

UNIVERSITE D'EVRY - VAL D'ESSONNE

THESE

*Présentée en vue de l'obtention du*

Doctorat de l'Université d'Evry Val d'Essonne

*Spécialité*

Robotique

PAR

**Mohamed El-Hadi TRABELSI**

TITRE

---

***Combinaison d'informations visuelles et ultrasonores  
pour la localisation d'un robot mobile et la saisie  
d'objets***

---

Soutenue le **7 décembre 2006** devant la commission d'examen :

Mr. Patrick BONNIN	Prof. Université Paris XIII	Rapporteur
Mr. Michel DEVY	Directeur de recherche LAAS	Rapporteur
Mr. Hichem MAARAF	Prof. Université d'Evry	Examinateur
Mme. Odile HORN	McF Université de Metz	Examinaterice
Mr. Christian BARAT	McF Université de Nice	Examinateur
Mme. Naïma AITOUFROUKH	McF Université d'Evry	Examinaterice
Mme. Sylvie LELANDAIS	McF Université d'Evry	Directrice de thèse

## Remerciements

Ce travail de recherche s'est déroulé au sein du laboratoire IBISC de l'Université d'Evry Val d'Essonne. Je tiens à remercier tous les responsables et le personnel du laboratoire et de l'UFR Science et Technologie de l'Université d'Evry Val d'Essonne.

Il est difficile d'exprimer en peu de mot tout ce que je dois à Sylvie Lelandais, ma directrice de thèse. Je la remercie beaucoup d'avoir accepté de diriger ma thèse et de m'avoir soutenu tout au long de ce travail. Grâce à ses idées, ses conseils, ses critiques, sa confiance et son côté humain j'ai pu avancer dans le bon sens.

Je tiens à remercier aussi Naima Aitoufroukh pour son encadrement et sa participation active dans la réalisation de cette thèse. Qu'elle trouve ici l'expression de mon respect et ma profonde reconnaissance.

Je remercie profondément Monsieur Patrick Bonnin, Professeur de l'Université Paris XIII d'avoir bien voulu accepter d'être rapporteur de ces travaux et d'avoir effectué une lecture minutieuse de ce mémoire. J'ai beaucoup apprécié sa grande compétence dans le domaine de la robotique mobile et les applications temps réel.

Je remercie vivement Monsieur Michel Devy, Directeur de recherche au LAAS/CNRS de Toulouse d'avoir accepté d'être rapporteur de ce travail et d'effectuer une lecture minutieuse de ce manuscrit.

Je tiens à remercier aussi Monsieur Hichem Maaraf, Professeur à l'Université d'Evry Val d'Essonne pour ses encouragements et d'avoir accepté d'examiner mon travail

J'adresse mes vifs remerciements à Madame Odile Horn, MCF à l'Université de Metz et Monsieur Christian Barat, MCF à l'Université de Nice pour avoir accepté d'examiner mon mémoire de thèse.

Je remercie l'ensemble des chercheurs du laboratoire IBISC et surtout Said Mammar et Christophe Montagne, et l'ensemble des doctorants que j'ai côtoyé pendant ces quatre dernières années.

Je remercie infiniment ma femme Fatima, qui a cru en moi et qui m'a soutenu et encouragé tout au long de cette thèse et dans la vie. Merci énormément.

Je ne peux oublier de remercier mes parents et tous mes frères et sœurs pour leur générosité et leur soutien.

Enfin, mes derniers remerciements vont à tous ceux qui de près ou de loin, ont contribué à l'aboutissement de ce travail de recherche. Que chacun d'entre eux trouve ici l'expression de toute ma gratitude.

---

<b>Table des matières .....</b>	01
<b>Liste des figures.....</b>	05
<b>Liste des tableaux.....</b>	07
<b>Introduction générale .....</b>	09
<b>Chapitre I</b>	
<b>Etat de l'art : Localisation et saisie d'objets en robotique mobile .....</b>	13
1- Assistance aux personnes handicapées .....	14
1-1 Projets existants .....	14
1-1-1 Le robot infirmier (the Nursing Robot) .....	14
1-1-2 Le projet RAID (Robotic Workstation for the Disabled) .....	15
1-1-3 Le projet RAIL (A Robotic-Arm to Aid Independence) .....	15
1-2 Le projet ARPH .....	16
1-2-1 La base mobile .....	17
1-2-2 Le bras Manus .....	18
2- Localisation en robotique mobile .....	18
2-1 Généralités .....	18
2-1-1 Localisation relative.....	18
2-1-2 Localisation absolue .....	19
2-2 Localisation d'intérieur.....	21
2-2-1 Types de navigation.....	22
2-2-2 Localisation utilisant une carte .....	22
2-2-3 Approches de localisation par carte.....	24
2-3 Quelques méthodes d'appariement 2D/3D .....	28
2-4 Quelques méthodes de calcul de la pose .....	29
2-5 Le SLAM, Cartographie et localisation simultanées .....	30
3- La saisie d'objets .....	31
3-1 Les différentes configurations .....	32
3-1-1 Caméra fixe.....	32
3-1-2 Caméra mobile.....	32
3-2 Stratégies d'approche .....	35
5- Conclusion .....	37
<b>Chapitre II</b>	
<b>Synthèse de travaux antérieurs effectués au laboratoire IBISC.....</b>	39
1- Travaux sur la localisation monoculaire .....	40
1-1 Phase de calibration de la caméra .....	41
1-2 Le modèle de l'environnement 3D.....	42
1-3 Les segments image 2D.....	42
1-3-1 Détection des contours.....	43
1-3-2 Représentation des contours dans l'espace de Hough.....	43

---

1-4	La mise en correspondance .....	44
1-4-1	Contexte de la mise en correspondance .....	45
1-4-2	Algorithme de mise en correspondance .....	45
1-5	La fonction de transformation et de choix .....	46
1-5-1	Représentation par quaternion unitaire .....	47
1-5-2	La fonction d'erreur .....	48
1-5-3	La méthodes des quaternions .....	50
1-5-4	Choix de la meilleure pose .....	52
1-6	Résultats de localisation .....	53
1-6-1	Les tests .....	53
1-6-2	Les données réelles .....	54
1-7	Analyse des résultats et conclusion .....	55
2-	Reconnaissance de l'environnement par ultrason .....	57
2-1	Dispositif expérimental .....	57
2-2	Base de données .....	58
2-3	Extraction et sélection de paramètres .....	59
2-4	Méthodes de classification .....	59
2-4-1	Réseaux de neurones .....	60
2-4-2	Le neurone formel .....	60
2-4-3	Le réseau de neurones global .....	61
2-4-4	Le réseau de neurones spécialisé .....	61
2-4-5	Classification des objets simples .....	62
2-4-6	Classification des objets complexes .....	63
2-5	Analyse de scène .....	63
2-6	Analyse des résultats et conclusion .....	65
3-	Nos objectifs .....	66
<b>Chapitre III</b>		
<b>Localisation du robot mobile par vision et ultrasons .....</b>		<b>67</b>
1-	Introduction .....	68
2-	Organigramme de la localisation .....	68
3-	Préparation des informations 2D et 3D .....	69
3-1	Création du modèle 3D enrichi .....	69
3-2	Informations extraites des ultrasons .....	70
3-3	Calibration de la caméra .....	72
3-4	Informations extraites de la vision .....	72
3-4-1	Détection des contours .....	72
3-4-2	Extraction des segments .....	75
3-4-3	Comparaison avec la méthode initiale .....	75
3-4-4	Exemple d'extraction de segments .....	76
4-	Appariement 2D/3D .....	76
4-1	Rappels de la méthode initiale .....	77
4-2	Amélioration 1 : Pré-combinaison .....	77
4-3	Amélioration 2 : Post-combinaison .....	79
4-4	Comparaison des trois approches .....	80
5-	Transformation repère caméra / repère modèle .....	80
5-1	Présentation de la rotation par les angles d'Euler .....	81
5-2	Méthode de Lowe .....	82

---

5-2-1 Fonction d'erreur .....	82
5-2-2 Formulation adaptée à la robotique mobile.....	86
5-2-3 Solution non linéaire .....	87
5-2-4 Solution linéaire .....	88
5-3 Comparaison des trois méthodes.....	89
<b>6- Choix de la meilleure solution .....</b>	<b>90</b>
6-1 Amélioration 1 : les trois meilleures hypothèses .....	91
6-2 Amélioration 2 : la fonction coût d'appariement.....	92
6-2-1 Distance entre deux segments dans l'espace .....	92
6-2-2 Distance entre deux groupes de segments.....	93
6-2-3 Comparaison des trois approches .....	94
<b>7- Résultats .....</b>	<b>95</b>
7-1 Environnements de travail.....	95
7-2 Réduction du nombre d'hypothèses .....	96
7-3 Réduction du temps de calcul de la transformation .....	99
7-4 Erreurs de localisation.....	100
7-5 Temps de calcul de la localisation.....	102
7-6 Résultats de localisation .....	103
7-7 Evaluation de la méthode de localisation.....	105
7-8 Comment augmenter la robustesse .....	106
7-8-1 Re-localisation.....	106
7-8-2 Manque de luminosité .....	107
7-8-3 Localisation avec obstacle.....	108
7-8-4 Evaluation de la localisation avec obstacle .....	111
<b>8- Conclusion.....</b>	<b>112</b>

**Chapitre IV****Stratégie d'approche d'objets par asservissement visuel et ultrasons . 114**

<b>1- Introduction .....</b>	<b>115</b>
<b>2- Asservissement visuel .....</b>	<b>116</b>
2-1 Asservissement 3D .....	116
2-2 Asservissement 2D .....	117
2-3 Notre méthode d'asservissement .....	118
2-3-1 Système de mesure.....	118
2-3-2 Calcul des corrections.....	120
2-3-3 Calcul de la largeur de l'objet .....	121
2-4 Stratégie d'approche .....	122
2-4-1 Organigramme de l'approche d'objets .....	123
2-4-2 Calcul de la distance.....	124
<b>3- Première approche .....</b>	<b>125</b>
3-1 Objets utilisés .....	126
3-2 Segmentation par contours .....	126
3-3 Algorithme.....	127
3-4 Résultats .....	128
<b>4- Deuxième approche .....</b>	<b>130</b>
4-1 Objets utilisés .....	130
4-2 Traitement d'images .....	130
4-2-1 Segmentation par région .....	131

Table des matières

---

4-2-2	Seuillage par les couleurs .....	131
4-2-3	Exemple de segmentation .....	133
4-3	Classification des objets .....	134
4-3-1	Réseau de neurones spécialisé .....	134
4-3-2	Taux de reconnaissance .....	135
4-4	Algorithme.....	135
4-5	Résultats .....	136
5-	Troisième approche .....	139
5-1	Objets utilisés .....	139
5-2	Choix de l'objet à partir de sa largeur.....	140
5-3	Algorithme.....	141
5-4	Résultats .....	142
6-	Conclusion .....	145
<b>Conclusion et perspectives .....</b>		<b>147</b>
<b>Annexe A.....</b>		<b>151</b>
<b>Annexe B.....</b>		<b>156</b>
<b>Annexe C.....</b>		<b>160</b>
<b>Annexe D.....</b>		<b>164</b>
<b>Références bibliographiques.....</b>		<b>173</b>

## Liste des figures

### **Chapitre I**

Figure 1-1. Le robot infirmier .....	14
Figure 1-2. La station de travail RAID .....	15
Figure 1-3. Le robot mobile Handy1 du projet RAID.....	16
Figure 1-4. Le système d'assistance ARPH.....	16
Figure 1-5. La structure du robot mobile .....	17
Figure 1-6. Localisation dans un environnement d'intérieur.....	18
Figure 1-7. Les étapes de traitement dans le système de localisation FINALE.....	26
Figure 1-8. Schéma de l'algorithme d'identification des amers.....	27
Figure 1-9. Caméra fixe externe.....	32
Figure 1-10. Le manipulateur sous-marin Victor 6000.....	33
Figure 1-11. Caméra embarquée sur le préhenseur.....	33
Figure 1-12. Caméra embarquée sur un deuxième bras .....	34
Figure 1-13. Site expérimental du LIFIA .....	34
Figure 1-14. Saisie d'un objet polyédrique .....	35
Figure 1-15. Exemple de positionnement.....	36

### **Chapitre II**

Figure 2-1. Synoptique du travail effectué sur la localisation.....	41
Figure 2-2. Les éléments du repère géométrique d'une caméra .....	41
Figure 2-3. Représentation de l'environnement par des segments 3D .....	42
Figure 2-4. Représentation d'une droite dans l'espace de Hough.....	43
Figure 2-5. Extraction de segments image .....	44
Figure 2-6. Schéma bloc de la mise en correspondance.....	45
Figure 2-7. Projection d'un segment modèle sur l'image .....	49
Figure 2-8. Environnement utilisé pour les tests de l'appariement 2D / 3D.....	53
Figure 2-9. Exemple d'images réelles utilisées dans l'appariement .....	54
Figure 2-10. Exemple d'écho reçu.....	57
Figure 2-11. Les objets simples.....	58
Figure 2-12. Les objets complexes.....	58
Figure 2-13. Structure d'un neurone artificiel .....	60
Figure 2-14. Réseau spécialisé composé de 5 classes .....	62
Figure 2-15. Exemple d'environnement expérimental .....	64
Figure 2-16. Classification de plusieurs objets dans l'environnement .....	65

### **Chapitre III**

Figure 3-1. Schéma bloc de la méthode de localisation.....	68
Figure 3-2. Modèle enrichi par les ultrasons.....	70
Figure 3-3. Structure du réseau de neurones spécialisé .....	71
Figure 3-4. Différents types de contours .....	73
Figure 3-5. La fonction d'intensité et ses dérivées première et seconde .....	73
Figure 3-6. Images des trois poses utilisées dans les tests .....	75
Figure 3-7. Extraction de segments 2D .....	76
Figure 3-8. Appariement avec la pré-combinaison .....	78

---

Figure 3-9. Schéma de l'appariement avec une post-combinaison.....	79
Figure 3-10. Cercle d'erreur autour du robot .....	90
Figure 3-11. Projection des segments modèle transformés sur l'image .....	92
Figure 3-12. Mesure de distance dans l'espace 3D .....	93
Figure 3-13. Premier environnement de travail : Pièce du laboratoire .....	95
Figure 3-14. Deuxième environnement de travail : Couloir .....	96
Figure 3-15. Troisième environnement de travail : Salle de classe .....	96
Figure 3-16. Les trois photos de l'environnement présentant la pièce.....	97
Figure 3-17. Les deux photos de l'environnement présentant le couloir .....	97
Figure 3-18. Les deux localisations dans le couloir .....	103
Figure 3-19. Les trois localisations dans la pièce.....	104
Figure 3-20. Re-localisation dans la pièce et dans le couloir.....	107
Figure 3-21. Cinq photos de l'obstacle table .....	109
Figure 3-22. Cinq photos de l'obstacle armoire.....	110

## Chapitre IV

Figure 4-1. Schéma bloc de l'asservissement visuel 3D.....	117
Figure 4-2. Schéma bloc de l'asservissement visuel 2D.....	118
Figure 4-3. Différentes éléments du bras Manu .....	119
Figure 4-4. Système expérimental utilisé pour l'acquisition.....	120
Figure 4-5. Les deux corrections de la pince $C_x$ et $C_y$ .....	120
Figure 4-6. Largeur de trois types d'objets .....	122
Figure 4-7. Schéma bloc de l'asservissement visuel .....	123
Figure 4-8. Mesure de distance avec l'écho de l'objet.....	125
Figure 4-9. Localisation de l'objet balle dans l'image .....	126
Figure 4-10. Localisation de l'objet canette dans l'image .....	127
Figure 4-11. Algorithme de la première approche d'asservissement .....	128
Figure 4-12. Séparation des trois couleurs d'une image RGB.....	131
Figure 4-13. Les trois histogrammes de la couleur de l'objet orange.....	132
Figure 4-14. Exemple de sélection d'objets.....	133
Figure 4-15. Structure du réseau de neurones spécialisé .....	134
Figure 4-16. Algorithme de la deuxième approche d'asservissement .....	135
Figure 4-17. Sélection d'un objet orange à partir de sa largeur.....	140
Figure 4-18. Sélection d'un objet rouge à partir de sa largeur .....	141
Figure 4-19. Algorithme de la troisième approche d'asservissement.....	142

## Annexes

Figure A-1. Projection d'un segment 3D .....	152
Figure A-2. La contrainte unaire .....	153
Figure A-3. Les angles de vue.....	154
Figure A-4. La contrainte binaire .....	155
Figure C-1. Contour sous forme exponentielle avec $s = 0,5$ .....	161
Figure D-1. Modèle géométrique d'une caméra .....	166
Figure D-2. Principe de calibrage d'une caméra.....	169

## Liste des tableaux

### **Chapitre II**

Table 2-1. Exemple d'une table d'hypothèses.....	44
Table 2-2. Structure de la table de transformations .....	46
Table 2-3. Résultats de localisation sur l'environnement de test .....	54
Table 2-4. Résultats de localisation sur l'environnement réel.....	55
Table 2-5. Paramètres des cinq classes du réseau spécialisé.....	62
Table 2-6. Classification des objets simples avec le réseau spécialisé .....	63
Table 2-7. Taux de classification des objets complexes .....	63

### **Chapitre III**

Table 3-1. Modèle de l'environnement enrichi par les ultrasons .....	70
Table 3-2. Taux de reconnaissance des trois formes de l'environnement .....	71
Table 3-3. Les trois poses utilisées dans les tests.....	75
Table 3-4. Nombre de segments.....	76
Table 3-5. Comparaison des trois méthodes d'appariement .....	80
Table 3-6. Comparaison des trois méthodes de transformation .....	89
Table 3-7. Comparaison des trois méthodes de choix.....	94
Table 3-8. Dimensions des environnements de travail .....	96
Table 3-9. Coordonnées des poses initiales utilisées .....	97
Table 3-10. Gain en nombre d'hypothèses avec la méthode pré-combinaison .....	98
Table 3-11. Gain en nombre d'hypothèses avec la méthode post-combinaison.....	99
Table 3-12. Gain en temps de traitement pour les trois méthodes de transformation.....	99
Table 3-13. Les erreurs suivant $x$ , $y$ et $\Phi$ donnés par la méthode de la pose initiale .....	101
Table 3-14. Les erreurs suivant $x$ , $y$ et $\Phi$ donnés par la méthode des trois hypothèses..	101
Table 3-15. Les erreurs suivant $x$ , $y$ et $\Phi$ donnés par la méthode des distances .....	102
Table 3-16. Temps de calcul des étapes de la localisation pour les cinq poses de l'environnement .....	103
Table 3-17. Choix des blocs de traitement utilisés pour l'évaluation .....	105
Table 3-18. Evaluation de la méthode de localisation développée .....	105
Table 3-19. Poses et distances par rapport à l'obstacle table.....	108
Table 3-20. Résultats de localisation avec l'obstacle table .....	109
Table 3-21. Poses et distances par rapport à l'obstacle armoire.....	110
Table 3-22. Résultats de localisation avec l'obstacle armoire.....	111
Table 3-23. Evaluation de la localisation avec un obstacle.....	111

### **Chapitre IV**

Table 4-1. Coordonnées réelles et calculées des centres d'objets .....	127
Table 4-2. Calcul des paramètres pour la canette rouge .....	129
Table 4-3. Calcul des paramètres pour la balle orange .....	129
Table 4-4. Objets utilisés dans la deuxième approche .....	130
Table 4-5. Seuils bas et haut des couleurs utilisées.....	133
Table 4-6. Calcul des coordonnées du centre objet.....	134
Table 4-7. Taux de reconnaissance des deux formes d'objets.....	135
Table 4-8. Calcul des paramètres de l'asservissement pour une balle orange.....	137

Table 4-9. Calcul des paramètres de l'asservissement pour une canette rouge .....	138
Table 4-10. Objets utilisés dans la troisième approche .....	139
Table 4-11. Caractéristiques des deux objets orange de la scène .....	140
Table 4-12. Caractéristiques des deux objets rouges de la scène .....	141
Table 4-13. Paramètres de l'asservissement pour la peluche orange .....	143
Table 4-14. Paramètres de l'asservissement pour la peluche noire .....	144
Table 4-15. Paramètres de l'asservissement pour la canette orange / bleue .....	144
Table 4-16. Paramètres de l'asservissement pour la canette verte / rouge .....	145

## Introduction générale

Dans le domaine de la robotique mobile, le problème de la localisation automatique a fait couler beaucoup d'encre et donné lieu à de nombreuses publications. Quels capteurs utiliser pour la perception ? Faut-il ou non un modèle de l'environnement ? Quelle méthode d'appariement doit-on développer entre modèle et réalité ? Utilise-t-on des amères naturels ou artificiels ?

Un autre défi de la robotique mobile est la saisie automatique d'objets. Il faut d'abord localiser l'objet voulu dans l'environnement, s'en approcher et enfin le saisir. Ce genre d'action fait appel à des théories diverses telles la reconnaissance de formes et la commande. Les premiers travaux développés dans ce domaine utilisent le plus souvent l'aide d'une caméra et sont connus sous le terme d'asservissement visuel.

Le travail de recherche que nous présentons dans ce manuscrit se situe dans le contexte que nous venons d'exposer. Il a été effectué au sein du laboratoire IBISC (ex. Laboratoire des Systèmes Complexes : LSC) de l'Université d'Evry Val d'Essonne et entre dans le cadre du projet ARPH : Assistance Robotisée aux Personnes Handicapées.

Ce projet a pour objectif de proposer une assistance à des handicapés moteurs pour leur conférer une autonomie de vie à leur domicile. Dans cette optique, il est important de ne pas trop contraindre l'environnement de la personne, à la fois pour ne pas empiéter sur son intimité et également pour viser un coût financier minimum. Notre travail consiste dans le développement d'un robot mobile équipé d'un bras manipulateur.

La base mobile d'aide aux handicapés doit pouvoir se localiser de façon autonome pour effectuer une tâche de la vie quotidienne telle la saisie ou la manipulation d'objets comme chercher un verre, ouvrir une porte...

Le projet ARPH est composé de deux équipes de recherche. La première travaille sur la partie perception, et particulièrement la localisation du robot et la saisie d'objets simples par le bras manipulateur en utilisant plusieurs types de capteurs comme les caméras et les capteurs ultrasons. La deuxième équipe est concentrée sur la partie automatique, et notamment la commande du bras et le contrôle de la base mobile. Le travail présenté ici crée donc un lien perception/action dans ce projet.

La première partie de ma thèse est destinée au développement d'un système de localisation d'intérieur efficace et rapide pour le robot mobile, basé sur le principe de la localisation incrémentale. Cette étude s'est inspirée d'un travail de recherche antérieur effectué au sein du laboratoire entre 1999 et 2002, utilisant la vision monoculaire et l'appariement 3D/2D entre le modèle filaire qui représente l'environnement et les attributs de l'image.

Dans un premier temps, j'ai dû analyser cette première approche et les résultats obtenus, tout en effectuant une recherche bibliographique exhaustive sur les différents travaux de recherche menés dans le domaine de la localisation robotique.

Dans un deuxième temps, et après avoir cerné les limites théoriques et techniques de la méthode proposée, j'ai pu apporter plusieurs améliorations à cette première approche en introduisant de nouvelles stratégies et en utilisant d'autres types de capteurs. Parmi ces améliorations nous pouvons citer les cinq points suivants :

- 1- L'introduction de l'information ultrasonore permettant de reconnaître les formes (plan, coin et arête) présentes dans l'environnement. Cette reconnaissance est faite au moyen d'un réseau de neurones spécialisé,
- 2- L'enrichissement du modèle 3D initial par l'étiquetage de chacun des segments. cette étiquette décrit la forme d'appartenance du segment en terme de plan, coin ou arête et sera comparée aux informations citées dans le premier point afin de réduire le nombre d'hypothèses pendant l'étape d'appariement,
- 3- Le remplacement de la transformation basée sur les quaternions permettant le passage du repère caméra vers le repère global par une autre transformation beaucoup plus rapide en temps de calcul,
- 4- Le développement d'une nouvelle méthode plus performante pour la sélection de la meilleure hypothèse de localisation, basée sur le calcul d'une fonction de coût indépendant de l'information odométrique,
- 5- La généralisation de ce travail sur plusieurs environnements réalistes tels une salle de classe ou un couloir, ce qui a permis de valider les réelles possibilités de cette méthode, de mesurer ses performances et d'évaluer sa robustesse.

La deuxième partie de ma thèse traite le problème de la saisie d'objets de forme simples comme le cylindre ou la sphère au moyen de la pince du bras manipulateur monté sur la base mobile. Ainsi, notre travail a été consacré au développement d'une succession de stratégies de complexités croissantes, pour approcher l'objet à saisir avec précision en utilisant le principe de l'asservissement visuel.

Une caméra et un capteur ultrasonore équipent la pince du bras manipulateur Manus. La combinaison des informations reçues de ces deux capteurs permet de reconnaître l'objet à saisir en utilisant à nouveau un réseau de neurones spécialisé. Cet objet est ensuite centré dans le champ visuel de la caméra grâce au traitement d'images. En combinant les informations visuelles et ultrasonores, la pince approche l'objet jusqu'à la saisie.

Cette méthode est basée sur plusieurs éléments qui participent ensemble à l'élaboration des stratégies d'approche d'objets, ainsi :

- 1- Le système ultrasonore composé de deux capteurs ultrasonores et d'un réseau de neurones pour la reconnaissance de la forme de l'objet et calcule sa distance par rapport à la pince du robot mobile,
- 2- La caméra et les traitements d'images associés permettent de localiser l'objet dans l'image, puis de calculer les distances (horizontale et verticale) entre les centres de l'objet et de l'image, fournissant ainsi les paramètres de l'asservissement visuel,
- 3- La commande cartésienne permet au bras manipulateur de bouger suivant les axes  $\vec{x}$ ,  $\vec{y}$  et  $\vec{z}$  à la pince. Une boucle de commande visuelle transmet au bras les trois types de déplacements ( $\Delta X$ ,  $\Delta Y$  et  $\Delta Z$ ) à effectuer pour approcher l'objet et amener la pince en capacité de le saisir.

Pour valider cette approche, nous avons d'abord utilisé des scènes très simples comportant un seul objet de couleur unique. Puis nous avons généralisé notre approche en utilisant des scènes complexes proposant plusieurs objets de couleurs et de formes variées. Nous avons ainsi pu valider l'efficacité des stratégies proposées.

Les différentes parties que nous avons exposées rapidement dans cette introduction sont détaillées dans les quatre chapitres qui composent ce manuscrit.

Le chapitre I est destiné à la partie bibliographie où nous présentons quelques exemples de projets d'aide aux personnes handicapées moteur utilisant des robots mobiles. Nous exposons ensuite de manière plus détaillée le projet ARPH développé au laboratoire IBISC. Après, nous donnons un état de l'art général sur les méthodes de localisation utilisées dans la robotique mobile d'intérieur. Nous finissons ce chapitre par une présentation de l'état des recherches menées pour saisir des objets au moyen d'un bras manipulateur et les stratégies d'approche proposées dans ce cadre.

La première partie du chapitre II est consacrée à la méthode de localisation incrémentale développée dans un travail de recherche mené précédemment au laboratoire. Nous donnons ainsi les points forts et les points faibles de la méthode, les choix effectués et les résultats obtenus. La perception intelligente de l'environnement à partir de capteurs ultrasonores et les réseaux de neurones ont donné lieu à d'autres travaux menés également il y a quelques années au laboratoire. Dans la seconde partie du chapitre II nous présentons le dispositif expérimental utilisé, les méthodes de classification développées et les résultats de reconnaissance obtenus sur des objets simples et complexes.

Le troisième chapitre représente la partie la plus importante de la thèse car il traite la méthode de localisation développée pour le robot mobile. Cette méthode nécessite une initialisation fournie par un système odométrique installé sur les deux roues motrices du robot et combine deux types d'informations capteurs, la vision monoculaire et les données ultrasonores. Deux approches d'appariement 2D / 3D sont développées ainsi qu'une transformation repère caméra / repère modèle basée sur la méthode de Lowe et une fonction de choix de la meilleure solution de localisation. Les résultats de cette méthode sont donnés à la fin du chapitre III et cela sur plusieurs environnements, dont certains avec présence d'un obstacle.

Le dernier chapitre a une mission prospective et ne prétend pas à l'exhaustivité. Nous nous sommes donnés pour mission de proposer des stratégies d'asservissement visuel basées sur les informations extraites de la vision et les données ultrasonores dans le but d'approcher avec précision des objets simples. Nous commençons par le rappel des différentes méthodes d'asservissement visuel existantes puis, nous exposons notre approche globale ainsi que les outils utilisés. Par la suite, nous exposons les trois stratégies développées dans ce cadre. Le but de cette partie est de fournir au système de commande les paramètres de l'asservissement visuel nécessaires pour approcher l'objet : les deux déplacements horizontal et vertical de la pince du robot et la distance à l'objet.

## Chapitre I

### Etat de l'art : Localisation et saisie d'objets en robotique mobile

*Dans ce chapitre nous allons présenter un état de l'art sur les travaux de recherche menés dans les domaines directement reliés au travail de thèse faisant l'objet de ce document.*

*Nous nous attacherons tout d'abord à présenter des projets faisant appel à la robotique pour l'aide aux handicapés moteurs. Nous détaillerons ensuite le projet ARPH (Assistance Robotique aux Personnes Handicapées) mené au laboratoire IBISC.*

*Nous nous intéressons ensuite au problème de localisation d'un robot mobile dans un environnement d'intérieur. Nous évoquons ainsi, les différentes approches qui existent, et notamment la localisation incrémentale basée sur un modèle de l'environnement et un appariement 2D/3D.*

*La dernière partie de ce chapitre est consacrée à la saisie d'objets simples par un bras manipulateur embarqué sur un robot mobile. Nous citons les différentes configurations et les stratégies d'approche qui existent actuellement.*

## 1- Assistance aux personnes handicapées

L'aide aux personnes handicapées moteur se développe depuis quelques années en France et dans le monde. En effet, l'utilisation d'un robot mobile équipé d'un bras manipulateur est une solution très intéressante pour remplacer les mouvements des jambes et des bras de la personne handicapée et ainsi, répondre à ses besoins de la vie courante. D'un coté cela va permettre d'accéder aux différents endroits de l'environnement réel, ainsi, saisir et manipuler des objets. Et d'un autre coté, faciliter l'exécution de tâches, comme l'ouverture d'une porte ou d'une fenêtre, aider à manger et à boire,...etc.

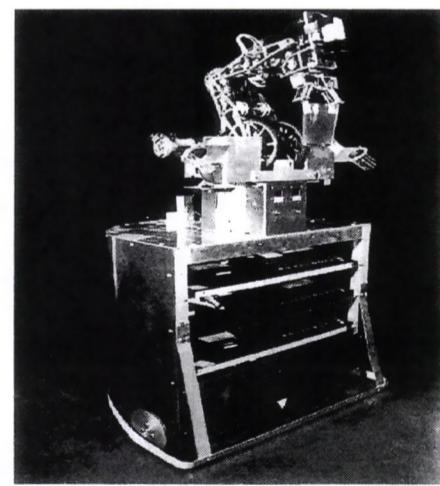
Depuis le début des années quatre-vingts plusieurs chercheurs se sont intéressés à développer des robots mobiles capables d'assister cette catégorie de personnes. De tels projets nécessitent la prise en compte des besoins réels de la personne handicapée, ainsi que le coût de l'assistance qui doit être suffisamment faible pour qu'elle puisse y accéder. Un autre point très important, l'interface de communication entre l'homme et la machine doit être simplifiée et facilement utilisable (écran tactile, commande vocale, casque et gants virtuels ...).

### 1-1 Projets existants

Il existe différents travaux dans le domaine de l'aide aux personnes handicapées. Dans la suite, nous allons citer trois importants projets, qui se basent principalement sur un robot équipé d'un bras manipulateur.

#### 1-1-1 Le robot infirmier (the Nursing Robot)

Le travail de J. Borenstein "Development of a Nursing Robot System" inclus le développement d'un système basé sur un robot mobile pour l'aide aux personnes handicapées moteur. Finalisé en 1986 [Bor.86], le robot infirmier était un des premiers robots mobiles opérationnels. Ce robot peut assurer de simples tâches, tel que chercher un verre d'eau, mettre en marche des appareils électriques ou remplacer une cassette dans un magnétoscope.



**Figure 1-1.** Le robot infirmier.

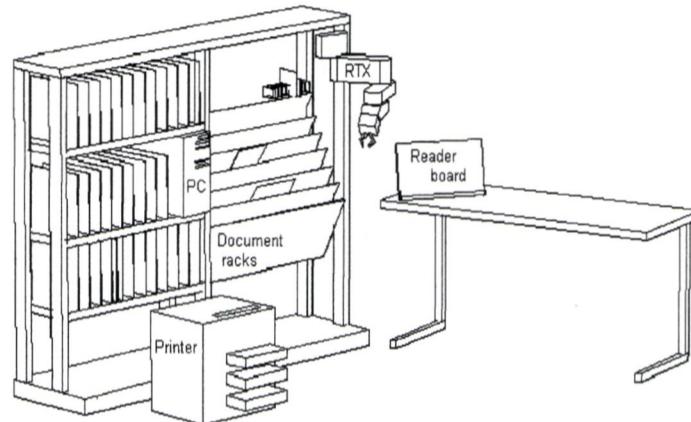
Le système sur la Figure.1-1 est composé d'un véhicule mobile, un bras robotique avec cinq degrés de liberté monté sur le véhicule, et un poste de contrôle automatisé.

Ce robot est équipé aussi de sept fonctions permettant d'effectuer les différentes tâches :

- 1- Deux capteurs à ultrasons pour la détection et l'évitement des obstacles,
- 2- Des micro-switches pour détecter les collisions avec les obstacles,
- 3- Des capteurs odométriques montés sur les roues pour avoir une position relative,
- 4- Trois sources de lumières pour mieux voir l'environnement,
- 5- Un capteur de force intégré sur la pince du bras manipulateur,
- 6- Une caméra vidéo permettant la détection des objets de cible,
- 7- Un système de reconnaissance vocal pour recevoir les ordres.

### 1-1-2 Le projet RAID (Robotic Workstation for the Disabled)

Le but du projet RAID est de développer un prototype de station de travail robotique pour personnes handicapées [Dal.92]. Ce projet fait partie du programme européen appelé "TIDE" pour (Technology for The Socio-economic Integration of disabled and Elderly people). Ce système robotique est prévu principalement pour l'usage professionnel dans un environnement bureautique (Figure.1-2).



**Figure 1-2.** La station de travail RAID.

Ce projet se base dans son fonctionnement sur la conception assistée par ordinateur (CAD). L'opérateur gère les différents mouvements du bras manipulateur pour effectuer plusieurs tâches comme :

- Manipulation d'un livre entre l'étagère et la table de lecteur,
- L'ouverture du livre à un point quelconque et rotation de page vers l'avant ou l'arrière,
- Transport du papier de l'imprimeur vers la table de lecteur,
- Manipulation d'un disque de l'étagère vers le PC,
- Servir les boissons.

Les essais techniques initiaux sur le robot en 1993 ont donné de très bons résultats en ce qui concerne la durée de cycle, la fiabilité et la fonctionnalité.

### 1-1-3 Le projet RAIL (A Robotic-Arm to Aid Independence)

Le but du projet RAIL est de développer un robot mobile appelé "Handy1" capable d'effectuer des tâches de la vie quotidienne (Figure.1-3). Parmi ces activités, donner à manger ou à boire, aider la personne handicapée à se laver, se raser, se rincer les dents et se maquiller.

La catégorie des handicapés qui pourraient bénéficier de ce système sont les paralysies cérébrales, les personnes atteintes de dystrophie musculaire ou de sclérose en plaques, et aussi les maladies dégénératives pour les personnes âgées.

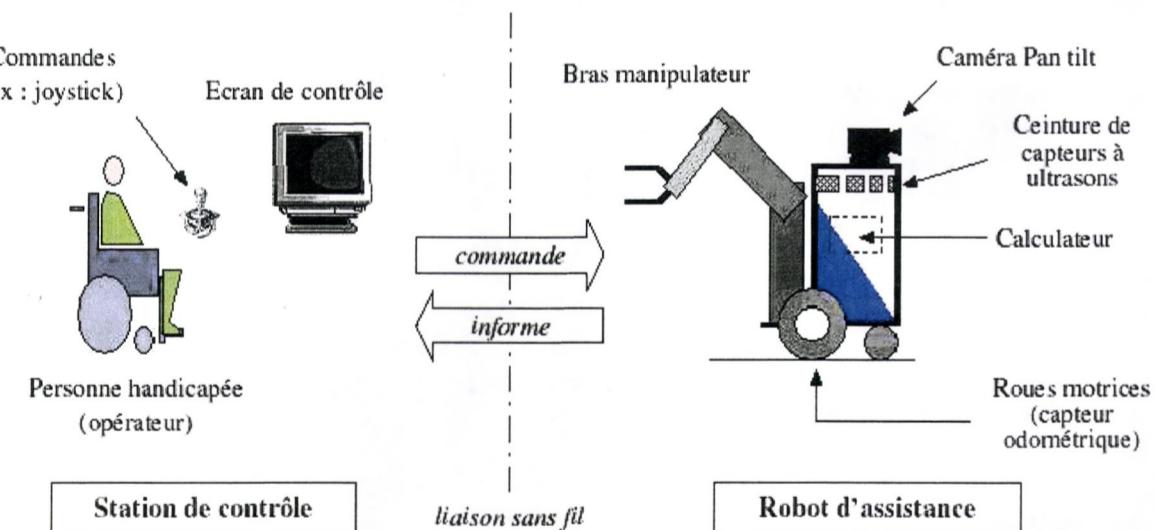


**Figure 1-3.** Le robot mobile Handy1 du projet RAIL.

Le robot "Handy1" développé en grande Bretagne par Mike Topping [Top.95] est le système de réadaptation le plus réussi au monde, avec plus de 120 unités vendues fin 1996. Son coût est le plus bas sur le marché.

