au tableau 6.2.

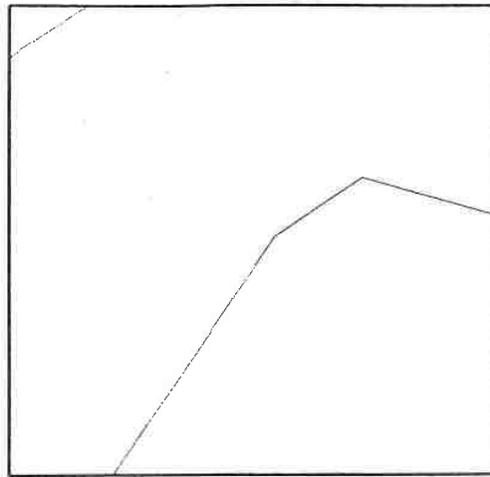


Figure 6.22. *test 2* : les segments acquis de la scène

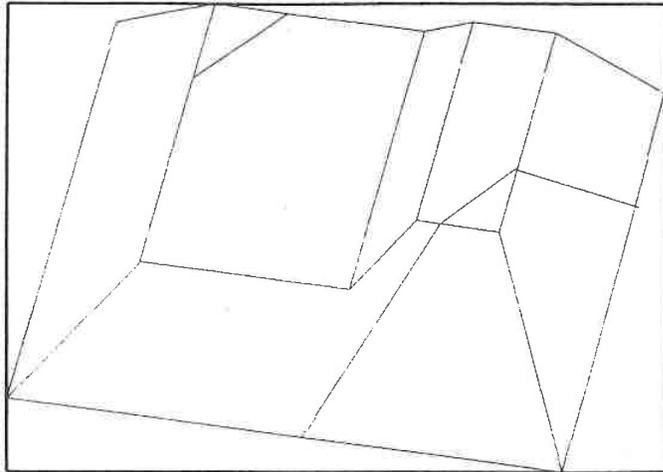


Figure 6.23. *test 2* : la superposition des segments acquis de la scène avec le *modèle 2* ayant été transformé dans la scène

TEST 2		ESTIMATION DE LA TRANSFORMATION			VARIANCES SUR LES ESTIMATIONS		
		Ang. (deg.)	Tx (mm)	Ty (mm)	V(ang) (deg.)	V(Tx) (mm)	V(Ty) (mm)
modèle 2	interprétation 1	146.8	140.4	90.4	1.7	0.6	1.7

Figure 6.24. Tab 6.2

Pour le test 3 (fig. 6.25), aucune interprétation possible n'est trouvée pour modèle 1, mais pour le modèle 2, comme le montre la figure 6.26, quatre couples de segments hypothétiques correspondant à un couple de segments acquis sont trouvés, chaque couple de segments hypothétiques dans le modèle correspond à une interprétation possible. Les résultats sur les estimations des transformations sont montrés au tableau 6.3. La figure 6.27 montre la superposition de segments acquis de la scène avec les quatre positions possibles (interprétations 1, 2, 3, 4) de modèle 2.

Ayant calculée l'intersection des polygones associés aux quatre interprétations, une nouvelle prise de vue est prise (l'équation de la ligne formée par l'intersection du plan laser avec la surface du plateau est $x = 61$). La figure 6.29 montre les segments prédits pour les quatre interprétations ainsi que les données de la nouvelle prise de vue. Lors du processus de vérification, pour les interprétations 3, 4, le nombre de points cohérents est peu devant le nombre total de nouveaux points acquis, ces deux interprétations sont donc rejetées. Le taux de qualité associé à l'interprétation 1 est légèrement plus grand que celui associé à l'interprétation 2, comme la variance sur la transformation associée à l'interprétation 2 est supérieure à celle associée à l'interprétation 1, l'interprétation 1 (correspondant à couple de segments (1, 1')) reste donc valide.

Le tableau 6.4 montre les différences entre la transformation réelle et la transformation estimée pour chaque test. Remarquons que pour un modèle très différent de l'objet de la scène, le modèle sera tout de suite rejeté par le processus d'appariement.

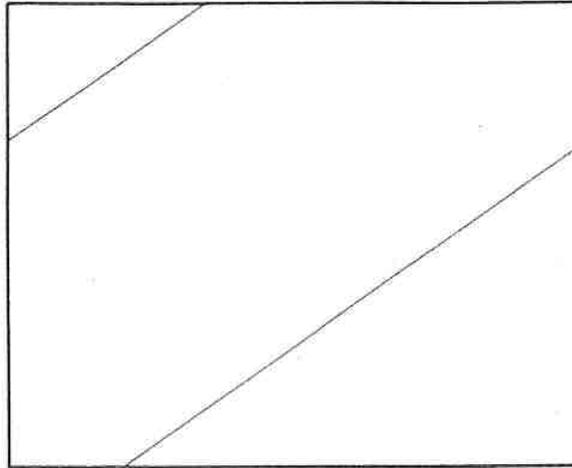


Figure 6.25. *test 3* : les segments acquis de la scène

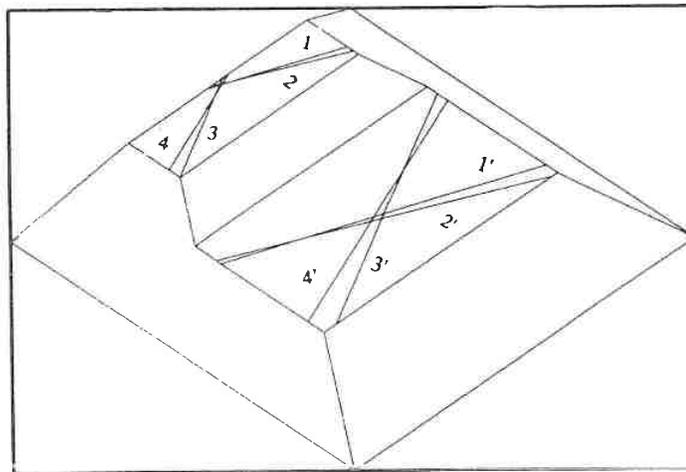


Figure 6.26. *test 3* : les segments de droite hypothétiques correspondant aux segments de droite acquis (1, 1'), (2, 2'), (3, 3'), (4, 4')

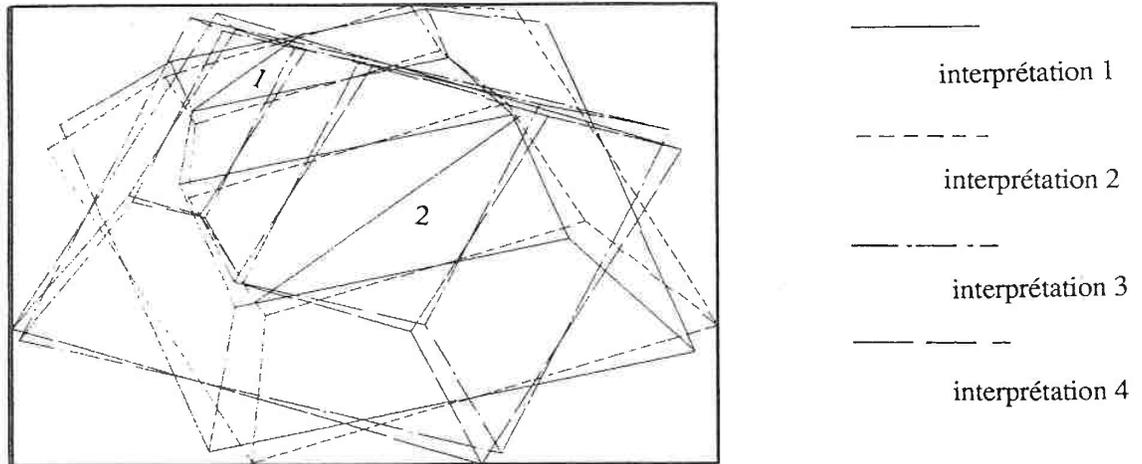


Figure 6.27. test 3 : les segments acquis de la scène et les quatre positions possibles du modèle 2 dans la scène