## 1-2 Le projet ARPH

Créé en collaboration avec l'Association Française contre les Myopathies (AFM), le projet ARPH (Assistance Robotisée pour les personnes handicapées) consiste à élaborer un robot mobile capable d'aider et d'assister les personnes handicapées moteurs dans les différentes tâches de la vie quotidienne [Hop.01].



**Figure 1-4.** Le système d'assistance ARPH.

Le projet est constitué de deux parties principales, un robot mobile et une interface de commande composée d'un écran et un joystick (Figure.1-4). Le robot est à son tour composé de deux éléments, une base mobile pour les déplacements dans l'environnement intérieur, et

un bras manipulateur de type Manus [Eve.01] embarqué sur la base mobile et dans le but de saisir et de transporter des objets de différentes formes entre l'opérateur humain et le reste de l'environnement (cuisine, salle de bain, séjour...).

Le robot est doté de plusieurs capteurs pour l'interaction avec l'environnement. D'un point de vue sensoriel, deux types de capteurs sont utilisés et implémentés sur le robot, les capteurs proprioceptifs (odométrie) et les capteurs extéroceptifs (ultrason et caméra). Les informations acquises des différents capteurs sont traitées de manière individuelle ou fusionnelle, ainsi nous avons :

- 1- **L'odométrie** : Installée sur les deux roues motrices de la base mobile. Elle sert à donner une position relative globale du robot dans l'environnement,
- 2- **Les capteurs ultrasonores** : Installés sur la base mobile et le bras manipulateur. Leur rôle est multiple, mesurer les distances, éviter les obstacles, reconnaître la forme de l'environnement et des objets à saisir,
- 3- **Deux caméras** : Une, installée sur la base mobile pour la localisation du robot et l'autre sans fil, installée sur la pince du bras manipulateur pour reconnaître la forme et la couleur des objets à saisir.

Actuellement, il y a deux équipes de recherche au niveau du laboratoire IBISC qui travaillent en collaboration pour la réalisation et le développement de ce projet :

- La première équipe travaille sur la localisation du robot dans un environnement d'intérieur connu, l'évitement d'obstacles ainsi que des travaux sur l'asservissement visuel sont en cours pour le développement de la saisie d'objets,
- La deuxième équipe se focalise sur la partie commande du robot (base mobile et bras) et sur la réalisation d'une interface homme / machine pour piloter le robot à distance.

### 1-2-1 La base mobile

La base mobile représentée dans la Figure.1-5 est inspirée d'un fauteuil roulant électrique pour des raisons de fiabilité et de maintenance. La motorisation utilisée est du type DX, très répandue sur les fauteuils roulants électriques. La commande classique est réalisée par un joystick. Le système DX est basé sur le bus CAN et permet de piloter l'ensemble des périphériques du fauteuil (moteurs, klaxon, feux...).

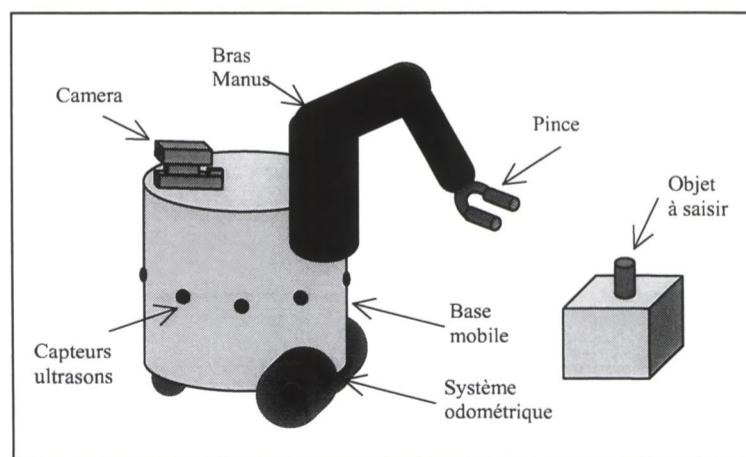


Figure 1-5. La structure du robot mobile.

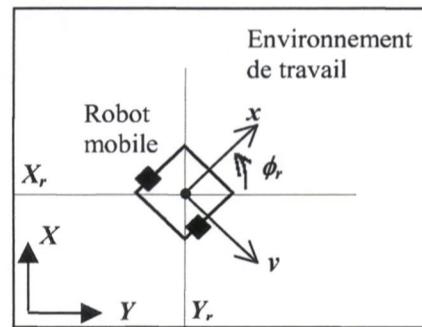
Les dimensions et la forme du bras Manus permettent l'accès à tous les points de l'espace autour du robot. Parmi les éléments du robot nous trouvons les différentes cartes électroniques d'acquisition et de traitement, les roues associées au système odométrique, la ceinture de capteurs ultrasons, la caméra, les batteries ainsi que l'ordinateur embarqué.

### 1-2-2 Le bras Manus

Développé aux Pays Bas par l'IRV Hoensbroek "Institute for Rehabilitation Development and Knowledge" en 1985, puis commercialisé depuis 1990 par la société néerlandaise Exact Dynamics b.v. Le bras Manus est spécialement conçu pour les personnes handicapées et vise à favoriser leur indépendance. C'est un bras robotisé de faible poids prévu initialement pour s'embarquer sur un fauteuil roulant électrique.

## 2- Localisation en robotique mobile

Un robot mobile doit répondre à trois questions de base, quelque que soit sa tâche et le nombre d'éléments qui le compose : "Où suis-je ?", "Où vais-je ?" et "Comment-y-aller ?". La première question concerne le problème de la localisation dans l'environnement. Les deux autres questions correspondent à la planification de trajectoire et à la navigation et leur exécution est fortement liée à la partie qui nous intéresse le plus : La localisation.



**Figure 1-6.** Localisation dans un environnement d'intérieur.

Le problème de localisation consiste à déterminer les coordonnées  $X_r$  et  $Y_r$  ainsi que l'orientation  $\phi_r$  du robot dans le repère de l'environnement "XY" et cela par rapport au repère relatif mobile "xy" lié au robot comme monté dans la Figure 1-6.

Dans la suite, nous allons donner des généralités sur les types de localisation, exposer la localisation d'intérieur, présenter quelques exemples de travaux sur les méthodes d'appariement et de calcul de la pose et finalement, nous donnons le principe de SLAM.

### 2-1 Généralités

Dans le cadre de la robotique mobile on peut citer deux types de localisation, relative et absolue. Chaque type comporte plusieurs systèmes basés sur des capteurs différents.

#### 2-1-1 Localisation relative

Dans ce type de localisation la position du robot est calculée d'une façon incrémentale par rapport à la position précédente. Un capteur proprioceptif permet de donner à chaque pas de

déplacement les variations  $\Delta X$ ,  $\Delta Y$  et  $\Delta \phi$  du robot et ainsi, calculer la position actuelle par addition des trois variations relatives. Le problème de ce type de localisation est l'erreur relative qui s'accumule au fur et à mesure des différents déplacements. Par contre, son avantage est le coût.

#### a) Système odométrique

Les systèmes odométriques mesurent les déplacements du robot en fonction des rotations effectives des roues motrices. L'idée fondamentale de ce système est l'intégration de l'incrément de la position calculée grâce à des encodeurs montés sur chaque roue par rapport au temps. Les déplacements des roues droite et gauche permettent de calculer la variation de l'orientation  $\Delta \phi$  ainsi que les variations de position  $\Delta X$  et  $\Delta Y$  [Bor.97].

Après quelques déplacements, une erreur non bornée apparaisse car une petite erreur d'orientation induit d'importante erreur en position. Les deux erreurs augmentent en fonction de la distance parcourue et à la fin d'une étape de navigation, il serait difficile de trouver la position exacte du robot. Ces erreurs peuvent être regroupées en deux catégories:

- 1- **Les erreurs systématiques** : Dues aux incertitudes dans les dimensions du robot et aux diamètres différents des deux roues),
- 2- **Les erreurs non systématiques** : Qui résultent de l'interaction du robot avec son environnement comme les glissements, la nature de la surface et autre.

#### b) Système inertiel

La localisation inertie est basée sur la double intégration des composantes accélérométriques du robot, exprimée selon les trois axes orthogonaux du repère de référence lié au robot. Cette opération nécessite la détermination de l'attitude du robot tout au long de son mouvement car en navigation inertie, localisation et détermination de l'attitude sont fortement liées [Bar.95]. Les deux capteurs utilisés sont :

- 1- **Les accéléromètres** : Ils mesurent des accélérations absolues. Selon l'orientation du capteur par rapport à la verticale terrestre, la mesure d'accélération peut inclure éventuellement une composante gravitationnelle qui doit être séparée de la composante accélérométrique utile avant le processus de double intégration,
- 2- **Les gyromètres** : Ils mesurent les variations angulaires du robot autour de ses axes. L'attitude est ensuite déterminée au long de son mouvement par simple intégration des mesures gyrométriques.

Contrairement à l'odométrie, la localisation basée sur des mesures inertielles a pour principaux avantages d'être indépendante des paramètres géométriques du robot. Elle est par ailleurs, peu sensible aux conditions de roulement, notamment le glissement ou la présence d'obstacles ponctuels sur le sol. En revanche, un recalage régulier à l'aide d'un dispositif de localisation absolue est nécessaire en raison des phénomènes de dérive affectant les capteurs.

#### 2-1-2 Localisation absolue

Elle se repose généralement sur la perception d'amers (naturels ou artificiels) présents ou disposés dans l'environnement de navigation du robot mobile. Ce dernier est alors localisé

grâce à des capteurs extéroceptifs dans un repère global et fixe lié à l'environnement et indépendant de sa configuration initiale.

Les techniques de localisation absolues ne présentent pas des phénomènes de dérive comme dans le cas des techniques relatives. Par contre, le robot doit avoir dans son domaine de visibilité les amers sur lesquels s'appuie le processus de localisation, ce qui n'est pas toujours le cas (obstacles, obscurité...).

De telles techniques nécessitent parfois que l'environnement soit aménagé et équipé de balises. En outre, le traitement d'informations issues de ces capteurs est plus lourd par rapport aux capteurs proprioceptifs.

#### a) Localisation par repères

Selon la nature des repères utilisés, les approches de localisation par repère se basent généralement sur des instruments de mesures et des balises actives ou passives pour identifier la position d'un robot mobile.

- 1- **Localisation par compas magnétique** : Un compas magnétique permet de donner une orientation absolue en mesurant la composante horizontale du champ magnétique terrestre (Nord géographique affecté d'une petite déclinaison). Il est basé sur le principe suivant "tout corps qui tourne sur lui-même tend à maintenir son axe de rotation dans une direction donnée", ainsi, l'axe de rotation du compas est stabilisé dans la direction du Nord. Un système d'amortissement (volant d'inertie) conserve cette direction quels que soient les mouvements du robot. Très facile à utiliser sur un robot ou un véhicule mobile, mais présent in un inconvénient majeur, car le champ magnétique terrestre peut être altéré surtout au voisinage de lignes à haute tension ou par des structures métalliques opaques.
- 2- **Localisation par balises actives et passives** : Selon leur nature active ou passive, les balises émettent ou réfléchissent des signaux qui peuvent être de nature optique, ultrasonore ou encore électromagnétique. Le robot étant, par ailleurs, équipé du dispositif adéquat pour la réception et l'émission de ces genres de signaux. Parmi les balises actives, on peut citer les satellites GPS qui émettent dans le domaine des hyperfréquences, les balises acoustiques utilisées dans la navigation sous-marine autonome ou encore les balises lumineuses. Pour les balises passives on peut citer les cibles catadioptiques utilisées pour renvoyer la lumière laser issue d'un émetteur à faisceau tournant solidaire du robot ou encore les balises triédriques en métal utilisées pour réfléchir dans leur direction d'arrivée les signaux radar à modulation de fréquence.
- 3- **Localisation par modèle** : Dans ce type de localisation le robot se base sur une mise en correspondance entre les observations acquises en ligne (image, signal ultrasonique ou laser...) et un modèle 3D virtuel de son environnement construit d'une manière hors ligne ou pendant la navigation. Le modèle peut être aussi une carte 2D de l'environnement représentant l'occupation ou non de l'espace de travail. S'il y a un appariement positif entre le modèle et l'observation, alors la position et l'orientation du robot peuvent être calculées.

#### b) Localisation par calcul de la position et de l'orientation

Dans ce cas de situation, la localisation se base sur la résolution d'un ou de plusieurs algorithmes de calcul mathématiques, ayant pour entrée les données capteurs.

- 1- **La trilateration** : Le calcul de la position dans cette méthode se fait à partir de mesures de distances aux balises. L'information de distance est généralement déduite de mesure de temps de vol (temps d'aller / retour de l'onde émise). Le positionnement des balises peut être dans un plan en 2D ou dans l'espace 3D. Il est à noter qu'il est nécessaire de disposer d'autant de balises dans l'environnement que d'inconnues à déterminer. D'un point de vue géométrique, la trilateration revient dans le cas 2D à déterminer le lieu d'intersection des cercles centrés sur les balises et passant par le robot. Pour le cas 3D, des sphères remplacent les cercles en gardant toujours le même principe. Un minimum de trois balises autour du robot permet de calculer la position de ce dernier. Un rayon laser est considéré comme le meilleur outil pour mesurer ces distances avec assez de précision et de vitesse. Le système américain GPS et russe GLONASS, fonctionnent selon un tel principe avec toutefois la particularité d'utiliser des balises satellitaires mobiles.
- 2- **La triangulation** : Dans le cas d'une localisation 2D (deux inconnues de position  $x$  et  $y$  et un cap  $\phi$ ), la position du robot peut être calculée à partir de trois mesures angulaires orientées. Ces trois angles sont à déterminer entre l'axe de l'avancement du robot et trois balises dans son environnement de travail. Ces mesures peuvent être effectuées grâce à un capteur laser rotatif travaillant en émission / réception et embarqué sur le robot ou encore, une caméra panoramique ou un sonar.

Nous pouvons prendre pour exemple, le système "Lasernet" développé par Vaganay sur le robot VEGA [Vag.93]. Il est constitué d'un faisceau laser tournant, qui après réflexion sur des cibles catadioptriques disposées dans l'environnement, permet de déterminer l'angle de gisement de chacune des cibles par rapport à l'axe d'avancement du robot. La précision en mode statique est de quelques centimètres en position et autour d'un degré en cap. La portée moyenne d'un tel système est de 6 m, il est destiné plutôt à la localisation en environnement d'intérieur.

- 3- **Localisation à deux points** : Aldon et al. présentent dans [Ald.94] une autre technique de calcul de la position basée sur une caméra et un plan laser. Une coupe de l'environnement est effectuée par le laser et deux points  $R$  ( $X_R, Y_R$ ) et  $L$  ( $X_L, Y_L$ ) sont extraits de cette coupe et représentés dans le référentiel  $R_r$  lié au robot. Un autre repère  $R_{ij}$  est défini grâce à ces deux points et la position absolue du robot est déterminée après que les points  $R$  et  $L$  sont mis en correspondance avec des points de l'environnement. Cependant, cette méthode ne tient pas compte des incertitudes de mesures.

## 2-2 Localisation d'intérieur

A partir des travaux de Giralt et al. en 1979 [Gir.97], suivit par Moravec en 1981 [Mor.81] et Nilsson en 1984 [Nil.84], les systèmes de vision sont devenus un outil impératif pour la navigation et la localisation de robots mobiles.

Ces premiers systèmes développés utilisent principalement la géométrie de l'environnement et autres informations métriques pour piloter le processus de vision et améliorer l'auto localisation dans un environnement d'intérieur plus au moins connu (appartement, laboratoire, usine...).

Les différents environnements d'intérieur sont représentés par des modèles de CAO (Conception Assistée par Ordinateur) de complexité variable. Dans certains travaux ces modèles de CAO sont remplacés par des modèles simples comme les cartes d'occupation et

les cartes topologiques ou par des séquences d'images. Un appariement entre le modèle choisi et les informations sensorielles permet de localiser le robot dans l'environnement.

### 2-1-1 Types de navigation

Un robot mobile a besoin d'une stratégie de déplacement dans son environnement pour effectuer des tâches précises, c'est la navigation. D'après le travail de synthèse de G. N. DeSouza et al. sur la navigation robotique par vision [Des.02], il y a trois types de navigation qui sont :

- 1- **La navigation basée sur une carte** : Ce sont les systèmes qui dépendent des modèles géométriques ou des cartes topologiques données auparavant par l'utilisateur,
- 2- **La navigation basée sur la construction d'une carte** : Ce sont les systèmes qui utilisent des capteurs pour construire leur propre modèle géométrique ou topologique de l'environnement (le SLAM). Après, ils utilisent ces modèles pour se localiser et naviguer dans l'environnement,
- 3- **La navigation sans carte** : Ce sont les systèmes qui utilisent des représentations non explicites de l'environnement de navigation, comme la reconnaissance d'objets, le tracking automatique grâce à des informations visuelles comme le flot optique ou encore la navigation par apparence.

### 2-1-2 Localisation utilisant une carte

La localisation basée sur une carte consiste à fournir le robot mobile d'un modèle de son environnement de travail. Ces modèles peuvent être du type CAO très complexe ou juste des simples graphes d'interconnexions entre les différents éléments de l'environnement. Dans la suite, nous allons voir trois types de cartes développées pour faciliter la localisation robotique.

#### a) La grille d'occupation

Dans certains systèmes de localisation, la reconnaissance de l'environnement se fait grâce à des représentations en cellules (grilles), dans lesquelles chaque objet est représenté par une projection 2D de son volume sur le plan horizontal. Ce type de grilles est construit hors ligne grâce à un balayage ultrasonore ou en ligne pendant la navigation du robot. Ce type d'approche est appelé grille d'occupation [Mor.85].

Quelques années après, l'idée de la carte d'occupation a été améliorée par l'incorporation des champs de forces virtuelles ou VFF [Bor.89]. Le VVF est une carte d'occupation où chaque cellule occupée exerce une force répulsive vers le robot et la destination exerce une force attractive. A la fin, toutes les forces sont combinées en utilisant les vecteurs d'addition et de soustraction, et la force résultante permet au robot d'avoir sa position ainsi que la meilleure trajectoire à suivre pour atteindre un but.

Une autre version élaborée de la grille d'occupation est la S-Map [Kim.94], qui consiste à projeter l'espace 3D dans une carte 2D. Cela exige une analyse spatiale des objets 3D dans l'environnement et la projection de leurs surfaces dans une carte 2D.

Une autre élaboration consiste à l'incorporation de l'incertitude dans les grilles d'occupation pour interpréter les erreurs de mesure sur la position des objets dans l'espace et pour la quantification de ses coordonnées en cellules occupées ou libres.

Parmi les techniques utilisées pour exprimer l'incertitude dans ces grilles :

- 1- **Les grilles d'histogramme** [Bor.91] : Où la détection d'un objet conduit à incrémenter ou à décrémenter respectivement la valeur de la cellule sur la grille. La nouvelle direction du robot va dépendre alors de la largeur des pics sur la grille d'histogramme et du chemin menant à au but souhaité,
- 2- **Les opérateurs flous** [Ori.95] : Où chaque cellule contient une valeur floue qui représente l'occupation ou non de l'espace correspondant.

#### b) La TBM (cartes orientées topologie)

Ce travail qui a succédé à la grille d'occupation propose la réalisation d'une carte orientée topologie ou TBM (Topological Based Map) comme une nouvelle représentation de l'espace de travail d'un robot mobile. La TBM consiste à structurer l'espace dans l'environnement avec une notion topologique (connectivité et adjacence). Elle peut être dérivée d'une grille d'occupation construite à partir des données capteurs, comme elle peut être extraite directement du plan de l'architecte d'une construction.

E. Fabrizi et A. Saffiotti [Fab.00] proposent une nouvelle méthode pour extraire la TBM à partir des données capteurs. Elle commence par l'acquisition d'une carte initiale appelée carte brouillée par l'intermédiaire d'un ensemble de capteurs à ultrasons. Ensuite, et à partir de deux techniques de traitement d'images très connues (morphologie mathématique et ligne de partage des eaux).

L'algorithme améliore cette carte initiale et ajoute des informations et des précisions par mise à jour successif de la carte par l'information capteurs. A la fin, un graphe topologique modélisant l'espace de travail est construit. Ce graphe contient généralement trois types d'éléments :

- 1- Des noeuds représentent les composants connectés,
- 2- Des arcs représentent les relations d'adjacence entre ces composants,
- 3- Des informations métriques liées à chaque noeud pour donner les dimensions de l'espace, ainsi que la longueur de chaque arc.

La TBM fournit une nouvelle notion de la carte topologie classique grâce aux informations géométriques de l'environnement et les données métriques qui les accompagnent et qui sont très utiles pour une auto localisation robotique et la navigation autonome.

#### c) Localisation orientée vision

Dans ce cas, le système de localisation n'utilise pas une carte précise de l'environnement, mais uniquement un appariement entre une image acquise au cours de la navigation et un ensemble d'images stocké dans une base de données construite hors ligne (principe d'indexation d'images). S'il y a un appariement positif entre les primitives d'une image et ceux d'une autre image dans la base de données, une position du robot est estimée. Les primitives les plus utilisées dans ce type de localisation sont les couleurs et les segments de droites. La localisation basée sur la vision peut être divisée en quatre étapes [Bor.96] :

- 1- **Acquisition des informations capteurs** : Pour la navigation basée vision, ceci signifie l'acquisition et la numérisation de l'image caméra.

- 2- **Détection de références** : Habituellement, cela signifie l'extraction de contours, lissage, filtrage et la segmentation en régions à partir du niveau de gris, la couleur, la profondeur ou le mouvement.
- 3- **Appariement entre l'observation et l'espérance** : Dans cette étape, le système essaye de chercher et d'identifier la référence observée sur l'image dans la base de données en prenant en compte certains nombre de critères de mesure.
- 4- **Calcul de la position** : Une fois un appariement obtenu, le système calcul la position en fonction de la référence observée et sa position dans la base de données.

La troisième étape (appariement entre l'observation et la base de données) représente la partie la plus difficile du travail. Elle passe par des recherches contrariées par des connaissances à priori sur la position des références dans l'environnement. Plusieurs approches sur la localisation robotique basée sur la vision se diffèrent par rapport à la prise en compte de ces contraintes.

#### d) Carte orientée fusion multicapteurs

En parallèle avec les grilles d'occupation et les carte topologiques obtenues par un simple balayage de capteurs à ultrasons, et la localisation orientée vision, d'autres approches de construction de carte utilisent la fusion de plusieurs types de données capteurs, essentiellement la vision, les ultrasons et les capteurs odométriques.

Le travail de R. Chatila et de J-P. Laumond [Cha.85] combine deux types d'informations données par une caméra et un ensemble de capteurs ultrasonores, et permet de produire un modèle basé sur l'occupation de l'environnement. L'idée centrale est de fournir au robot, directement ou indirectement, une séquence de références (ou balises) à trouver dans l'environnement durant la tâche de navigation, et ainsi pouvoir se localiser par rapport à la position des ces références.

Les données ultrasonores permettent de construire une carte initiale de l'environnement. La mise à jour de cette carte se fait au fur et à mesure que le système de vision cherche et d'identifie des références sur les images acquises en les comparants à la liste de référence dans la base de données. A la fin de cette procédure, une carte contenant les références identifiées et l'environnement est construite. Le robot utilise cette carte pour estimer sa position par auto-localisation grâce à un appariement entre l'observation (référence dans l'image acquise) et l'espérance (les références se trouvant dans la base de données).

#### 2-2-3 Approches de localisation par carte

Il existe trois approches de localisation utilisant une carte 2D ou un modèle 3D de l'environnement. Ces approches se basent principalement sur deux éléments, la vision monoculaire ou stéréo et l'appariement entre les primitives de l'observation et les informations stockées dans la base de données construite hors ligne.

##### a) Localisation globale (absolue)

La position initiale et l'orientation du robot sont inconnues dans ce type de localisation comme lors de la mise en marche ou lors d'un dysfonctionnement. Dans ce genre de situation, le système de navigation fait un appariement entre l'observation et la base de données pour fournir une estimation globale du robot. À cause des incertitudes liées aux observations, il est

possible qu'un même ensemble d'observations donne plusieurs solutions de localisation. Les ambiguïtés peuvent être résolues par d'autres méthodes comme la triangulation déterministe, le filtrage de Kalman basé sur des mixtures de gaussiennes ou encore, l'utilisation des intervalles d'incertitude.

En 1988, Sugihara [Sug.88] a présenté un des travaux de base sur l'auto localisation des robots utilisant une caméra monoculaire. Dans son travail, une fois l'extraction de paramètres dans l'image est effectuée, le problème de la localisation est réduit à un ensemble de contraintes géométriques.

Après, en 1989, Krotkov [Kro.89] a étendu les travaux de Sugihara par la proposition d'un modèle où les observations incertaines sont aussi prises en considération.

Quelques années après, en 1993 Atiya et Hager [Ati.93] ont développé un algorithme temps réel d'auto-localisation basé sur la vision. Dans ce papier, les auteurs mentionnent le manque de traitement de l'erreur d'observation dans le travail de Sugihara et la négligence de détection des faux caractéristiques qui ne correspondent pas aux références connues dans les travaux de Krotkov. Atiya et Hager.

Ils proposent de représenter les erreurs des capteurs par une tolérance qui mène à un algorithme basé sur les ensembles pour résoudre le problème de l'appariement et ainsi calculer la position absolue du robot mobile pour la navigation d'intérieur. Nous nous référerons à ces approches par localisation absolue ou globale.

### b) Localisation incrémentale

Ce type de méthodes utilise le plus souvent une représentation géométrique de l'espace de travail. Elle suppose que la position et l'orientation du robot sont connues d'une manière approximative au début de la session de navigation, et dont le rôle du système de vision est de raffiner les coordonnées de position. Le choix final de la pose du robot se base globalement sur un ensemble de contraintes géométriques et sur la recherche de primitives entre le modèle et l'observation dans l'intervalle de position et d'orientation considérée au départ.

Dans de tels cas, une vue générale est prise au départ à partir de la position initiale donnée par un système odométrique. Un appariement entre les primitives de l'image et du modèle de l'environnement intervient après pour corriger cette approximation initiale, et donne la position réelle du robot. Le résultat est ainsi plus cohérent et précis.

Plusieurs exemples de travaux utilisant cette approche incrémentale combinée avec la stéréo vision pour définir la position d'un robot mobile. Le système de Kriegman et al. [Kri.89] utilise la stéréo vision pour la navigation et la construction de carte de l'environnement. Matthies et Shafer modélisent l'erreur de localisation grâce à un système de vision stéréo utilisé en conjonction avec un modèle CAO de l'espace [Mat.87].

Dans notre travail de thèse, nous nous intéressons plus aux systèmes utilisant une seule caméra (la vision monoculaire) et un modèle géométrique représentant l'environnement de travail. Les deux exemples suivants illustrent cette approche.

**1- Le système FINALE :** Le système FINALE (Figure.1-7) de Kosaka et Kak [Kos.92] est un exemple très intéressant pour représenter la localisation incrémentale. Il utilise une

représentation géométrique de l'espace et un modèle d'incertitude statistique dans la localisation d'un robot mobile. Nous pouvons le composer en quatre éléments clés :

- 1- Représentation de l'incertitude dans la position du robot ( $x, y, \phi$ ) par une distribution gaussienne et ainsi, caractériser l'incertitude chaque position du robot par une moyenne  $\bar{p}$  et une covariance  $\Sigma_p$ ,
- 2- Caractérisation et paramétrisation des mouvements exécutés en réponse aux commandes de translation et de rotation dans l'environnement,
- 3- Détermination des changements de l'incertitude de position du robot quand les commandes de translation et de rotation sont effectuées (mise à jour de la moyenne et de la covariance),
- 4- Conception d'un modèle basé sur le filtre de Kalman.

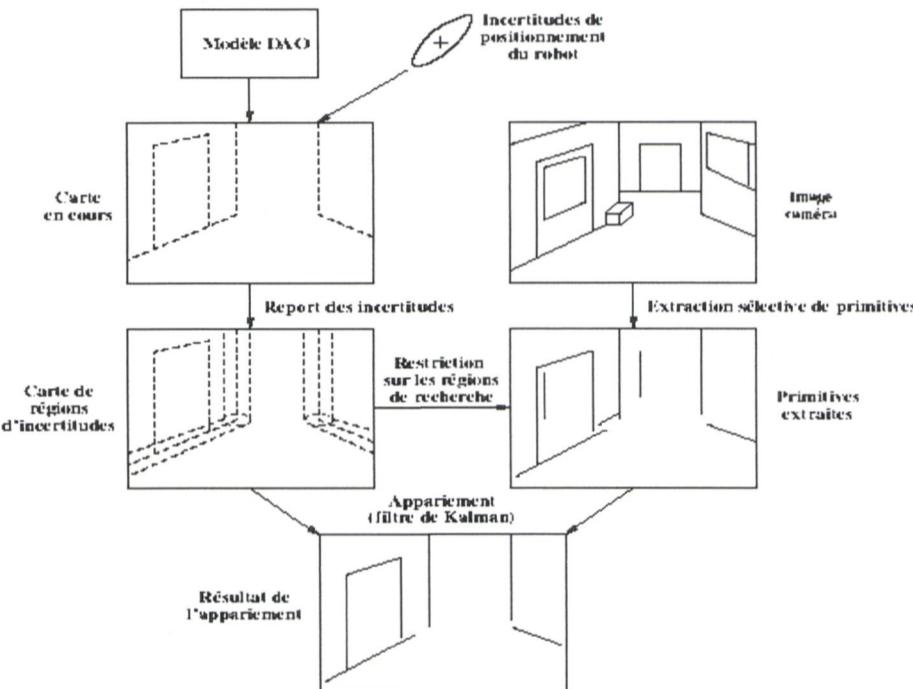


Figure 1-7. Les étapes de traitement dans le système de localisation FINALE.

Les résultats obtenus grâce au système FINALE montrent l'importance de la mise à jour de l'incertitude sur la position du robot après l'exécution d'un mouvement de translation ou de rotation. En générale, un autre point important qui n'a pas été cité, c'est l'incertitude des capteurs utilisés et les compensations qui existent entre l'incertitude des capteurs et la décision prise pour la navigation.

- 3- **Le système de Chenavier** : Un autre travail sur la localisation incrémentale dans un environnement connu est présenté par Chenavier [Che.95]. Ce système ne nécessite pas l'équipement du site en balises et la position du robot est estimée à chaque instant par l'odométrie. Un recalage périodique à l'aide d'une caméra montée sur le robot est effectué pour permettre au système de corriger la position.

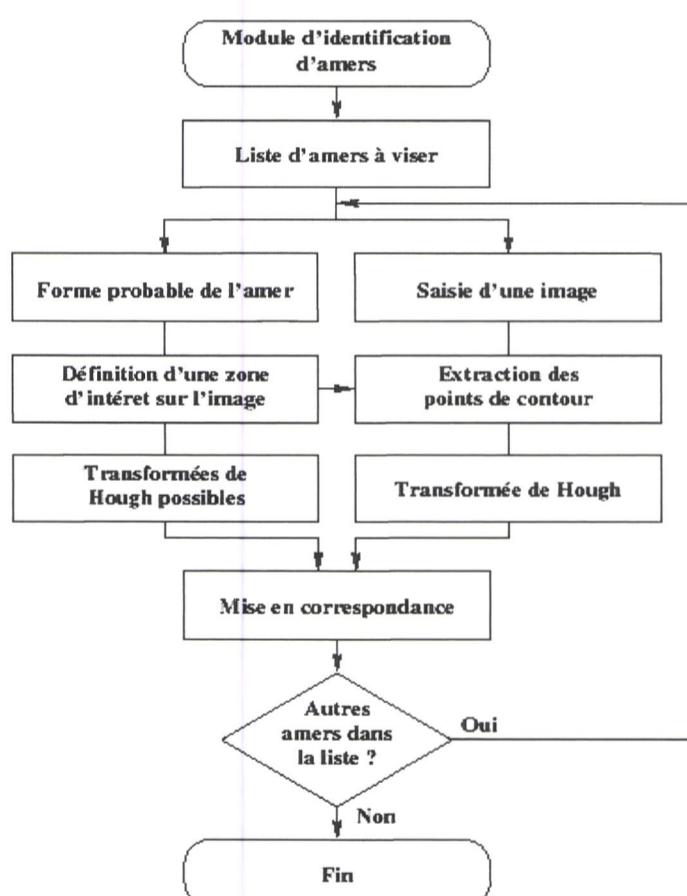
Le principe de base de cette méthode est le repérage par la caméra des objets de l'environnement jouant le rôle d'amer. Ces objets sont répertoriés dans une base de

données constituée au préalable. Un appariement entre l'image acquise par la caméra et les images dans la base de données permet la reconnaissance de l'objet dans l'image. Chaque objet dans la base correspond à une position connue dans le repère de l'environnement, ainsi, l'identification d'un ou de plusieurs objets permet la localisation du robot dans le repère de l'environnement.

D'un autre côté, l'identification d'un amer nécessite de connaître la l'incertitude sur la position de la caméra qui équipe le robot, ainsi, connaissant la relation entre repère robot et caméra le calcul de cette incertitude est immédiat.

La Figure.1-8 résume la technique utilisée pour identifier les amers, basée principalement sur l'utilisation des informations disponibles à priori. Ainsi, depuis la position courante du robot l'algorithme établit une liste d'amers à identifier, puis choisit les amers dont le centre est visible par la caméra.

Une fois cette étape achevée, l'algorithme tente d'identifier tour à tour chacun des amers de la liste. Grâce aux connaissances disponibles, le système synthétise l'image qu'il s'attend à voir avant de la mettre en correspondance avec l'image réellement acquise.



**Figure 1-8.** Schéma de l'algorithme d'identification des amers.

Compte tenu de l'incertitude sur la position, un ensemble d'images possibles doit être considéré. Pour ne pas perdre beaucoup de temps dans les calculs, l'auteur choisit d'utiliser les segments de l'image au lieu de prendre les régions et les contours (transformée de

Hough). Il sélectionne la meilleure image par rapport au calcul de la distance de Mahalanobis entre les segments de l'image acquise et ceux des images synthétisées.

Ainsi, pour conclure avec cette méthode, l'identification d'un ou de plusieurs amers sur une image permet de localiser le robot avec une exactitude de 7 cm en  $X$  et  $Y$  et de 1° selon le cap. Cette précision peut être accrue de 50 % en accumulant les mesures visuelles.

- 3- **Système NEURO-NAV** : Dans la suite, on va exposer une autre approche de localisation incrémentale complètement différente des deux autres précédemment présentées. Elle est basée sur les techniques Neuro-floue et utilise la plupart du temps un modèle topologique [Men.93]. Dans le système NEURO-NAV un graphe topologique représentant la disposition des pièces de l'environnement est utilisé pour piloter le processus de vision.

Une version plus performante a été développée dans [Pan.98] où le système de localisation est contrôlé par un contrôleur de supervision flou basé sur les règles de la logique floue (FUZZY-NAV).

#### c) Localisation basée sur la poursuite de primitives

Cette méthode de localisation se base sur la localisation de primitives dans l'environnement du robot puis le suivi de certaines d'entre elles. Sa mise en œuvre est possible à deux conditions, la connaissance de la position approximative du robot et la possibilité d'identifier et de suivre les primitives sur les images données par une caméra.

Les primitives utilisées peuvent être artificielles comme des codes barres ou des bandes réfléchissantes à différentes couleurs. Elles peuvent être aussi naturelles comme les portes, les fenêtres ou meubles. Dans la majorité de cas, la poursuite des primitives est effectuée par un appariement entre les primitives observées avec celles d'un modèle conçu d'avance. Ainsi, le système visuel pointe vers une des primitives (couleur ou forme) et calcule la position courante.

Hashima et al. dans [Has.97] ont présenté un système qui utilise des primitives naturelles et la technique de corrélation pour poursuivre des primitives consécutives et pour détecter des obstacles. La pose 3D d'une primitive est calculée par stéréo-vision et la position du robot est estimée à travers la comparaison entre cette pose et la position de la primitive sur la carte ou le modèle de l'environnement. Ce type de localisation est dérivé de la poursuite d'objets.

#### 2-3 Quelques méthodes d'appariement 2D/3D

Le processus d'appariement 2D/3D permet de trouver des correspondances entre les attributs de l'image et ceux du modèle de l'environnement. Si une correspondance est validée, il va donner une hypothèse de localisation. Plusieurs équipes de recherches se sont focalisés sur l'appariement 2D/3D.

Grimson et al. [Gri.90] ont proposé une méthode de recherche arborescente. L'arbre d'interprétation est constitué de tous les appariements possibles entre les primitives de l'image et du modèle. Un nœud de l'arbre contient une hypothèse de correspondance entre une primitive 2D et une primitive 3D. Le problème est de rechercher dans l'arbre la suite d'hypothèses qui permettent une mise en correspondance correcte.

Dans [Gri.84] Grimson utilise des contraintes géométriques pour élaguer l'arbre d'interprétation. Les contraintes géométriques les plus utilisées sont les contraintes unaire ou binaire.

Ayache et Faugeras [Aya.86] ont proposé l'utilisation d'un triplet de segments ayant un sommet commun. Cette technique permet de réduire considérablement le nombre d'appariement possible.

Lowe dans [Low.87] a proposé d'améliorer le traitement d'image. Il a établi des critères de regroupement des indices 2D afin d'obtenir des primitives consistantes qui facilitent l'appariement. Par ailleurs, un degré de probabilité est attribué à une hypothèse d'appariement afin de réduire l'espace des solutions. Enfin, une reconstruction 3D est effectuée afin de vérifier les hypothèses d'appariement.

Lamdan et Wolfson [Lam.88] ont introduit une méthode géométrique. Celle-ci consiste en la création hors ligne d'une table de hachage et à une reconnaissance en ligne. La table de hachage contient les différents aspects des objets du modèle 3D exprimés dans des bases différentes (une base peut être représentée par 2, 3 points ou arêtes). La reconnaissance en ligne est obtenue par mise en correspondance de points 2D, exprimés à leur tour dans différentes bases, avec le contenu de la table de hachage.

Pampagnin et Devy [Pam.88] ont introduit le graphe de compatibilité qui regroupe dans ses nœuds toutes les hypothèses d'appariement entre une chaîne de segments 2D et une face d'un modèle 3D et dont les arcs représentent les compatibilités entre hypothèses d'appariement.

Les travaux de Mallem et al. [Mal.96] consistent à développer deux méthodes d'appariement d'objets polyédriques basées sur la connaissance de leur modèle 3D et de leur image caméra 2D. La première est particulière, car elle suppose une connaissance géométrique à priori. Elle utilise la technique de la recherche arborescente et le raisonnement par l'évidence.

L'originalité de cette méthode est qu'elle prend en compte les contraintes physiques (liaisons) entre les objets. La deuxième méthode est plus générale, elle utilise les méthodes du graphe d'appariement et celle du hachage géométrique. La méthode du graphe d'appariement permet d'apparier des faces du modèle 3D avec des chaînes de segments dans une image.

La méthode du hachage géométrique permet de mettre en correspondance des sommets du modèle 3D avec des sommets de segments dans une image. L'originalité réside dans la l'utilisation de l'une de ces méthodes, selon la qualité du traitement de l'image, afin de permettre une reconnaissance de l'objet même si le traitement de l'image ne fournit pas suffisamment de chaînes cohérentes de segments.

#### 2-4 Quelques méthodes de calcul de la pose

Dans cette partie nous présentons quelques solutions pour résoudre le problème de calcul de la pose qui vient après le processus d'appariement. Ce problème est basé sur l'optimisation d'une fonction d'erreur exprimée en fonction des paramètres de localisation. Cette fonction exprime généralement la somme de distance entre les primitives vues sur l'image et les estimations obtenues en projetant les primitives correspondantes sur le modèle.

S. Ganapathy a présenté dans [Gan.84] une solution linéaire pour résoudre le problème de transformation dans le domaine de la vision robotique. Il a représenté la rotation par une matrice  $3 \times 3$  sans exprimer explicitement les contraintes d'orthonormalité de cette matrice. Parmi les inconvénients de cette méthode la sensibilité au bruits.

Yuan a proposé dans [Yua.89] de séparer la rotation et la translation et il a concentré plus ses efforts sur la détermination de la rotation. Les six contraintes d'orthonormalité de la matrice de rotation donnent naissance à six contraintes quadratiques. La solution commune à ces six contraintes est trouvée grâce à la méthode de descente de gradient de Newton.

Liu et al. dans [Liu.90] ont utilisé les angles d'Euler et une méthode itérative pour résoudre le problème de la rotation. D'abord, ils constatent que si on met en correspondance des droites et non des points, la rotation est séparée de la translation et une fois que l'on a déterminé rotation, la détermination de la translation devient un problème linéaire.

Dans ce même contexte que Liu, Kumar et Hanson ont examiné dans [Kum.89] deux techniques de minimisation. Une technique itérative qui linéarise la fonction d'erreur et une technique combinatoire qui permet d'éliminer des mesures aberrantes mais elle reste très lente, car elle considère tous les triplets possibles de correspondance.

## 2-5 Le SLAM, Cartographie et localisation simultanées

Quel que soit le type de robot mobile utilisé, les différentes approches classiques de navigation se base globalement sur trois étapes, la construction d'une carte de l'environnement, la localisation et la planification et l'exécution des mouvements. Depuis les années quatre vingt, une nouvelle technique de navigation appelé SLAM pour (Simultaneous Localization and Mapping) ou CML pour (Concurrent Mapping and Localization) est apparue.

Smith et Cheeseman présentent dans [Smi.85] un cadre statistique pour résoudre simultanément le problème de la localisation d'un robot mobile et de la construction d'une carte de l'environnement. Le plus de cette nouvelle méthode est de permettre la compensation des erreurs d'estimation de la position et de l'orientation à chaque nouvelle acquisition de données.

En 2002, Thrun donne un aperçu global sur la localisation et la cartographie dans la robotique mobile d'intérieur [Thr02]. Il décrit et compare les diverses techniques utilisées.

Chatila a présenté dans [Cha.03] un modèle global de fonctionnement d'un système SLAM appelé "Architecture LAAS" permettant au robot mobile de naviguer dans un environnement d'intérieur non connu d'avance. Elle est basée sur quatre éléments :

- 1- **Perception** : Acquisition d'un ensemble de données mono ou multicapteurs concernant l'environnement du travail et de navigation du robot mobile (mono et stéréo vision, laser, ultrasons, ...),
- 2- **Construction d'une carte locale** : A partir de la combinaison de ces données, une carte locale qui représente ce que le robot voit est construit. La carte peut être 2D ou 3D suivant le type de localisation choisi et les capteurs utilisés,
- 3- **Appariement et calcul de la pose** : Au fur et à mesure de ses déplacements, le robot construit aussi une carte globale de tout l'environnement par l'ajout de l'ensemble de cartes

locales. Un appariement entre cette carte et les observations permet de calculer la position et l'orientation du robot,

- 4- **Mettre à jour la carte globale** : Une fois la pose est connue et validée, l'algorithme mis à jour la carte globale de l'environnement par l'addition ou la suppression de nouveaux éléments.

### 3- La saisie d'objets

Pour un robot mobile manipulateur, la navigation représente la tâche la plus importante qui lui permet de se déplacer dans un environnement d'intérieur. L'autre tâche aussi importante, c'est la capacité du robot de saisir des objets simples ou complexes et de les transporter entre deux points de l'environnement ou entre un point et l'opérateur. L'utilité de la saisie d'objets se résume dans ces points :

- 1- Explorer l'environnement qui entoure le robot,
- 2- Manipuler des objets à distance,
- 3- Transporter des outils,
- 4- Apporter de l'assistance à l'opérateur.

Le retour d'informations de capteurs externes peut être une nécessité pour localiser et saisir un objet statique, notamment si la connaissance de la situation de l'objet dans son environnement et le positionnement absolu du robot ne sont pas suffisamment précis.

Le positionnement par rapport à un objet ou la saisie de cet objet présuppose une connaissance du système global :

- 1- Les capteurs de mesure intégrés,
- 2- Les paramètres de l'objet à saisir,
- 3- Les actionneurs utilisés.

La réalisation de la commande de saisie doit ensuite tenir compte des imprécisions inhérentes à toute modélisation. Il peut par exemple s'agir d'erreurs sur un modèle initial ou sur les hypothèses de bruits, d'évolution temporelle de paramètres supposés constants, ou de modifications importantes du comportement du système. Il est alors nécessaire d'adapter le modèle et la commande [Gha.04].

Dans la théorie on ne trouve pas de solution générale et unique pour la manipulation et la saisie d'objets. Celle-ci dépend de la structure du robot utilisé et du type de capteurs installés dessus. Un autre facteur très important c'est la grande variabilité de la forme et de la couleur des objets réels, nécessitant des capacités pour les reconnaître et les localiser en se basant sur des apprentissages à priori et sur des traitements spécifiques.

De la même façon que l'homme, le système robotique doit être muni d'une capacité sensorielle pour obtenir des informations sur son environnement lui permettant de réaliser chaque étape de la manipulation.

Dans la suite, nous allons donner les différentes configurations possibles entre un bras manipulateur et le système de vision. Puis on s'intéresse à la tâche de positionnement du bras par rapport à l'objet qu'on veut saisir et au tracking (suivi de cible). Il est important de noter que dans notre travail nous nous intéressons uniquement aux objets fixes, c'est le robot qui est

mobile et qui avance vers ces cibles pour les saisir grâce à une stratégie basée sur l'asservissement visuel.

### 3-1 Les différentes configurations

La vision monoculaire reste l'outil le plus utilisé dans cette application et qui apporte des informations riches et variées sur la nature de l'objet, sa forme et sa couleur. La technologie ultrasonore ou laser ajoute une information de distance qui manquante à la vision monoculaire. Toutefois dans la plupart des applications alliant caméra et robot on distingue généralement trois types de configurations entre le bras manipulateur et le système de vision utilisé représenté par une caméra.

#### 3-1-1 Caméra fixe

Dans ce cas une caméra fixe observe la scène dans sa globalité (Figure.1-9). La relation entre la vision et la commande est alors unilatérale dans le sens où les informations visuelles permettent de commander le robot mais les déplacements du robot n'interagissent pas sur la position de la caméra.



Figure 1-9. Caméra fixe externe.

Ce type de configuration présente l'avantage de voir la scène de l'extérieur, mais a l'inconvénient de ne proposer qu'une vue unique de la scène considérée. Nous citons les travaux d'Allen dans [All.91] qui développe cette configuration.

#### 3.1.2 Caméra mobile

La seconde possibilité exploite la mobilité de la caméra. Trois situations sont possibles :

##### a) Caméra embarquée sur une base mobile

La caméra suit les mouvements de la base mobile et filme les déplacements du bras manipulateur depuis le même angle de vue. Comme exemple de cette configuration, les travaux de Marchand et al. [Mar.01] qui présentent une méthode basée sur la vision pour commander le déplacement d'un bras manipulateur monté sur un ROV sous-marin (Remote Operating Véhicule), utilisé pour l'exploration des fonds marins.

Un système fonctionnant en boucle fermée et basé sur une approche d'asservissement visuel a été conçu pour réaliser la tâche de manipulation et la récupération des objets.

Victor 6000 est équipé de deux manipulateurs : un manipulateur appelé Maestro avec 6 degrés de liberté et un manipulateur appelé Sherpa de 4 degrés de liberté (Figure.1-10). Le manipulateur Sherpa n'est pas équipé avec des capteurs internes et il est commandé en boucle ouverte avec un joystick. Par conséquent, il n'y a aucune manière de mesurer les mouvements du manipulateur et les commandes seront imprécises si aucun capteur externe n'est utilisé.

Pour faire face à ce problème l'asservissement visuel s'est avéré être une méthode très efficace pour commander le manipulateur dans les environnements hostiles.

Le but n'est pas de commander le ROV lui-même, mais de commander le mouvement de son manipulateur non-équipé de capteurs Sherpa, en employant des informations fournies par une caméra montée sur le ROV et observant la pince du manipulateur. Il est donc important de commander le site et l'azimut de la caméra afin de s'assurer que le terminal du manipulateur reste dans le champ visuel de la caméra.

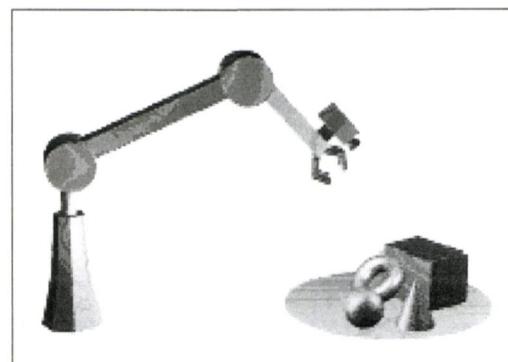


**Figure 1-10.** Le manipulateur sous-marin Victor 6000.

Pour effectuer une commande, le système calcule en ligne la position articulaire du manipulateur et réalise le déplacement indiqué en utilisant une loi de commande à partir de l'asservissement visuel.

#### b) Caméra embarquée sur l'effecteur du robot

Dans cette situation la caméra est montée sur une partie du bras manipulateur, ainsi, l'asservissement de la caméra va permettre l'asservissement du bras (Figure.1-11).

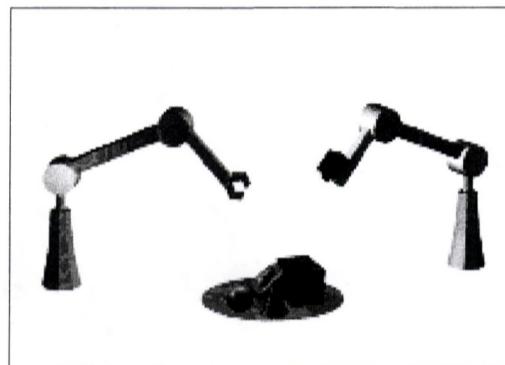


**Figure 1-11.** Caméra embarquée sur le préhenseur.

C'est la configuration que nous allons utiliser dans notre approche de saisie d'objets et qui sera développée plus en détail dans le chapitre IV. Matsikis dans [Mat.03] et Andreff dans [And.99] utilisent cette même configuration.

### c) Caméra embarquée sur un deuxième manipulateur

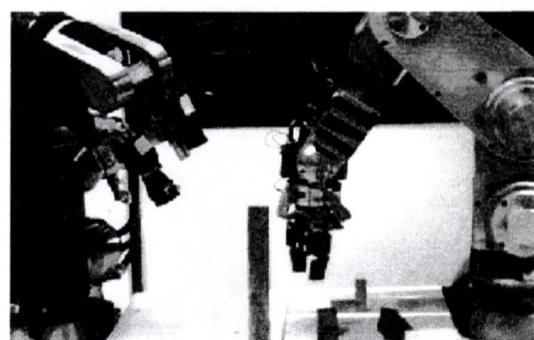
La troisième possibilité a les avantages des deux précédentes possibilités en embarquant la caméra sur un second manipulateur (Figure.1-12). La caméra observe alors la scène de l'extérieur tout en ayant la possibilité de se mouvoir. Ce genre d'approche est largement employé dans les problématiques de placement optimal de caméra.



**Figure 1-12.** Caméra embarquée sur un deuxième bras.

Parmi les travaux utilisant cette configuration, nous citons le laboratoire LIFIA (IMAG - INRIA Rhône-Alpes) qui dispose d'un système expérimental composé de deux robots manipulateurs (Figure.1-13).

Le premier robot effectue les tâches d'asservissement et le deuxième maintient une ou plusieurs caméras CCD pour visualiser la scène. L'utilisation d'une caméra embarquée sur le bras d'un robot permet de positionner, d'une manière automatique ou non, la caméra en un endroit optimal : la pince du robot et les objets d'intérêts doivent être dans le champ visuel de la caméra [Hor.98].



**Figure 1-13.** Site expérimental du LIFIA.

L'asservissement visuel nécessite l'utilisation des primitives 3D liées à la pince. Ces primitives peuvent être des amers naturels aussi bien qu'artificiels. Dans le cadre de ce travail, ces amers sont constitués de quatre cibles blanches imprimées sur une plaque noire qui est

collée à la pince du robot. Ainsi, les primitives géométriques sont les points 3D constitués par les centres de ces quatre cibles circulaires.

La tâche de localisation de la pince dans l'image devient simple. Cependant, il faut que les coordonnées 3D de ces quatre points soient connues dans le repère de la pince. Autrement dit, il faut que la transformation rigide plaque / pince soit connue. Ce problème est résolu grâce à la technique du calibrage caméra / pince où une caméra fixe observe les déplacements du système plaque / pince.

Bien que l'asservissement proposé dans ce travail puisse être appliqué à des différentes tâches robotiques, la saisie d'objets représente une application très intéressante car c'est un cas particulier des tâches nécessitant le positionnement relatif du robot par rapport à son environnement.

L'approche développée est divisée en deux grandes étapes :

- 1- **Etape hors ligne** : Dans cette étape, la planification de la saisie automatique doit tenir compte de plusieurs,
- 2- **Etape en ligne** : Cette étape comprend le calcul de la position 2D finale de la pince dans l'image et l'asservissement visuel du robot pour approcher et saisir l'objet. L'exemple de la Figure.1-14 illustre cette deuxième étape.



Figure 1-14. Saisie d'un objet polyédrique.

Le calcul de la position finale de la pince dans l'environnement de travail s'effectue suivant les trois étapes suivantes :

- 1- Extraction des contours de l'image et approximation par des segments de droite,
- 2- Prédiction du modèle 3D de l'objet grâce aux paramètres géométriques de la caméra,
- 3- Appariement 2D/3D afin de déterminer la matrice de projection entre l'objet et l'image.

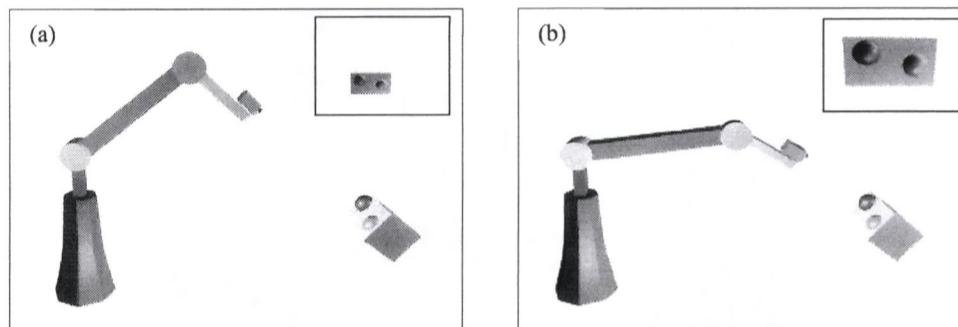
### 3-2 Stratégies de saisie

Le but d'une "stratégie de saisie", appelée aussi "tâche de positionnement" est d'amener le robot dans une situation d'équilibre donnée pour lui permettre d'effectuer d'autres opérations comme la saisie d'un objet, l'ouverture d'une porte ou autre. Ce type de tâche trouve aussi de nombreuses applications dans le secteur industriel. Cependant, les premières tâches de positionnement couplant un robot à une caméra se révélaient plus être des études de faisabilité plutôt que de réelles recherches [Cep.04].

Actuellement nous parlons beaucoup de l'asservissement visuel, c'est l'outil qui permet au robot manipulateur d'atteindre l'objet d'intérêt avec une grande précision et aussi de le suivre s'il est mobile. Grâce à l'image de l'objet acquise par la caméra, un ensemble de paramètres de déplacement en "X" et en "Y" sont calculés en pixel, transformés en valeurs métriques par le modèle géométrique de la caméra puis à la fin, transmis au bras robotique sous forme d'une commande de déplacement dans l'espace cartésien.

La Figure.1-15 montre un exemple de positionnement en deux temps d'un bras manipulateur par rapport à une cible fixe constituée de deux balles, une bleue et l'autre orange. Pour chaque position une image de l'objet est prise par la caméra embarquée sur la pince. Nous remarquons que la taille de l'objet sur les deux images change, elle augmente au fur et à mesure que le bras avance vers l'objet.

C'est à partir de "l'aire de l'objet" sur l'image que le système de vision détermine la distance à cet objet, puis il calcule les paramètres de l'asservissement visuel suivant les axes *X* et *Y*.



**Figure 1-15.** Exemple de positionnement, (a) : Etat initial et (b) Etat final.

Chronologiquement, on trouve les travaux de Shirai et Inoue en 1973 qui ont été les premiers à utiliser les données d'une caméra fixe pour améliorer le positionnement d'un préhenseur [Shi.73]. Le système fonctionne à une cadence 0,1 Hz et doit saisir un objet cubique afin de le déposer dans une boîte.

En 1983, Coulon et Nougaret utilisent des fenêtres d'intérêt afin de réduire la quantité d'information à traiter [Cou.83]. Leur manipulation consistait à localiser un objet cible dans l'image acquise, puis à réaliser un asservissement en *X* et en *Y* afin de s'approcher de l'objet visé et le saisir.

Toutefois, ce n'est qu'en 1984, avec les travaux de Lee E. Weiss que le contrôle de robot par retour visuel est réellement formalisé [Wei.84]. Dans ces travaux, Weiss couple une boucle d'asservissement visuel à un processus adaptatif lui permettant d'évaluer en ligne les relations non linéaires entre les primitives de l'image (forme, couleurs) et la position de l'effecteur dans le repère de travail. La particularité de l'approche de Weiss est de ne pas utiliser une boucle interne pour la commande des différents actionneurs.

Contrairement à Weiss, Feddema et Mitchell ont réussi à intégrer une simple caméra dans un système robotique pour commander la position et l'orientation relatives entre la pince du robot et une pièce mobile en temps réel [Fed.89]. Seules, des techniques de vision monoculaires sont employées en raison de la limitation du temps de réponse par rapport à la vitesse de traitement. Cette approche emploie le modèle géométrique de la pièce ainsi que celui de la

caméra et un nombre de primitives de l'image pour générer des commandes appropriées à la poursuite de l'objet en mouvement dans l'environnement.

Dans la même optique, Chaumette adapte dans ces travaux l'approche "fonction de tâche" au cas du capteur visuel et formalise ainsi "la commande référencée vision", [Chau.90]. Cette commande relativement performante trouve son originalité en considérant la caméra comme un capteur particulier associé à une tâche donnée et directement insérée dans la boucle de commande.

## 5- Conclusion

Dans cette première partie nous avons présenté quelques projets qui développent des systèmes pour l'assistance des personnes handicapées moteur. Dans notre laboratoire IBISC, le projet ARPH est considéré comme l'un des thèmes les plus importants qui ouvre plusieurs perspectives de recherche.

Le développement d'un tel robot mobile facilite l'interaction de la personne handicapée avec son environnement et le rend plus autonome. Les deux principaux objectifs de notre travail de thèse sont :

- 1- Intégrer dans le robot un système de localisation puissant, rapide et performant,
- 2- Apporter des informations aux gens qui font la commande et qui concerne les objets à saisir. Notamment la distance et les paramètres de l'asservissement visuel pour les déplacements horizontaux et verticaux de la pince.

Comme le projet ARPH est dédié aux personnes handicapées, toutes les tâches du robot et notamment la saisie d'objets, doivent respecter un temps de réponse un peu supérieur à la moyenne (2 à 3 secondes). La base mobile du robot se déplace à une vitesse de 0,01 m/s ce qui nécessite une localisation rapide mais pas nécessairement dans la norme temps réel.

La deuxième partie de ce chapitre est consacrée à la localisation de la base mobile et particulièrement la localisation d'intérieur. Nous avons évoqué aussi les différentes approches de l'appariement 2D / 3D et les méthodes de calcul de la pose du robot dans un environnement d'intérieur structuré. Pour finir cette partie, nous avons parlé brièvement de la méthode de localisation et de cartographie SLAM.

Dans mon travail sur la localisation, je me base sur une approche de localisation par carte qui est la méthode incrémentale. Elle est basée sur un appariement 2D / 3D entre un modèle de l'environnement et des attributs de l'image.

La troisième partie met la lumière sur la saisie d'objets simples par un bras manipulateur et les différentes configurations existantes. La plupart des travaux effectués dans ce cadre de recherche utilisent des techniques basées sur l'asservissement visuel (commande orientée vision). Nous avons présenté quelques stratégies utilisées pour approcher des objets simples dans l'environnement de navigation du robot.

La configuration qui nous intéresse le plus dans notre travail c'est le cas où la caméra est embarquée sur le préhenseur du robot. Les mouvements de la pince vont suivre les paramètres de l'asservissement visuel donnés par la caméra.

Dans le chapitre suivant, nous allons présenter deux travaux effectués au sein du laboratoire IBISC. Le premier se focalise sur la localisation incrémentale et utilise l'information odométrique et la vision monoculaire. Le deuxième étudie la reconnaissance des formes de l'environnement du robot par analyse d'échos ultrasonores.

Nous allons nous inspirer des ces travaux et les combiner pour développer dans la suite un système de localisation fiable et des nouvelles stratégies pour approcher les objets simples.

## Chapitre II

### Synthèse de travaux antérieurs effectués au laboratoire IBISC

*Deux travaux de recherche indépendants qui entrent dans le développement du projet ARPH ont été effectués au sein du laboratoire LSC.*

*Le premier travail concerne la localisation d'un robot mobile par vision monoculaire, et utilise un appariement entre un modèle fil de fer 3D et une image de l'environnement 2D. Cet appariement permet de donner une position et une orientation du robot dans son espace de travail.*

*Le deuxième travail permet d'avoir la forme de l'environnement du robot à partir d'un écho ultrasonore et d'un réseau de neurones. Certains paramètres de l'écho acquis obtenus par traitement de signal, comme l'énergie, la fréquence et l'enveloppe permettent au réseau de neurone de classifier l'obstacle rencontré en cinq familles différentes, le plan, le coin, l'arête, le cylindre et la forme complexe.*

*Nous allons présenter les techniques retenues dans ces deux approches, ce qui nous permettra de situer notre travail dans ce contexte et de fixer des objectifs à atteindre.*

## 1- Travaux sur la localisation monoculaire

Parmi les différentes approches de localisation absolue existantes, une technique très répandue est celle basée sur l'utilisation d'un modèle de l'environnement du robot (Map-Matching ou Model Based-Localization). Le principe général est que le robot perçoit en ligne et grâce à ses capteurs extéroceptifs un modèle local de son espace de travail qu'il met en rapport avec le modèle global, puis qu'il estime sa position et son orientation courante.

Comme nous nous l'avons vu dans le chapitre bibliographique, un modèle peut se présenter sous diverses formes, grilles d'occupation, primitives géométriques ou encore modèles topologiques. Le modèle utilisé le plus souvent dans le cas d'une localisation d'intérieur est le modèle géométrique donné par l'architecte, qui présente l'environnement par un ensemble de primitives géométriques comme la longueur, la largeur et la hauteur.

Ainsi, le type des primitives dépend des capteurs utilisés. En effet, les informations issues des capteurs et celles stockées dans la mémoire du système doivent être compatibles.

En supposant que le modèle géométrique interne de la caméra est connu et que le modèle de l'environnement est disponible, la localisation grâce à la vision monoculaire suit un schéma général composé de quatre étapes de base :

- 1- **Acquisition de l'information capteur** : Le robot saisit une image de son environnement de travail à partir de sa position courante,
- 2- **Détection de primitives** : Cette étape correspond aux opérations de traitement d'image telles que l'extraction de contours et la segmentation,
- 3- **Mise en correspondance des observations avec le modèle** : Recherche des appariements entre les primitives de l'image et celles du modèle si ces appariements existent,
- 4- **Calcul de la position et de l'orientation de la caméra** : Se fait grâce aux appariements établis dans l'étape précédente et grâce à un choix par rapport à la position initiale.

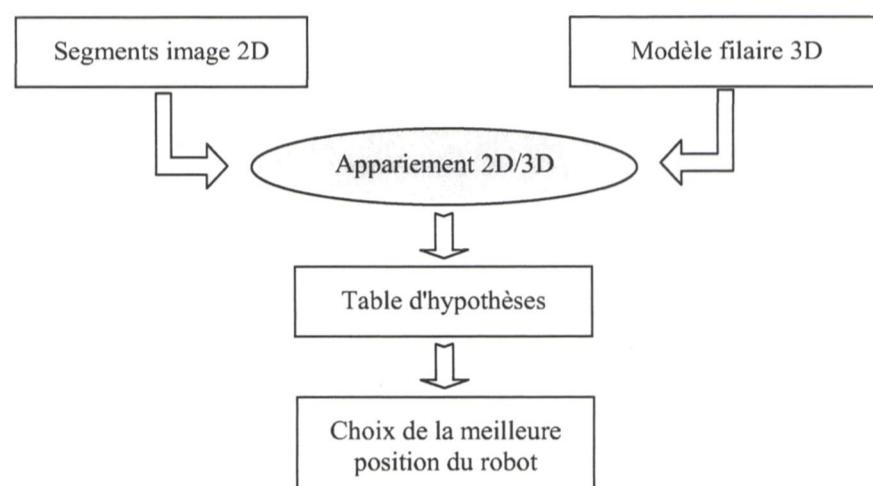
La succession de ces traitements, qui généralement sont complexes ou mettent en jeu de grandes quantités de données, rend les méthodes basées sur le modèle coûteuses en temps de calcul. Les problèmes de mise en œuvre en ligne sont donc un aspect majeur du problème de localisation. La plupart des systèmes adoptent une approche qui consiste à combiner ces calculs avec une localisation relative (généralement odométrique).

Cette dernière permet de maintenir à jour la connaissance de la position et de l'orientation avec un petit pas d'échantillonnage, mais sa précision se détériore au fil des déplacements à cause de l'accumulation d'erreurs. La performance du système de localisation final en terme de temps de calcul et de précision dépend donc de celle de chacune des étapes de calcul en ligne mais aussi des choix effectués quant aux modèles de la caméra et celui de l'environnement.

Le travail sur la localisation du robot mobile effectué par Ait Aider au laboratoire IBISC a pour but la mise en œuvre d'un système de localisation incrémentale, basé principalement sur la vision monoculaire [Aid.02]. Dans ce travail de recherche présenté par la Figure.2-1, le modèle de l'environnement est un modèle filaire qui remplace l'entourage du robot par des segments de droites 3D. Ce modèle respecte les dimensions données par le plan de l'architecte et il est construit d'une manière hors ligne.

Lors de la procédure de localisation, une caméra donne une image de l'environnement depuis une pose estimée par le système odométrique appelée pose initiale. Un traitement d'image permet d'obtenir les segments 2D de cette image et ainsi, une mise en correspondance entre ces segments image 2D et les segments du modèle 3D génère une table de correspondance appelée "table d'hypothèses de localisation".

Le choix de la meilleure solution dans cette table se fait par une comparaison de la pose donnée par chaque hypothèse avec la pose initiale donnée par le système odométrique. Le choix final de la position et de l'orientation correspond à la pose la plus proche de la valeur odométrique en  $x$ ,  $y$  et à l'angle d'orientation  $\Phi$ .

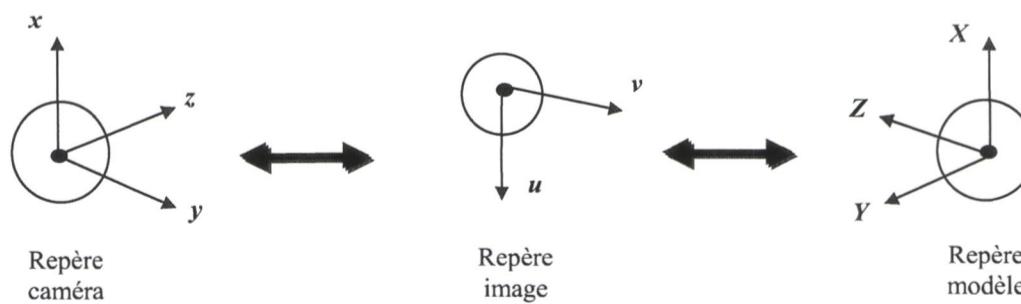


**Figure 2-1.** Synoptique du travail effectué sur la localisation.

L'objectif du travail présenté ici consiste à optimiser chacune des quatre étapes de ce traitement, pour qu'elles répondent au mieux aux exigences du contexte réel de la robotique mobile. Nous allons donner par la suite plus de détails sur chaque étape de ce travail, les choix effectués et les résultats obtenus.

### 1-1 Phase de calibration de la caméra

Cette étape est très importante en vision. Son objectif est de modéliser le capteur de vision qui est la caméra par un modèle géométrique composé de trois repères qui sont : le repère 3D de la caméra, le repère 2D de l'image et le repère 3D de l'objet se trouvant dans l'environnement (Figure 2-2).



**Figure 2-2.** Les éléments du repère géométrique d'une caméra.

Dans ce travail, la méthode utilisée pour calibrer la caméra se base sur l'approche de Zhang illustrée dans [Zha.98]. La calibration nous permet ainsi, d'avoir les paramètres des matrices de transformations entre ces trois repères qui sont, la matrice des paramètres intrinsèques et la matrice des paramètres extrinsèques.

### 1-2 Le modèle de l'environnement 3D

Le type de modèle adopté est un modèle à base de primitives géométriques détectables par des capteurs extéroceptifs tels que caméras, capteurs à ultrasons ou télémètres laser. Les exemples de primitives les plus courants sont les lignes droites, les coins, les plans et les points d'intérêts. Ces primitives ont des dispositions connues et peuvent être paramétrées par des attributs tels que la longueur et l'orientation d'un segment, la couleur des objets dans l'environnement ou encore la texture de régions.

Dans la plupart des applications orientées vision, les segments de droites sont les primitives les plus utilisées car d'une part ils sont bien adaptés à la représentation d'un environnement d'intérieur qualifié généralement de semi-structuré, et d'autre part, ils sont facilement détectables par traitement d'images.

Ainsi, les primitives utilisées dans le travail d'Ait Aider sont les segments de droites. Comme le montre la Figure 2-3, chaque élément de forme de l'environnement (coin, plan et arête) est remplacé par un segment numéroté de  $1$  jusqu'à  $N$ .

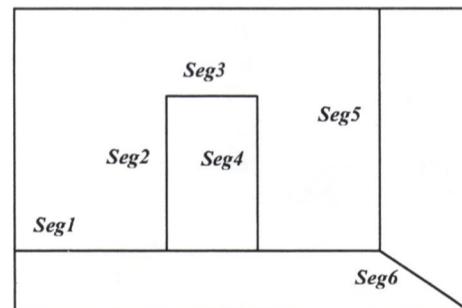


Figure 2-3. Représentation de l'environnement par des segments 3D.

Tous les segments qui représentent une partie de l'environnement sont regroupés dans une table qui contient le numéro du segment et les coordonnées des points de début et de fin.

Ce type de modèle filaire représente plusieurs avantages par rapport au modèle représenté par des surfaces ou des volumes de l'environnement. La phase de mise en correspondance entre les segments image 2D et les segments modèle 3D se fait plus rapidement. L'utilisation des segments est plus adaptée aux opérations de changement de repères entre la caméra, l'image et le modèle.

### 1-3 Les segments image 2D

Une image représentant une scène de l'environnement contient un très grand nombre de données, néanmoins, une grande partie ne comporte aucune information pertinente. Dans le cadre d'une application en robotique mobile, il est indispensable de pouvoir identifier les

zones de l'image qui véhiculent des informations importantes en fonction du type de primitives.

La segmentation de l'image est la base de tout système de vision. Elle a pour but de localiser et de caractériser par des attributs les zones d'intérêts. Il s'agit essentiellement des contours qui correspondent généralement aux frontières et arêtes des objets présents dans l'image et les régions qui correspondent aux surfaces. Dans la suite, nous présentons les méthodes nécessaires pour l'extraction des primitives géométriques, à savoir les segments de droites de l'image.

### 1-3-1 Détection des contours

La détection de contours se traduit par une opération de lissage suivi d'une dérivation de l'image par convolution avec un filtre optimal. La méthode utilisée dans cette partie consiste à appliquer un filtrage médian sur l'image d'origine. Ceci a pour effet d'éliminer le bruit de façon plus efficace qu'avec un filtre linéaire.

La méthode de Canny est ensuite utilisée pour la détection des contours [Can.86]. Elle est basée sur une approche gradient dans laquelle le gradient de l'image lissée est estimé grâce à la dérivée d'un filtre gaussien. La détection des extréma locaux est effectuée par un seuillage par hystérosis.

### 1-3-2 Représentation des contours dans l'espace de Hough

Après l'extraction des contours de l'image, c'est au tour des segments de droite. Une droite de l'image peut être représentée par la transformation de Hough [Gon.02] grâce à son orientation " $\rho$ " par rapport à l'axe horizontal du repère image et sa distance à l'origine " $d$ " comme l'illustre la Figure.2-4, avec le respect des deux contraintes :

$$1- \rho \in [-\pi/2, +\pi/2]$$

$$2- d \in [0, d_{\max}]$$

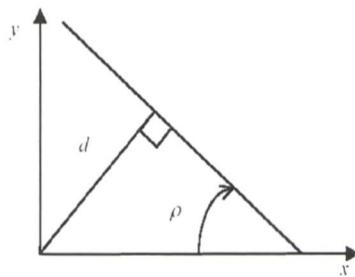


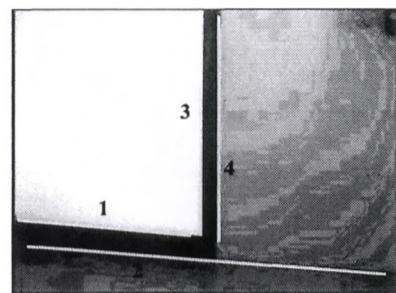
Figure 2-4. Représentation d'une droite dans l'espace de Hough.

La valeur  $d_{\max}$  représente la moitié de la longueur de la diagonale de l'image. Tout point de l'image  $(u, v)$  appartenant à une droite de paramètres  $(\rho, d)$  vérifie l'équation suivante :

$$d = u \cdot \cos(\rho) + v \cdot \sin(\rho) \quad (2-1)$$

L'espace paramétrique à deux dimensions défini par  $\rho$  et  $d$  est discrétré sous forme de tableau. Le nombre de cellules de ce tableau dépend des pas de discrétrisation  $\Delta\rho$  et  $\Delta d$  de chacune des deux dimensions.

La Figure.2-5 montre un exemple de l'extraction des segments 2D image par la succession deux méthodes exprimées auparavant, la détection des contours de l'image et la transformation de Hough qui donne les segments de droite.



**Figure 2-5.** Extraction de segments image.

Ces segments sont regroupés dans une table, où chaque segment est étiqueté par un numéro ainsi que les coordonnées du point de début et du point de fin. Nous ne détaillons pas cette partie sur laquelle nous ne sommes pas intervenus. Les détails de l'implémentation de la transformée de Hough sont donnés dans la thèse de Ait Aider [Aid.02].

#### 1-4 La mise en correspondance

Après la modélisation de l'environnement en segments 3D et l'extraction des segments images 2D, l'appariement 3D / 2D a pour objectif de créer des listes de correspondance entre ces deux groupes de segments, en se basant sur des contraintes géométriques. La validation d'une correspondance (appelée aussi hypothèse) nécessite au moins trois appariements entre trois segments modèle et trois segments image [Sha.99].

La table d'hypothèses est générée à la fin de cet appariement. Elle contient toutes les correspondances possibles entre les deux groupes de segments (2D / 3D). La Table.2-1 montre la structure d'une table d'hypothèses de dimension  $M \times N$ .