TEST 3		ESTIMATION DE LA TRANSFORMATION			VARIANCES SUR LES ESTIMATIONS		
		Ang. (deg.)	Tx (mm)	Ty (mm)	V(ang) (deg.)	V(Tx) (mm)	V(ty) (mm)
modèle 2	interprétation 1	21.6	39.5	15.09	0.08	0.075	0.2
	interprétation 2	28.18	46.6	13.28	0.01	0.01	0.87
	interprétation 3	- 25.4	- 4.7	61.2	0.014	0.27	4.6
	interprétation 4	- 21.5	- 1.82	57.08	0.015	0.083	3.85

Figure 6.28. Tab 6.3

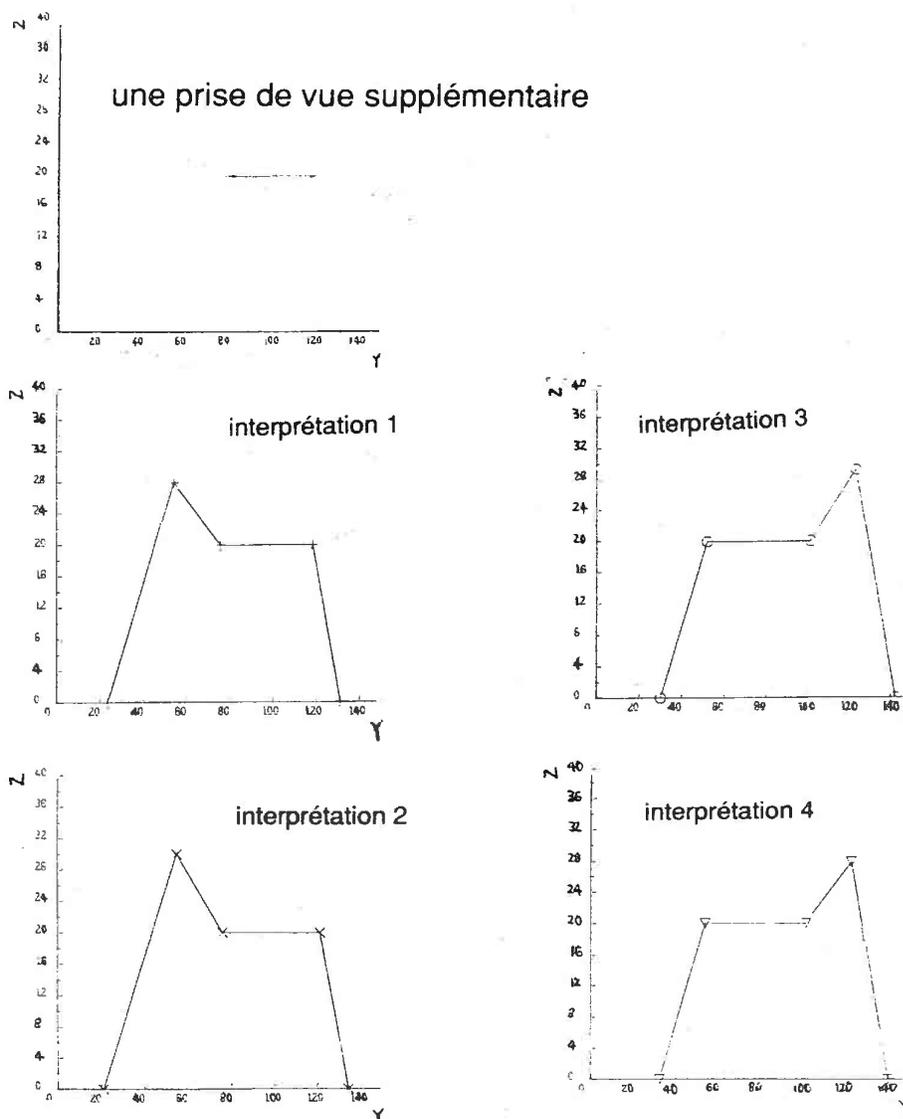


Figure 6.29. test 3 : les segments prédits pour les deux interprétations à la nouvelle position de prise de vue

	Angle de rotation (degré)	translation	
		Tx (mm)	Ty (mm)
test 1	-2.76	-2.54	-1.50
test 2	1.09	0.40	0.36
test 3	-0.89	-0.50	-0.09

Figure 6.30. Tab 6.4 : La différence entre la transformation réelle et la transformation estimée

7

CONCLUSION

Nous avons présenté un système d'acquisition, de modélisation, de localisation et d'identification d'objets 3D. Les recherches que nous avons effectuées dans cette thèse se résument comme suit :

- **calibrage**

Nous avons mise en œuvre une méthode de calibrage de l'ensemble de système d'acquisition permettant de calculer les coordonnées 3D des points à partir de leurs images 2D. L'utilisation de la projection d'un plan lumineux comme moyen d'acquisition des données 3D s'est avérée bien apte à notre objectif : reconnaître et localiser un objet polyédrique à partir de peu de données mesurées. Il faut toutefois citer que le problème de parties cachées pose de sérieux problème pour ce type de système d'acquisition dans certain cas (par exemple, objets fortement concaves). Pour résoudre ce problème, il est facile d'étendre notre système d'acquisition en ajoutant une autre caméra. Dans ce cas, le processus de calibrage reste le même, il suffit de l'effectuer respectivement pour les deux caméras. Une coupe de la scène est donc observée par les deux caméras, en balayant toute la scène, on obtient ainsi deux images de distance pour une même scène, l'image de distance unique de la scène peut être reconstruite en regroupant les deux images de distance de vues différentes. Un point (x, y) sera un point caché seulement s'il n'est vu par aucune de ces deux caméras. On peut également résoudre le problème de parties cachées en ajoutant une autre source laser. mais cela demande une identification de différentes projections lasers (coupes) perçues simultanément [Lelandais 84].

- **modélisation**

Nous avons également mise en œuvre une méthode de construction de modèle d'objet polyédrique à partir de plusieurs vues de l'objet. Les différentes vues

sont acquises par une rotation connue de l'objet. Pour chaque vue de l'objet, la méthode de croissance de régions est utilisée pour segmenter l'image de distance. Grâce à la signification physique des prédicats, le processus de la croissance de régions est assez stable. La partition en régions de chaque vue est représentée par un graphe d'adjacence. Le modèle complet de l'objet est finalement obtenu par fusion de tous les graphes correspondant aux différentes vues de l'objet.

- localisation et identification

Pour accélérer le processus d'identification d'objets, contrairement à la plus part des travaux publiés dans ce domaine qui utilisent les données denses de la scène comme données de départ, nous avons proposé une méthode d'identification d'objet en réduisant au maximum les données à acquérir, montrant ainsi l'intérêt de la méthode (le temps d'acquisition est largement réduit, et le processus de segmentation est supprimé). Le problème délicat de la correspondance entre les données éparses mesurées et le modèle d'objet est résolu sous forme d'un algorithme de prédiction et de vérification.