	$SM_1$	$SM_2$	$SM_3$	$SM_4$	$SM_5$	...	$SM_N$
$SI_1$	0	1	0	1	0	...	0
$SI_2$	1	0	0	0	0	...	1
$SI_3$	0	0	1	0	1	...	0
$\vdots$	0	1	0	0	0	...	0
$SI_M$	1	0	0	0	0	...	1

**Table 2-1.** Exemple d'une table d'hypothèses.

où  $SM_i$  représente un segment du modèle d'ordre " $i$ " et qui participe à l'appariement. De même,  $SI_i$  est un segment image qui participe aussi à l'appariement. Le chiffre " $1$ " dans la table d'hypothèses signifie que l'appariement entre le segment modèle et le segment image

est accepté, par contre "0" signifie que l'appariement est rejeté. Chaque hypothèse est composée d'au moins trois appariements entre les deux groupes de segments.

#### 1-4-1 Contexte de la mise en correspondance

La transformation entre le repère modèle et le repère caméra est très importante car elle permet de trouver la pose du robot dans l'environnement. Elle doit générer 6 variables (3 translations et 3 rotations), puisque la position et l'orientation de la caméra sont définies dans un espace 3D.

Trois de ces variables sont connues, les deux rotations  $\theta$  et  $\Psi$  qui sont nulles parce que la caméra est fixée sur le robot et la translation  $z$  qui représente la hauteur de la caméra. Le vecteur des inconnues est donc à trois composantes  $x$  et  $y$  pour la position et  $\Phi$  pour l'orientation.

#### 1-4-2 Algorithme de mise en correspondance

L'algorithme développé pour la recherche des bonnes correspondances entre l'ensemble des segments de l'image et l'ensemble des segments du modèle de l'environnement est basé sur une approche de recherche arborescente.

Cette première phase dans l'approche de localisation consiste à la génération d'hypothèses de correspondance cohérentes par l'élagage de l'arbre d'interprétation composé d'une part des segments du modèle filaire 3D et d'autre part, des segments image 2D (Figure.2-6).

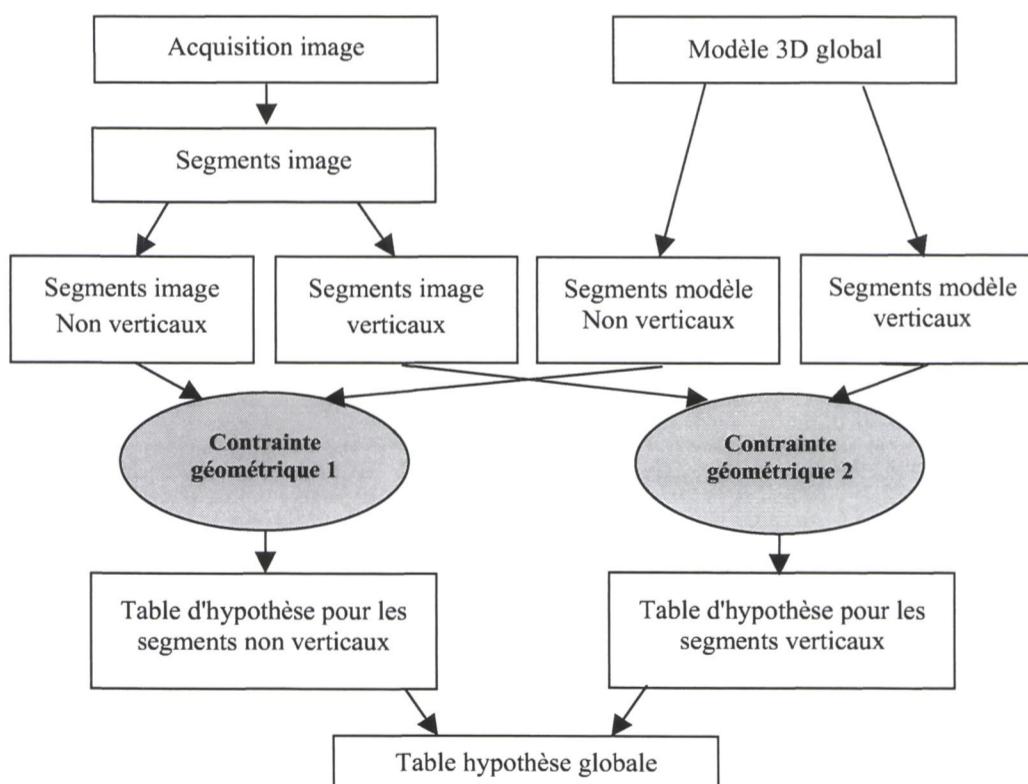


Figure 2-6. Schéma bloc de la mise en correspondance.

Ces deux ensembles de segments sont, au début du processus, subdivisés chacun en deux sous-ensembles de segments, les segments (image et modèle) non-verticaux et les segments verticaux. Ainsi, les deux arbres d'interprétation de dimensions inférieures vont remplacer le seul grand arbre de départ.

Deux contraintes géométriques sont ensuite définies. La première sert à vérifier les appariements entre les segments non-verticaux du premier arbre d'interprétation appelée première contrainte géométrique ou "contrainte unaire".

La deuxième quant à elle sert à vérifier les appariements entre les segments verticaux dans le deuxième arbre d'interprétation appelée deuxième contrainte géométrique ou "contrainte binaire".

Ces deux contraintes géométriques sont bien détaillées dans l'Annexe A. Elles sont directement applicables sur des appariements 2D / 3D. L'intérêt de ces deux contraintes géométriques est de limiter le nombre de correspondances générées (hypothèses de localisation).

### 1-5 Phase de transformation et de choix

La phase de transformation et de sélection consiste à calculer la transformation correspondant à chaque hypothèse retenue à l'issue de la mise en correspondance 2D / 3D à appliquer cette transformation sur les segments du modèle.

L'hypothèse la plus vraisemblable dans la table est celle qui représente une pose plus proche de la pose odométrique que nous appellerons par la suite "pose initiale".

La Table.2-2 donne la structure de la transformation repère caméra / repère image.

	$X_i$ calculé	$Y_i$ calculé	$\Phi_i$ calculée
<b>Hypothèse 1</b>	$x_1$	$y_1$	$\Phi_1$
<b>Hypothèse 2</b>	$x_2$	$y_2$	$\Phi_2$
$\vdots$	$\vdots$	$\vdots$	$\vdots$
<b>Hypothèse N</b>	$x_N$	$y_N$	$\Phi_N$

**Table 2-2.** Structure de la table de transformations.

Les valeurs de coordonnées  $X_i$ ,  $Y_i$  et  $\Phi_i$  calculées, présentent la position et l'orientation du robot pour une hypothèse d'appariement d'ordre " $i$ ". La transformation repère caméra / repère modèle utilisée se base sur le principe des quaternions unitaires. Elle est inspirée des travaux de Phong et Horaud [Pho.95].

Nous allons donner dans le paragraphe suivant comment présenter une rotation par des quaternions unitaires. Puis, nous allons donner l'aspect mathématique de cette méthode de transformation.

### 1-5-1 Présentation de la rotation par les quaternions

Une des présentations possibles de la matrice de rotation utilisée dans ce travail est celle utilisant les quaternions unitaires. Les propriétés de cet outil mathématique sont détaillés dans l'Annexe B.

Ils sont utilisés par la suite pour décrire une rotation, à condition de trouver une transformation qui change un quaternion purement imaginaire en un simple vecteur. Cette transformation préserve la longueur du vecteur transformé ainsi que le produit scalaire et le signe du produit vectoriel.

Si on suppose  $r$  est un quaternion purement imaginaire et  $q$  un quaternion unitaire, le quaternion purement imaginaire  $r'$  peut s'écrire de la façon suivante :

$$r' = q * r * \bar{q} \quad (2-2)$$

De plus nous avons :

$$q * r * q' = (Q(q)r) * \bar{q} = (W(q)^t Q(q)) r \quad (2-3)$$

avec  $Q(r)$  donnée par :

$$Q(r) = \begin{pmatrix} r_0 & -r_x & -r_y & -r_z \\ r_x & r_0 & -r_z & r_y \\ r_y & r_z & r_0 & -r_x \\ r_z & -r_y & r_x & r_0 \end{pmatrix} \quad (2-4)$$

et  $W(r)$  donnée par :

$$W(r) = \begin{pmatrix} r_0 & -r_x & -r_y & -r_z \\ r_x & r_0 & r_z & -r_y \\ r_y & -r_z & r_0 & r_x \\ r_z & r_y & -r_x & r_0 \end{pmatrix} \quad (2-5)$$

par ailleurs :

$$(-q) * r * (-\bar{q}) = q * r * \bar{q} \quad (2-6)$$

Cela implique que les deux quaternions unitaires  $q$  et  $(-q)$  représentent la même rotation. La matrice  $W(q)^t Q(q)$  est une matrice orthonormée donnée par :

$$W(q)^t Q(q) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & q_0^2 + q_x^2 - q_y^2 + q_z^2 & 2(q_x q_y - q_0 q_z) & 2(q_x q_z + q_0 q_y) \\ 0 & 2(q_x q_y + q_0 q_z) & q_0^2 - q_x^2 + q_y^2 - q_z^2 & 2(q_y q_z - q_0 q_x) \\ 0 & 2(q_x q_z - q_0 q_y) & 2(q_y q_z + q_0 q_x) & q_0^2 - q_x^2 - q_y^2 + q_z^2 \end{pmatrix} \quad (2-7)$$

Maintenant, on peut établir l'expression d'une rotation  $R$  en fonction d'un quaternion unitaire  $q$  comme suivant :

$$R = \begin{pmatrix} q_0^2 + q_x^2 - q_y^2 + q_z^2 & 2(q_x q_y - q_0 q_z) & 2(q_x q_z + q_0 q_y) \\ 2(q_x q_y + q_0 q_z) & q_0^2 - q_x^2 + q_y^2 - q_z^2 & 2(q_y q_z - q_0 q_x) \\ 2(q_x q_z - q_0 q_y) & 2(q_y q_z + q_0 q_x) & q_0^2 - q_x^2 - q_y^2 + q_z^2 \end{pmatrix} \quad (2-8)$$

La matrice  $W(q)^t Q(q)$  peut s'écrire également sous la forme suivante :

$$W(q)^t Q(q) = \begin{pmatrix} 1 & 0 \\ 0 & R \end{pmatrix} \quad (2-9)$$

L'expression donnée par l'équation (2-8) représente la matrice de rotation  $R$ . Elle va faciliter le calcul de la fonction d'erreur qui correspond à la transformation repère caméra / repère modèle obtenue par une correspondance de droites image et modèle.

### 1-5-2 La fonction d'erreur

Si nous considérons un modèle géométrique d'une caméra, point  $m$  de l'image qui a pour coordonnées  $(u_m, v_m)$  dans le repère image aura  $(x_m, y_m, z_m)$  comme coordonnées dans le repère caméra (Figure.2-7), ainsi nous avons :

$$\begin{cases} x_m = (u_m - u_0) / \alpha_u \\ y_m = (v_m - v_0) / \alpha_v \\ z_m = f \end{cases} \quad (2-10)$$

où  $u_0, v_0, \alpha_u$  et  $\alpha_v$  sont les paramètres intrinsèques de la caméra et  $f$  la distance focale. Un point du modèle  $M$  appartenant à un segment modèle a comme coordonnées dans le repère caméra  $(x, y, z)$ . Sa projection sur le plan image est ce même point  $m$ , alors on peut écrire :

$$\begin{cases} x_m = x / z \\ y_m = y / z \end{cases} \quad (2-11)$$

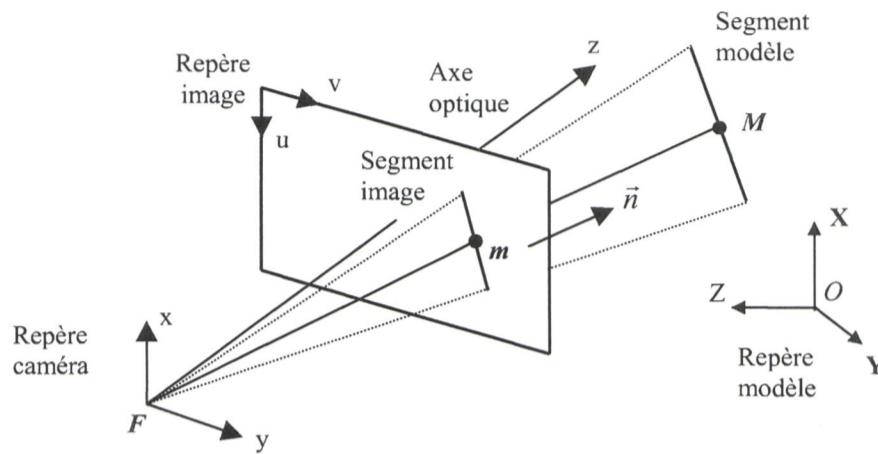


Figure 2-7. Projection d'un segment modèle sur l'image.

Le point  $m$  appartient à un segment de l'image dont l'équation s'écrit dans le repère de la caméra sous la forme :

$$ax_m + by_m + c = 0 \quad (2-12)$$

où  $a$ ,  $b$  et  $c$  sont les paramètres de la droite image. En remplaçant les deux formules de l'équations (2-11) dans la formule (2-12) nous obtenons :

$$ax + by + cz = n \cdot \overrightarrow{FM} = 0 \quad (2-13)$$

Ceci donne l'équation d'un plan contenant le centre de projection, la droite de l'image et le point  $M$ . Le vecteur  $n$  est le vecteur normal à ce plan. Si on décrit la droite du modèle contenant le point  $M$  par sa représentation paramétrique, où  $V_c$  est son vecteur directeur unitaire et  $P_c$  son vecteur position.

Cette droite peut également être exprimée dans le repère de la caméra comme :

$$\overrightarrow{FM} = P_c + \lambda V_c \quad (2-14)$$

La relation entre les paramètres de la droite dans deux repères différents qui sont le repère caméra et le repère modèle est donnée par :

$$\begin{cases} V_c = R V \\ P_c = RP + t \end{cases} \quad (2-15)$$

où  $R$  et  $t$  sont respectivement les matrices de rotation et de translation entre le repère modèle et le repère caméra. Elles contiennent les paramètres extrinsèques liés à la caméra et à l'objet du modèle qui contient la droite.

La contrainte de correspondance exprime le fait que la droite du modèle appartient au plan défini auparavant par l'équation (2-13), alors nous avons :

$$\begin{cases} n \cdot V_c = 0 \\ n \cdot P_c = 0 \end{cases} \quad (2-16)$$

Des équations (2-15) et (2-16) nous obtenons :

$$\begin{cases} n \cdot (R \cdot V) = 0 \\ n \cdot (R \cdot P + t) = 0 \end{cases} \quad (2-17)$$

Par conséquent, chaque correspondance "droite modèle / droite image" fournit deux contraintes. Le nombre de paramètres inconnus est de six (trois rotations et trois translations). La résolution de ce problème consiste à effectuer trois correspondances de droite. En général, trois droites peuvent être construites grâce à 3, 4, 5 ou 6 points.

La solution présentée dans cette partie est donc valable pour des correspondances de droites et de points et il faut avoir au moins trois correspondances "droite / droite" ou "point / point". La solution revient à minimiser la fonction d'erreur suivante obtenue à partir des deux équations de (2-17) :

$$f(R, t) = \sum_{i=1}^N (n_i (R \cdot V_i))^2 + (n_i (R \cdot P_i + t))^2 \quad (2-18)$$

Cette dernière équation représente la fonction d'erreur obtenue par correspondance de droites. Sa résolution représente la solution à la transformation repère caméra / repère modèle. Pour cela, plusieurs types d'approches existent.

La solution utilisée dans ce travail est basée sur la méthode de Phong-Horaud et les quaternions. Nous allons l'exposer rapidement dans le paragraphe suivant.

### 1-5-3 La méthodes des quaternions

L'équation (2-18) obtenue dans la section précédente exprime la fonction d'erreur. Elle peut être décomposée en deux parties, la rotation  $R$  et la translation  $t$  comme :

$$f(R, t) = \sum_{i=1}^N (n_i (R \cdot V_i))^2 + \sum_{i=1}^N (n_i (R \cdot P_i + t))^2 \quad (2-19)$$

Une fois la rotation connue, le problème de la translation  $t$  devient linéaire en remplaçant  $R$  dans la deuxième partie de l'équation (2-19). L'intérêt du couplage est que l'on remplace la résolution d'un système à six inconnues par la résolution de deux systèmes à trois inconnues chacun.

En effet, le coût calculatoire lors de l'optimisation numérique de fonction avec un nombre élevé d'inconnues est considérable.

Grâce à la présentation des rotations par des quaternions unitaires vue auparavant, Horaud transforme la fonction d'erreur obtenue (2-19) en une somme de carrés de contraintes quadratiques en utilisant les quaternions.

Les vecteurs de dimension trois vont être traités comme des quaternions purement imaginaires, ainsi la première contrainte de l'équation (2-17) peut s'écrire de la façon suivante :

$$n(RV) = n^t(W(r)^t Q(r)V) = r^t Q(n)^t W(n)r \quad (2-20)$$

Où  $r$  est le quaternion unitaire associé à la matrice de rotation  $R$ . La seconde contrainte de l'équation (2-17) devient comme la suite :

$$n(RP + t) = n^t(W(r)^t Q(r)P + t) = r^t Q(n)^t W(P)r + n^t t \quad (2-21)$$

En d'autres termes, chaque correspondance  $i$  entre une droite image et droite modèle fournit deux contraintes quadratiques de la forme suivante :

$$\begin{cases} r^t A_i r = 0 \\ r^t B_i r + n_i^t t = 0 \end{cases} \quad (2-22)$$

Avec la contrainte  $r^t r = 1$  et  $A_i, B_i$  étant deux matrices  $4 \times 4$  définies par :

$$\begin{cases} A_i = Q(n_i)^t W(V_i) \\ B_i = Q(n_i)^t W(P_i) \end{cases} \quad (2-23)$$

Ainsi, une nouvelle expression de la fonction d'erreur peut être donnée par l'équation :

$$f(r, t) = \sum_{i=1}^N ((r^t A_i r)^2 + (r^t B_i r + n_i^t t)^2) + \lambda(r^t r - 1)^2 \quad (2-24)$$

Où les paramètres à estimer sont les éléments du quaternion unitaire  $r$  et du vecteur de translation  $T$ . Le dernier terme de la fonction d'erreur est un terme de pénalisation avec  $\lambda$  positif. Nous pouvons faire la même remarque que pour le cas de la fonction d'erreur initiale donnée par (2-18), où la rotation et translation peuvent être estimées séparément.

La rotation est tout d'abord estimée par la fonction :

$$f(r) = \sum_{i=1}^N ((\hat{r}^t A_i \hat{r})^2 + \lambda(r^t r - 1)^2) \quad (2-25)$$

Une fois la rotation estimée, le calcul de la translation optimale est un problème linéaire. Ici  $\hat{r}$  représente le quaternion unitaire optimal qu'on cherche à déterminer, avec :

$$Q(n_i) = \begin{vmatrix} 0 & -n_{i1} & -n_{i2} & -n_{i3} \\ n_{i1} & 0 & -n_{i3} & n_{i2} \\ n_{i2} & n_{i3} & 0 & -n_{i1} \\ n_{i3} & -n_{i2} & n_{i1} & 0 \end{vmatrix} \quad (2-26)$$

et :

$$W(V_i) = \begin{vmatrix} 0 & -V_{i1} & -V_{i2} & -V_{i3} \\ V_{i1} & 0 & V_{i3} & -V_{i2} \\ V_{i2} & -V_{i3} & 0 & V_{i1} \\ V_{i3} & V_{i2} & -V_{i1} & 0 \end{vmatrix} \quad (2-27)$$

où :

- $n_i(n_{i1}, n_{i2}, n_{i3})$  : Vecteur normal au plan d'interprétation contenant le segment,
- $V_i(V_{i1}, V_{i2}, V_{i3})$  : Vecteur directeur d'un segment.

Une fois la rotation calculée, le critère à minimiser pour la translation est donnée par :

$$f(r, t) = \sum_{i=1}^N (r^t B_i r + n_i^t t)^2 + \lambda (r^t r - 1)^2 \quad (2-28)$$

avec :

- $P_i(P_{i1}, P_{i2}, P_{i3})$  : Vecteur position d'un segment.

Cette méthode d'analyse donne des résultats de localisation acceptables, par contre un temps d'exécution très important.

#### 1-5-4 Choix de la meilleure pose

Au début de la localisation du robot, le système odométrique installé sur les deux roues motrices fournit une approximation initiale de la pose du robot dans l'environnement par rapport à la distance parcourue dans l'étape précédente.

La pose donnée par l'odométrie est donnée par  $[X_0, Y_0, \Phi_0]$  et  $[X_i, Y_i, \Phi_i]$  représente une pose d'ordre "i" dans la table de transformations. La meilleure hypothèse de localisation est sélectionnée par la fonction coût  $F$  donnée par :

$$F = \text{Min} \left( \sum_{i=1}^N (X_i - X_0)^2 + (Y_i - Y_0)^2 + (\Phi_i - \Phi_0)^2 \right) \quad (2-29)$$

Pour une valeur de  $F$  minimale qui correspond à une position "i" dans la table de transformation, l'hypothèse numéro "i" est sélectionnée et considérée comme la meilleure solution. Cette fonction d'erreur choisie directement la meilleure pose du robot dans la table de transformations en calculant la somme des erreurs en  $X$ ,  $Y$  et  $\Phi$ .

La solution de cette méthode peut donner une petite erreur sur un ou deux axes, et une grande erreur sur le troisième axe, bien que la somme globale est la meilleure par rapport aux autres hypothèses de localisation. L'inconvénient majeur de cette méthode est qu'elle est dépendante de la pose initiale donnée par l'odométrie.

### 1-6 Résultats de localisation

Les résultats de localisation utilisant la méthode exposée se divisent en deux parties, localisation dans un environnement de test et localisation dans l'environnement réel. Pour chaque type de localisation, plusieurs paramètres sont calculés pour évaluer les résultats.

Parmi ces paramètres nous trouvons :

- L'erreur de translation  $\Delta T = (\Delta x + \Delta y)/2$  entre la position initiale et la position calculée,
- L'erreur de rotation  $\Delta \Phi$  entre le cap initial et l'orientation calculée,
- Le nombre de segments participant à l'appariement,
- Le nombre d'hypothèses de localisation générée.

#### 1-6-1 Les tests

Pour valider la partie tests, un objet simple composé de deux planches formant un coin vertical et des motifs carrés noir et blanc (Figure 2-8) a été utilisé. Le modèle filaire de cet objet est mis en correspondance avec des segments 2D des images de cet objet.

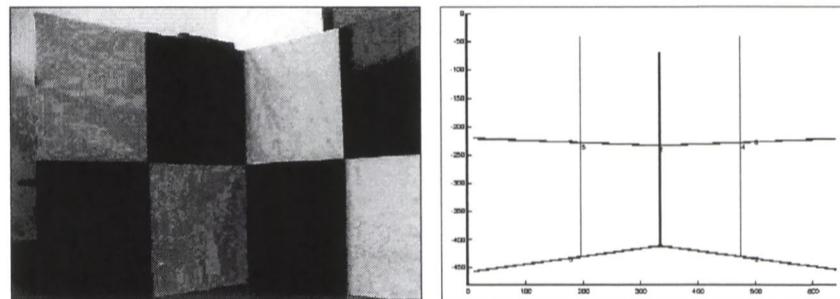


Figure 2-8. Environnement utilisé pour les tests de l'appariement 2D / 3D.

Les résultats obtenus pour cette partie de test sont donnés dans la Table 2-3. Neuf poses connues dans l'environnement de travail sont utilisées pour valider l'algorithme de localisation. Nous présentons dans cette table l'erreur de position  $\Delta T$  en mètre et l'erreur d'orientation  $\Delta \Phi$  en degré.

Nous présentons aussi le nombre de segments participants à l'appariement et le nombre d'hypothèses de localisation générée.

L'erreur de translation obtenue par cette méthode de localisation dans cet environnement de test varie entre 8 cm et 20 cm, et l'erreur de rotation varie entre 0.5° et 2.6°. Ces valeurs d'erreurs sont assez importantes par rapport à la dimension de l'environnement de test utilisé (1,5 x 1,5 x 1,5 mètre).

Le nombre de segments entrant dans l'appariement varie entre 4 et 6 segments et le nombre d'hypothèses de localisation varie entre 24 et 620 hypothèses. Il faut noter dans cette partie de

test que l'auteur n'a pas mesuré le temps de calcul nécessaire à l'exécution d'une étape de localisation.

	Erreur calculée		Nombre d'appariement	Nombre d'hypothèse
	$\Delta T (m)$	$\Delta \Phi (^{\circ})$		
<b>Pose 1</b>	0.16	2.1	5	24
<b>Pose 2</b>	0.11	1.8	4	65
<b>Pose 3</b>	0.20	1.5	5	63
<b>Pose 4</b>	0.13	2.5	6	120
<b>Pose 5</b>	0.17	2.0	5	65
<b>Pose 6</b>	0.09	2.3	4	180
<b>Pose 7</b>	0.13	2.6	4	121
<b>Pose 8</b>	0.08	0.5	4	580
<b>Pose 9</b>	0.20	1.8	5	620

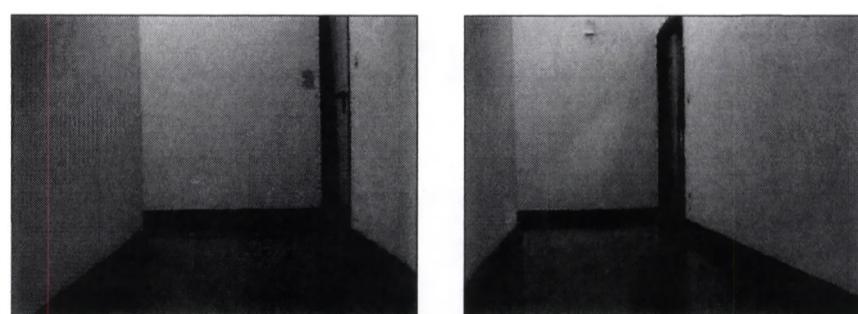
**Table 2-3.** Résultats de localisation sur l'environnement de test.

Pour bien évaluer cette méthode de localisation développée, nous allons présenter maintenant les résultats obtenus dans un environnement de travail réel.

### 1-6-2 Les données réelles

Ce travail à utiliser le couloir du laboratoire comme environnement. Neuf poses ont été utilisées pour valider la méthode de localisation sur des cas réels (Figure.2-9).

Le modèle 3D du couloir a été construit hors ligne et il contient 58 segments 3D représentants les différents coins et arêtes du couloir. Il est structuré de la même manière que la table donnée dans la Figure.2-2.



**Figure 2-9.** Exemple d'images réelles utilisées dans l'appariement.

Pour cette deuxième partie de résultats, l'auteur a calculé les mêmes paramètres que dans le cas précédent. Pour mieux évaluer cette méthode de différents points de vue, nous avons ajouté une boucle dans le programme principal qui calcule le temps de traitement correspondant à l'exécution d'une localisation.

Les erreurs de translation et de rotation, le nombre d'appariement, le nombre d'hypothèses générées ainsi que le temps de calcul nécessaire pour effectuer une localisation sont donnés dans la Table.2-4, et cela pour les neuf poses fixées au départ.

	Erreur calculée		Nombre d'appariement	Nombre d'hypothèse	Temps de calcul (s)
	$\Delta T$ (m)	$\Delta \Phi$ (°)			
<b>Pose 1</b>	0.60	11.28	6	275	35.52
<b>Pose 2</b>	0.24	6.88	6	68	12.80
<b>Pose 3</b>	*	*	*	Pas de résultat	*
<b>Pose 4</b>	*	*	*	Pas de résultat	*
<b>Pose 5</b>	*	*	*	Pas de résultat	*
<b>Pose 6</b>	*	*	*	Pas de résultat	*
<b>Pose 7</b>	*	*	*	Pas de résultat	*
<b>Pose 8</b>	*	*	*	Pas de résultat	*
<b>Pose 9</b>	0.81	3.07	6	42	11.59

**Table 2-4.** Résultats de localisation sur l'environnement réel.

La première remarque à faire concerne le taux de réussite du processus de mise en correspondance sur les neuf poses de l'environnement utilisées. En effet, sur ces neuf poses la localisation s'effectue uniquement pour 3 situations, c'est à dire un taux de réussite de 33 %. L'erreur de translation varie entre 24 et 81cm et l'erreur de rotation quant à elle varie entre 3° et 11°.

Ces valeurs sont très importantes par rapport à une localisation d'intérieur, pour laquelle on admet généralement une erreur de l'ordre de 10 % par rapport à la pose réelle. Le nombre d'appariement nécessaire pour valider une hypothèse est de six, ainsi que le nombre d'hypothèses générées varie entre 42 et 275 hypothèses.

Et pour finir, le temps d'exécution moyen pour une seule localisation est très important et très loin de la norme "temps réel". Quand la mise en correspondance s'effectue, ce temps varie entre 12 et 35 secondes.

### 1-7 Analyse des résultats et conclusion

Ait Aider a posé les bases théoriques d'une méthode de mise en correspondance entre segments modèle 3D et segments image 2D dans le but de localiser un robot mobile dans un environnement d'intérieur structuré. Le principe de cette méthode est basé sur une recherche arborescente par génération d'hypothèses de localisation.

L'objectif de ce travail était d'apporter une contribution à la localisation robotique d'intérieur grâce à la vision monoculaire.

La validation proposée par Ait Aider a été réalisée sur un environnement de test simple. Les résultats obtenus montrent que la méthode permet dans certaines configurations d'obtenir une précision de l'ordre de 3° en orientation et de 20 cm en position. Dans le cas d'un

environnement de travail réel, la valeur de ces erreurs augmente considérablement aussi bien pour la translation que pour la rotation.

Les temps de traitement moyens obtenus avec le processus de localisation développé sont considérables. La partie qui occupe la majorité de ce temps de calcul est sans doute la transformation des coordonnées de la pose du robot entre le repère caméra et le repère modèle.

Le choix de la meilleure pose utilise une comparaison entre les poses calculées dans la table de transformations et la pose initiale donnée par le système odométrique. Puisque l'erreur donnée par un capteur odométrique est non bornée, la dépendance de cette méthode à la pose initiale augmente les erreurs de position et d'orientation.

Les objectifs d'une partie du travail présenté ici consistent à reprendre cette méthode et à l'améliorer de manière à atteindre un bon niveau de résultats sur des environnements plus complexes que celui de test.

Parmi les éléments à prendre en compte pour atteindre cet objectif :

- 1- Travailler la qualité de l'estimation odométrique donnée au départ,
- 2- Améliorer la méthode de calibration de la caméra utilisée, et ainsi le modèle interne,
- 3- Améliorer le modèle filaire 3D de l'environnement et le rendre plus précis.

Mais dans l'objectif de rendre cette méthode de localisation opérationnelle, beaucoup de pistes sont à explorer afin d'améliorer notoirement les résultats obtenus et baisser le taux d'échec, et ainsi :

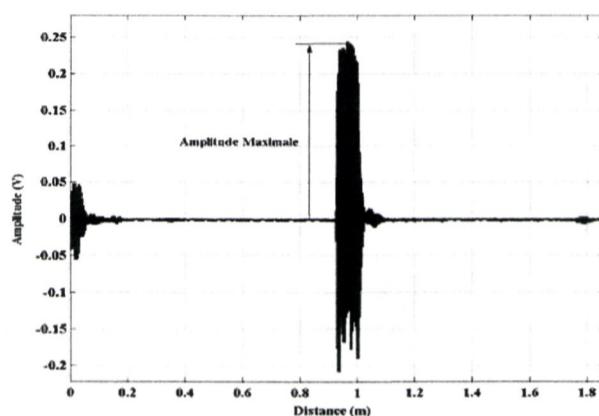
- Augmentation du taux de réalisation,
- Accroissement de la précision (position et rotation),
- Diminution notoire du temps de calcul,
- Indépendance par rapport à l'odométrie,
- Généralisation des résultats sur plusieurs environnements réels.

## 2- Travaux sur la reconnaissance d'objets par ultrason

La navigation d'un robot mobile dans un environnement d'intérieur doit être complétée par un processus secondaire d'évitement d'obstacles. La majorité de travaux dans ce domaine très important se base sur les capteurs à ultrason, appelés aussi sonars.

Un système de perception basé sur les ultrasons doit être capable de mesurer les distances entre le robot et les obstacles, mesurer les orientations et aussi reconnaître la forme et la position des objets situés en face du capteur [Ham.98].

La Figure.2-10 montre un exemple d'écho ultrasonore reçu après l'impacte du signal envoyé avec l'objet de l'environnement que nous voulons connaître. Un tel signal véhicule des informations concernant le type de l'objet. Ces données peuvent être obtenues par des traitements spécifiques et notamment le seuillage, la transformée de Fourier, mesure de l'énergie ou encore, l'extraction de l'enveloppe.



**Figure 2-10.** Exemple d'écho reçu.

Les travaux de N. Aitoufroukh au laboratoire IBISC entrent dans ce dernier cadre qui consiste à analyser l'écho reçu de l'interaction de l'onde ultrasonore et l'objet, et déterminer ensuite le type de l'objet et sa position dans l'environnement [Ouf.02].

L'objectif de ces travaux était de développer un système de perception multimodal basé sur deux éléments :

- 1- L'acquisition et le traitement du signal reçu,
- 2- La reconnaissance et la localisation de l'objet situé dans un environnement structuré.

Dans la suite de cette deuxième partie, nous allons exposer les différentes étapes de ce travail en commençant par le système de mesure, ensuite la base de données et enfin, le processus de reconnaissance.

### 2-1 Dispositif expérimental

Le système expérimental développé est composé de deux capteurs Polaroid 6500 avec un rayon de 1,4 cm et une fréquence de résonance de 50 kHz. Le premier capteur fonctionne en

émission / réception et le deuxième fonctionne uniquement en réception. Ces deux capteurs sont placés sur un plan fixé sur une unité de rotation réglable au demi-degré.

La distance inter-capteurs de 8 cm est la distance horizontale séparant les deux transducteurs du système de mesure. Elle est calculée par rapport à l'angle d'ouverture des capteurs et la distance capteur / cible.

L'ordinateur gère l'ensemble du système expérimental et la carte DSP permet de contrôler le fonctionnement des deux transducteurs (génération et émission du signal). Enfin, la carte d'acquisition numérise l'enveloppe du signal reçu avec une résolution de 12 bits et une fréquence d'échantillonnage de 150 kHz.

## 2-2 Base de données

La base de données est composée de trois sous-ensembles, une base d'apprentissage, une base de généralisation et une base de test. Chacune des ces trois bases comporte un nombre important d'exemples représentant quatre types d'objets simples et deux types d'objets complexes. Ce nombre dépend généralement de la forme et des dimensions des objets, la distance séparant l'objet et le système de mesure, et aussi la qualité des échos reçus.

Les objets simples utilisés représentent en général les différents éléments qui composent un environnement d'intérieur tels que le plan, le coin, l'arête et le cylindre (Figure.2-11). Ces objets peuvent avoir des dimensions différentes.

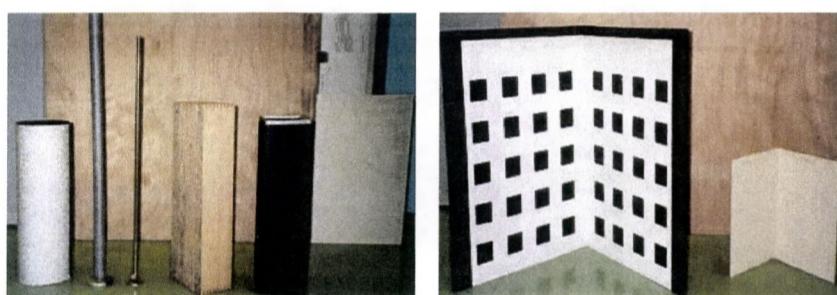


Figure 2-11. Les objets simples.

Les objets complexes utilisés ici sont une chaise et un robot (Figure.2-12).

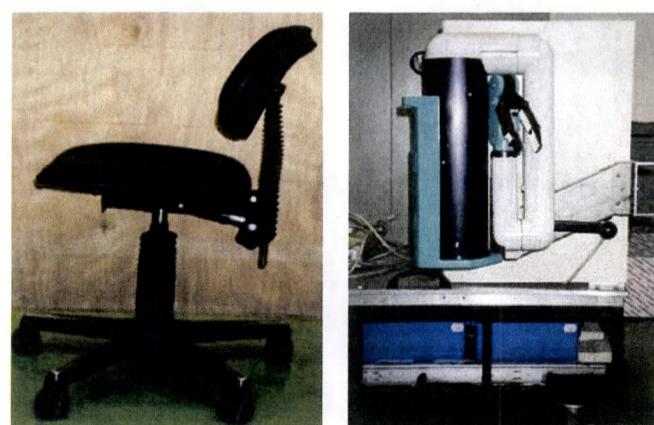


Figure 2-12. Les objets complexes.

### 2-3 Extraction et sélection des paramètres

Les signaux sont caractérisés par un ensemble de paramètres descriptifs, qui constitue le vecteur d'entrée d'un classifieur. Dans ce cas, le choix de ces paramètres se fait par la méthode heuristiques. Seule l'expertise axée sur des observations expérimentales permet de collecter et de juger de la capacité des ces paramètres à pouvoir caractériser correctement les signaux ultrasonores.

Les performances d'un classifieur quelconque dépendent fortement de la qualité de représentation des formes à reconnaître, ce qui implique généralement l'obligation de représenter ces derniers au moyen d'un nombre élevé de primitives.

Le processus connu sous le nom de "sélection de paramètres" a pour rôle de filtrer le vecteur des primitives de base, de manière à en extraire l'information discriminante et à présenter celle-ci au classifieur de manière pertinente.

Parmi les paramètres utilisés dans ce travail de recherche nous citons :

- L'amplitude maximale,
- La longueur de l'écho,
- L'énergie de l'écho,
- Le gain (correspondant à 50 valeurs de fréquence autour de la fréquence de résonance qui est 50 kHz),
- La distance de vol,
- La différence entre les amplitudes issues de chaque capteur,
- La différence entre les longueurs issues de chaque capteur,
- La différence entre les distances issues de chaque capteur.