Les principales caractéristiques de notre méthode sont les suivantes:

- au lieu d'être une liste des primitives, la description utilisée pour le processus d'identification est simplement un ensemble de segments de droite répartis sur la surface de l'objet;
- déduire les informations sur la position et l'orientation de l'objet en localisant directement les segments de droite acquis;
- identifier l'objet en utilisant l'algorithme de recherche de clique maximale;
- stratégie de prise de vue supplémentaire pour identifier l'interprétation correcte parmi plusieurs interprétations possibles.

Remarquons que, pour les expérimentations, les modèles choisis sont semblables, le problème d'identification ne pose aucune difficulté pour des objets fondamentalement différents. Par exemple, pour chaque modèle, on sauvegarde :

- la distance maximale entre tous les couples d'arêtes se situant sur la même facette : DA_{max} ;

- la distance maximale entre tous les couples de sommets du modèle : DS_{max} ;
- la plus grande valeur de la coordonnée Z d'un sommet : VZ_{max} .

Parmi tous les segments de droite mesurés, s'il y en a un dont la longueur est supérieure à DA_{max} , ou si la distance maximale entre les extrémités de deux segments de droite mesurés est supérieure à DS_{max} , ou bien si la valeur de la coordonnées Z d'une extrémité d'un segment de droite mesuré est supérieure à VZ_{max} , alors ce modèle est immédiatement écarté.

Les expérimentations montrent des résultats encourageants dans le cas d'objet isolé qui se déplace sur un plan. Dans le cas où l'objet a un déplacement quelconque (i.e. l'objet ayant six degrés de liberté), comme chaque segment de droite mesuré fournit des contraintes sur deux degrés de liberté: la translation perpendiculaire à la surface et la rotation autour d'un axe perpendiculaire à la fois à la normale de la surface et au segment de droite, trois segments de droite mesurés non parallèles se situant sur trois facettes différentes (non parallèles) de l'objet fournissent donc des informations suffisantes pour déterminer la position et l'orientation de l'objet. Cependant, comme nous l'avons mentionné, la principale difficulté réside dans la recherche des correspondances. Ce phénomène est dû essentiellement à l'absence des primitives pertinentes. Pour exploiter efficacement l'espace de recherche, réduire le nombre de solutions à tester dès le début de la recherche, il faut acquérir des contraintes plus rigoureuses. Par exemple, en utilisant un couple de plans lasers non parallèles, l'intersection de deux plans lumineux avec un objet polyédrique forme deux courbes intersectées. On peut calculer la position du point d'intersection et le vecteur normale sur ce point. On obtient ainsi des contraintes sur les normales des facettes. Nous avons constaté que les contraintes sur les normales sont des contraintes efficaces pour réduire l'espace de recherche.

La technique d'identification d'objet à partir de peu de données mesurées peut s'étendre dans le cas d'objet non polyédrique. Par exemple, dans le cas d'un cylindre, l'intersection du plan lumineux avec l'objet forme une courbe. En approximant chaque courbe par une ellipse, on obtient une hypothèse sur l'axe de cylindre. En identifiant un ensemble de hypothèses mutuellement compatibles, on peut déterminer la position et l'orientation de l'objet. Malheureusement cette technique ne peut pas s'appliquer au cas d'objet plus complexe (l'objet dont le type de surfaces est quelconque, c'est à dire, les surfaces ne sont ni de type polyédrique ni de

type régulière), dans ce cas, il est préférable d'utiliser la technique de reconnaissance à partir de données denses mesurées.

Une autre technique d'identification d'objet qui nous paraît intéressante est d'utiliser conjointement les informations acquises par un capteur 2D et celles acquises par un capteur 3D [Magee 83]. Mais le problème d'exploiter efficacement des informations 2D et 3D reste un problème ouvert. Nous pensons que cette technique ouvre une voie prometteuse.

Annexe A

Reconstitution d'une image de distance à partir des données 3D

Problème :

construire une image de forme $z = f(x, y)$ à partir des données 3D obtenues.

Avec notre système d'acquisition, les données 3D de sortie sont de forme suivante:

$$x_1, (y_{11}, z_{11}), (y_{12}, z_{12}), \dots, (y_{1m_1}, z_{1m_1});$$

$$x_2, (y_{21}, z_{21}), (y_{22}, z_{22}), \dots, (y_{2m_2}, z_{2m_2});$$

⋮

$$x_n, (y_{n1}, z_{n1}), (y_{n2}, z_{n2}), \dots, (y_{nm_n}, z_{nm_n});$$

m_i ($i = 1, n$) étant le nombre de points pour chaque profil, et n le nombre de coupes acquises. Remarquons que, en raison de l'effet de la projection perspective, les coordonnées ne sont pas nécessairement régulièrement espacées suivant la direction de y .

On peut créer une image de distance en choisissant arbitrairement l'axe de vue, il suffit de transformer les données dans le repère approprié:

$$X = x \cos L \cos b + y \sin L \cos b - z \sin b$$

$$Y = -x \sin L + y \cos L$$

$$Z = x \sin b \cos L + y \sin b \sin L + z \cos b$$

dont L est l'angle entre la projection de la normal de l'axe de vue avec l'axe x , et b l'angle entre la normal de l'axe de vue avec l'axe z (figure A.1).

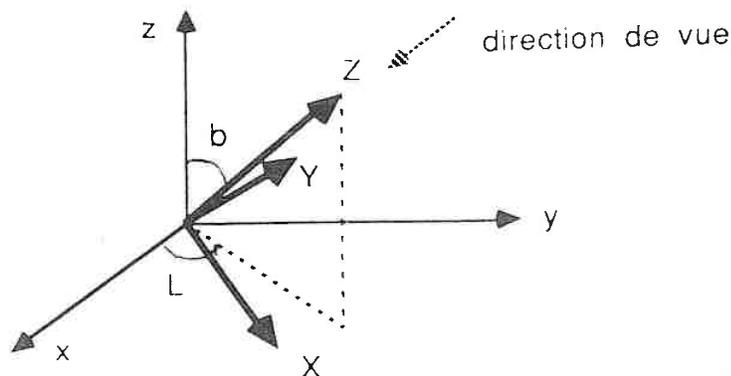


Figure A.1. détermination d'une vue

Les données sont alors sauvegardées dans trois tableaux:

$$X[u][v], Y[u][v], Z[u][v]$$

où u ($u = 0, n$) est l'indice associé aux projections successives de laser (u -ème coupe), et v est l'indice de la ligne du plan image.

Ensuite le problème consiste à calculer Z_{ij} en chaque point d'une grille régulièrement espacée (X_i, Y_j) .

Pour ce faire, un échantillonnage bidimensionnel en X et en Y est effectué. On obtient $X_i = X_{i-1} + \Delta x$ et $Y_j = Y_{j-1} + \Delta y$, où Δx est le pas d'échantillonnage suivant l'axe X , et Δy le pas d'échantillonnage suivant l'axe Y .

Nous utilisons la technique de l'interpolation bilinéaire pour calculer Z_{ij} en point (X_i, Y_j) [SaintMarc 88] [Rosenfeld 82]:

étant données quatre points adjacents:

$$(X[u_0][v_0], Y[u_0][v_0], Z[u_0][v_0]), (X[u_0][v_0 + 1], Y[u_0][v_0 + 1], Z[u_0][v_0 + 1]),$$

$$(X[u_0 + 1][v_0], Y[u_0 + 1][v_0], Z[u_0 + 1][v_0]), (X[u_0 + 1][v_0 + 1], Y[u_0 + 1][v_0 + 1], Z[u_0 + 1][v_0 + 1]),$$

la surface continue interpolée par ces points sont:

$$X(\alpha, \beta) = (1-\alpha)(1-\beta)X[u_0][v_0] + (1-\alpha)\beta X[u_0][v_0+1] + \alpha(1-\beta)X[u_0+1][v_0] + \alpha\beta X[u_0+1][v_0+1]$$

$$Y(\alpha, \beta) = (1-\alpha)(1-\beta)Y[u_0][v_0] + (1-\alpha)\beta Y[u_0][v_0+1] + \alpha(1-\beta)Y[u_0+1][v_0] + \alpha\beta Y[u_0+1][v_0+1]$$

dont $\alpha \in [0, 1]$, $\beta \in [0, 1]$.