Les méthodes de sélection de paramètres utilisées réduisent la dimension de l'espace de représentation des objets à reconnaître. Ceci permet d'accélérer le processus de reconnaissance.

Le nombre de paramètres issus des deux capteurs est relativement élevé (113). Il est donc nécessaire de procéder à une phase de sélection de paramètres afin de réduire cet ensemble en gardant les paramètres les plus discriminants.

### 2-4 Méthodes de classification

Le choix du type de classifieur est très important, car il constitue l'élément de décision dans un système de reconnaissance des formes. Le but des traitements antérieurs était d'extraire l'information essentielle concernant l'objet et de la présenter sous forme concentrée au classifieur.

C'est un problème de classification supervisée dans lequel l'utilisateur connaît l'information que le système utilise pour l'apprentissage.

Plusieurs méthodes de classification ont été développées au cours de ce travail, telles que l'estimation non paramétrique de la densité de probabilité par la fenêtre de Parzen (**FP**) ou par les K plus proche voisins (**KPPV**), les Méthodes d'analyse discriminante quadratique (**QDA**) et linéaire (**LDA**), ou encore les réseaux de neurones (global et spécialisé).

Les différentes expérimentations menées dans ce travail ont montré que le meilleur résultat est obtenu par le classifieur basé sur les réseaux de neurones que nous allons les présenter dans les paragraphes suivants.

#### 2-4-1 Réseaux de neurones

Un réseau de neurones est un modèle de calcul dont la conception est très schématiquement inspirée du fonctionnement des vrais neurones humains. Ce nouveau système est adapté au traitement des problèmes complexes comme la reconnaissance, la classification et l'intelligence artificielle.

Il sont appelés aussi réseaux connexionnistes, car ils sont structurés en couches successives capables d'échanger des informations au moyen des connexions qui les relient.

Nous appelons les processeurs de traitement dans les réseaux de neurones artificiels cellules ou unités. L'intérêt des réseaux de neurones réside dans les propriétés qui résultent de leur association en réseaux, c'est-à-dire de la composition des fonctions non linéaires réalisées par chacun des neurones.

#### 2-4-2 Le neurone formel

La première étude systématique du neurone formel est due au neuropsychiatre McCulloch et au logicien Pitts qui en s'inspirant de leurs travaux sur les neurones biologiques, proposent en 1943 le modèle d'un neurone formel illustré sur la Figure.3-13.

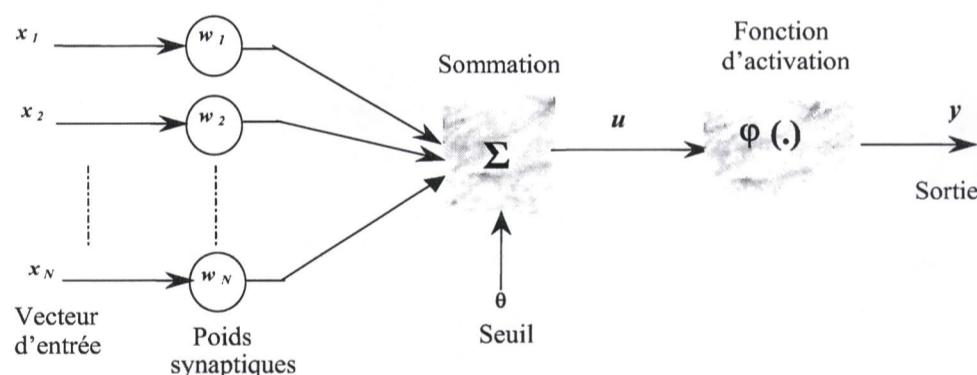


Figure 2-13. Structure d'un neurone artificiel.

Ce neurone formel est un processeur élémentaire qui réalise une somme des signaux d'entrées  $x_i$  pondérées par des poids synaptiques  $w_i$ . La valeur de cette sommation est ensuite comparée à un seuil défini par  $\theta$  et fixé généralement à un.

La sortie du réseau est une fonction non linéaire, dont le calcul passe par deux étapes :

1- Une combinaison linéaire  $u$  des entrées pondérées données par :

$$u = \sum_{j=1}^N w_j x_j - \theta \quad (2-30)$$

2- Une fonction d'activation du neurone  $y$ , exprimée par :

$$y = \varphi(u) \quad (2-31)$$

Le choix de la fonction d'activation dépend de l'application à traiter. Par exemple, pour un problème de classification, nous utilisons généralement des fonctions sigmoïdes qui ont l'avantage d'être dérивables.

La fonction d'activation des neurones doit absolument être non linéaire pour permettre au réseau de générer des fonctions discriminantes non linéaires. Nous distinguons deux types de réseaux de neurones : le réseau global et les réseaux spécialisés.

#### 2-4-3 Le réseau de neurones global

Un problème de classification à  $M$  classes peut être résolu avec un perceptron multicouches (PMC). Ce réseau se présente comme la succession de plusieurs couches. La définition du réseau de neurones global utilisé comporte plusieurs paramètres :

##### a) Architecture du réseau

Le réseau employé est non bouclé et constitué d'une seule couche cachée. La fonction d'activation de cette couche cachée est de type sigmoïde. Elle est de type Softmax pour la couche de sortie, ce qui a permis de calculer les probabilités à posteriori des classes.

Le nombre de neurones en sortie est équivalent au nombre de classes. Le nombre de neurones de la couche d'entrée est égal à la dimension du vecteur d'entrée.

##### b) L'apprentissage

Le but de la rétro-propagation du gradient est de minimiser la distance entre les réponses du réseau et les réponses désirées selon un certain critère. Dans ce travail une fonction d'erreur quadratique moyenne est utilisée.

#### 2-4-4 Le réseau de neurones spécialisé

Dans cette approche le problème global est divisé en sous-problèmes ainsi, les surfaces de séparation sont plus simples à approcher. Ce réseau est donc composé de plusieurs perceptrons multicouche (PMC). Leur nombre est équivalent au nombre de classes. Les paramètres de chaque PMC spécialisé sont identiques à ceux du réseau global.

A partir des sorties des  $M$  classificateurs il faut affecter une classe à chaque observation, à l'aide d'une règle de décision. Elle consiste à affecter l'observation à la classe correspondant à la sortie maximale.

Cette approche de réseaux spécialisés a l'avantage de construire des surfaces de séparations locales. Elle peut néanmoins conduire à un nombre relativement important de classificateurs quand le nombre de classes est élevé.

De même que pour les autres classificateurs cités auparavant, l'espace de représentation propre à chaque sous-classificateur est optimisé par les méthodes de sélection de paramètres.

Le classifieur adopté dans ce travail de recherche (Figure.2-14) est composé de cinq réseaux spécialisés de type PMC, donc un réseau spécialisé pour chaque classe simple.

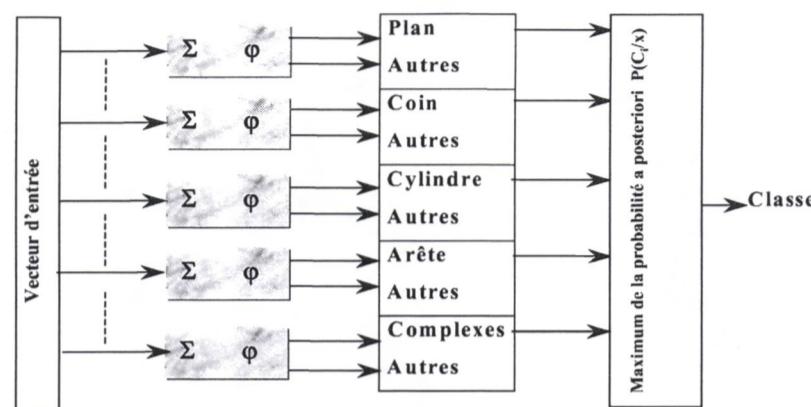


Figure 2-14. Réseau spécialisé composé de 5 classes.

Deux importants paramètres sont à prendre en considération pour chaque réseau :

- 1- Le nombre de neurones de la couche cachée,
- 2- La dimension du vecteur d'entrée pour chaque réseau spécialisé.

Les valeurs de ces paramètres sont différentes pour chaque réseau spécialisé utilisé. Le nombre de neurones dans la couche cachée ainsi que le nombre de paramètres dans le vecteur d'entrée de chaque réseau spécialisé sont données dans la Table.2-5.

Objets	Neurones en couche cachée	Paramètres en entrée
Plan	20	30
Coin	14	30
Cylindre	20	12
Arête	20	15
Complexes	11	68

Table 2-5. Paramètres des cinq classes du réseau spécialisé.

Pour la classe plan par exemple, il faut avoir 20 neurones en couche cachée et le vecteur d'entrée doit contenir 30 paramètres pour avoir le meilleur taux de classification.

Ces deux paramètres sont déterminés par comparaison de différentes architectures, en variant le nombre de neurones et la dimension du vecteur d'entrée. Les réseaux retenus sont ceux qui offrent le meilleur résultat en classification.

#### 2-4-5 Classification des objets simples

La classification des différents objets à l'aide de la méthode des réseaux de neurones pour un vecteur d'entrée comportant cinq paramètres est illustrée dans la table Table.2-6.

Les objets autres que ceux définis comme simples sont regroupés dans une classe appelée "complexe". Nous notons que le réseau de neurone utilisé contient 18 neurones dans la couche cachée.

Objets	Plan	Coin	Arête	Cylindre	Complexes
Taux de classification	70 %	74 %	59 %	61 %	82 %

**Table 2-6.** Classification des objets simples avec le réseau spécialisé.

La méthode neuronale donne une valeur de taux moyen de bonne classification d'environ 70 % pour un vecteur d'entrée de cinq paramètres. L'augmentation du nombre de paramètres à l'entrée va automatiquement améliorer le taux de classification.

#### 2-4-6 Classification des objets complexes

Cette partie est consacrée à la distinction entre différents objets complexes à l'aide des réseaux de neurones utilisant 3 paramètres discriminants déterminés lors de la phase de sélection. Le réseau de neurones utilisé dans cette étape est composé de 10 neurones dans la couche cachée. Les résultats de classification sont présentés dans la Table 2-7.

Objets	Chaise	Robot	Simples
Taux de classification	100 %	98 %	95 %

**Table 2-7.** Taux de classification des objets complexes.

Ces résultats montrent que le réseau de neurones global est suffisant pour différentier des objets complexes.

#### 2-5 Analyse de scène

Les performances du classifieur neuronal ont été testées dans le cas d'un environnement réel composé d'une succession d'éléments simples. Le but de cette procédure est de classer les différents objets se trouvant dans l'environnement du robot après avoir effectué un balayage circulaire.

L'environnement expérimental est conçu d'un nombre d'objets simples comme des plans, des coins et des cylindres de diamètres égaux à 2, 4 et 20 cm.

Le balayage a été effectué dans un intervalle angulaire de  $[-90^\circ, +90^\circ]$  avec un pas de balayage variable entre  $2^\circ$  et  $5^\circ$ .

L'emplacement de ces différents objets dans l'environnement de travail est représenté sur la Figure 2-15. La position des objets a été déterminée à l'aide des méthodes dédiées au calcul de distance réel et de l'orientation système / cible.

Il est à noter que le réseau spécialisé utilisé traite les deux premiers échos à chaque pas de balayage. C'est une des caractéristiques du système de perception de pouvoir reconnaître plusieurs objets situés le long de l'axe de mesure.

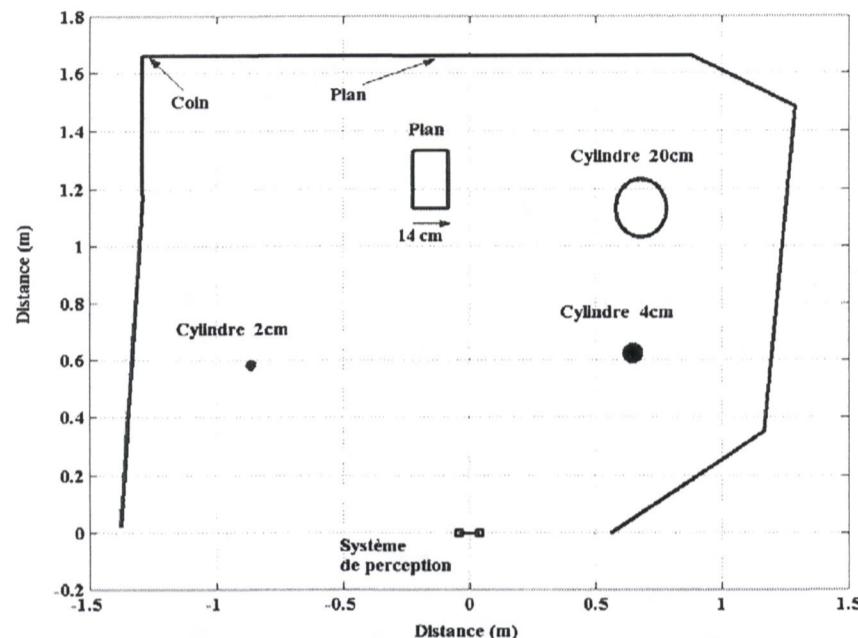


Figure 2-15. Exemple d'environnement expérimental.

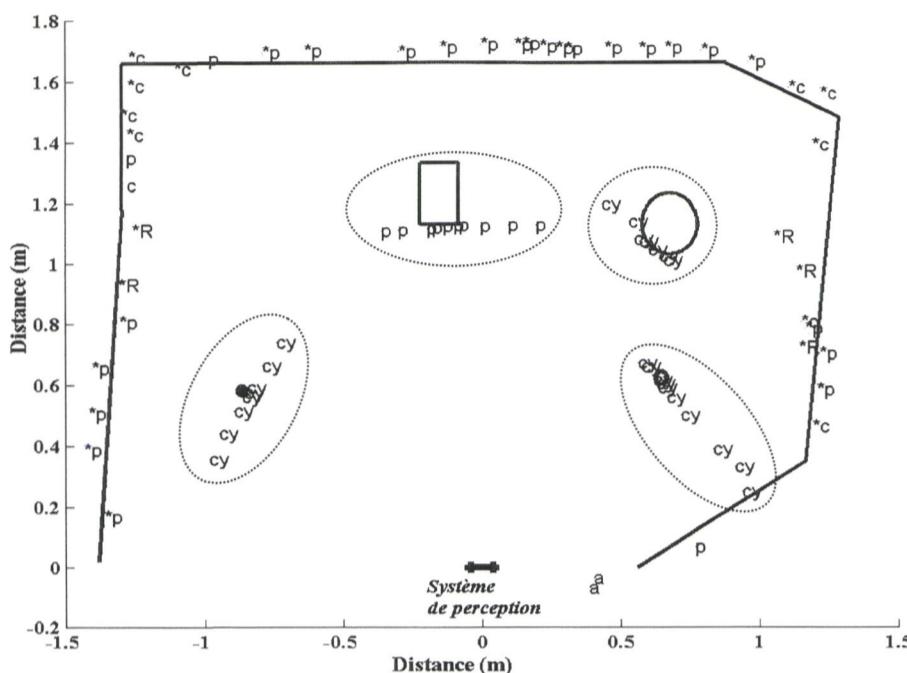
Une classe de rejet a été ajoutée au réseau pour réduire l'erreur de classification durant la phase de reconnaissance des objets dans l'environnement. Cette classe intervient lors d'une ambiguïté entre les différentes classes.

La notation suivante a été adoptée lors de la classification :

- **p** : pour la classe plan,
- **c** : pour la classe coin,
- **cy** : pour la classe cylindre,
- **R** : pour la classe rejet,
- **\*** : pour les objets occultés.

Les résultats de classification font apparaître les problèmes dus à l'ouverture du cône d'émission et à l'occultation. Le cône d'émission induit des zones de reconnaissance élargies, ainsi, les objets occultant (les trois cylindres et le plan de forme cubique) sont reconnus sur une assez large zone (Figure.2-16).

Ces zones élargies entraînent certaines zones de rejet au niveau des objets occultés si l'axe de l'onde ultrasonore n'est pas perpendiculaire à ces objets. Le taux d'erreurs obtenu pour la classification des objets de l'environnement est de 4,9 % et le taux de rejet est de 6,2 %, ce qui représente des valeurs faibles.



**Figure 2-16.** Classification de plusieurs objets dans l'environnement.

## 2-6 Analyse des résultats et conclusion

Dans cette deuxième partie du chapitre, on a établi un résumé des travaux de thèse de N. Aitoufroukh, qui consiste à étudier théoriquement et expérimentalement la reconnaissance d'objets simples et complexes dans un environnement de travail réaliste. Ce travail utilise principalement, des informations extraites d'un écho reçu pour la reconnaissance d'objets.

En premier lieu, une extraction de paramètres de l'écho est effectuée pour permettre de caractériser les différents objets. Ensuite, pour accélérer le processus de classification, l'espace de représentation des paramètres est réduit par une sélection de paramètres.

Après l'extraction et la sélection de paramètres, il reste la dernière étape du processus de reconnaissance des objets, à savoir la classification avec un dispositif composé de deux sonars. Cette phase a été réalisée à l'aide de différentes méthodes statistiques et neuronales.

Les résultats obtenus à l'aide des méthodes statistiques sont classés dans l'ordre croissant en termes du taux de classification : la LDA, la QDA, la KPPV et la fenêtre de Parzen. Les deux dernières méthodes s'avèrent très coûteuses en temps de calcul. Elles offrent respectivement un taux moyen de classification entre 68 % et 70 % avec 5 paramètres à l'entrée.

Les méthodes linéaires (LDA) et quadratique (QDA), quant à elles, offrent respectivement un taux moyen de classification entre 50 % et 53 % aussi avec 5 paramètres à l'entrée.

Ces méthodes ont l'avantage d'être rapide en traitement mais ne donnent pas un résultat satisfaisant en terme de taux de classification. Les réseaux de neurones ont été utilisés aussi pour la classification avec deux types, le réseau global et le réseau spécialisé. Leur taux moyen de reconnaissance est d'environ 85 % et un temps de traitement négligeable.

Cette dernière méthode de classification est la plus robuste et la plus rapide par rapport aux autres méthodes existantes.

Les résultats de classification obtenus dans un l'environnement réaliste sont très satisfaisants et prometteurs. Cela prouve la possibilité de reconnaître plusieurs objets dans un environnement de travail intérieur grâce à un balayage du robot.

Ce travail de recherche illustre donc en général, l'intérêt de l'information ultrasonore même si le résultat de 85 % montre que cette information ne peut pas être suffisante pour prendre des décisions dans n'importe quelle situation.

### **3- Nos objectifs**

A partir des travaux que nous venons de présenter, le défi pour mon travail est de combiner ces deux approches pour construire un système de localisation robuste et efficace.

L'idée emise consiste à proposer un enrichissement du modèle filaire 3D par des attributs du type "plan", "coin" et "arête" dans l'objectif de diminuer le nombre d'hypothèses de localisation, de réduire le temps de calcul et d'augmenter la précision sur le calcul de la position et de l'orientation du robot.

Dans un premier temps, nous allons détailler la mise en œuvre de la combinaison de ces deux travaux pour présenter une nouvelle approche de localisation dans le chapitre III.

Dans un deuxième temps, nous allons utiliser cette combinaison "vision / ultrason" pour proposer une méthode originale de reconnaissance d'objets simples. L'objectif de cette étude est d'approcher et de saisir des objets à l'aide du bras manipulateur par un asservissement visuel. Ce travail sera détaillé dans le chapitre IV.

## Chapitre III

### Localisation du robot mobile par vision et ultrasons

*Le troisième chapitre présente la méthode de localisation développée pour le robot mobile basée sur l'approche incrémentale. Cette méthode est composée de deux parties principales et utilise deux types de capteurs.*

*La première partie utilise une caméra permettant d'obtenir la structure de l'environnement. La deuxième partie utilise quant à elle, un système ultrasonore dont le rôle est de reconnaître les différentes formes.*

*La combinaison de ces deux types de données aboutit à un modèle filaire 3D enrichi par l'information ultrasonore. Elle permet aussi de générer une table d'hypothèses de localisations qui contient probablement la position réelle du robot.*

*Plusieurs tests et mesures ont été effectués sur différents types d'environnement pour valider cette méthode de localisation. La précision obtenue ainsi que les temps de traitement enregistrés prouvent la fiabilité et la robustesse de cette méthode.*

## 1- Introduction

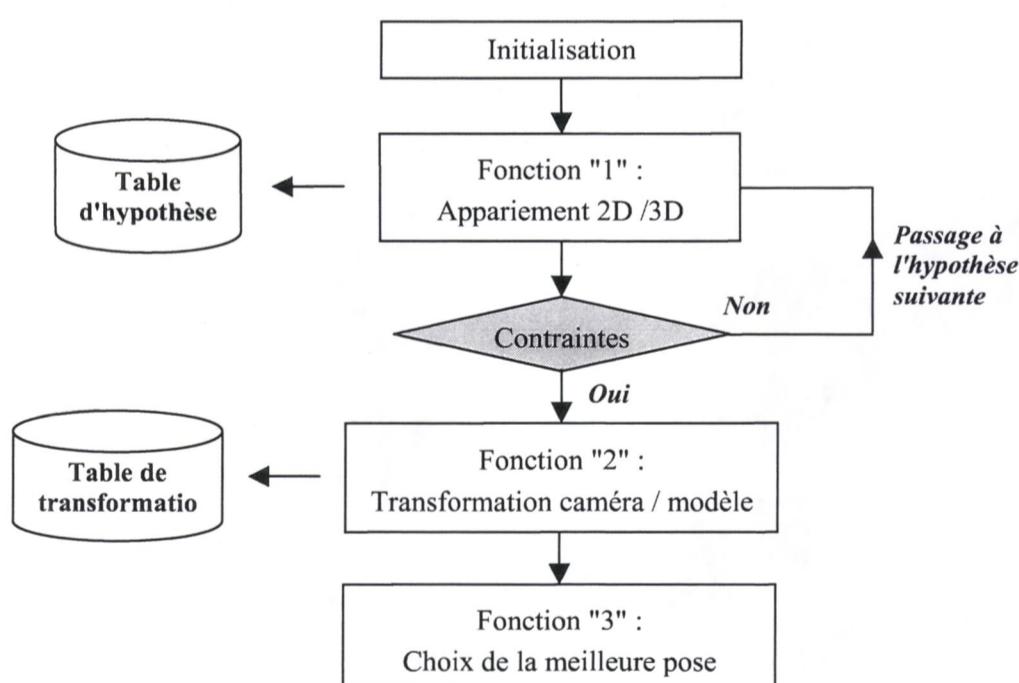
Dans ce chapitre, nous allons traiter le problème de localisation vu dans la première partie du chapitre précédent et apporter des améliorations à chaque étape de l'algorithme. Nous allons utiliser aussi l'information ultrasonore pour représenter les différentes formes de l'environnement et l'intégration de cette nouvelle donnée dans le processus de mise en correspondance 2D / 3D [Tra.a.06].

Le schéma bloc de notre stratégie de localisation va être donné dans un organigramme général. Nous préparons dans un premier temps les deux types d'informations que nous allons utiliser dans le processus de localisation, et notamment les données ultrasonores et visuelles, tous les deux nécessaires au processus de la mise en correspondance.

Nous développerons ensuite les deux méthodes d'appariement 2D / 3D utilisant la contrainte ultrasonore. Nous présenterons également la transformation "repère caméra / repère image" basée sur l'approche de Lowe et ses deux solutions linéaire et non-linéaire. Enfin, nous présenterons les différents résultats obtenus sur les trois types d'environnements.

## 2- Organigramme de la localisation

La méthode de localisation que nous allons présenter dans ce troisième chapitre est composée de trois fonctions principales et un bloc d'initialisation des données comme la Figure.3-1.



**Figure 3-1.** Schéma bloc de la méthode de localisation.

L'étape d'initialisation consiste à récupérer trois types d'informations :

- 1- La donnée odométrique permettant de connaître la position probable du robot,
  - 2- L'image de la scène qui servira de base au calcul des segments de droite utilisés dans la phase d'appariement.

### 3- L'écho ultrasonore qui donne une indication sur la forme de l'environnement.

La fonction "1" permet d'effectuer l'appariement 2D / 3D entre les segments de l'image et les segments 3D du modèle de l'environnement de travail. Si cet appariement est cohérent, la fonction test va le valider et l'hypothèse d'appariement générée sera acceptée et transmise à l'étape suivante. Si la réponse au test d'appariement est négative, on incrémentera le compteur de la table des hypothèses pour étudier la proposition d'appariement suivante.

La fonction "2" permet de passer des coordonnées des segments 3D du modèle participants à l'appariement aux coordonnées de la caméra, et par conséquence à celle du robot dans ce même repère 3D de l'environnement. Ainsi, chaque appariement cohérent va se traduire par une pose du robot dans le modèle de l'environnement.

La fonction "3" quant à elle, calcule la pose du robot la plus plausible parmi les différentes autres poses issues de la fonction "2". Ce calcul utilise une fonction de coût qui sera développée dans la suite du chapitre.

## 3- Préparation des informations 2D et 3D

Pour l'amélioration de la méthode initiale de localisation présentée dans le chapitre précédent, nous commençons par la présentation du nouveau modèle 3D qui contient en plus des segments 3D, des informations concernant la forme de l'environnement. Puis, nous présentons le classifieur neuronal utilisé dans la reconnaissance des formes de l'environnement à partir des échos ultrasonores.

Ensuite, nous exposons le modèle géométrique d'une caméra ainsi que la méthode de calibration utilisée. Enfin, nous présenterons les informations extraites du système de vision et le traitement d'images développées [Tra.d.05].

Pour comprendre les différentes améliorations et quantifier les résultats apportés à chaque étape de localisation, nous allons choisir trois exemples de poses (position et orientation du robot) dans l'environnement qui correspondent aux trois formes de base (plan, coin et arête).

Pour chaque étape de localisation, nous allons comparer les résultats obtenus avec les nouvelles approches proposées avec les résultats donnés par les approches initiales développées dans le chapitre précédent.

### 3-1 Crédit du modèle 3D enrichi

Le système ultrasonore décrit dans la deuxième partie du chapitre II nous permet de reconnaître les différentes formes de l'environnement. Ces formes peuvent être un plan, un coin, une arête ou encore une forme inconnue. La réduction du nombre d'hypothèses de localisation consiste à enrichir le modèle filaire classique décrit au chapitre II par des informations ultrasonores.

Nous associons à chaque segment une étiquette comportant des informations à savoir la forme d'appartenance, le numéro du segment et ses coordonnées en 3D. Si la forme d'appartenance d'un segment ne correspond pas à une des trois classes (plan, coin et arête), il sera donc considéré comme un segment de liaison avec une étiquette ultrasonore "\*\*".

Le rôle de ce genre de segments est de relier les autres segments étiquetés du modèle et de participer à la mise en correspondance "2D / 3D". La Figure.3-2 donne un exemple de ce modèle enrichi.

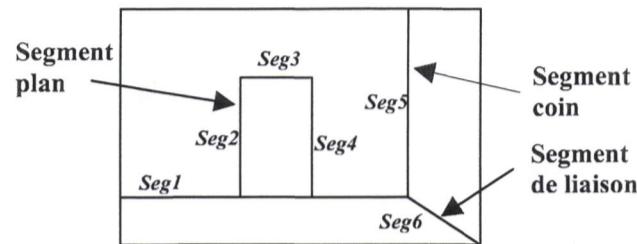


Figure 3-2. Modèle enrichi par les ultrasons.

La Table.3-1 représente la structure des données dans le modèle enrichi développé, où chaque segment du modèle de l'environnement est caractérisé par son numéro, ses coordonnées 3D et la forme d'appartenance.

	<b>X1</b>	<b>Y1</b>	<b>Z1</b>	<b>X2</b>	<b>Y2</b>	<b>Z2</b>	<b>Contrainte US</b>
<b>Segment 1</b>	$x_{11}$	$y_{11}$	$z_{11}$	$x_{21}$	$y_{21}$	$z_{21}$	<b>Plan</b>
<b>Segment 2</b>	$x_{12}$	$y_{12}$	$z_{12}$	$x_{22}$	$y_{22}$	$z_{22}$	*
<b>Segment 3</b>	$x_{13}$	$y_{13}$	$z_{13}$	$x_{23}$	$y_{23}$	$z_{23}$	<b>Coin</b>
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
<b>Segment N</b>	$x_{1N}$	$y_{1N}$	$z_{1N}$	$x_{2N}$	$y_{2N}$	$z_{2N}$	<b>arête</b>

Table 3-1. Modèle de l'environnement enrichi par les ultrasons.

L'objectif de l'utilisation des informations ultrasonores est de réduire le nombre d'hypothèses lors du processus de mise en correspondance et de rendre la tâche de localisation plus performante et plus rapide.

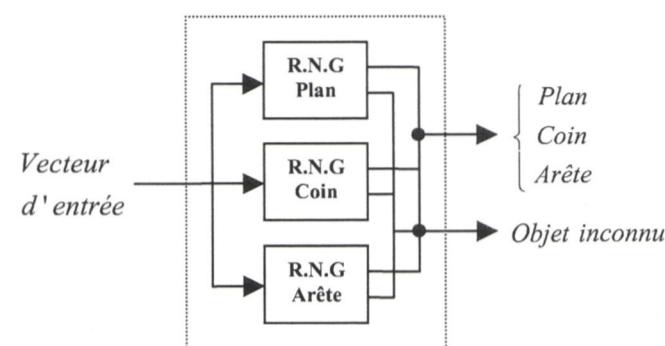
### 3-2 Informations extraites des ultrasons

La méthode de classification d'objets que nous avons choisis se base sur les réseaux de neurones spécialisés. Ce choix est motivé par deux avantages, le bon taux de classification qui peut atteindre les 85 % et la rapidité dans la prise de décision comme présenté au chapitre précédent.

L'environnement qui entoure le robot est constitué généralement de trois formes de base, le plan, le coin et l'arête. Nous avons réduit le nombre de classes de sortie du réseau de cinq à trois, pour qu'il soit adapté à notre application. Nous avons regroupé les différentes modifications dans ses trois points :

- 1- La réduction du nombre de classes à reconnaître de cinq à trois, car ici nous nous focalisons plus sur les formes de l'environnement que sur les formes d'objets. Ainsi, le nouveau réseau spécialisé conçu est composé de trois réseaux globaux : un R.N.G plan, un R.N.G coin et un R.N.G arête, comme présenté dans la Figure.3-3,

- 2- Le nombre d'échantillons dans la base de données est réduit, car nous supprimons tous les échos à faible énergie,
- 3- Le vecteur d'entrée qui alimente le réseau de neurones spécialisé contient les paramètres de l'écho les plus discriminants parmi tous les paramètres présentés auparavant dans le chapitre précédent.



**Figure 3-3.** Structure du réseau de neurones spécialisé.

Les trois R.N.G sont présentés chacun par un perceptron multicouche avec une couche d'entrée, une couche cachée et une couche de sortie. La fonction d'activation est une probabilité à posteriori  $P_r(k/x)$  évaluée par une fonction Softmax donnée par l'équation de probabilité (3-1) suivante :

$$\Pr(k/x) = \frac{\exp(y_k)}{\sum_{j=1}^K \exp(y_j)} \quad (3-1)$$

Où  $y_j$  est la sortie " $j$ " du réseau de neurones et  $K$  est le nombre de classes à reconnaître. La probabilité maximum de  $P_r(k/x)$  donne la classe de sortie. Le classifieur neuronal est conçu après deux étapes : L'apprentissage et la généralisation, toutes les deux se basent sur les données ultrasonores reçues.

Les taux de bonne classification obtenus avec le réseau de neurones spécialisé utilisé ainsi que le nombre de neurones en couche cachée sont présentés sur la Table.3-2.

Nous trouvons aussi le nombre de paramètres du vecteur d'entrée du réseau pour chaque forme de l'environnement : plan, coin et arête.

	Neurones en couche cachée	Paramètres en entrées	Taux de classification
<b>Plan</b>	17	07	99 %
<b>Coin</b>	10	05	92 %
<b>Arête</b>	13	10	91 %

**Table 3-2.** Taux de reconnaissance des trois formes de l'environnement.

Les taux de reconnaissance obtenus pour chaque forme de l'environnement sont toujours supérieurs à 90 %. Les nombres de neurones dans la couche cachée ainsi que les paramètres

dans le vecteur d'entrée pour chaque forme sont fixés après le test de plusieurs configurations possibles.

La conception de ce classifieur neuronal a nécessité la préparation de deux bases de mesures. Une première base d'apprentissage contenant 2000 échantillons et une deuxième de généralisation contenant aussi 2000 échantillons (un total de 4000 échos).

Les bons résultats de reconnaissance obtenus nous permettent d'utiliser le système ultrasonore (capteurs et classifieur) dans le travail de localisation robotique. Cette intégration se fait sur les deux niveaux suivants :

- 1- **Hors ligne** : Lors de la construction du modèle 3D, nous attribuons à chaque segment une étiquette,
- 2- **En ligne** : Durant le processus de localisation, l'information ultrasonore reçue est utilisée dans l'appariement.

Chaque échantillon de ces deux bases représente un écho d'objet pris lors d'un balayage horizontal à différentes distances bien définies. L'objet peut avoir une des trois formes de l'environnement, plan, coin ou arête.

### 3-3 Calibration de la caméra

En vision, la calibration ou l'étalonnage d'une caméra revient à modéliser le processus de formation des images, c'est-à-dire trouver les relations qui lient les coordonnées spatiales d'un point de l'espace avec ses coordonnées dans l'image et la pose de la caméra.

Notre méthode de localisation repose sur un appariement 3D / 2D et utilise des coordonnées 3D dans l'espace et 2D dans l'image. D'où la nécessité d'une calibration du capteur de vision utilisé dans les mesures.

Les détails de la méthode de calibration que nous avons utilisé dans notre travail sont présentés dans l'Annexe D.

### 3-4 Informations extraites de la vision

Pour finir cette partie, nous allons expliquer comment extraire les segments 2D de l'image nécessaire au processus d'appariement qui est le cœur de la méthode de localisation incrémentale.

Nous utilisons des techniques de traitement d'images appropriées pour l'extraction de ces segments. La première consiste à filtrer les contours de l'image en utilisant un détecteur optimal de Deriche, puis choisir de ces contours les segments de droites ayant une longueur supérieure à un seuil donné en pixel.

#### 3-4-1 Détection des contours

La détection des contours est une étape préliminaire à de nombreuses applications de l'analyse d'images. Les contours constituent en effet, des indices riches, au même titre que les points d'intérêts.

Les contours dans une image proviennent généralement des :

- 1- Discontinuités de la fonction de réflectance (texture, ombre),
- 2- Discontinuités de profondeur (bords de l'objet).

Les différents types de contours, marche, toit et pointe sont présentés sur la Figure 3-4 :

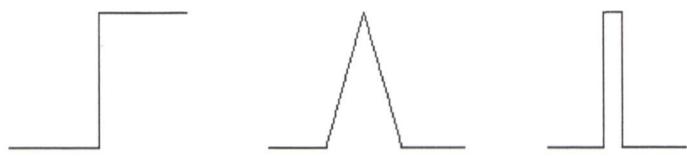


Figure 3-4. Différents types de contours.

Le principe de la détection des contours repose donc sur l'étude des dérivées de la fonction d'intensité dans l'image : Les extréma locaux du gradient de la fonction d'intensité ou les passages par zéro du laplacien (Figure 3-5). La méthode qui nous intéresse dans la suite c'est celle basée sur le gradient.

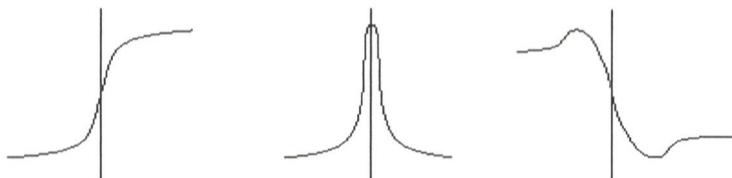


Figure 3-5. La fonction d'intensité et ses dérivées première et seconde.

Le gradient d'une image  $I(x, y)$  est donné par le vecteur  $\nabla I(x, y)$  défini par :

$$\nabla I(x, y) = \left( \frac{\partial I(x, y)}{\partial x}, \frac{\partial I(x, y)}{\partial y} \right) \quad (3-2)$$

Il est donc caractérisé par un module  $M$  et une direction  $\phi$  dans l'image avec :

$$M = \sqrt{\left( \frac{\partial I(x, y)}{\partial x} \right)^2 + \left( \frac{\partial I(x, y)}{\partial y} \right)^2} \quad (3-3)$$

$$\phi = \arctan \left( \frac{\partial I(x, y)}{\partial x} / \frac{\partial I(x, y)}{\partial y} \right) \quad (3-4)$$

En traitement d'image, il existe deux approches pour estimer la dérivée de la fonction d'intensité d'une image :

- 1- **Les différences finies** : Qui consistent à approximer les dérivées par des différences finies. Une convolution est effectuée entre l'image d'origine et un masque de différences.

Il existe plusieurs types de masques, chacun a ses avantages et ses inconvénients comme les masques de Sobel, de Prewitt, de Roberts et de Robinson.

- 2- **Le filtrage optimal** : Dans ce cas, les dérivations consistent à convoluer l'image par des masques de petites dimensions. Ces approches sont donc dépendantes de la taille des objets traités et elles sont très sensibles au bruit. Ce type d'approches repose sur la définition de critères d'optimalité de la détection de contours.

Ces critères débouchant sur des filtres de lissage optimaux et l'implantation se fait de manière récursive (la valeur de sortie du filtre en un point est déterminée en fonction de celles des voisins) par deux balayages de l'image (ligne et colonne) en utilisant la propriété de séparabilité.

Le travail initial sur la localisation utilisait un détecteur de contour du type Canny, où le gradient de l'image lissée est estimé par la dérivée d'un filtre gaussien avec une moyenne et un écart type donnés.

Dans notre travail, nous avons choisi d'utiliser le détecteur optimal de Deriche, qui répond exactement aux mêmes critères de qualité que celui de Canny, mais qui possède une réponse impulsionnelle infinie (filtre RII).