Si les deux plan $X = X_i$ et $Y = Y_j$ intersectent avec la surface interpolée délimitant par ces points, en substituant $X(\alpha, \beta)$ par X_i et $Y(\alpha, \beta)$ par Y_j dans les équations ci-dessus, on a donc :

$$X_i = (1-\alpha)(1-\beta)X[u_0][v_0] + (1-\alpha)\beta X[u_0][v_0+1] + \alpha(1-\beta)X[u_0+1][v_0] + \alpha\beta X[u_0+1][v_0+1]$$

$$Y_j = (1-\alpha)(1-\beta)Y[u_0][v_0] + (1-\alpha)\beta Y[u_0][v_0+1] + \alpha(1-\beta)Y[u_0+1][v_0] + \alpha\beta Y[u_0+1][v_0+1]$$

Ceci nous conduit à une équation du seconde degré:

$$(a_x c_y - a_y c_x) \alpha^2 + (b_x c_y - b_y c_x + a_x d_y - a_y d_x) \alpha + (b_x d_y - b_y d_x) = 0$$

avec

$$a_x = X[u_0][v_0] - X[u_0+1][v_0],$$

$$b_x = X_i - X[u_0][v_0],$$

$$c_x = X[u_0][v_0] - X[u_0][v_0+1] - X[u_0+1][v_0] + X[u_0+1][v_0+1],$$

$$d_x = X[u_0][v_0+1] - X[u_0][v_0],$$

$$a_y = Y[u_0][v_0] - Y[u_0+1][v_0],$$

$$b_y = Y_j - Y[u_0][v_0],$$

$$c_y = Y[u_0][v_0] - Y[u_0][v_0+1] - Y[u_0+1][v_0] + Y[u_0+1][v_0+1],$$

$$d_y = Y[u_0][v_0+1] - Y[u_0][v_0].$$

La valeur de β peut être facilement obtenue en substituant la valeur de α dans l'un des deux équations ci-dessus.

Soit α_0, β_0 les valeurs des paramètres α, β , on peut alors calculer la valeur de Z_{ij} :

$$Z_{ij} = (1-\alpha_0)(1-\beta_0)Z[u_0][v_0] + (1-\alpha_0)\beta_0 Z[u_0][v_0+1] + \alpha_0(1-\beta_0)Z[u_0+1][v_0] + \alpha_0\beta_0 Z[u_0+1][v_0+1]$$

Ce processus est effectué pour tous les combinaisons de quatre points adjacents et pour toutes les combinaisons possibles des valeurs X_i et Y_j . Il se peut que pour un point (X_i, Y_j) il y ait plusieurs valeurs de Z_{ij} , dans ce cas, la valeur la plus grande est conservée (car l'on suppose que l'axe de vue est opposé à l'axe Z).

Une vue orthogonale des données 3D d'une scène est montrée en figure A.2. La figure A.3 montre les images de distance (de deux vues différentes) de même données 3D.

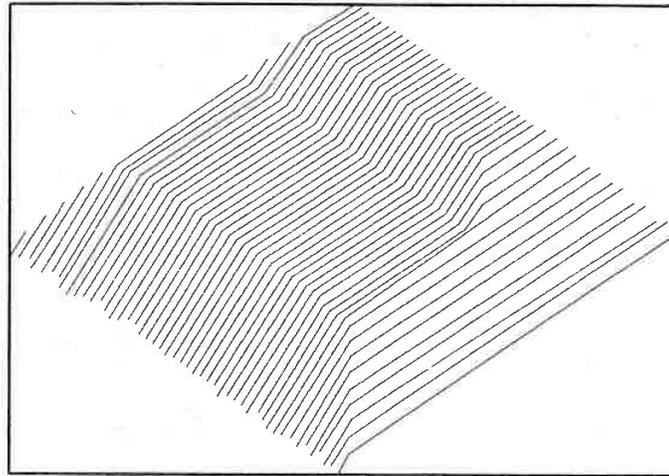


Figure A.2. une vue orthogonale

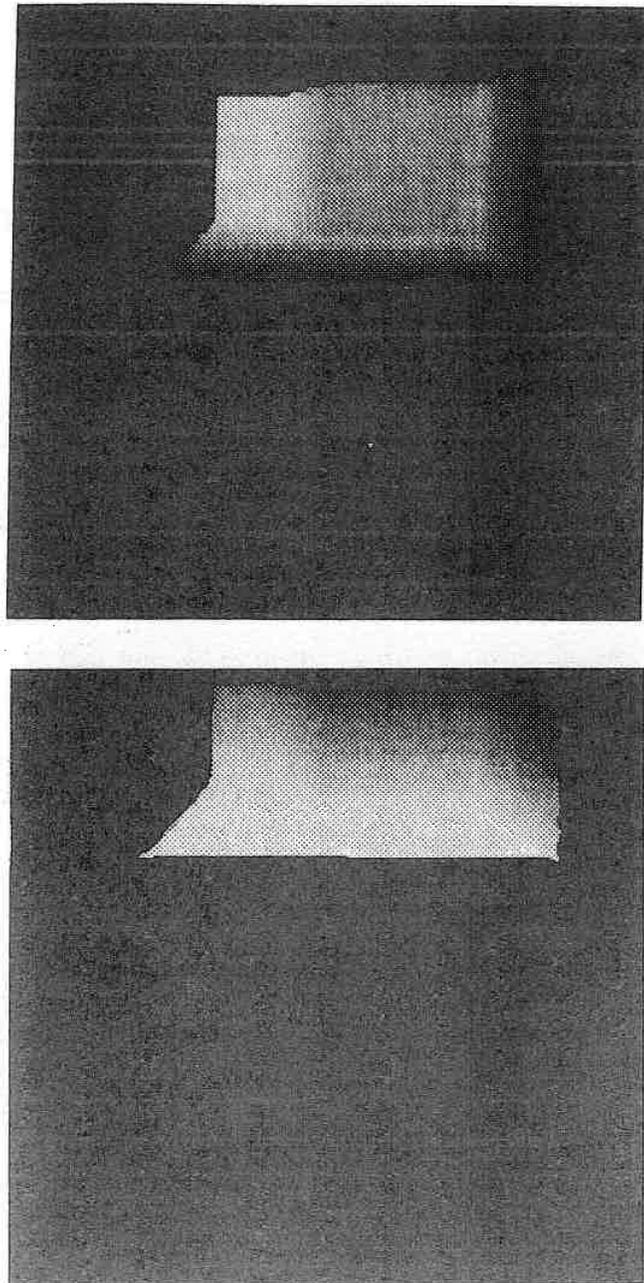


Figure A.3. les images de distance reconstruites

Annexe B

Paramétrisation d'un segment de droite de longueur fixe dont les deux extrémités se situent sur deux arêtes connues

Problème:

Etant donnés la longueur du segment de droite recherché et l'angle formé par un couple d'arêtes entre lesquelles le segment de droite recherché doit se placer, le problème est de localiser ce segment de droite sur ce couple d'arêtes.

Considérons le problème différemment:

Au lieu de chercher le segment, on cherche où devrait se situer le sommet formé par le couple d'arêtes.

En fixant le segment de longueur L , l'angle entre le couple d'arêtes étant θ_0 , on cherche le sommet M (figure B.1):

L'angle entre deux arête étant θ_0 , le sommet $M(x, y)$ satisfait donc l'équation suivante:

$$\tan(\theta_0) = \frac{k_2 - k_1}{1 + k_1 k_2}$$

avec $k_1 = \frac{y}{x + \frac{L}{2}}$ et $k_2 = \frac{y}{x - \frac{L}{2}}$.

Par simple modification, l'équation ci-dessus devient la forme suivante:

$$x^2 + \left(y - \frac{L}{2 \tan(\theta_0)}\right)^2 = \left(\frac{L}{2 \sin(\theta_0)}\right)^2$$

Le lieu du sommet est donc un cercle de rayon $R = L/2 \sin(\theta_0)$ qui passe par les deux extrémités de segment de droite.

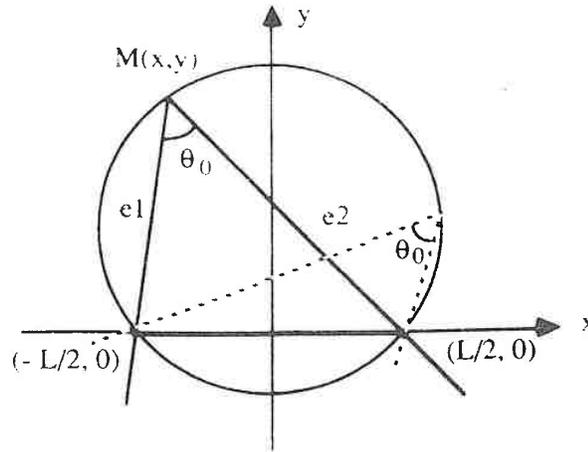


Figure B.1. le lieu du sommet recherché est un cercle

Maintenant on associe un repère 2D au sommet, et cherche l'équation du cercle qui passe par les deux extrémités du segment recherché et le sommet (figure B.2):

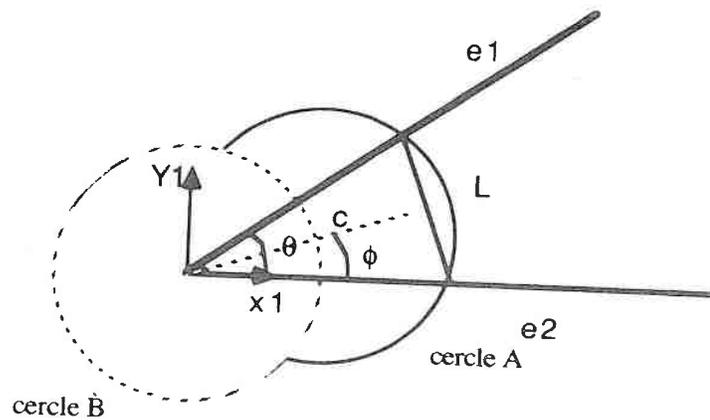


Figure B.2. les segments recherchés définis par le cercle:
cercle A: cercle recherché; cercle B: le lieu du centre de cercle A

supposons que l'équation du cercle avec rayon $L/2\sin(\theta_0)$ dans ce repère est la suivante:

$$(x - a)^2 + (y - b)^2 = \left(\frac{L}{2\sin(\theta_0)}\right)^2$$

le cercle passe par le sommet (0, 0). on a donc :

$$a^2 + b^2 = \left(\frac{L}{2\sin(\theta_0)}\right)^2$$

Le lieu du centre du cercle recherché est donc un cercle de même rayon et centré à l'origine de ce repère.

Dans le repère polaire:

$$a = \frac{L}{2 \sin(\theta_0)} \cos \phi$$

$$b = \frac{L}{2 \sin(\theta_0)} \sin \phi$$

où ϕ est l'angle polaire du centre de cercle recherché.

On trouve ainsi l'équation du cercle recherché en fonction de ϕ :

$$\left(x - \frac{L}{2 \sin(\theta_0)} \cos(\theta_0)\right)^2 + \left(y - \frac{L}{2 \sin(\theta_0)} \sin(\theta_0)\right)^2 = \left(\frac{L}{2 \sin(\theta_0)}\right)^2$$

Les équations des deux arêtes dans le repère associé sont respectivement: $y = 0$ et $y = x \tan(\theta_0)$. Les extrémités du segment de droite recherché sont donc les points d'intersection de ce cercle avec les deux arêtes, l'équation du segment de droite recherché en fonction de ϕ est donc la suivante:

$$y = \tan(\phi - \theta_0) \left(x - \frac{L}{\sin(\theta_0)} \cos(\phi)\right)$$

Comme le cercle doit passer par l'origine de ce repère (le sommet) et également par les deux extrémités de segments de droite recherchés, l'angle ϕ doit satisfaire la condition suivante (figure B.3):

$$\left(\theta_0 - \frac{\pi}{2}\right) < \phi < \frac{\pi}{2}$$

Remarquons que lorsque l'angle θ_0 est supérieur à $\pi/2$, l'intervalle de ϕ est très restreint.

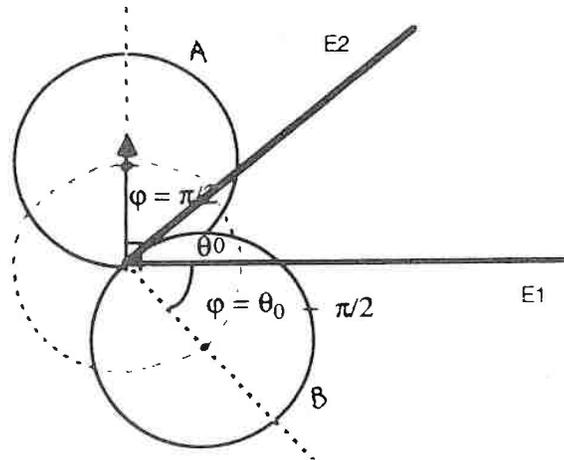


Figure B.3. la condition pour qu'il y ait trois points d'intersection entre le cercle A et les deux arêtes:

$\phi = \frac{\pi}{2}$: le cercle A n'intersecte l'arête $e1$ qu'à l'origine:

$\phi = \theta_0 - \frac{\pi}{2}$: le cercle B n'intersecte l'arête $e2$ qu'à l'origine

Annexe C

Courbes différentielles

Les courbes $|dx_L/d\beta|$, $|dy_L/d\beta|$, $|dz_L/d\beta|$, $|dP/d\beta|$, sont tracées pour les quatre valeurs différentes de β en fonction de j , avec $i_0 = 162$, $j_0 = 127$, $\alpha_u = 863.8$, $\alpha_v = 1310.5$, $t_x = 429.1$, $t_y = 97.2$, $t_z = 543.6$, $i = 150$ et $\theta = 50$ degrés