Ce filtre a pu être synthétisé de façon récursive particulièrement efficace [Der.93]. Il a une expression générale mono-dimensionnelle de la forme :

$$h'(x) = -k' x e^{-\alpha|x|} \quad (3-5)$$

Avec :

$$k' = \frac{(1 - e^{-\alpha})^2}{e^{-\alpha}} \quad (3-6)$$

Le facteur  $\alpha$  définit la largeur du filtre, donc le compromis entre la détection et la localisation. Plus  $\alpha$  est grand, plus on localise précisément le contour plus nous détectons facilement la présence des contours. Le filtre de lissage  $h$  s'obtient par intégration du filtre de dérivation  $h'(x)$ , ainsi :

$$h(x) = k(\alpha|x| + 1) e^{-\alpha|x|} \quad (3-7)$$

Avec :

$$k = \frac{(1 - e^{-\alpha})^2}{(1 + 2\alpha e^{-\alpha} - e^{-2\alpha})} \quad (3-8)$$

Ce genre de filtres est mono-dimensionnel, ainsi le filtrage 2D est obtenu par l'action de deux filtres croisés, l'un suivant les lignes et l'autre suivant les colonnes. La propriété de séparabilité, l'implantation récursive. Les détails de ce filtre optimal sont donnés dans l'Annexe C.

### 3-4-2 Extraction des segments

L'image binaire des points contours est obtenue avec un seuillage par hystérisis des maxima locaux après le calcul de l'image des gradients. L'obtention d'une carte contours n'est pas suffisante pour avoir les segments de droites utilisés dans le processus d'appariement. Une transformation de Hough est nécessaire pour extraire ces segments de droites 2D [Han.97].

Chaque point de l'image appartenant à une droite est sélectionné puis stocké dans une table à deux dimensions appelée table de Hough. Les points n'appartenant pas à une droite sont supprimés. A la fin de cette transformation, des segments de différentes longueurs seront stockés dans cette table.

Nous appliquons un seuillage sur la longueur pour ne garder que les segments de taille significative. Nous avons gardé la valeur de 20 pixels pour éliminer les petits segments et les différents bruits.

### 3-4-3 Comparaison avec la méthode initiale

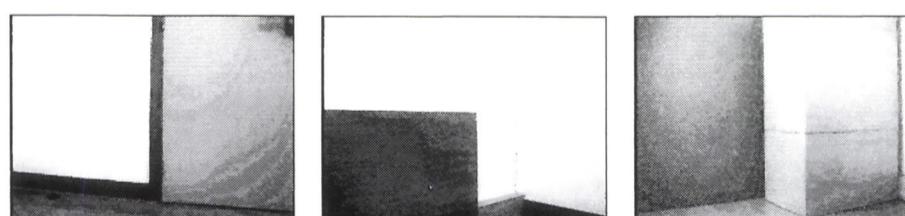
Nous avons quantifié l'amélioration apportée en utilisant un détecteur de contours de type Deriche au lieu d'un filtrage de Canny sur trois exemples d'images. Nous avons calculé pour chacune le nombre de segments extrait avec les deux méthodes.

Ces trois exemples correspondent à trois poses du robot dans son environnement de travail (Table.3-3). Ces poses correspondent aux différentes formes de base de l'environnement (plan, coin et arête).

Poses	Formes	X (m)	Y (m)	Φ (°)
1	Plan	0.52	-0.25	15
2	Coin	1.00	0.30	45
3	Arête	1.72	-0.25	150

**Table 3-3.** Les trois poses utilisées dans les tests.

La Figure.3-8 donne les trois images des poses de l'environnement utilisées pour faire les différents tests.



**Figure 3-6.** Images des trois poses utilisées dans les tests.

A partir de ces images, nous extraireons les contours par le filtre de Deriche, puis nous appliquons la transformée de Hough. Et enfin, nous comparons le nombre de segments obtenu avec ces deux méthodes d'extraction de contours.

Les résultats sont présentés dans la Table.3-4.

Poses	Segments extraits à partir de l'approche de Canny	Segments extraits à partir de l'approche de Deriche
1	4	5
2	6	8
3	8	11

**Table 3-4.** Nombre de segments.

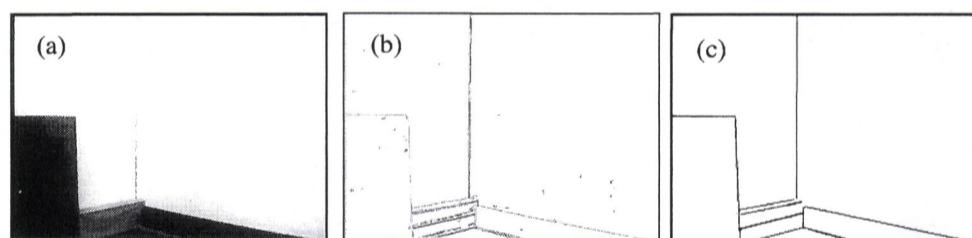
Ces résultats montrent que l'utilisation de la méthode basée sur le détecteur de Deriche permet d'obtenir un peu plus de segments qu'avec celle utilisant le filtrage de Canny.

Cette amélioration est intéressante, car un nombre plus grand de segments 2D facilite le processus de mise en correspondance 2D / 3D et diminue le taux d'échec de localisation dû au manque de segments image.

#### 3-4-4 Exemple d'extraction de segments

La Figure.3-7 montre un exemple de résultats obtenu avec les méthodes de traitement d'images appliquées (filtre de Deriche et transformation de Hough). En (a) l'image originale couleur présentant une partie de l'environnement de travail.

En (b) la carte contours obtenue par le filtrage optimal de Deriche. et en (c) les segments retenus après extraction par la table de Hough et seuillage sur la longueur de segment.



**Figure 3-7.** Extraction de segments 2D, (a) : Image originale, (b) : Carte contours, (c) : Extraction et seuillage des segments (seuil de 20 pixels).

Cet exemple donne une idée sur les résultats obtenus avec la méthode que nous avons développée. Le choix du seuil concernant la longueur des segments est très important. Il permet de réduire considérablement le bruit sur la carte contours, en conséquence, réduire la quantité de données à traiter pour bien préparer l'appariement.

### 4- Appariement 2D / 3D

La mise en correspondance est l'étape la plus complexe dans les applications orientées vision et qui permet de faire le lien entre deux représentations différentes en se basant sur des primitives connues. Elle peut se faire entre les primitives de deux images (2D / 2D) comme dans la vision stéréoscopique ou l'indexation d'images.

Dans certaines applications de la vision par ordinateur, l'environnement de travail est représenté par un modèle 3D, ainsi un appariement 2D / 3D entre les primitives de l'image et du modèle.

Il existe plusieurs travaux de localisation basés sur la combinaison de plusieurs types d'informations capteurs.

R. Meier présente dans [Mei.99] un algorithme de fusion de données pour réduire l'incertitude de localisation et détecter les obstacles dans l'environnement, dans un but d'améliorer la télé-opération d'un véhicule mobile "Pioneer AT". Cette plate-forme qui utilise trois types de capteurs : Un système de vision stéréo, une ceinture de capteurs ultrasonores et un système odométrique.

Nous pouvons citer aussi les travaux d'O. Horn sur le positionnement d'un fauteuil électrique autonome "VAHM" dans un environnement d'intérieur connu [Horn.00]. Deux approches distinctes sont adoptées selon l'état du fauteuil en arrêt ou en mouvement : "Localisation statique" et "Localisation dynamique".

Dans la première approche, le système utilise des mesures ultrasonores pour localiser le robot dans une grille d'occupation, ce qui permet d'initialiser le système odométrique nécessaire à la deuxième approche. Cette dernière utilise en plus en caméra et un appariement de segments pour calculer la pose du fauteuil au cours du mouvement.

Dans notre travail, nous nous intéressons à l'appariement 2D / 3D avec trois types de capteurs, la vision monoculaire, un système ultrasonore et l'odométrie. Dans la suite, nous allons rappeler rapidement la méthode de mise en correspondance initiale, puis nous allons présenter les deux méthodes que nous avons développé et qui utilisent en plus de la vision, la forme de l'environnement donnée par le système ultrasonore.

#### 4-1 Rappel de la méthode initiale

C'est la méthode incrémentale classique présentée dans le chapitre II et qui utilise le segment de droite comme primitive de base. Elle est basée sur l'information vision et sur une recherche arborescente utilisant deux contraintes géométriques et un modèle filaire 3D de l'environnement.

#### 4-2 Première amélioration : Pré-combinaison

La méthode d'appariement proposée rajoute la contrainte ultrasonore aux deux contraintes géométriques. Cette nouvelle contrainte représente la forme de l'environnement du robot mobile. Nous rappelons que le système ultrasonore est composé de deux sonars pour l'acquisition des échos et d'un réseau de neurones pour la classification de forme.

Le modèle 3D représentant l'environnement dans cette méthode est enrichi par l'information ultrasonore comme nous l'avons présenté dans la partie 2-1.

Cette étape d'appariement entre les segments modèle et image est suivie d'un deuxième appariement pour la contrainte ultrasonore. Cet appariement va comparer les étiquettes ultrasonores des segments du modèle participant à l'appariement avec la contrainte ultrasonore donnée en ligne par le réseau de neurones.

S'il y a au moins un segment modèle dans l'appariement qui possède une étiquette semblable à celle de la contrainte ultrasonore, l'hypothèse comportant cet appariement sera dans ce cas acceptée. Sinon, cette hypothèse est rejetée de la table finale appelée table d'hypothèses réduite (Figure.3-8).

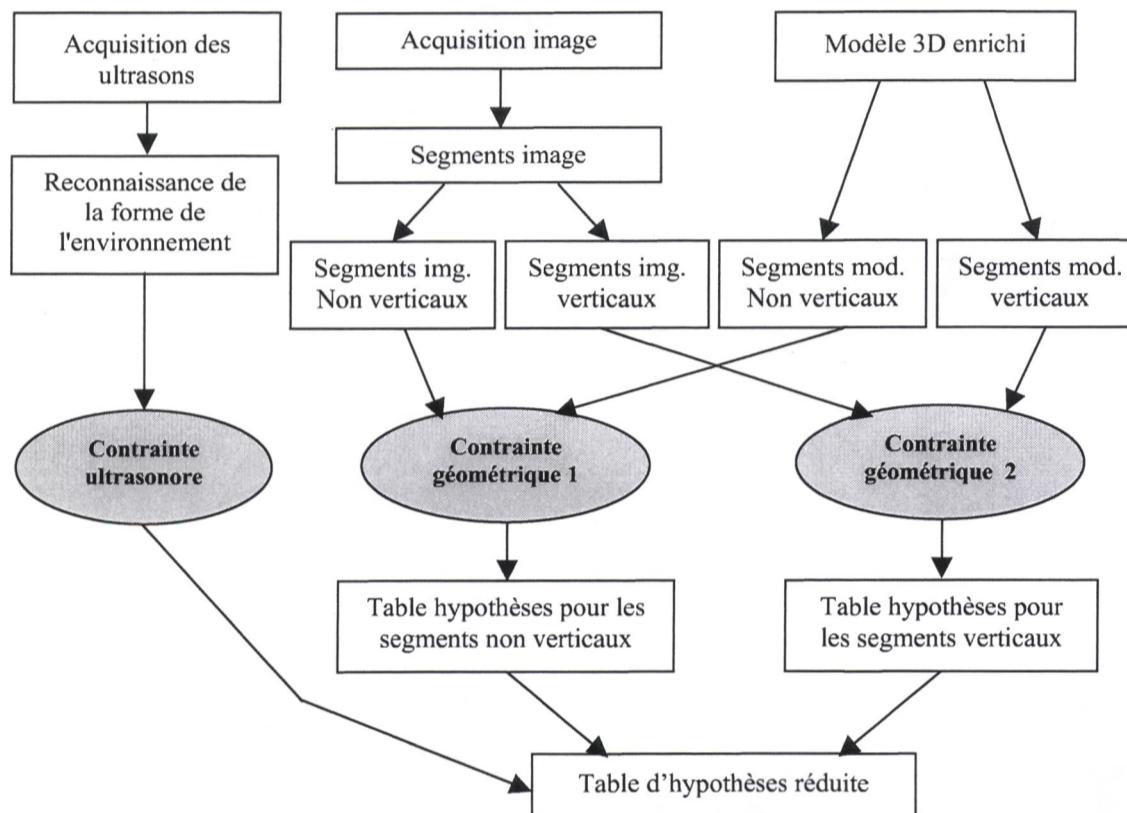


Figure 3-8. Appariement avec la pré-combinaison.

Pour éclaircir cette méthode, nous allons donner deux exemples, le premier pour une hypothèse acceptée et l'autre pour une hypothèse non acceptée. Nous allons utiliser les notations suivantes :

- $L_1, L_2, L_3, L_4, L_5$  : cinq segments modèle associés aux étiquettes ultrasonores  $US_1 = \text{plan}$ ,  $US_2 = \text{coin}$ ,  $US_3 = *$ ,  $US_4 = \text{arête}$  et  $US_5 = *$ ,
- $l_1, l_2, l_3, l_4$  : quatre segments image.

a) Exemple d'une hypothèse acceptée :

- Contrainte ultrasonore obtenue : coin,
- Correspondances :  $(L_2 = \text{coin avec } l_1)$ ,  $(L_3 = * \text{ avec } l_2)$  et  $(L_5 = * \text{ avec } l_4)$ .

Dans cette situation, l'hypothèse est acceptée car il y a une correspondance entre la contrainte ultrasonore obtenue et le premier appariement entre  $L_2$  et  $l_1$ .

b) Exemple d'une hypothèse non acceptée :

- Contrainte ultrasonore obtenue : arête,
- Correspondances :  $(L_1 = \text{plan avec } l_2)$ ,  $(L_2 = \text{coin avec } l_3)$  et  $(L_5 = * \text{ avec } l_4)$ .

Dans ce cas, l'hypothèse n'est pas acceptée car il n'y a pas de correspondance entre la contrainte ultrasonore obtenue et un des trois appariements.

#### 4-3 Deuxième amélioration : Post-combinaison

La deuxième approche est appelée Post-fusion car la contrainte ultrasonore est utilisée d'emblée à l'entrée de l'algorithme d'appariement (Figure 3-9). Le modèle global enrichi contient l'ensemble de segments représentant l'environnement, ce qui représente un nombre important de segments à traiter.

L'idée de base de cette méthode est de garder uniquement les segments intéressants du modèle global et de les utiliser dans le processus d'appariement. Ces segments sont regroupés dans un modèle partiel.

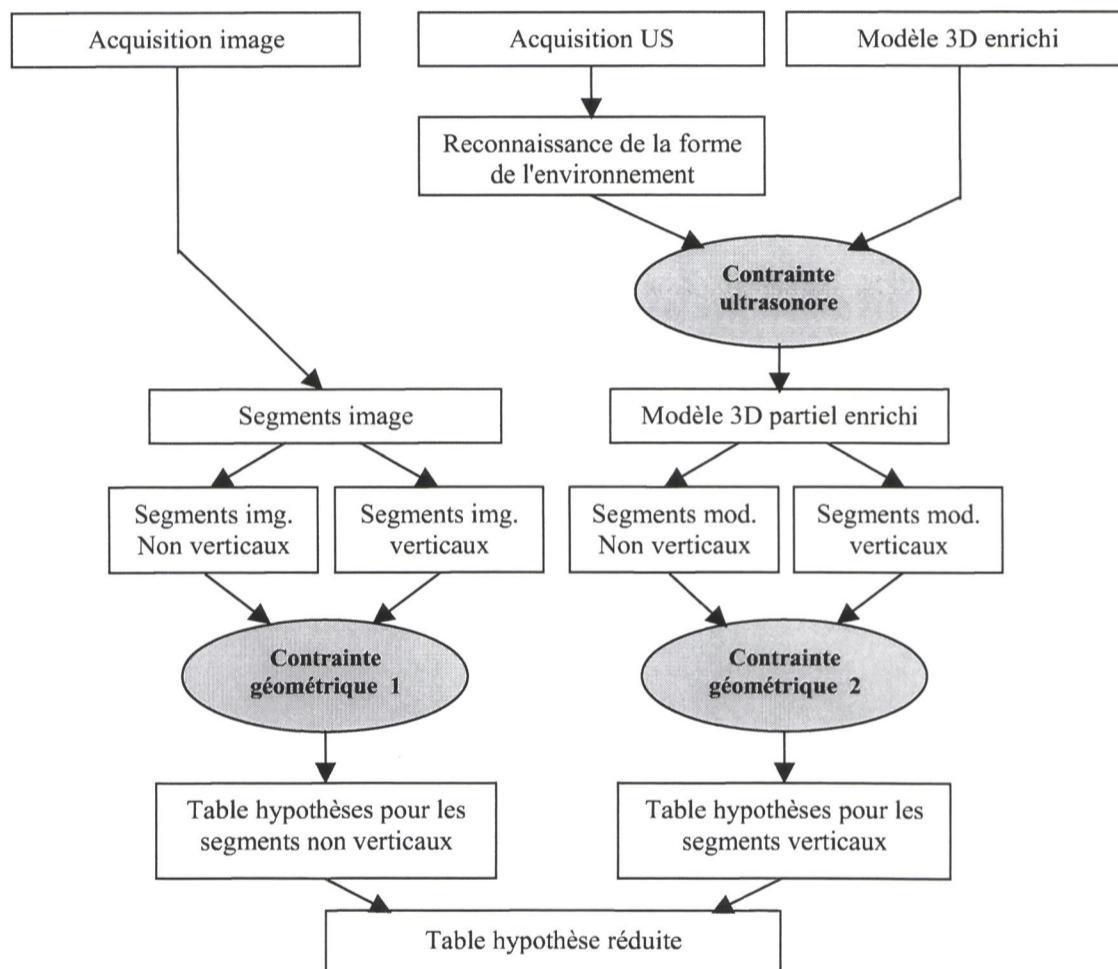


Figure 3-9. Schéma de l'appariement avec une post-combinaison.

Il y a deux catégories de segments que nous allons garder pour l'appariement. La première catégorie contient les segments comportant une étiquette ultrasonore qui correspond à celle obtenue en ligne par le système ultrasonore. La deuxième catégorie contient les segments appelés de liaison, notés par un numéro et une étiquette "\*\*".

Ces deux catégories de segments sont regroupées dans un nouveau modèle appelé modèle 3D partiel enrichi et les segments du modèle global non retenus sont supprimés. Le modèle

partiel obtenu nous a permis d'avoir moins d'hypothèses dans la table réduite à la fin du processus de mise en correspondance.

#### 4-4 Comparaison des trois approches

Dans cette partie, nous allons calculer le nombre d'hypothèses obtenu par chaque approche avec les trois poses de l'environnement fixées auparavant. Chaque pose correspond à un des trois formes : plan, coin et arête.

Nous calculons aussi le gain en nombre d'hypothèses de la deuxième (Pré-combinaison) et la troisième méthode (Post-combinaison) par rapport à la méthode initiale. Tous ces résultats sont présentés dans La Table.3-5. A la fin, on va comparer les différents résultats obtenus et choisir la meilleure approche d'appariement 2D / 3D.

Poses	Formes	Méthode sans ultrasons	Pré-combinaison		Post-combinaison	
		N° d'hypothèses	N° d'hypothèses	Gain	N° d'hypothèses	Gain
1	Plan	55	28	49.1 %	19	65.4 %
2	Coin	224	98	56.2 %	76	66.1 %
3	Arête	688	340	50,6 %	218	68.3 %

**Table 3-5.** Comparaison des trois méthodes d'appariement.

L'utilisation de la contrainte ultrasonore dans l'appariement réduit considérablement le nombre d'hypothèses généré. En effet, avec la deuxième méthode nous obtenons un gain moyen de 52 % par rapport au nombre d'hypothèses généré par la méthode initiale.

La troisième méthode quant à elle, le gain moyen atteint une valeur de 66 %. Plus de résultats concernant la réduction du nombre d'hypothèses d'appariement sont donnés dans la partie résultat dans la fin de ce chapitre.

Nous constatons que l'approche basée sur la méthode post-combinaison donne les meilleurs résultats d'appariement, de plus cette méthode supplanté les autres car elle permet :

- D'éliminer dès le départ les segments inutiles,
- De garder des segments ayant la même étiquette que la contrainte ultrasonore afin de rendre les hypothèses d'appariements plus fiables,
- De réduire considérablement le temps de traitement global de l'appariement puisque le nombre d'hypothèses à traiter est plus faible.

#### 5- Transformation repère caméra / repère modèle

Dans cette partie, nous nous intéressons au problème de la détermination de la position et de l'orientation de la caméra et ainsi du robot dans l'environnement de localisation. Plus formellement étant donné :

- Un ensemble de points ou de droites 3D décrits dans un repère global de l'environnement,

- Les projections 2D de ces points ou droites sur l'image décrites dans le repère caméra,
- Les paramètres intrinsèques de la caméra.

L'objectif est de déterminer la transformation rigide qui permet de passer des coordonnées de ces points ou droites dans le repère global vers les coordonnées du robot dans ce même repère appelé aussi repère objet ou repère monde. Cette transformation est composée d'une rotation et d'une translation.

Le choix du formalisme mathématique pour définir cette transformation est alors déterminant pour les performances de chaque méthode. Dombre présente dans son livre [Dom.88] plusieurs solutions pour résoudre cette transformation, et ainsi la représentation la plus naturelle définit la rotation par ses angles d'Euler  $\theta$ ,  $\phi$  et  $\psi$  et la translation par un vecteur de l'espace.

Le travail initial exposé dans le chapitre précédent utilise une solution basée sur les quaternions unitaires inspirée des travaux de Phong et Horaud. Dans notre travail on s'intéresse plus à l'approche présentée par Lowe dans [Low.87].

Dans la suite de ce chapitre on va exposer les formalismes mathématiques de ces deux méthodes. Pour cela, on va donner tout d'abord la présentation de la rotation par les angles d'Euler et par les quaternions unitaires, puis la fonction d'erreur à minimiser pour une approche basée sur une correspondance de droites.

### 5-1 Présentation de la rotation par les angles d'Euler

Il existe plusieurs façons pour représenter une rotation dans le domaine de la vision par ordinateur. Le choix se fait suivant le contexte de l'application et la complexité des calculs.

Les deux manières d'exprimer la rotation qui nous intéresse pour définir les deux types de transformations repère caméra / repère modèle sont la représentation par les quaternions unitaires que nous avons présentée dans le chapitre II, et la représentation par les angles.

Cette dernière, utilise la décomposition de la matrice de rotation en trois matrices de rotation élémentaires, chacune autour d'un des trois axes de l'espace  $X$ ,  $Y$ ,  $Z$ . Ces trois matrices sont :

$$R_x = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\psi & -\sin\psi \\ 0 & \sin\psi & \cos\psi \end{pmatrix} \quad (3-9)$$

$$R_y = \begin{pmatrix} \cos\phi & 0 & \sin\phi \\ 0 & 1 & 0 \\ -\sin\phi & 0 & \cos\phi \end{pmatrix} \quad (3-10)$$

$$R_z = \begin{pmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad (3-11)$$

La matrice de rotation  $R$  est le produit des trois angles d'Euler. Elle est donnée généralement par le produit des trois matrices :

$$R = R_z \ R_y \ R_x \quad (3-12)$$

### 5-2 Méthode de Lowe

Soit un repère modèle 3D d'origine  $O$  lié à l'environnement et un repère 3D d'origine  $F$  lié à une caméra montée sur un robot mobile (Figure.2-7). L'axe des  $z$  du repère caméra est confondu avec l'axe optique.

Soit un point  $P$  du modèle 3D exprimé dans le repère lié à ce dernier par les coordonnées  $X$ ,  $Y$  et  $Z$ , et exprimé dans le repère caméra par les coordonnées  $x_c$ ,  $y_c$  et  $z_c$ , ainsi on a :

$$(x_c, y_c, z_c) = R(P - T) \quad (3-13)$$

Où  $R$  et  $T$  sont respectivement la matrice de rotation et le vecteur de translation, tous les deux relient le repère caméra et le repère modèle. Les coordonnées  $u$  et  $v$  du pixel de projection du point  $P$  sur le plan image sont données de la façon suivante, où  $f$  est la distance focale :

$$(u, v) = (f \frac{x_c}{z_c}, f \frac{y_c}{z_c}) \quad (3-14)$$

#### 5-2-1 Fonction d'erreur

Le problème à résoudre dans ce cas est le suivant :

Étant donné un ensemble de  $N$  points  $P_i$  ( $i = 1, 2, \dots, N$ ) appartenant au modèle de l'environnement 3D et leurs projections théoriques respectives sur le plan image sont  $(u_i, v_i)$  et l'ensemble des pixels réels correspondants à ces points et observés sur l'image sont  $(\hat{u}_i, \hat{v}_i)$ , calculer la rotation optimale  $R$  et la translation optimale  $T$  minimisant la fonction d'erreur suivante :

$$E = \sum_{i=1}^N (u_i - \hat{u}_i)^2 + (v_i - \hat{v}_i)^2 \quad (3-15)$$

Une droite observée sur l'image peut être caractérisée par l'angle  $\hat{\rho}$  et la distance perpendiculaire à l'origine du repère  $\hat{d}$  avec la présentation suivante :

$$\hat{d}_i = \cos(\hat{\rho}_i) \cdot \hat{u}_i + \sin(\hat{\rho}_i) \cdot \hat{v}_i \quad (3-16)$$

Dans laquelle  $(\hat{u}_i, \hat{v}_i)$  est un point appartenant à cette droite. Pour un autre point quelconque appartenant à cette droite  $(u_i, v_i)$ .

Sa distance à l'origine dans la même direction  $\hat{\rho}$  est donnée par :

$$d_i = \cos(\hat{\rho}_i)u_i + \sin(\hat{\rho}_i)v_i \quad (3-17)$$

La distance entre le point  $(u_i, v_i)$  et la droite considérée est donnée par  $|\hat{d}_i - d_i|$ .

Si on considère ce point comme la projection théorique d'un point 3D,  $P_i$  appartenant à un segment du modèle, la distance théorique  $d_i$  peut être alors exprimée en fonction des six paramètres d'orientation et de position en remplaçant dans (3-17)  $u_i$  et  $v_i$  par leurs expressions issues de (3-14).

En prenant deux points arbitraires de deux segments, la fonction d'erreur dans le cas de correspondance des droites peut être écrite comme suite :

$$E = \sum_{i=1}^N (d_{1i} - \hat{d}_i)^2 + (d_{2i} - \hat{d}_i)^2 \quad (3-18)$$

Avec  $d_{1i}$  et  $d_{2i}$  sont les distances à l'origine des deux points arbitraires données par l'expression (3-17), on peut écrire encore :

$$E = \sum_{i=1}^N ((\cos(\hat{\rho}_i)u_{1i} + \sin(\hat{\rho}_i)v_{1i}) - \hat{d}_i)^2 + ((\cos(\hat{\rho}_i)u_{2i} + \sin(\hat{\rho}_i)v_{2i}) - \hat{d}_i)^2 \quad (3-19)$$

Chaque correspondance entre un segment modèle et un segment image fournit deux équations, ainsi, il faut donc au minimum trois correspondances pour déterminer les six inconnues.

Le robot opère dans un environnement 3D, le déplacement de la caméra quant à lui, se fait dans un plan horizontal situé à une hauteur fixe (déplacement site / azimut). Cette contrainte permet de réduire le nombre de degrés de liberté de six à trois : une rotation  $\Phi$  et deux translations suivant  $x$  et  $y$ .

La rotation entre les deux repères est constituée d'une rotation connue  $\theta$  autour de l'axe  $y$  du repère caméra et d'une rotation inconnue  $\Phi$  autour du nouvel axe  $z$  obtenu après la première rotation.

A partir de ces données et la présentation de la matrice de rotation donnée par les angles d'Euler dans (3-45), la matrice de rotation s'exprime comme la suit :

$$R = \begin{pmatrix} \cos(\theta)\cos(\phi) & -\cos(\theta)\sin(\phi) & \sin(\theta) \\ \sin(\phi) & \cos(\phi) & 0 \\ -\sin(\theta)\cos(\phi) & \sin(\theta)\sin(\phi) & \cos(\theta) \end{pmatrix} \quad (3-20)$$

De la même façon, le vecteur translation est exprimé dans le repère modèle avec  $z$  constant et connu. Le problème consiste donc à calculer trois paramètres pour la localisation qui sont, une rotation  $\Phi$  et deux translations  $x$  et  $y$ .

L'approche de Lowe peut être formulée ainsi en exprimant  $R$  en fonction de ces paramètres donnés dans (3-20). Pour chaque point  $P_i(x_i, y_i, z_i)$  l'équation de projection donnée par (3-14) s'écrit donc comme suit :

$$\begin{cases} u_i = f \cdot \frac{(x_i - x) \cos(\theta) \cos(\phi) - (y_i - y) \cos(\theta) \sin(\phi) + (z_i - z) \sin(\theta)}{-(x_i - x) \sin(\theta) \cos(\phi) + (y_i - y) \sin(\theta) \sin(\phi) + (z_i - z) \cos(\theta)} \\ v_i = \frac{(x_i - x) \sin(\phi) + (y_i - y) \cos(\phi)}{-(x_i - x) \sin(\theta) \cos(\phi) + (y_i - y) \sin(\theta) \sin(\phi) + (z_i - z) \cos(\theta)} \end{cases} \quad (3-21)$$

Les coordonnées du vecteur translation  $T(x, y, z)$  sont exprimées dans l'expression (3-21) directement dans le repère modèle, et  $z$  a une valeur connue (hauteur de la caméra).

Ainsi, la réduction du nombre de variables permet de réduire le nombre minimal de primitives nécessaires à la résolution des équations.

Par contre cette formulation reste complexe, car si on exprime la translation dans le repère caméra "t" plutôt que dans le repère modèle "T", l'équation (3-12) devient plus simple :

$$(x_c, y_c, z_c) = RP + t \quad (3-22)$$

Ce qui donne des équations de projections plus simples que dans (3-56), ainsi :

$$\begin{cases} u_i = f \cdot \frac{x_i \cos(\theta) \cos(\phi) - y_i \cos(\theta) \sin(\phi) + z_i \sin(\theta) + x_c}{-x_i \sin(\theta) \cos(\phi) + y_i \sin(\theta) \sin(\phi) + z_i \cos(\theta) + z_c} \\ v_i = \frac{x_i \sin(\phi) + y_i \cos(\phi) + y_c}{-x_i \sin(\theta) \cos(\phi) + y_i \sin(\theta) \sin(\phi) + z_i \cos(\theta) + z_c} \end{cases} \quad (3-23)$$

La rotation et la translation sont mieux découplées mais le nombre de degrés de liberté (4 DDL) reste plus élevé que le nombre d'inconnues d'un système réel à 3 DDL.

La composante  $x_c$  n'est pas indépendante de  $z_c$  grâce à la relation qui lie la translation dans le repère caméra et la translation dans le repère modèle donnée par :

$$t = -RT \quad (3-24)$$

D'où les deux expressions suivantes :

$$\begin{cases} x_c = -x \cos(\theta) \cos(\phi) + y \cos(\theta) \sin(\phi) - z \sin(\theta) \\ z_c = x \sin(\theta) \cos(\phi) - y \sin(\theta) \sin(\phi) - z \cos(\theta) \end{cases} \quad (3-25)$$

L'expression de  $x_c$  en fonction de  $z_c$  est donnée alors par :

$$x_c = -z_c \cot(\theta) - \frac{z}{\sin(\theta)} \quad (3-26)$$

Maintenant, nous revenons à un système à trois inconnues :  $\phi$ ,  $x_c$  et  $y_c$ . Nous pouvons à présent réécrire les expressions de  $u_i$  et de  $v_i$  sous la représentation suivante :

$$\begin{cases} u_i = \alpha_u \frac{x_i \cos(\theta) \cos(\phi) - y_i \cos(\theta) \sin(\phi) + z_i \sin(\theta) - \cot(\theta)z_c - z/\sin(\theta)}{-x_i \sin(\theta) \cos(\phi) + y_i \sin(\theta) \sin(\phi) + z_i \cos(\theta) + z_c} + u_0 \\ v_i = \alpha_v \frac{x_i \sin(\phi) + y_i \cos(\phi) + y_c}{-x_i \sin(\theta) \cos(\phi) + y_i \sin(\theta) \sin(\phi) + z_i \cos(\theta) + z_c} + v_0 \end{cases} \quad (3-27)$$

Ainsi, l'équation (3-52) est donnée par l'expression suivante :

$$d_i = \frac{\alpha_u \cos(\hat{\rho})(x_i \cos(\theta) \cos(\phi) - y_i \cos(\theta) \sin(\phi) + z_i \sin(\theta) - \cot(\theta)z_c - z/\sin(\theta)) + \alpha_v \sin(\hat{\rho})(x_i \sin(\phi) + y_i \cos(\phi) + y_c)}{-x_i \sin(\theta) \cos(\phi) + y_i \sin(\theta) \sin(\phi) + z_i \cos(\theta) + z_c} + d_{0i} \quad (3-28)$$

avec le terme additif  $d_{0i}$  donné par :

$$d_{0i} = \cos(\hat{\rho}_i) \cdot u_0 + \sin(\hat{\rho}_i) \cdot v_0 \quad (3-29)$$

Pour deux points  $P_{1i}(x_{1i}, y_{1i}, z_{1i})$  et  $P_{2i}(x_{2i}, y_{2i}, z_{2i})$  appartenant à un segment modèle 3D, la fonction d'erreur donnée par l'équation (3-27) devient :

$$\begin{aligned} E = \sum_{i=1}^N & \left( \frac{A_{1i} \sin(\phi) + B_{1i} \sin(\phi) + C_{1i} y_c + D_{1i} z_c + E_{1i}}{F_{1i} \cos(\phi) + G_{1i} \sin(\phi) + H_{1i} z_c + I_{1i}} - \hat{d}_i \right)^2 \\ & + \left( \frac{A_{2i} \sin(\phi) + B_{2i} \sin(\phi) + C_{2i} y_c + D_{2i} z_c + E_{2i}}{F_{2i} \cos(\phi) + G_{2i} \sin(\phi) + H_{2i} z_c + I_{2i}} - \hat{d}_i \right)^2 \end{aligned} \quad (3-30)$$

avec pour  $i = 1, 2$  et  $j = 1, 2, \dots, N$ , les termes de  $A_{ij}$ ,  $B_{ij}, \dots, I_{ij}$  sont donnés par les expressions suivantes :

$$\begin{cases} A_{ij} = \alpha_u x_{ij} \cos(\hat{\rho}_i) \cos(\theta) + \alpha_v y_{ij} \cos(\hat{\rho}) \\ B_{ij} = -\alpha_u y_{ij} \cos(\hat{\rho}_i) \cos(\theta) + \alpha_v x_{ij} \sin(\hat{\rho}_i) \\ C_{ij} = \alpha_v \sin(\hat{\rho}_i) \\ D_{ij} = \alpha_u \cos(\hat{\rho}_i) \cot(\theta) \\ E_{ij} = \alpha_u \cos(\hat{\rho}_i) (z_{ij} \sin(\theta) - z/\sin(\theta)) \\ F_{ij} = -x_{ij} \sin(\theta) \\ G_{ij} = y_{ij} \sin(\theta) \\ H_{ij} = 1 \\ I_{ij} = z_{ij} \cos(\theta) \end{cases} \quad (3-31)$$

En posant comme vecteur inconnu :  $q = [\cos(\Phi) \quad \sin(\Phi) \quad y_c \quad z_c]$  qui représente les trois paramètres de la localisation inconnus, la fonction d'erreur devient (3-30) :

$$E = \sum_{i=1}^N \left( \frac{A_{1i}q_1 + B_{1i}q_2 + C_{1i}q_3 + D_{1i}q_4 + E_{1i}}{F_{1i}q_1 + G_{1i}q_2 + H_{1i}q_3 + I_{1i}} - \hat{d}_i \right)^2 + \left( \frac{A_{2i}q_1 + B_{2i}q_2 + C_{2i}q_3 + D_{2i}q_4 + E_{2i}}{F_{2i}q_1 + G_{2i}q_2 + H_{2i}q_3 + I_{2i}} - \hat{d}_i \right)^2 + \lambda(q_1^2 + q_2^2 - 1)^2 \quad (3-32)$$

Le terme  $\lambda(q_1^2 + q_2^2 - 1)^2$  a été ajouté pour garantir l'orthogonalité de la matrice de rotation obtenue, le scalaire  $\lambda$  permet de pondérer l'influence de cette contrainte.

Ainsi, en calculant les dérivées partielles de  $d_{ij}$  par rapport à  $q_1, q_2, q_3$  et  $q_4$  et en utilisant la méthode de Gauss-Newton ou de Levenberg-Marquardt, nous pouvons minimiser la fonction d'erreur donnée par (3-32).

### 5-2-2 Formulation adaptée à la robotique mobile

Nous allons exprimer à présent les deux contraintes dans le cas réel avec trois degrés de liberté en adoptant l'équation (3-20) pour la rotation et en exprimant la translation dans le repère de la caméra. En remplaçant  $R$  par son expression, les contraintes (3-28) et (3-30) deviennent :

$$\begin{cases} A \cos(\phi) + B \sin(\phi) + C = 0 \\ D \cos(\phi) + E \sin(\phi) + Fy_c + Gz_c + H = 0 \end{cases} \quad (3-33)$$

Ainsi, la fonction d'erreur finale obtenue avec cette méthode en posant toujours  $q$  comme un vecteur inconnu a quatre variables :  $[\cos(\Phi) \quad \sin(\Phi) \quad y_c \quad z_c]$  :

$$E(R, t) = \sum_{i=1}^N (A_i q_1 + B_i q_2 + C_i)^2 + (D_i q_1 + E_i q_2 + F_i q_3 + G_i q_4 + H_i)^2 + \lambda(q_1^2 + q_2^2 - 1)^2 \quad (3-34)$$

avec :

$$\begin{cases} A_i = (n_{1i} - n_{3i}) V_{1i} \cos(\theta) + n_{2i} V_{2i} \\ B_i = n_{2i} V_{1i} - (n_{1i} - n_{3i}) V_{2i} \sin(\theta) \\ C_i = n_{1i} V_{3i} \sin(\theta) + n_{3i} V_{3i} \sin(\theta) \\ D_i = (n_{1i} - n_{3i}) P_{1i} \cos(\theta) + n_{2i} P_{2i} \\ E_i = n_{2i} P_{1i} - (n_{1i} - n_{3i}) P_{2i} \sin(\theta) \\ F_i = n_{2i} \\ G_i = -\cot(\theta) n_{1i} + n_{3i} \\ H_i = n_{1i} (P_{3i} \sin(\theta) - z / \sin(\theta)) + n_{3i} P_{3i} \sin(\theta) \end{cases} \quad (3-35)$$

et :

$n_i(n_{1i}, n_{2i}, n_{3i})$  : Vecteur normal au plan d'interprétation contenant le segment,

$P_i(P_{1i}, P_{2i}, P_{3i})$  : Vecteur position d'un segment,

$V_i(V_{1i}, V_{2i}, V_{3i})$  : Vecteur directeur d'un segment.

La résolution de la fonction d'erreur donnée par l'équation (3-34) peut être fait par deux méthodes. La première se base sur une solution non linéaire et la seconde sur une solution linéaire.

### 5-2-3 Solution non linéaire

Le système (3-33) est surdéterminé avec au moins deux correspondances. Il peut être résolu par la méthode de Gauss-Newton ou de Levenberg-Marquardt.

Le calcul des dérivées partielles des termes de l'équation par rapport à  $q[q_1 \ q_2 \ q_3 \ q_4]$  est simple à déterminer car les dérivées des termes de la forme  $(A_i q_1 + B_i q_2 + C_i)$  sont :

$$\begin{cases} \frac{\partial(A_i q_1 + B_i q_2 + C_i)}{\partial q_1} = A_i \\ \frac{\partial(A_i q_1 + B_i q_2 + C_i)}{\partial q_2} = B_i \end{cases} \quad (3-36)$$

Et les dérivées des termes de la forme  $(D_i q_1 + E_i q_2 + F_i q_3 + G_i q_4 + H_i)$  sont :

$$\begin{cases} \frac{\partial(D_i q_1 + E_i q_2 + F_i q_3 + G_i q_4 + H_i)}{\partial q_1} = D_i \\ \frac{\partial(D_i q_1 + E_i q_2 + F_i q_3 + G_i q_4 + H_i)}{\partial q_2} = E_i \\ \frac{\partial(D_i q_1 + E_i q_2 + F_i q_3 + G_i q_4 + H_i)}{\partial q_3} = F_i \\ \frac{\partial(D_i q_1 + E_i q_2 + F_i q_3 + G_i q_4 + H_i)}{\partial q_4} = G_i \end{cases} \quad (3-37)$$

Il est ainsi possible dans cette situation de découpler la rotation et la translation dans la formule d'erreur obtenue (3-34).

Ceci revient à trouver dans un premier temps l'angle d'orientation en minimisant le système d'équation formé par la première contrainte pour chaque mise en correspondance. Ainsi la fonction d'erreur concernant la rotation devient :

$$f(R) = \sum_{i=1}^N (A_i q_1 + B_i q_2 + C_i)^2 + \lambda(q_1^2 + q_2^2 - 1) \quad (3-38)$$

Le résultat obtenu est injecté dans la deuxième équation pour calculer la translation :

$$f(t) = \sum_{i=1}^N (D_i q_1 + E_i q_2 + F_i q_3 + G_i q_4 + H_i)^2 + \lambda (q_1^2 + q_2^2 - 1)^2 \quad (3-39)$$

Comme pour la méthode des quaternions, les résultats de localisation obtenus avec cette méthode sont de bonne qualité par contre cette méthode est très coûteuse en terme de temps de calcul.

La solution linéaire du critère (3-34) présentée par la suite va pallier cet inconvénient.

#### 5-2-4 Solution linéaire

Cette solution a l'avantage d'être très rapide à condition d'avoir une bonne précision sur l'initialisation (pose initiale). Dans notre travail nous utilisons un système odométrique qui fournit une pose initiale de qualité acceptable s'il n'y a pas de glissement.

Le critère à minimiser prend une forme linéaire représentée par l'expression :

$$M \cdot q = L \quad (3-40)$$

où  $q [q_1 \ q_2 \ q_3 \ q_4]$  est le vecteur optimal que nous cherchons à calculer et qui nous donne l'orientation du robot  $\Phi$  et ses positions  $x$  et  $y$ . Ainsi, avec cette méthode il y a un seul critère à résoudre donné par :

$$f(q) = \sum_{i=1}^N (M_i \cdot \hat{q} - L_i)^2 = (L - M\hat{q})^T (L - M\hat{q}) \quad (3-41)$$

Avec la matrice  $M$  donnée par :

$$M = \begin{bmatrix} A_1 & B_1 & 0 & 0 \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ A_N & B_N & 0 & 0 \\ D_1 & E_1 & F_1 & G_1 \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ D_N & E_N & F_N & G_N \end{bmatrix} \quad (3-42)$$

et le vecteur  $L$  donné par :

$$L = \begin{bmatrix} C_1 \\ \vdots \\ C_N \\ H_1 \\ \vdots \\ H_N \end{bmatrix} \quad (3-43)$$

La solution au sens des moindres carrés du critère (3-41) est obtenue quand la dérivée par rapport à  $q$  est nulle, ainsi la solution globale est de la forme :

$$\hat{q} = (M^T M)^{-1} M^T L \quad (3-44)$$

Cette méthode linéaire nous offre un gain temps de calcul très important par rapport à la première méthode. C'est la méthode que nous allons adopter par la suite.

La partie résultats va permettre de quantifier le gain en temps de traitement obtenu avec cette méthode par rapport à la méthode initiale basée sur les quaternions et par rapport aussi à la méthode de résolution non linéaire présentée précédemment.

### 5-3 Comparaison des trois méthodes

Dans la Table.3-6, nous allons comparer les temps de traitement obtenus avec les trois méthodes de transformation présentées sur les trois poses de l'environnement vu auparavant.

Nous allons calculer aussi le pourcentage de gain en temps de calcul entre les deux solutions non-linéaire et linéaire de la méthode de Lowe par rapport à la méthode initiale basée sur l'approche de Phong-Horaud et les quaternions. La méthode d'appariement utilisée dans ces calculs est la post-combinaison.

Poses	Formes	Méthode des quaternions	Méthode de Lowe non linéaire		Méthode de Lowe linéaire	
		Temps (s)	Temps (s)	Gain	Temps (s)	Gain
1	Plan	28	12	57.1 %	0.40	98.5 %
2	Coin	69	27	60.9 %	0.82	98.8 %
3	Arête	116	49	57.8 %	1.04	99.1 %

**Table 3-6.** Comparaison des trois méthodes de transformation.

La méthode basée sur les quaternions donne des temps de calcul très importants dépassant la minute. La méthode basée sur la solution non-linéaire de l'approche de Lowe offre quant à elle, un gain en temps de l'environ 50 % par rapport à la première.

La solution linéaire de Lowe offre un temps de calcul de l'ordre d'une seconde pour chaque localisation et apporte un gain d'environ 98 % par rapport à la méthode initiale. Nous choisissons cette dernière solution comme méthode de transformation.

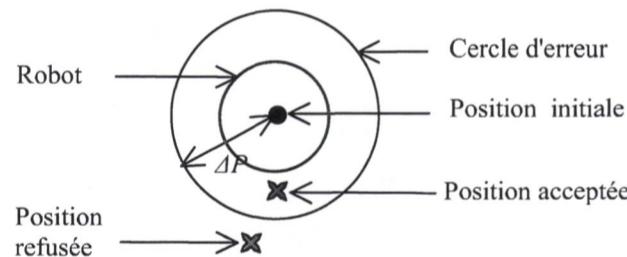
## 6- Choix de la meilleure pose

L'application de la transformation "repère caméra / repère modèle" sur la table d'hypothèses issue de la phase d'appariement 2D / 3D va donner une nouvelle table contenant pour chaque hypothèse de localisation, une pose d'ordre  $i$  donnée par les coordonnées  $X_i$ ,  $Y_i$ , et  $\Phi_i$ , où l'élément  $i$  est le numéro de l'hypothèse dans la table.

Dans un premier temps et avant de choisir la meilleure pose dans la table de transformations, nous allons considérer une erreur de position  $\Delta P$  présentée par un cercle d'erreur autour de la position initiale donnée par le système odométrique (Figure.3-10).

Le centre de ce cercle est la position initiale et son rayon  $\Delta P$  correspond au double du rayon du robot (50 cm). Dans un but de réduire le temps de calcul et rendre le choix de la meilleure pose plus fiable, nous avons décidé de ne pas traiter les hypothèses improbables.

Une hypothèse sera dite improbable si elle n'appartient pas au cercle d'erreur défini autour de la pose supposée du robot.



**Figure 3-10.** Cercle d'erreur autour du robot.

Avec cette procédure, nous allons s'assurer que la pose finale serait acceptée comme solution pour la localisation. Donc, si la position  $i$  dans la table de transformations appartient à ce cercle d'erreur elle va satisfaire les deux expressions :

$$\begin{cases} X_i \leq \Delta P \\ Y_i \leq \Delta P \end{cases} \quad (3-45)$$

Ainsi, elle va être acceptée et considérée comme une solution potentielle. Si elle ne satisfait pas les deux expressions de l'équation (3-45) elle sera rejetée de la table des transformations.

Ainsi, à la fin, le choix de la meilleure position va se faire par rapport aux solutions qui appartiennent uniquement au cercle d'erreur.

Pour obtenir la meilleure hypothèse dans cette table, qui correspond probablement à la meilleure position et orientation du robot dans son environnement, nous avons besoin d'une fonction de choix.

La méthode de localisation initiale se basait sur la comparaison des poses de la table de transformations avec la pose initiale donnée par le système odométrique, comme expliqué dans le chapitre précédent.

Pour mon travail, deux améliorations ont été développées pour garantir le choix de la meilleure solution. La première utilise aussi une comparaison entre la pose du robot dans la table et la pose initiale, par contre dans cette approche, la différence se fait suivant l'axe des  $X$ , l'axe des  $Y$  et l'axe des  $\Phi$ .

La deuxième approche utilise une stratégie différente, la distance euclidienne entre les segments de l'image et du modèle correspondant à une pose donnée dans la table assure le choix de la meilleure pose. Dans la suite, nous allons voir en détail ces trois méthodes de sélection.

### 6-1 Amélioration 1 : Les trois meilleures hypothèses

Dans ce premier cas, on utilise également la position initiale donnée par le système odométrique. Par contre l'erreur est mesurée par rapport à chacun des axes  $X$ ,  $Y$  et  $\Phi$  séparément. Ainsi nous aurons trois équations d'erreurs, une pour chaque axe. Ainsi pour l'axe  $X$  nous obtenons :

$$F_x = \text{Min} \left( \sum_{i=1}^N (X_i - X_0)^2 \right) \quad (3-46)$$

pour l'axe  $Y$  nous avons :

$$F_y = \text{Min} \left( \sum_{i=1}^N (Y_i - Y_0)^2 \right) \quad (3-47)$$

et pour l'axe  $\Phi$  nous avons :

$$F_\phi = \text{Min} \left( \sum_{i=1}^N (\phi_i - \phi_0)^2 \right) \quad (3-48)$$

La solution que nous gardons à la fin de ce processus correspond automatiquement à la meilleure pose possible dans la table de transformations suivant  $X$ ,  $Y$  et  $\Phi$ . C'est pour cette raison que nous avons appelé cette approche la méthode des trois hypothèses car, le meilleur  $X$ , le meilleur  $Y$  et la meilleure orientation  $\Phi$  peuvent se trouver dans la même pose de la table de transformations, dans deux poses différentes ou dans trois poses.

Ainsi, le premier inconvénient de cette amélioration est que la solution gardée comporte au moins, trois angles de prise de vue différents.

Cette deuxième méthode donne de bons résultats de localisation. L'inconvénient est sa dépendance de l'information de pose donnée par le système odométrique. Une valeur erronée entraîne des erreurs considérables dans des mesures et des choix effectués.

Dans la partie suivante nous proposons une méthode qui présente l'avantage d'être indépendante de la pose initiale.

### 6-2 Amélioration 2 : La fonction coût d'appariement

Chaque pose dans la table de transformations correspond à une mise en correspondance contenant un nombre d'appariement supérieur ou égal à trois, entre les segments image 2D et les segments modèle 3D.

Nous ne pouvons pas mesurer directement la distance entre ces deux groupes de segments car chaque groupe est exprimé dans un repère différent. Repère image pour les segments image et repère modèle pour les segments du modèle.

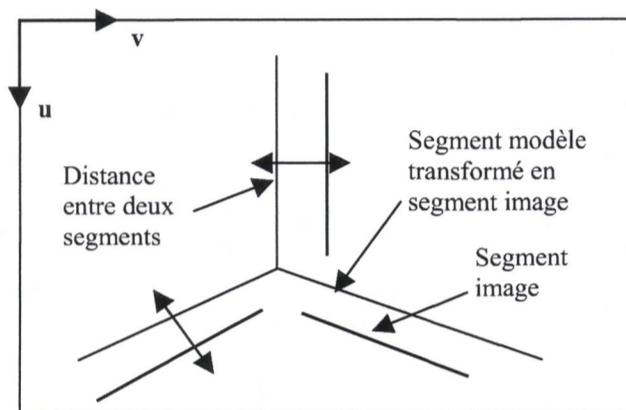


Figure 3-11. Projection des segments modèle transformés sur l'image.

C'est pour cette raison que nous allons transformer les segments du modèle pour chaque pose de la table de transformations en segments image par l'intermédiaire du modèle géométrique de la caméra.

Une fois cette transformation effectuée, nous projetons ces segments sur l'image (Figure 3-11) puis nous mesurons la distance euclidienne dans l'espace entre ces deux ensembles de segments.

A la fin, on choisit la plus petite distance qui correspond à la somme de toutes les distances entre un segment de l'image et un autre du modèle projeté sur l'image. Cette plus petite distance correspond automatiquement à la meilleure hypothèse de localisation.

Nous allons établir dans la suite la fonction qui mesure la distance entre deux ensembles de segments dans l'image en rappelant tout d'abord la notion de distance entre deux segments dans l'espace.

#### 6-2-1 Distance entre deux segments dans l'espace

La métrique généralement utilisée pour exprimer la distance entre deux segments de droites dans l'espace liés par un déplacement rigide est obtenue en calculant la somme des carrés des distances entre des points appartenant aux segments.

Dans notre cas, pour déterminer la distance entre deux segments dans un environnement 3D, il suffit de connaître la distance séparant les centres des deux segments comme présentée sur la Figure.3-12.

Chaque centre de segment est caractérisé par trois coordonnées données par les trois équations suivantes suivant les axes  $X$ ,  $Y$  et  $Z$  :

$$\begin{cases} X_c^1 = \frac{X_1 + X_2}{2} \\ Y_c^1 = \frac{Y_1 + Y_2}{2} \\ Z_c^1 = \frac{Z_1 + Z_2}{2} \end{cases} \quad (3-49)$$

où  $(X_1, Y_1, Z_1)$  sont les coordonnées du point de début du segment et  $(X_2, Y_2, Z_2)$  sont les coordonnées du point de fin du segment. Ainsi, on détermine les coordonnées du centre de chaque segment. La distance entre ces deux centres est donnée par l'équation suivante :

$$D = \sqrt{(X_c^1 - X_c^2)^2 + (Y_c^1 - Y_c^2)^2 + (Z_c^1 - Z_c^2)^2} \quad (3-50)$$

où  $X_c^2$ ,  $Y_c^2$  et  $Z_c^2$  sont les coordonnées du centre du deuxième segment dans l'espace.

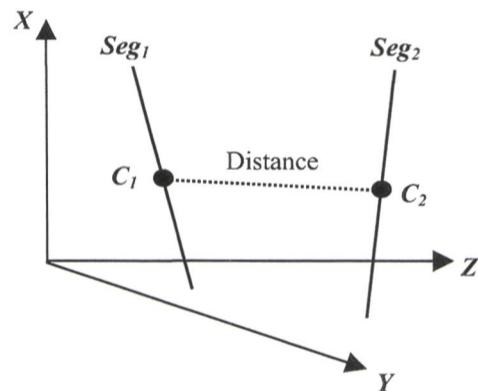


Figure 3-12. Mesure de distance dans l'espace 3D.

Cette même formule est valable pour mesurer les distances entre deux segments de l'image 2D, en prenant en considération les axes du repère image  $u$  et  $v$ .

#### 6-2-2 Distance entre deux ensembles de segments

Dans le cas de deux groupes de segments dans l'espace, la distance séparant ces deux groupes correspond à la somme de toutes les distances séparant chaque segment de l'image et le segment du modèle qui lui correspond dans la table d'hypothèses. Nous obtenons ainsi la distance globale séparant les deux ensemble  $Dist_g$  donnée par :

$$Dist_g = \sum_{i=1}^N D_i \quad (3-51)$$

Où  $D_i$  est la distance entre un segment modèle et un segment image et  $N$  est le nombre d'appariements dans une hypothèse de mise en correspondance.

La fonction choix dans ce cas correspond à la fonction suivante :

$$F = \min(Dist_g) \quad (3-52)$$

Ainsi, avec cette méthode l'initialisation donnée par le système odométrique n'intervient pas dans le choix de la meilleure solution et ne sert qu'à la détermination du cercle d'erreur autour de la position initiale. Nous sommes sûrs que l'hypothèse sélectionnée par la fonction (3-52) correspond à la meilleure solution de localisation pour le robot.

Pour ces deux raisons, cette dernière méthode basée sur la fonction de coût est considérée comme la plus fiable par rapport aux deux autres.

### 6-2-3 Comparaison des trois approches

La Table 3-7 donne l'erreur de position et l'erreur d'orientation, toutes les deux obtenues sur les trois poses de l'environnement présentées auparavant. Ces erreurs sont calculées avec les trois méthodes de choix de la meilleure pose présentées par rapport à la pose réelle mesurée au début de la phase d'acquisition, avec :

$$\Delta P = \left| \frac{x_0 + y_0}{2} - \frac{x_i + y_i}{2} \right| \quad (3-53)$$

et :

$$\Delta \phi = |\phi_0 - \phi_i| \quad (3-54)$$

Où  $[x_0 \ y_0 \ \phi_0]$  sont les coordonnées de la pose réelle et  $[x_i \ y_i \ \phi_i]$  sont les coordonnées de la meilleure pose choisie avec l'une des trois méthodes de choix.

Poses	Formes	Méthode de la pose initiale		Méthode des 3 hypothèses		Méthode de la fonction de coût	
		$\Delta P$ (cm)	$\Delta \phi$ (°)	$\Delta P$ (cm)	$\Delta \phi$ (°)	$\Delta P$ (cm)	$\Delta \phi$ (°)
1	Plan	8.26	2.64	3.97	1.20	1.36	0.12
2	Coin	12.1	1.88	5.71	0.93	2.78	0.35
3	Arête	23.9	3.91	8.04	1.77	4.32	0.58

Table 3-7. Comparaison des trois méthodes de choix.

Ces trois exemples de localisation montrent la précision apportée par la troisième méthode par rapport à la méthode qui fait référence à l'odométrie.

D'un point de vue quantitatif et comme nous l'avions constaté dans le chapitre II, la première méthode donne une erreur de position supérieure à 20 cm et une erreur de rotation supérieure à 3° par rapport à la pose réelle.

La méthode basée sur les trois meilleures solutions donne de meilleures valeurs de l'erreur mais qui reste un peu élevées par rapport à l'application dans un environnement d'intérieur restreint. Sans oublier que ces deux premières méthodes utilisent l'information odométrique initiale, et donc une dépendance du résultat obtenu avec la qualité de l'initialisation.

La troisième méthode contourne cet inconvénient par l'utilisation des distances entre les groupes de segments, et ainsi choisir la meilleure pose par rapport au coût le plus faible. Les erreurs obtenues avec cette méthode sont meilleures que celles obtenues avec les deux autres. La partie résultats va confirmer cette remarque.

## 7- Résultats

Dans cette partie nous présentons différents résultats associés à la procédure de localisation. Tout d'abord, nous montrons les environnements de travail utilisés pour la localisation d'intérieur du robot. Puis, nous illustrons la réduction du nombre d'hypothèses d'appariement grâce à l'utilisation de la contrainte ultrasonore et la réduction du temps de traitement nécessaire pour effectuer une localisation complète.

Nous donnons après les erreurs de précision en position et en orientation et enfin, nous présentons quelques exemples réels de localisation dans les environnements développés.

Nous évoquons également le cas où le système n'arrive pas à se localiser à cause de manque de segments appariés et nous étudions aussi le cas de localisation avec présence d'un obstacle dans l'environnement. Une évaluation globale de la méthode de localisation développée est donnée à la fin de la partie résultats.

### 7-1 Environnements de travail

Pour généraliser notre approche de localisation, nous avons choisi d'utiliser trois environnements de travail, une pièce du laboratoire (Figure.3-13), le couloir qui relie les différentes pièces du laboratoire (Figure.3-14) et enfin, une salle de classe (Figure.3-15).

Chaque environnement est représenté par un modèle 3D enrichi par les informations ultrasonores. Les obstacles dans le premier environnement, représentés par une armoire et une planche, sont répertoriés dans le modèle 3D. Par contre, dans le troisième environnement la table et l'armoire ne sont pas répertoriés.

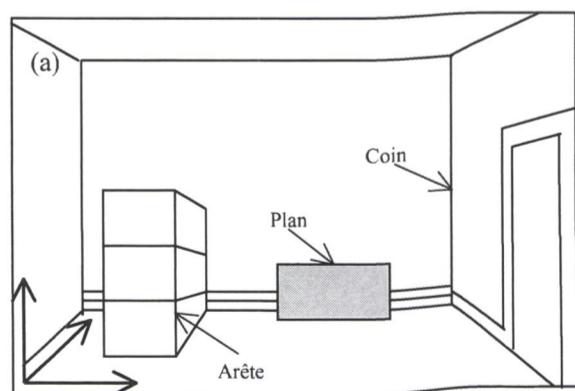


Figure 3-13. Premier environnement de travail : Pièce du laboratoire.

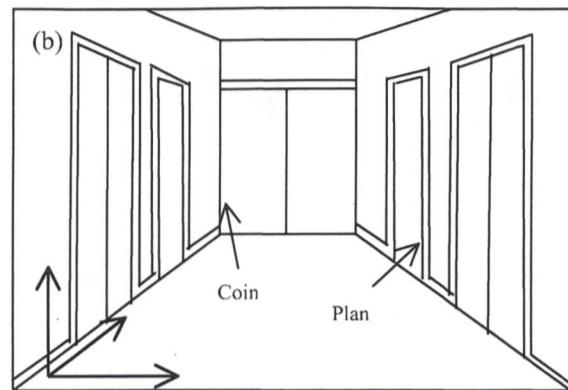


Figure 3-14. Deuxième environnement de travail : Couloir.

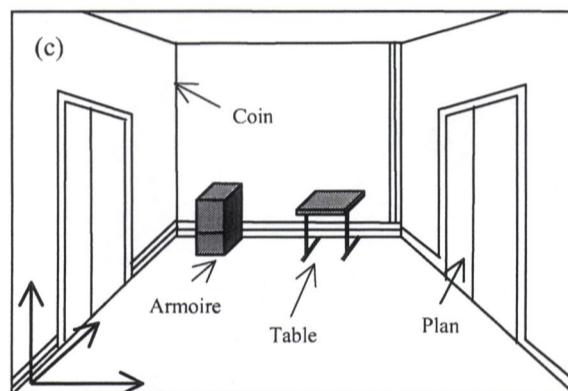


Figure 3-15. Troisième environnement de travail : Salle de classe.

Le choix du couloir comme environnement de travail du robot mobile est crucial car, pour passer d'une pièce à une autre dans un appartement, il est nécessaire de le traverser. Les dimensions des différentes pièces utilisées comme environnement de travail du robot sont données en 3D dans la Table 3-8, où la longueur correspond à l'axe des  $X$ , la largeur à l'axe des  $Y$  et la hauteur à l'axe des  $Z$  :

Les dimensions	Longueur (m)	Largeur (m)	Hauteur (m)
Pièce du laboratoire	2.32	1.20	2.8
Couloir du laboratoire	1.61	5.67	3.0
Salle de classe	3.78	4.24	3.0

Table 3-8. Dimensions des environnements de travail.

La partie localisation avec obstacles non répertoriés dans le troisième environnement va être traitée indépendamment des deux autres environnements dans la partie 5-7.

## 7-2 Réduction du nombre d'hypothèses

Avant de calculer le gain en nombre d'hypothèses apporté par les méthodes d'appariement développées, nous allons fixer pour chaque environnement de travail deux à trois poses qui

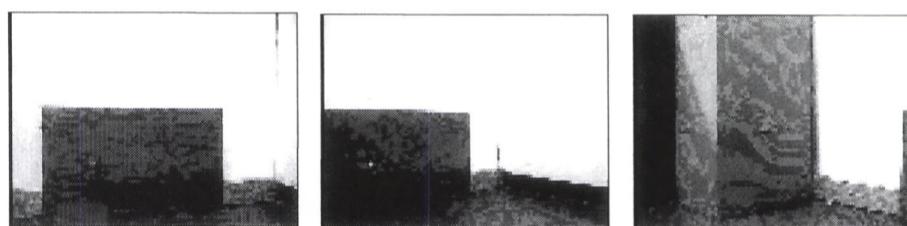
correspondent chacune à la position initiale donnée par le système odométrique et qui présentent une des trois formes de l'environnement (plan, coin ou arête). On note que dans le couloir on ne trouve pas d'arête, juste des plans et des coins.

Les coordonnées de position et de rotation pour chaque pose, ainsi que le nombre de segments image 2D détecté par la vision sont données dans la Table.3-9 au-dessous.

Modèle de l'environnement	Les coordonnées des poses initiales			Segments image détectés
	<i>x</i> (m)	<i>y</i> (m)	$\Phi$ (°)	
Pièce	Pose 1	1.15	-0.30	75
	Pose 2	0.52	-0.25	45
	Pose 3	1.72	-0.25	135
Couloir	Pose 1	1.20	1.00	135
	Pose 2	0.38	2.50	45

**Table 3-9.** Coordonnées des poses initiales utilisées.

Sur la Figure.3-16 nous présentons les trois photos de l'environnement associées aux trois poses de la pièce du laboratoire données sur la table précédente. Chacune des ces photos correspond à une des trois formes de l'environnement, plan, coin et arête.



**Figure 3-16.** Les trois photos de l'environnement présentant la pièce.

La Figure.3-17 donne les deux autres photos qui représentent les deux poses du couloir du laboratoire associées respectivement à un plan et un coin.



**Figure 3-17.** Les deux photos de l'environnement présentant le couloir.

Nous allons mesurer dans les deux tables suivantes le pourcentage de gain en nombre d'hypothèses apporté par rapport à la méthode d'appariement sans contrainte ultrasonore, et cela pour les deux méthodes d'appariement développées.

Les mesures se font sur les trois poses de la pièce du laboratoire et les deux poses du couloir qui relie les différentes pièces.

La première méthode d'appariement utilisée est la méthode pré-combinaison. Dans ce cas l'utilisation de la contrainte ultrasonore se fait au niveau de la table d'hypothèses générée après le processus d'appariement 2D / 3D, pour réduire les fausses hypothèses comportant une étiquette ultrasonore différente de celle donnée par la contrainte ultrasonore mesurée auparavant.

La Table.3-10 contient les résultats obtenus avec cette première méthode d'appariement, ainsi que le pourcentage du gain en nombre d'hypothèses apporté par rapport à la méthode initiale sans ultrasons.

Modèle de l'environnement		Hypothèses sans ultrason	Contrainte ultrasonore	Hypothèses avec pré-combinaison	Gain en hypothèses avec pré-combinaison
Pièce	Pose 1	97	Plan	39	60 %
	Pose 2	984	Coin	433	56 %
	Pose 3	446	Arête	218	51 %
Couloir	Pose 1	66	Plan	32	52 %
	Pose 2	416	Coin	174	58 %
Gain moyen en hypothèses avec la pré-combinaison					55.40 %

**Table 3-10.** Gain en nombre d'hypothèses avec la méthode pré-combinaison.

La contrainte ultrasonore permet de réduire le nombre d'hypothèses de localisation pour la première pose dans la pièce du laboratoire de 97 à 39 hypothèses, avec un gain de 60 %. La même chose pour la deuxième et la troisième pose de la pièce, où le gain en nombre d'hypothèses est respectivement de 56 % et 51 %.

Pour les deux poses du couloir, le gain varie aussi entre 52 % et 58 %. Ainsi pour cette méthode d'appariement nous avons un gain moyen en nombre d'hypothèse de **55.4 %**, ce qui est très intéressant.

La deuxième méthode d'appariement est la post-combinaison. Dans ce cas, la contrainte ultrasonore est injectée directement dans le modèle 3D enrichi de l'environnement pour réduire le nombre de segments du modèle et garder que les segments intéressants (segments correspondant à la contrainte ultrasonore plus les segments de liaison).

Dans ce cas, la combinaison se fait au début du processus d'appariement. Les résultats obtenus avec cette méthode d'appariement sont présentés dans la Table.3-11.

La réduction du nombre d'hypothèses de localisation dans cette étape est plus significative que dans la première méthode d'appariement. Par exemple pour la pose numéro deux de la pièce du laboratoire, le nombre d'hypothèses est réduit à 66 %.

Nous pouvons dire la même chose pour les autres poses de la pièce et du couloir, ainsi pour cette méthode nous avons un gain moyen de **63 %**, ce qui représente 7.6 % de plus qu'avec la méthode pré-combinaison.

Modèle de l'environnement		Hypothèses sans ultrason	Contrainte ultrasonore	Hypothèses avec post-combinaison	Gain en hypothèses avec post-combinaison
Pièce	Pose 1	97	Plan	38	61 %
	Pose 2	984	Coin	334	66 %
	Pose 3	446	Arête	131	71 %
Couloir	Pose 1	66	Plan	30	55 %
	Pose 2	416	Coin	160	62 %
Gain moyen en hypothèses avec la post-combinaison					<b>63.00 %</b>

**Table 3-11.** Gain en nombre d'hypothèses avec la méthode post-combinaison.

Ces résultats montrent l'intérêt d'utiliser la contrainte ultrasonore dans les deux méthodes d'appariement et cela par rapport au cas sans ultrasons. La deuxième méthode d'appariement qui est la post-combinaison, donne un meilleur gain en nombre d'hypothèses que la première méthode, et sera donc retenue par la suite.

### 7-3 Réduction du temps de calcul de la transformation

Le temps nécessaire pour effectuer la transformation entre le repère caméra et le repère modèle dans une procédure de localisation représente environ 90 % du temps global d'une localisation. Cela est du à la complexité des calculs mathématiques à effectuer pour résoudre les fonctions d'erreurs générées dans chaque étape.

Dans la Table 3-12 nous calculons les temps de traitement pour les deux méthodes de transformation implémentées dans le cadre de cette thèse (méthode de Lowe avec ces deux solutions, non-linéaire et linéaire), ainsi que la méthode développée dans le travail initiale basée sur les quaternions.

Modèle de l'environnement		Temps de la méthode 1 (s)	Méthode 2		Méthode 3	
			Temps (s)	Gain	Temps (s)	Gain
Pièce	Pose 1	44	19.8	55 %	0.73	98 %
	Pose 2	347	145.7	58 %	1.56	99 %
	Pose 3	156	84.2	46 %	1.05	99 %
Couloir	Pose 1	19	9.32	51 %	0.72	97 %
	Pose 2	71	41.2	42 %	1.34	98 %
Moyenne		127,4	60.0	50.4 %	1,08	98.2 %

**Table 3-12.** Gain en temps de traitement pour les trois méthodes de transformation.

Nous calculons également le pourcentage de gain en temps de calcul entre la transformation initiale appelée Méthode 1 et les deux autres transformations, méthode 2 pour la méthode de Lowe non-linéaire et Méthode 3 pour la méthode linéaire. La méthode d'appariement utilisée dans ces résultats est la Post-combinaison.

La première remarque très importante à faire est que le temps de calcul pour la méthode des quaternions est toujours largement supérieur à une seconde, et peut atteindre plusieurs minutes parfois. Il varie entre 19 secondes pour la pose numéro un du couloir et 347 secondes pour la pose numéro deux de la pièce du laboratoire (à-peu-près six minutes ce qui est énorme pour une localisation).

Le temps moyen calculé pour cette première méthode de transformation sur les cinq poses est d'environ 127 secondes.

La méthode de Lowe avec solution non-linéaire donne de meilleurs résultats par rapport à la méthode initiale, un pourcentage de gain moyen en temps de traitement qui atteint les 50 % mais qui reste un peu élevé par rapport au type de l'application nécessitant une réaction efficace et rapide à la fois.

Par contre avec la méthode de Lowe avec la solution linéaire, le temps de calcul est d'environ 1 seconde en moyenne. Ce temps varie entre 0.72 secondes pour la pose un du couloir et 1.56 seconde pour la pose deux de la pièce du laboratoire. Le temps moyen estimé pour cette méthode est d'environ 1.08 secondes.

En ce qui concerne le temps moyen d'une localisation, nous allons le mesurer dans la partie 3-7. Tout d'abord, nous allons calculer la précision de la localisation et cela pour les trois méthodes de choix de la meilleure pose. Cette précision s'exprime par les trois erreurs suivant les axes  $X$ ,  $Y$  et  $\Phi$ .

#### 7-4 Erreurs de localisation

Pour choisir la meilleure hypothèse de localisation dans la table de transformations, trois méthodes ont été proposées. Les deux premières méthodes se basent sur la comparaison des poses calculées dans la table de transformations avec la pose initiale donnée par le système odométrique.

Ainsi, la première utilise directement la différence entre les deux poses suivant les axes  $X$ ,  $Y$  et  $\Phi$  instantanément, par contre la deuxième utilise les trois meilleures poses dans la table de transformations pour sélectionner les trois meilleures valeurs des coordonnées de localisation suivant les trois axes  $X$ ,  $Y$  et  $\Phi$ . La troisième méthode une fonction de coût basée sur la mesure de distances entre les segments de l'image et les segments du modèle.

Dans les trois tables suivantes on va présenter les erreurs de localisation données par ces trois méthodes suivant les trois paramètres de localisation  $x$ ,  $y$  et  $\Phi$ . Ces erreurs sont mesurées par rapport aux poses réelles du robot données dans la Table.3-9 lors de l'acquisition des données.

Ainsi, la Table.3-13 donne l'erreur calculée suivant l'axe  $X$ , l'axe  $Y$  et l'axe des rotations  $\Phi$ , et cela pour les trois poses de la pièce du laboratoire et les deux poses du couloir. On calcul aussi, la moyenne sur cinq poses des erreurs suivant  $x$ ,  $y$  et  $\Phi$ .

Ces résultats sont obtenus avec un appariement du type Poste-combinaison et une transformation basée sur la méthode de Lowe avec la solution linéaire.

Modèle de l'environnement		Erreur en $X$ (cm)	Erreur en $Y$ (cm)	Erreur en $\Phi$ (°)
Pièce	Pose 1	5.55	18.9	2.30
	Pose 2	17.2	4.97	4.22
	Pose 3	21.0	12.3	4.13
Couloir	Pose 1	13.1	6.46	1.85
	Pose 2	16.3	12.6	3.99
Moyenne		<b>14.6</b>	<b>11.1</b>	<b>3.29</b>

**Table 3-13.** Les erreurs suivant  $x$ ,  $y$  et  $\Phi$  donnés par la méthode de la pose initiale.

Les valeurs de l'erreur obtenues avec cette première méthode sur les cinq poses de l'environnement utilisées sont de l'ordre de 12.85 cm pour la position et de 3.3° pour la rotation. Si on compare ces résultats avec les dimensions du robot et de l'environnement on constate que ces erreurs sont importantes et pas facile à négliger. La Table.3-14 donne quant à elle les valeurs de l'erreur sur les axes  $X$ ,  $Y$  et  $\Phi$  obtenues avec la méthode des trois meilleures hypothèses.

Modèle de l'environnement		Erreur en $X$ (cm)	Erreur en $Y$ (cm)	Erreur en $\Phi$ (°)
Pièce	Pose 1	7.00	2.00	0.98
	Pose 2	3.13	16.50	3.37
	Pose 3	9.60	15.50	1.14
Couloir	Pose 1	2.18	3.63	0.82
	Pose 2	1.04	2.28	1.59
Moyenne		<b>4.59</b>	<b>7.98</b>	<b>1.57</b>

**Table 3-14.** Les erreurs suivant  $x$ ,  $y$  et  $\Phi$  donnés par la méthode des trois hypothèses.

Pour la pose numéro un de la pièce du laboratoire par exemple, l'erreur obtenue pour cette première méthode suivant l'axe  $x$  est de 7 cm et suivant l'axe  $y$  est de 2 cm, ainsi que l'erreur de rotation est de 0.98°.

Le calcul des erreurs moyennes donne une valeur de 4.59 cm suivant l'axe  $x$ , une valeur de 7.98 cm suivant l'axe  $y$  et une valeur de 1.57° suivant l'axe de rotation  $\Phi$ . Ainsi, pour cette méthode, la valeur globale de l'erreur en position est inférieure à 7 cm et pour l'erreur globale de rotation, elle est inférieure à 2°.

Cette méthode ainsi que la méthode précédente se basent principalement sur la pose initiale donnée par le système odométrique pour choisir la meilleure solution dans la table de

transformation, ce qui donne des résultats biaisés. Pour résoudre ce problème, une méthode basée sur une fonction de coût a été développée.

La Table.3-15 donne l'erreur obtenue avec la méthode de mesure des distances entre les deux ensembles de segments image et modèle qui participent à l'appariement. Cette erreur est exprimée aussi sur trois axes, l'axe des  $X$ , l'axe des  $Y$  et l'axe des rotations  $\Phi$ . La valeur moyenne de ces erreurs est aussi calculée.