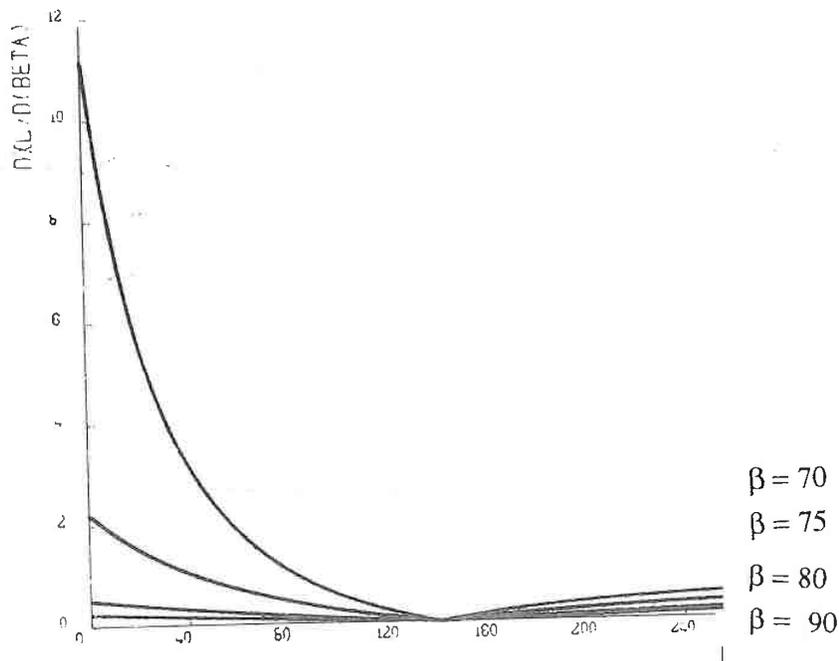


Figure C.1. la courbe $|dx_L/d\beta|$

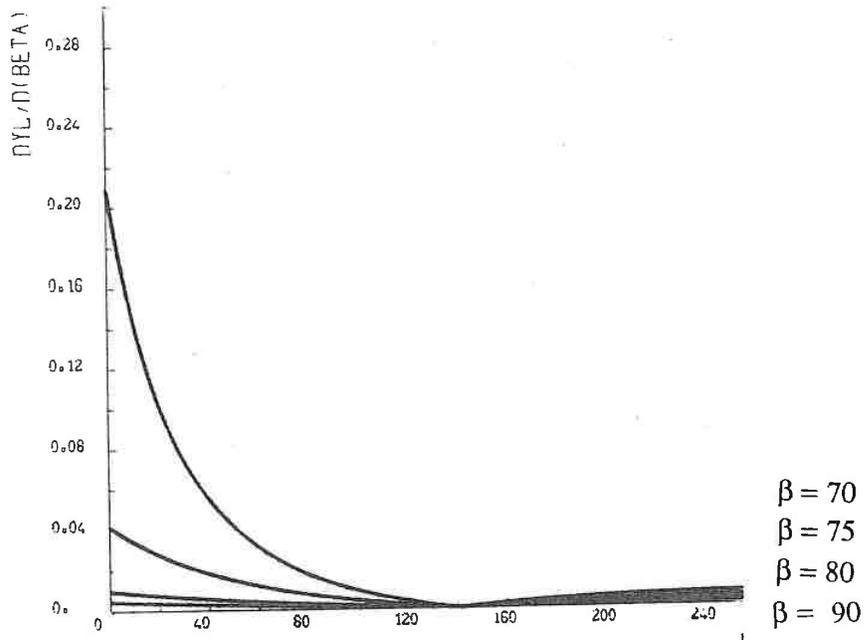


Figure C.2. la courbe $|dy_L/d\beta|$

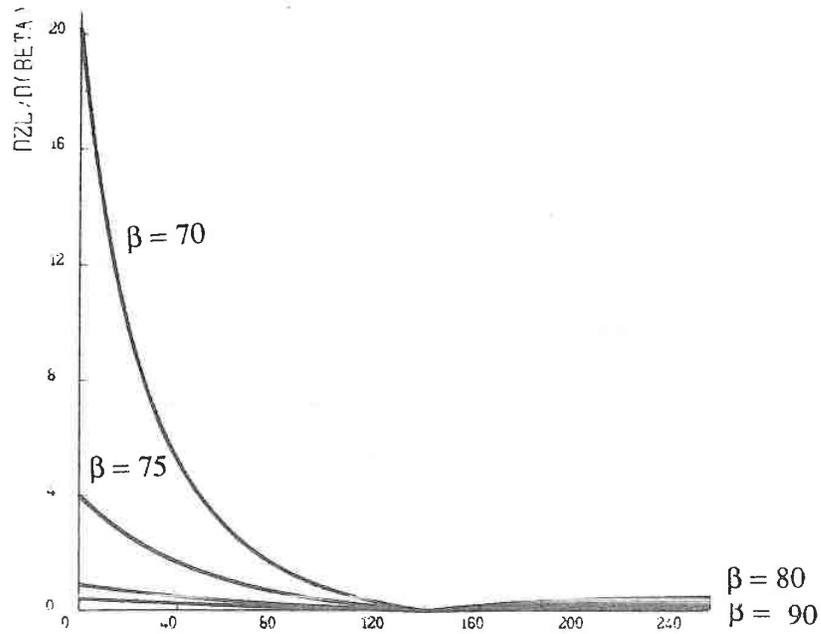


Figure C.3. la courbe $|dz_L/d\beta|$

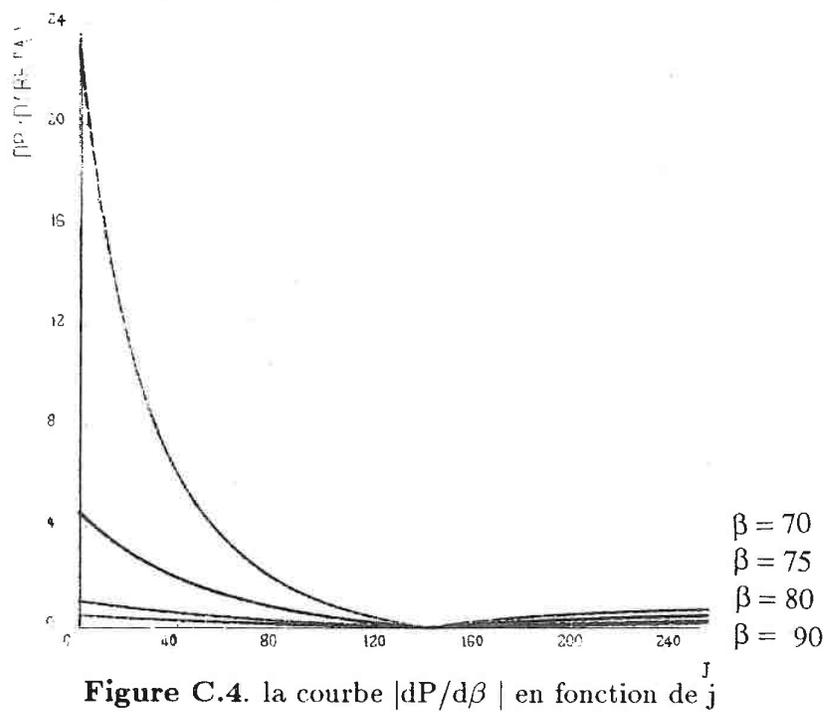


Figure C.4. la courbe $|dP/d\beta|$ en fonction de j

Bibliographie

- [Agin 87] G. J. Agin. Calibration of a Light Stripe Range Sensor. *IEEE international Conference on Robotic and Automation*, pages 807–815, 1987.
- [Arimoto 87] S. Arimoto. H. Noborio, et S. Fukuda. A Feasible Approach to Automatic Planning of Collision-free Robot Motion. *proc. of the 4th International Symposium of Robotique Research*, pages 163–172, California, USA, 1987.
- [Ayache 87] N. Ayache et F. Lustman. Fast and Reliable Passive Stereovision Using Three Cameras. *Proc. International workshop on MACHINE Vision and Machine Intelligence*, Tokyo, 1987.
- [Ballard 82] D. H. Ballard et C. M. Brown. *Computer Vision*. Prentice-Hall, Englewood Cliffs, New Jersey, 1982.
- [Besl 85] P. J. Besl et R. C. Jain. Three-Dimensional Object Recognition. *Computing Surveys*, 17(1), March 1985.
- [Besl 86] P. J. Besl et Ramesh C. Jain. Invariant Surface Characteristics For 3D Object Recognition in Range Images. *Computer Graphics. Vision and Image Processing*, (33):33–80, 1986.
- [Besl 88] P. J. Besl et Ramesh C. Jain. Segmentation Through Variable-Order Surface Fitting. *IEEE trans. on pattern analysis and machine intelligence*, 10(2):167–191, 1988.
- [Bhanu 84] B. Bhanu. Shape Matching and Image Segmentation Using Stochastic Labeling. 1984. Dep. of Electrical Engineering

- Image Processing Institute, University of Southern California, Los Angeles.
- [Bolles] R. C. Bolles et M. A. Fischler. A RANSAC-based Approach to Model Fitting and its Application to Finding Cylinders in Range Data. *SRI International*.
- [Bolles 82] Robert C. Bolles et Ronald A. Cain. Recognizing and Locating Partially Visible Objects: The Local-Feature-Focus Method. *The International Journal of Robotics Research*, 1(3), 1982.
- [Bolles 86] Robert C. Bolles et Patrice Horaud. 3DPO: A Three Dimensional Part Orientation System. *The International Journal of Robotics Research*, 5(3):3-26, 1986.
- [Borienne 84] P. L. Borienne. *Contribution à Vision Par Ordinateur Tridimensionnelle*. Thèse de Doctorat, Institut National Polytechnique de Grenoble, 1984.
- [Boyer 87] K. L. Boyer et A. C. Kak. Color-encoded Structured Light For Rapid Active Ranging. *IEEE trans. on Pattern Analysis and Machine Intelligence*, PAMI-9(1), January 1987.
- [Brady 88] J. P. Brady, N. Nandhakumar, et J. K. Aggarwal. Recent Progress in the Recognition of Objects from Range Data. *9th International Conference on Pattern Recognition*, pages 85-92. Rome. Italy, 1988.
- [Chou 86] H. L. Chou et W. H. Tsai. A New Approach to Robot Location by House Corners. *Pattern Recognition*, 19(6):439-451, 1986.
- [Connolly 86] C. I. Connolly et J. Ross Stenstrom. Construction of Polyhedral Models from Multiple Range Views. *8th International Conference on Pattern Recognition*, pages 85-87, Paris. 1986.
- [Dhome 87] M. Dhome et T. Kasvand. Polyhedral Recognition by Hypothesis Accumulation. *IEEE Trans. on Pattern Recognition and Machine Intelligence*. PAMI-9(3):429-438, 1987.

- [Duda 73] R. O. Duda et P. E. Hart. *Pattern Classification and Scene Analysis*. Wiley-interscience, New York, 1973.
- [Fan 88] T. J. Fan, G. Medioni, et R. Nevatia. *Recognizing 3-D Objects Using Surfaces Description*. 2th International Conference on Computer Vision, pages 474-481, 1988.
- [Faugeras 86] O. D. Faugeras et M. Hebert. *The Representation, Recognition and Localization of 3-D objects*. *The International Journal of Robotics Research*, 5(3):27-52, 1986.
- [Faugeras 88] O.D. Faugeras. *Vision artificielle en dimension trois. Deuxième Atelier Scientifique: Traitement d'images: du pixel à l'interprétation*. Aussois, Savoie, 1988.
- [Gaston 84] P.C. Gaston et T. Lozano-Perez. *Tactile Recognition and Localisation Using Object models: The case of Polyhedra on a Plan*. *IEEE trans. on Pattern Analysis and Machine Intelligence*. PAMI-6(3). May 1984.
- [Gerhard 86] A. Gerhard, H. Platzer, J. Steurer, et R. Lenz. *Depth Extraction by Stereo Triples and a Fast Correspondence Estimation Algorithm*. 8th International Conference on Pattern Recognition, pages 512-515, Paris, 1986.
- [Gordon 87] Steven J. Gordon et Warren P. Seering. *Locating Polyhedral Features from Sparse Light-stripe Data*. *IEEE conference on Robotics and Automation*, 1987.
- [Gordon 88] Steven J. Gordon et Warren P. Seering. *Real-Time Part Position Sensing*. *IEEE Trans. on Pattern Recognition and Machine Intelligence*, 1988.
- [Grimson 81] W. E. L. Grimson. *From Images to Surfaces*. MIT Press édition, 1981.
- [Grimson 84] W. E. L. Grimson et Tomas Lozano-Perez. *Model-based Recognition and Localization from Sparse Range or Tactile Data*. *The International Journal of Robotics Research*, 3(3). 1984.

- [Grimson 86] W. Eric L. Grimson. *Disambiguating Sensory Interpretation Using Minimal Sets of Sensory Data*. IEEE conference on Robotics and Automation, pages 286-292. April 1986.
- [Guisser 87] L. Guisser. *Système de vision active tridimensionnelle à paramétrisation directe de surface*. Thèse de Doctorat. Université Paul Sabatier de Toulouse, 1987.
- [Gurewitz 86] E. Gurewitz, I. Dinstein, et B. Sarusi. *More on the Benefit of a Third Eye for Machine Stereo Perception*. in Proceedings of 8th International Conference on Pattern Recognition, pages 966-968. Paris, 1986.
- [Han 87] J. Han, R. A. Volz, et T. N. Mudge. *Range Image Segmentation and Surface Parameters Extraction For 3D Object Recognition of Industrial Parts*. IEEE international Conference on Robotic and Automation, 1987.
- [Haralick 80] R. M. Haralick. *Using Perspective Transformations in Scene Analysis*. Computer Graphics and Image Processing, (13):191-221, 1980.
- [Hebert 83] M. Hebert. *Reconnaissance de Formes tridimensionnelles*. Thèse de Doctorat. INRIA, 1983.
- [Hegron] G. Hégron. *Synthese d'image: algorithmes élémentaires*.
- [Hoffman 87] R. Hoffman et A. K. Jain. *Segmentation and Classification of Range Image*. IEEE trans. on Pattern Analysis and Machine Intelligence. PAMI-9(5):608-620. 1987.
- [Jarvis 83] R. T. Jarvis. *A Perspective on Range Finding Techniques for Computer Vision*. IEEE Trans. on Pattern Analysis and Machine Intelligence. PAMI-5(2):122-139. 1983.
- [Lelandais 84] S. Lelandais. *Réalisation et Première Exploitation d'un Système de Numérisation de Formes Tridimensionnelles*. Thèse de Doctorat. Université de Compiègne. Avril 1984.
- [Lenz 87] R. K. Lenz et R. Y. Tsai. *Techniques For Calibration of The Scale Factor and Image Center For High Accuracy 3D*

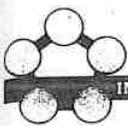
- Machine Vision Metrology. IEEE international Conference on Robotic and Automation. 1987.*
- [Limozin 86] P. Long Limozin et G. Giraudon. *Stereo Matching Using Contextual Line Region Primitives. in Proceedings of 8th International Conference on Pattern Recognition. Paris, 1986.*
- [LozanoPerez 87] T. Lozano-Perez et W. E. L. Grimson. *Finding Cylinders in Range Data. IEEE international Conference on Robotic and Automation. 1987.*
- [Magee 83] M. J. Magee, F. A. Boyer, C.H. Chien, et J. K. Aggarwal. *Experiments in Intensity Guided Range Sensing Recognition of Three Dimensional Objects. Conf. on Computer Vision and Pattern Recognition, pages 550-552, June 1983.*
- [Mansbach 86] P. Mansbach. *Calibration of a Camera and Light Source by Fitting to a Physical Model. Computer Vision, Graphics, and Image Processing. (35):200-219, 1986.*
- [Markowsky 80] G. Markowsky et M. A. Wesky. *Fleshing Out Wire Frames. IBM J. Res. Develop. (24):582-597, 1980.*
- [McCalla] Thomas Richard McCalla. *Introduction to Numerical Methods and Fortran Programming.*
- [Mitiche 83] A. Mitiche et J. K. Aggarwal. *Detection of Edges Using Range Information. IEEE trans. on Pattern Analysis and Machine Intelligence. PAMI-5(2), March 1983.*
- [Monga 86] O. Monga et N. Keskes. *A Hierarchical Algorithm for the Segmentation of 3-D Images. in Proceedings of 8th International Conference on Pattern Recognition, pages 735-737. Paris, 1986.*
- [Muller 84] Y. Muller. *Détection des surfaces par transformée de Hough en vision tridimensionnelle. Thèse de Doctorat, Université de Nancy 1. Centre de Recherche en Informatique de Nancy, mai 1984.*

- [Nitzan 88] D. Nitzan. *Three-Dimensional Vision Structure for Robot Application*. *IEEE Trans. on Pattern Recognition and Machine Intelligence*, 1988.
- [Ohta 85] Y. Ohta et T. Kanade. *Stereo by Two-level Dynamic Programming*. in *Proceedings of 9th International Joint Conference on Artificial Intelligence*, 1985.
- [Ohta 86] Y. Ohta, M. Watanabe, et K. Ikeda. *Improving Depth Map by Right-Angled Trinocular Stereo*. in *Proceedings of 8th International Conference on Pattern Recognition*, pages 519-521. Paris, 1986.
- [Oshima 83] Masaki Oshima et Yoshiaki Shirai. *Object Recognition Using Three Dimensional Information*. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, PAMI-5(4):353-361, 1983.
- [Pauchon 83] E. Pauchon. *Numérisation et Modélisation Automatique d'Objet 3D*. *Thèse de Doctorat, Université de Paris-sud, septembre 1983*.
- [Paulidis 82] Theo Paulidis. *Algorithms for Graphics and Image Processing*. SPRINGER-VERLAG, Berlin, 1982.
- [Ponce 85] J. Ponce et M. Brady. *Toward a Surface Primal Sketch*. in *Proceedings of 2nd IEEE International Conference on Robotics*, pages 420 - 425. 1985.
- [Potmesil 83] M. Potmesil. *Generating Models of Solid Objects by Matching 3D Surface Segments*. in *Proceedings of 8th International Joint Conference on Artificial Intelligence*, pages 8-12, Karlsruhe. West Germany, 1983.
- [Qiang 88] S. Qiang, R. Mohr, et K. Tombre. *Recognizing and locating polyhedral objects from sparse range data*. in *Proceedings of 9th International Conference on Pattern Recognition*, page . Rome, Italy, 1988.

- [Rao 88] K. Rao et R. Nevatia. *Computing Volume Descriptions from Sparse 3D Data. International Journal of Computer Vision*, (2):33-50, 1988.
- [Rioux 84] M. Rioux. *Laser Range Finder Based on Synchronized Scanners. Applied Optics*, 23(21):3837-3844, 1984.
- [Rosenfeld 82] A. Rosenfeld et A. C. Kak. *Digital Picture Processing. ACADEMIC PRESS. INC.*, 1982.
- [SaintMarc 88] P. Saint-Marc. *Outils fondamentaux pour l'analyse d'images en robotique. Thèse de Doctorat, Université Blaise Pascal, juin 1988.*
- [Silverman] G. Silverman, R. Tsai, et M. Lavin. *Locating Polyhedral Objects from Edge Point Data. IBM Los Angeles Scientific Center.*
- [Sobel 74] I. Sobel. *On Calibrating Computer Controlled Cameras For Perceiving 3D Scenes. Artificial Intelligence*, (5):185-198, 1974.
- [Thorpe 88] C. Thorpe, M. H. Hebert, T. Kanade, et S. A. Shafer. *Vision and Navigation for the Carnegie-Mellon Navlab. IEEE Trans. on Pattern Recognition and Machine Intelligence*, 1988.
- [Tombre 87] K. Tombre. *La saisie automatisée de documents composites: reconnaissance, codage et interprétation des parties graphiques. Thèse de Doctorat, L'Institut National Polytechnique de Lorraine. Centre de Recherche en Informatique de Nancy. Nancy, 1987.*
- [Toscani 87] G. Toscani. *Systèmes de Calibration Optique et Perception du Mouvement en Vision Artificielle. Thèse de Doctorat, Université de Paris-sud, centre d'orsay, 1987.*
- [Wall 84] K. Wall et P. E. Danielsson. *A Fast Sequential Method for Polygonal Approximation of Digitized Curves. Computer Vision, Graphics and Image Processing*, (28):220-227, 1984.

- [Wang 87] Y. F. Wang et J. K. Aggarwal. *3D Object Description From Stripe Coding and Multiple Views. in Proceedings of 5th Scandinavian Conference on Image Analysis, 1987.*
- [WrobelDautcourt 88] B. Wrobel-Dautcourt. *Perception de la distance par mise en correspondance de régions entre images stéréoscopiques. Thèse de Doctorat. L'Institut National Polytechnique de Lorraine. Centre de Recherche en Informatique de Nancy, 1988.*
- [Yachida 86] M. Yachida, Y. Kitamura, et M. Kimachi. *Trinocular Vision: New Approach for Correspondence Problem. in Proceedings of 8th International Conference on Pattern Recognition, pages 1041-1044. Paris, 1986.*
- [Yakimovsky 78] Y. Yakimovsky et R. Cunningham. *A system For Extracting Three-dimensional Measurements From a Stereo Pair of Cameras. Computer Graphics and Image Processing, (7):195-121, 1978.*
- [Yamamoto 86] H. Yamamoto, K. Sato, et S. Inokuchi. *Range Imaging System Based on Binary Image Accumulation. in Proceedings of 8th International Conference on Pattern Recognition, pages 233-235. Paris, 1986.*
- [Yang 85] H. S. Yang et A. C. Kak. *Determination of The Identity, Position and Orientation of The Topmost Object in a Pile. 3rd IEEE Comput. Soc. Workshop Comput. Vision, Representation and Control, pages 38-48, 1985.*





INPLI
INSTITUT NATIONAL POLYTECHNIQUE DE LORRAINE

**AUTORISATION DE SOUTENANCE DE THESE
DU DOCTORAT DE L'INSTITUT NATIONAL POLYTECHNIQUE DE LORRAINE**

VU LES RAPPORTS ETABLIS PAR :

Monsieur le Professeur CROWLEY, LIFIA Grenoble,
Monsieur le Professeur MARCHAND, CRIN/Université de Nancy I.

Le Président de l'Institut National Polytechnique de Lorraine,
autorise :

Madame QIANG Shijia

à soutenir devant l'INSTITUT NATIONAL POLYTECHNIQUE DE
LORRAINE, une thèse intitulée :

"Stratégie de localisation et identification d'objet à partir de
quelques mesures tridimensionnelles"



en vue de l'obtention du titre de :

DOCTEUR DE L'INSTITUT NATIONAL POLYTECHNIQUE DE LORRAINE

Spécialité : "Informatique"

Fait à Vandoeuvre le, 26 Juin 1989
Le Président de l'I.N.P.L.,

M. GANTOIS

2, avenue de la Forêt de Haye - B.P. 3 - 54501 VANDŒUVRE/CEDEX
Téléphone : 83. 57. 48. 48 - Télécopie : 83. 57. 49. 55