Modèle de l'environnement		Erreur en $X$ (cm)	Erreur en $Y$ (cm)	Erreur en $\Phi$ (°)
Pièce	Pose 1	3.60	5.10	0.74
	Pose 2	5.50	3.78	0.82
	Pose 3	7.33	4.50	0.98
Couloir	Pose 1	1.60	1.14	0.38
	Pose 2	2.40	3.60	0.86
Moyenne		4.09	3.63	0.76

**Table 3-15.** Les erreurs suivant  $x$ ,  $y$  et  $\Phi$  donnés par la méthode des distances.

Si nous prenons par exemple la pose numéro deux de la pièce du laboratoire, l'erreur de localisation suivant l'axe  $x$  est de 5.5 cm, suivant l'axe  $y$  est de 3.78 cm et suivant l'axe des rotations  $\Phi$ , elle est de 0.82°. Pour le deuxième environnement, c'est à dire le couloir, pour la pose numéro un par exemple, l'erreur suivant  $x$  est de 1.6 cm, suivant  $y$  est de 1.14 cm et suivant  $\Phi$ , elle est de 0.38°.

Pour les valeurs moyennes de l'erreur sur les trois axes, on a une valeur moyenne de 4.09 cm sur  $x$ , une valeur moyenne de 3.63 cm sur  $y$  et une valeur moyenne de 0.76° pour l'angle  $\Phi$ . Ainsi, pour cette méthode de sélection de la meilleure hypothèse de localisation, l'erreur de position globale est inférieure à 5 cm et l'erreur d'orientation globale est inférieure à un degré.

Une simple comparaison entre les résultats obtenus avec ces trois méthodes montre rapidement la supériorité de la méthode basée sur la mesure des distances. D'un côté, cette méthode ne prend pas en considération la pose initiale donnée par le système odométrique, et d'un autre côté, les valeurs de l'erreur sont meilleures que celles obtenues avec les deux méthodes précédentes.

Dans les deux méthodes de choix, les données vision et ultrasons servent à confirmer et conforter la pose initiale donnée par l'odométrie. Par contre dans la troisième méthode, les informations visuelles et ultrasonores corrigent la pose donnée par l'odométrie, car nous sommes certain de choisir la pose la plus plausible.

### 7-5 Temps de calcul de la localisation

Dans cette partie des résultats nous allons donner le temps de calcul nécessaire pour effectuer une tâche complète de localisation. La Table.3-16 détaille les temps nécessaires pour effectuer chacune des trois fonctions de la localisation : appariement, transformation et choix. On donne également le temps total atteint pour chacune des cinq poses de l'environnement, ainsi

que les valeurs des différentes moyennes. Notons que la méthode d'appariement utilisée dans ces mesures est la post-combinaison et la fonction de choix retenue est celle basée sur les mesures de distances.

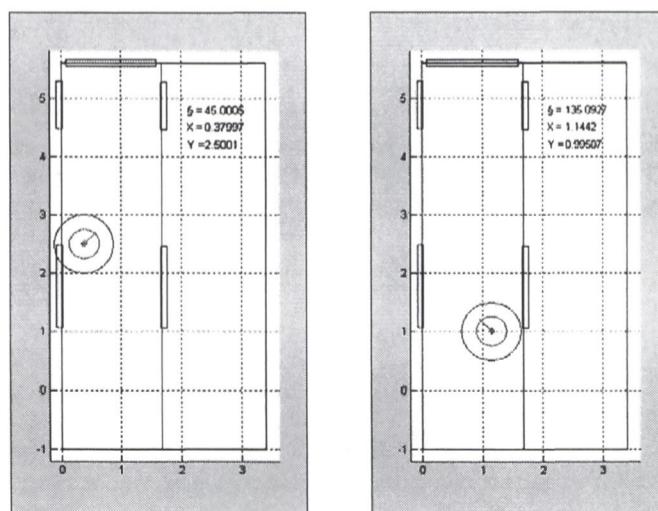
Modèle de l'environnement		Temps d'appariement	Temps de Transformation	Temps du choix	Temps de localisation
Pièce	Pose 1	1.06	0.73	0.01	1.80
	Pose 2	1.07	1.56	0.01	2.64
	Pose 3	0.92	1.05	0.01	1.98
Couloir	Pose 1	0.59	0.72	0.01	1.32
	Pose 2	1.17	1.34	0.01	2.52
Moyenne		<b>0.96</b>	<b>1.08</b>	<b>0.01</b>	<b>2.05</b>

**Table 3-16.** Temps de calcul des étapes de la localisation pour les cinq poses de l'environnement (temps donnés en seconde).

Nous constatons que sur les deux secondes nécessaires en moyenne pour obtenir une localisation, les fonctions appariement et transformation représentent chacune la moitié de ce temps de calcul. Le temps dédié à la fonction de choix étant négligeable.

### 7-6 Résultats de localisation

Dans les figures suivantes nous allons tester l'algorithme de localisation sur les cinq poses de l'environnement choisies pour la partie résultats et représenter ces résultats dans un environnement graphique 2D correspondants au modèle 3D. Ainsi dans la Figure 3-18, la localisation se fait dans le couloir avec les deux contraintes ultrasonores coin et plan.

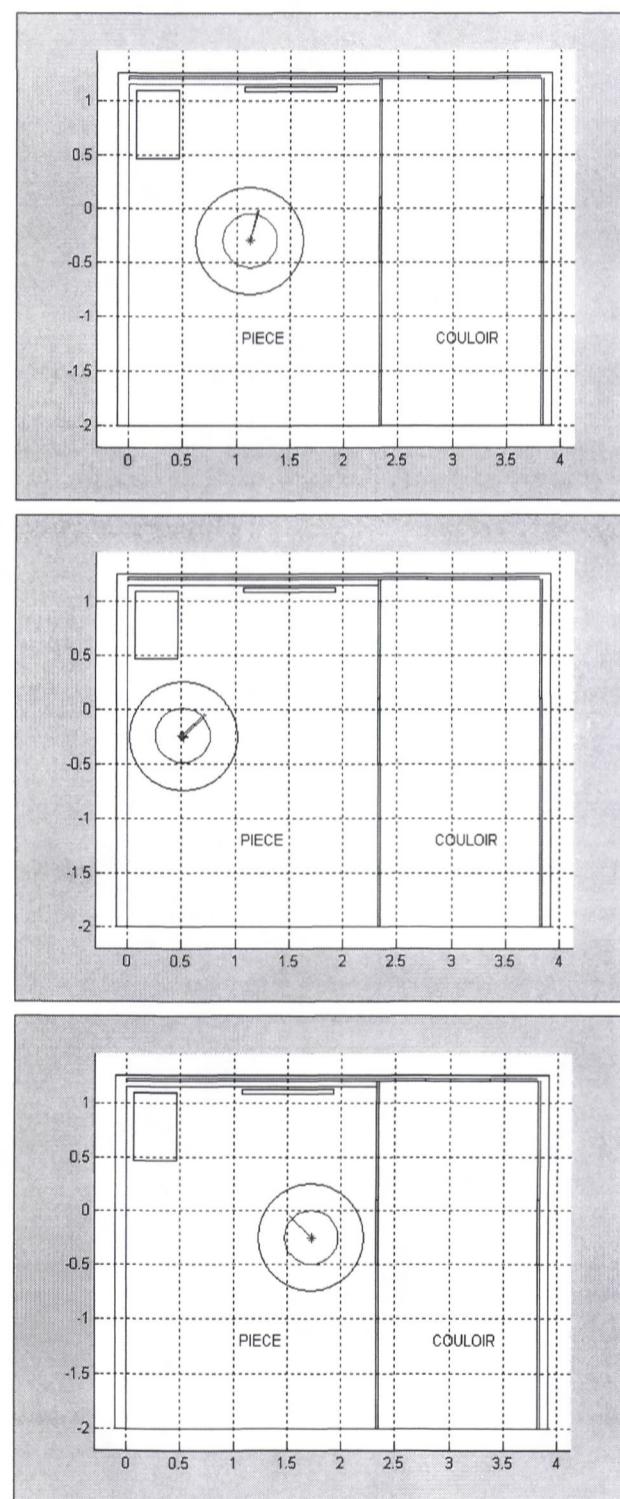


**Figure 3-18.** Les deux localisations dans le couloir.

Dans ces deux cas le robot est bien localisé car d'un côté, le cercle bleu le représentant se trouve à l'intérieur du cercle d'erreur tracé en noir. Cela signifie que la pose calculée est acceptée comme la meilleure solution.

D'un autre côté, la position et l'orientation du robot calculées données par une croix et un segment rouge sont très proches de la position et l'orientation réelles données par une croix et un segment bleu, ce qui correspond à la meilleure solution dans la table de transformations.

La Figure.3-19 donne les trois autres localisations dans la pièce du laboratoire. Nous constatons là aussi que le robot est bien localisé dans son environnement.



**Figure 3-19.** Les trois localisations dans la pièce.

### 7-7 Evaluation de la méthode de localisation

Pour généraliser les performances de notre système de localisation sur plusieurs situations nous avons acquis 66 images et échos ultrasonores correspondant à 66 poses dans le premier modèle de l'environnement (pièce du laboratoire) et 24 images et échos correspondant à 24 poses dans le couloir.

Le but de ces différentes mesures est de localiser le robot 90 fois dans ces deux environnements pour quantifier les différents paramètres de localisation. L'algorithme de localisation utilisé comporte les meilleurs blocs de traitement développés auparavant. Ils sont cités dans la Table 3-17, ainsi que les blocs utilisés dans la méthode initiale.

Blocs de traitement	Méthode initiale	Méthode utilisée
Appariement	Sans ultrasons	Post-combinaison
Transformation	Quaternions	Solution linéaire
Choix	Pose initiale	Fonction de coût basée sur la distance

**Table 3-17.** Choix des blocs de traitement utilisés pour l'évaluation.

Nous avons choisi d'utiliser les huit indicateurs de localisation suivants pour évaluer la méthode de localisation utilisée. Dans la Table 3-18 nous donnons une moyenne générale de ces indicateurs pour les 90 poses (66 dans la pièce et 24 dans le couloir) qui sont :

- 1- Gain en nombre d'hypothèses par rapport au cas sans ultrasons :  $G_H$
- 2- Gain en temps de calcul par rapport à la méthode des quaternions :  $G_T$
- 3- Gain en temps de calcul pour une localisation :  $G_L$
- 4- Temps de calcul moyen d'une localisation :  $T_{CL}$
- 5- Erreur moyenne de position sur 90 erreurs :  $\mu_{Pos}$
- 6- Ecart type de position :  $\sigma_{Pos}$
- 7- Erreur moyenne de rotation sur 90 erreurs :  $\mu_{Rot}$
- 8- Ecart type de rotation :  $\sigma_{Rot}$

Nous présentons aussi dans cette table le taux d'échec de localisation en pourcentage sur les 90 poses de l'environnement pour notre méthode,  $T_{EL}$ . En rappelant que le taux d'échec de la méthode présentée dans le travail initiale dans le chapitre précédent correspondait à 66 % en utilisant le modèle de l'environnement et les images du travail initial.

$G_H$	$G_{TR}$	$G_{TL}$	$T_{CL}$ (s)	$\mu_{Pos}$ (cm)	$\sigma_{Pos}$ (cm)	$\mu_{Rot}$ (°)	$\sigma_{Rot}$ (°)	$T_{EL}$
66 %	98 %	99 %	2.0	5.0	4.78	1.0	0.91	15 %

**Table 3-18.** Evaluation de la méthode de localisation développée.

En utilisant nos modèles d'environnements enrichis et nos images, nous obtenons un taux d'échec de la méthode initiale de l'ordre de 25 %.

Avec les différentes améliorations apportées à la méthode de localisation initiale, nous gagnons une moyenne de 66 % sur le nombre total d'hypothèses de localisation générées par le processus de d'appariement post-combinaison.

Nous gagnons aussi 98 % sur le temps de calcul nécessaire à la transformation repère caméra / repère image. Ces deux améliorations permettent à leurs tours de gagner un pourcentage de 99 % sur le temps globale d'une localisation. Ce dernier élément permet une localisation plus rapide du robot mobile avec une moyenne de 2 secondes.

L'erreur de position moyenne est de 5 cm suivant  $X$  et  $Y$  et l'erreur de rotation moyenne est de l'ordre de 1°. La comparaison de ces résultats avec les valeurs de l'erreur obtenues dans la méthode initiale laisse constater directement une amélioration très significative.

Les valeurs de l'écart type de position qui est inférieur à 5 cm et celui de la rotation qui est inférieur à 1° confirment ces bons résultats obtenus.

Les différentes améliorations apportées à la méthode de localisation ainsi que la généralisation de la localisation sur plusieurs environnements ont permis d'avoir un taux de réussite d'environ 85 %.

Tous ces éléments montrent l'efficacité de l'approche développée. La robustesse a été considérablement améliorée comme le montre la baisse du taux d'échec. Malgré tout, pour conclure cette partie nous avons illustré ce problème de robustesse à travers différents exemples de localisation développés dans la suite.

### 7-8 Comment augmenter la robustesse

Dans un certain nombre de cas, le système de localisation n'arrive pas à localiser le robot mobile. Cette situation est due au manque de segments image pour l'appariement. Le robot a besoin ainsi de changer de position dans l'environnement pour obtenir un nombre suffisant de segments image et effectuer à bien le processus d'appariement.

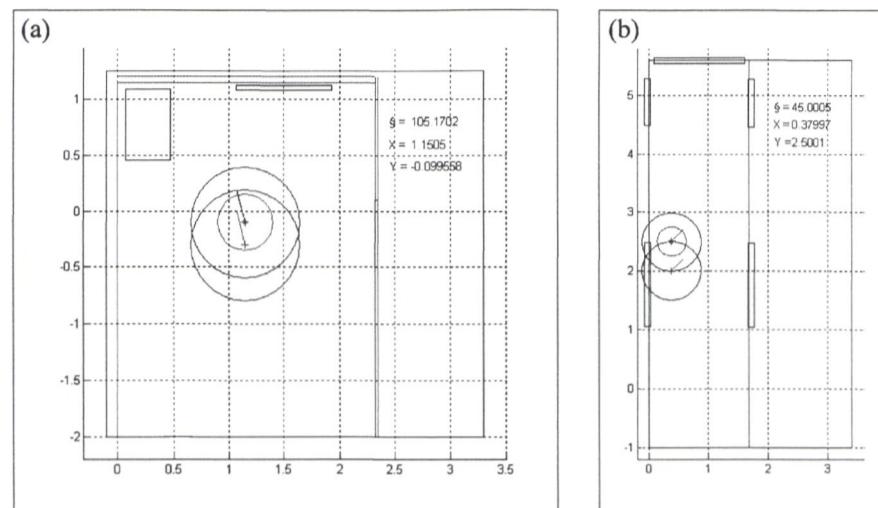
Cette solution est appelée Re-localisation. Nous allons voir dans la suite deux exemples de Re-localisation. Nous allons voir aussi le cas de manque de luminosité et le cas de localisation avec présence d'un obstacle non répertorié dans le modèle 3D.

#### 7-8-1 Re-localisation

Le nombre de segments du modèle ou de l'image, nécessaire à la mise en correspondance peut être insuffisant car il faut au moins trois correspondances "image / modèle" pour valider une localisation. Dans ce cas de figure, le système de localisation affiche le terme : "pas de résultat" par manque de segments.

Ainsi, nous obligeons le robot soit à avancer vers l'avant d'une petite distance suivant l'axe  $X$  ou  $Y$ , soit à effectuer une petite rotation suivant l'axe  $\Phi$ . Puis, il prend une nouvelle image et une nouvelle acquisition ultrasonore, et lance à nouveau, la procédure de localisation. C'est le principe de la "Re-localisation".

La Figure.3-20 donne deux exemples de Re-localisation dans la pièce du laboratoire en (a) et dans le couloir en (b).



**Figure 3-20.** Re-localisation dans la pièce et dans le couloir.

La pose "N° 1" est présentée par une croix rouge et un segment rouge pour l'orientation, elle correspond à la première localisation non effectuée à cause du manque de segments. La pose "N° 2" présentée par une étoile bleue et un segment bleu correspond à la re-localisation, qui donne un résultat positif et calcul la pose du robot.

Dans la pièce, suite au message "pas de résultat" le robot avance de 20 cm dans la direction de l'axe  $Y$ , prend de nouvelles mesures et lance de nouveau la procédure de localisation qui correspond à cette nouvelle pose. Pour le cas du couloir, le robot avance de 50 cm toujours dans la direction de l'axe  $Y$ .

Le cercle d'erreur noir avec une croix rouge et un segment rouge au centre présente la première localisation qui ne donne pas de résultat. Le deuxième cercle d'erreur noir avec le cercle bleu au milieu, qui présente le robot mobile, montre que la localisation a été effectuée cette fois.

Le système de localisation va afficher les coordonnées de la nouvelle pose donnée par l'étoile bleue et le segment bleu au centre du cercle d'erreur, ainsi que les coordonnées de l'ancienne pose qui a le même  $x$  et le même angle  $\theta$ , mais avec une valeur de  $y$  décalée de 20 cm.

### 7-8-2 Manque de luminosité

La deuxième situation est liée toujours au problème de manque de segments. Par exemple lors de la navigation, une panne électrique rend la localisation impossible la nuit. Alors pour résoudre cette situation, une source de lumière est installée sur le robot à côté de la caméra, pour éclaircir la scène de travail du robot.

Cette lumière va permettre d'améliorer la qualité des images acquises et ainsi augmenter le nombre de segments image détecté pour l'appariement.

### 7-8-3 Localisation avec obstacles

Le modèle de l'environnement utilisé dans cette partie de résultats est le troisième modèle qui correspond à la salle de classe. Ce dernier diffère des deux autres modèles par la présence de deux obstacles, une chaise et une armoire qui devraient gêner le déplacement et la localisation du robot mobile.

L'acquisition des images des échos ultrasonores nécessaires à la localisation se font à des distances différentes aux obstacles. Nous notons aussi que ces deux obstacles ne sont pas répertoriés dans le modèle 3D enrichi de la salle de classe qui contient uniquement les murs et les portes.

Dans la suite, nous allons tester la localisation dans cinq poses de la salle de classe avec la présence d'une table puis d'une armoire. Après, nous allons évaluer la localisation avec obstacle non répertorié dans le modèle de l'environnement sur 24 poses.

La méthode de localisation que nous utilisons se base sur l'appariement par Post-combinaison, la fonction de transformation utilisant la solution de Lowe linéaire et la fonction choix basée sur la distance.

#### a) Présence d'une table

Pour tester la localisation avec la présence d'une table, nous allons choisir cinq poses dans le troisième environnement. Pour chaque pose, le système d'acquisition est placé à une distance fixe de la table.

Les cinq poses ainsi que ces distances sont données dans la Table.3-19. Dans cette partie ainsi que dans la suivante, nous avons choisi de tester la localisation avec les poses dont les photos montrent la table d'une façon complète ou partielle.

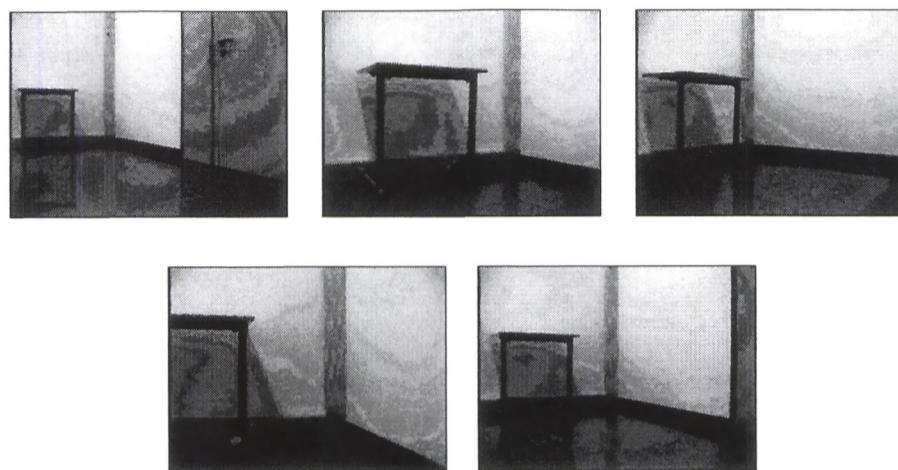
Poses	X (m)	Y (m)	$\Phi$ (°)	Distance à l'objet (m)
Pose 1	1.96	0.00	60	4.15
Pose 2	1.96	2.12	60	2.05
Pose 3	0.92	2.12	30	2.80
Pose 4	2.88	2.12	75	1.70
Pose 5	1.96	1.06	60	2.95

Table 3-19. Poses et distances par rapport à l'obstacle table.

Les photos de l'environnement correspondant à ces cinq poses sont données dans la Figure.3-21. Chaque photo donne le positionnement de la table dans la salle de classe par rapport au système d'acquisition. Rappelons que dans cette manipulation, la table est fixe dans la salle et c'est le robot qui est mobile.

Les résultats de localisation pour ces cinq poses sont donnés dans la Table.3-20. En plus des erreurs de localisation suivant les axes  $X$ ,  $Y$  et  $\Phi$  nous mesurons aussi pour chaque pose le

nombre de segments de l'image participant à l'appariement et la contrainte ultrasonore correspondante à la forme de l'environnement vue depuis cette pose.



**Figure 3-21.** Photos de l'obstacle table.

A la fin, nous donnons la moyenne de l'erreur suivant les trois axes de la pose sur les situations avec une localisation positives. Si la localisation n'est pas possible, le système affiche "pas de résultat".

Poses	Nombre de segments	Contrainte ultrasonore	Erreur en X (cm)	Erreur en Y (cm)	Erreur en $\Phi$ (°)
<b>Pose 1</b>	11	<b>Coin</b>	7.20	4.22	1.08
<b>Pose 2</b>	9	<b>Plan</b>	3.16	1.98	0.56
<b>Pose 3</b>	10	<b>Coin</b>	5.03	4.44	0.72
<b>Pose 4</b>	7	<b>Plan</b>	Pas de résultat de localisation		
<b>Pose 5</b>	10	<b>Coin</b>	5.52	3.01	0.77
<b>Moyenne de l'erreur</b>			<b>5.22</b>	<b>3.41</b>	<b>0.78</b>

**Table 3-20.** Résultats de localisation avec l'obstacle table.

Les résultats obtenus montrent qu'il est possible d'effectuer une localisation dans un environnement avec la présence d'une table. Les erreurs moyennes suivant les axes  $X$ ,  $Y$  et  $\phi$  sont aussi acceptables par rapport aux erreurs obtenues sans obstacle.

Dans certaines situations, le système n'arrive pas à se localiser comme le cas de la pose 4. Cela est dû principalement au manque de segments image nécessaire à l'appariement.

### b) Présence d'une armoire

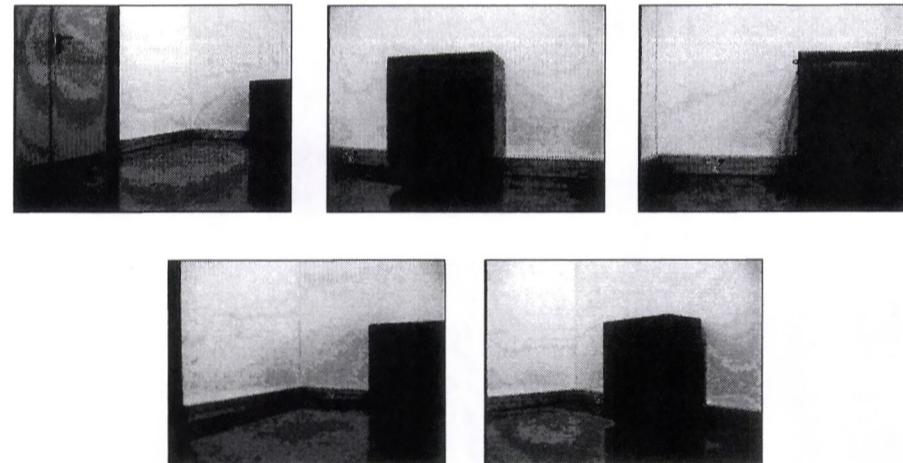
Pour le cas de l'armoire on a choisi les mêmes cinq poses de l'environnement utilisées dans le cas de la table. Le système d'acquisition est placé à une distance fixe de l'armoire (Table.3-21).

Dans ce cas aussi, on a choisi de tester la localisation avec les poses dont les photos montrent l'armoire d'une façon complète ou partielle.

Poses	Distance à l'objet (m)	X (m)	Y (m)	$\Phi$ (°)
<b>Pose 1</b>	4.10	1.96	0.00	120
<b>Pose 2</b>	2.00	1.96	2.12	105
<b>Pose 3</b>	1.65	0.92	2.12	90
<b>Pose 4</b>	2.75	2.88	2.12	135
<b>Pose 5</b>	2.90	1.96	1.06	120

**Table 3-21.** Poses et distances par rapport à l'obstacle armoire.

Les photos de l'environnement correspondant à ces cinq poses sont données dans la Figure.3-22. Chaque photo donne le positionnement de la table dans la salle de classe par rapport au système d'acquisition. Nous rappelons que dans cette manipulation, la table est fixe dans la salle et c'est le robot qui est mobile.



**Figure 3-22.** Photos de l'obstacle armoire.

Les résultats de localisation pour ces cinq poses sont donnés dans la Table.3-22. En plus des erreurs de localisation suivant les axes  $X$ ,  $Y$  et  $\Phi$  on mesure aussi pour chaque pose le nombre de segments de l'image participant à l'appariement et la contrainte ultrasonore correspondante à la forme de l'environnement vue depuis cette pose.

A la fin, on donne la moyenne de l'erreur suivant les trois axes de la pose sur les situations positives. Ces résultats montrent la possibilité de se localiser dans un environnement avec la présence d'une armoire non répertoriée dans le modèle 3D enrichi de la salle de classe.

L'erreur moyenne suivant l'axe des  $X$  est d'environ 6 cm, celle sur l'axe des  $Y$  est d'environ 5 cm et enfin sur l'axe des rotations  $\Phi$  l'erreur est de l'ordre un degré. Ces valeurs sont acceptables par rapport aux erreurs obtenues dans le cas sans obstacle.

Poses	Nombre de segments	Contrainte ultrasonore	Erreurs en X (m)	Erreurs en Y (m)	Erreurs en $\Phi$ (°)
Pose 1	11	Coin	8.11	9.20	0.98
Pose 2	7	Plan	4.82	3.55	0.46
Pose 3	5	Plan		Pas de résultat de localisation	
Pose 4	6	Coin	3.21	1.99	0.95
Pose 5	10	Arête	6.76	4.87	0.53
<b>Moyenne de l'erreur</b>			<b>5.72</b>	<b>4.90</b>	<b>0.73</b>

**Table 3-22.** Résultats de localisation avec l'obstacle armoire.

Comme dans le cas d'une table, le système de localisation n'arrive pas dans certaines configurations à se localiser correctement comme pour le cas de la pose 3. Cela est dû aussi au manque de segments de l'image nécessaire pour exécuter la mise en correspondance.

Ces segments sont souvent cachés par l'obstacle ou coupés en petits segments de longueur négligeable. Il faut rappeler ici que les forme et les dimensions de l'obstacle vont influencer l'appariement, car le système de vison va générer des segments images qui correspondent à l'obstacle non répertorié et ainsi fausser la localisation.

#### 7-8-4 Evaluation de la localisation avec un obstacle

Dans cette partie on va évaluer la procédure de localisation avec la présence d'un obstacle non répertorié dans le modèle de l'environnement sur 24 poses. 12 des ces poses correspondent à une localisation avec une table et les 12 autres correspondent à une localisation avec une armoire.

Dans la Table.3-23 nous allons quantifier le taux de localisation correcte avec un obstacle sur ces 24 poses de l'environnement.

	Oui	Non
Obstacle : Table	8	4
Obstacle : Armoire	7	5
Total	15	9
Taux de localisation	<b>62.50 %</b>	<b>37.50 %</b>

**Table 3-23.** Evaluation de la localisation avec un obstacle.

Les résultats obtenus montre qu'on peut se localiser avec un taux de 62.50 % dans un environnement d'intérieur avec la présence d'un obstacle non répertorié dans le modèle 3D. Le taux d'échec quant à lui, est de 37.50 %. La conclusion qu'on peut tirer de cette partie, qui étudie la localisation avec obstacle est résumée dans les deux points suivants :

- 1- S'il y a assez de segments (image et modèle) pour effectuer la procédure de mise en correspondance, le système peut effectuer la tâche de localisation et cela malgré la présence d'un obstacle non répertorié dans l'environnement de travail du robot,
- 2- Si l'obstacle empêche le système de vision d'avoir un nombre suffisant de segments image nécessaire à l'appariement, et cela à cause de son positionnement ou des ces dimensions, dans ce cas, on ne peut pas avoir une localisation.

## 8- Conclusion

Dans ce chapitre, nous avons présenté les différentes améliorations apportées à une méthode de localisation basée sur l'approche incrémentale utilisée pour un robot mobile dans un environnement d'intérieur. Cette méthode utilise trois types d'informations sensorielles, vision, ultrasons et odométrie.

La première amélioration concerne la partie vision. L'image subit deux types de traitement pour extraire les segments, un filtrage optimal de Deriche suivi par une transformation de Hough. Le filtre de Deriche utilisé permet d'avoir un maximum de segments par rapport à celui de Canny implémenté initialement. Un nombre important de segments 2D image diminue le taux d'échec en générant plus d'hypothèses vraisemblable.

Nous utilisons la fusion entre les données vision et l'information ultrasonore pour réduire le nombre d'hypothèses initiales d'appariement, générées par un processus de mise en correspondance 2D / 3D. Deux nouvelles méthodes d'appariement ont été développées pour utiliser la donnée ultrasonore.

La première méthode est la pré-combinaison, où la donnée ultrasonore est injectée au niveau de la table d'hypothèses. Dans la deuxième méthode, la donnée ultrasonore est utilisée dès le début pour réduire directement le nombre de segments du modèle participant à la mise en correspondance.

Nous notons aussi que les formes de l'environnement (plan, coin et arête) sont reconnues grâce à un réseau de neurones spécialisé. Le taux moyen de la réduction d'hypothèse atteint les 66 %, ce qui est très significatif.

Pour la partie transformation repère caméra / repère image, l'implémentation de la méthode de Lowe avec ces deux types de solution linéaire et non-linéaire, nous a permis de réduire le temps de calcul de la transformation de plus de 50 % pour la solution non-linéaire, et d'environ 98 % pour la solution linéaire, et cela par rapport à la méthode initiale basée sur le principe de Phong-Horaud et les quaternions.

Ces deux gains (nombres d'hypothèses et temps de traitement) influe directement sur le temps global de localisation et permet enfin, d'envisager une application réelle de la méthode dans la mesure où le temps de localisation devient de l'ordre de 2 secondes. Cette moyenne en temps de localisation correspond à la vitesse du robot mobile et au temps de réponse d'une personne handicapée.

Dans la partie choix de la meilleure hypothèse deux nouvelles approches ont été développées. La première utilise les trois meilleures hypothèses dans la table de transformations pour à la fin choisir le meilleur  $x$ , le meilleur  $y$  et la meilleure orientation  $\Phi$ . La deuxième approche utilise une fonction coût calculée sur les distances entre les deux groupes de segments

d'appariement, pour choisir la plus petite distance qui correspond automatiquement à la meilleure hypothèse de localisation.

Cette deuxième voie permet en outre une grande indépendance par rapport à la position fournie par l'odométrie qui servait de référence unique pour le choix de la meilleure pose dans la méthode présentée dans le travail initial.

L'ensemble de la méthode avec les nombreuses améliorations apportées a été testée dans trois environnements différents pour des positions variées du robot. Une centaine de situations ont permis d'évaluer la robustesse de la méthode, de quantifier sa précision et de mesurer les gains apportés par ces améliorations.

Pour finir remarquons qu'une solution a été proposée pour résoudre les échecs de localisation et que des tests de localisation ont été effectués avec un bon pourcentage de succès dans des environnements comportant des obstacles non répertoriés dans le modèle 3D de l'environnement.

L'ensemble de ces résultats permet de montrer que les travaux réalisés dans le cadre de cette thèse ont permis de faire évoluer l'algorithme initialement proposé par O. Ait Aider vers une méthode de localisation opérationnelle pour un robot d'intérieur utilisant des informations multicapteurs.

## Chapitre IV

### Stratégie d'approche d'objets par asservissement visuel et ultrasons

*Le deuxième objectif de ma thèse est également consacré à la combinaison d'information vision et ultrason pour la saisie d'objets simples.*

*Le principe est d'appliquer une stratégie pour approcher les objets à saisir avec une bonne précision. Dans ce cadre, nous utilisons d'une part, une petite caméra HF pour calculer les paramètres de l'asservissement visuel suivant les axes X et Y de la caméra et d'une autre part, nous utilisons l'information ultrasonore pour avoir la forme de l'objet et sa distance Z du système de saisie.*

*Dans la partie vision, plusieurs méthodes qui permettent de localiser l'objet d'intérêt dans l'image ont été développées. Elles sont basées principalement sur des approches contours ou régions.*

*Pour la deuxième partie, un nouveau réseau de neurones est utilisé pour classifier les objets simples en deux formes de base, le cylindre et la sphère.*

## 1- Introduction

La saisie et le transport d'objets simples à l'aide d'un bras robotisé présente une tâche primordiale dans le cadre de la robotique mobile d'intérieur. Pour le projet ARPH, cette procédure permet à une personne handicapée moteur d'interagir facilement avec son environnement grâce à une interface homme / machine, sans avoir besoin d'une assistance ou intervention humaine.

Une fois un algorithme de localisation efficace de la base mobile est obtenu, la deuxième partie de ma thèse est consacrée au développement d'une stratégie pour approcher des objets de forme simple. Le but de cette stratégie est de saisir ces objets avec la pince du bras manipulateur MANUS par l'intermédiaire d'un système utilisant une caméra sans fil et des capteurs ultrasons.

La combinaison des informations reçues de ces deux capteurs permet de reconnaître tout d'abord la forme de l'objet, de calculer sa distance, puis de le localiser et de le centrer dans le champ visuel de la caméra afin de l'approcher. Cette méthode est basée sur plusieurs éléments, chacun d'entre eux participe à cette stratégie de type "asservissement visuel".

Un robot mobile est un système mécanique capable de déplacer son effecteur afin de réaliser une tâche qui lui est assignée. Une cellule robotisée se forme généralement de deux composantes essentielles :

- 1- **La perception** : Qui permet de gérer les relations entre le robot et son environnement grâce à des capteurs et des systèmes de traitement spécifiques,
- 2- **La commande** : Qui synthétise les consignes des asservissements pilotant les actionneurs. A partir de la fonction de perception et des ordres de l'utilisateur, elle permet d'engendrer les mouvements du robot.

Notre travail entre dans le cadre de la première fonction. Elle permet au système de commander d'avoir les données visuelles nécessaires au bon fonctionnement et notamment, les valeurs des différents déplacements dans l'espace pour approcher l'objet à saisir. Ainsi dans ce chapitre, nous n'allons pas nous intéresser aux lois de commande qui gèrent le robot et le bras Manus. Notre objectif est de fournir à ce système les différents paramètres qui vont lui permettre d'assurer la tâche de saisie.

Pour valider notre approche, nous avons étudié trois situations, chacune utilise une quantité d'informations différentes.

Nous avons commencé par des scènes simples comportant un seul objet avec une couleur unique. Le système ultrasonore utilisé dans cette partie fournit la distance de l'objet. Un algorithme de traitement d'images basé sur les contours horizontaux et verticaux permet de localiser l'objet dans l'image [Tra.a.05], [Tra.b.05].

La deuxième situation traite des scènes complexes composées de plusieurs objets de couleurs et de formes variées. Le réseau de neurones utilisé dans la partie localisation a été modifié pour pouvoir reconnaître la forme des objets utilisés à partir de leur signature ultrasonore. Par ailleurs, l'algorithme de segmentation "Ligne de partage des eaux" a été utilisé pour séparer les différentes régions et étiqueter les différentes zones de l'image au moyen d'un label

couleur. Cette partie a été suivie par un filtrage couleur pour identifier et sélectionner l'objet d'intérêt [Tra.c.05].

La dernière situation est consacrée à l'étude des objets réels composés d'une ou de plusieurs couleurs et se trouvant à des distances différentes [Tra.b.06].

## 2- Asservissement visuel

Dans cette partie, nous allons aborder certains aspects de l'utilisation de la vision pour la commande des robots mobiles. Comme son nom l'indique, l'organe de perception de base utilisé est la caméra.

Le principe de l'asservissement visuel consiste à prendre en compte des informations visuelles issues d'une ou plusieurs caméras dans la boucle de commande d'un robot afin d'en contrôler son mouvement, ainsi que les différentes tâches qu'il peut effectuer.

La première utilisation de la vision en boucle fermée a été réalisée par Shirai et Inoue en 1973 dans [Shi.73]. Ils décrivent comment un capteur de vision peut augmenter la précision du positionnement d'un robot mobile. On parlait alors de retour par vision (visual feed-back). Mais, c'est à Hill et Park [Hil.79] que l'on doit l'apparition du terme asservissement visuel (visual servoing) en 1979.

Grâce à cette apparition, plusieurs approches ont ensuite vu le jour. Elles ont été séparées par Sanderson et Weiss dans [San.80] en deux grandes classes selon la nature des mesures utilisées à l'entrée du module de commande :

- Une première approche, nommée asservissement en position ou asservissement visuel 3D (position-based control),
- Une seconde approche connue sous le nom d'asservissement basé sur l'image ou asservissement visuel 2D (image-based control).

Ces deux approches d'asservissement ont fait l'objet de plusieurs travaux depuis des années. Hashimoto a présenté dans [Has.93] une approche 3D optimale pour contrôler à distance un bras manipulateur. Espiau expose dans [Esp.96] une nouvelle approche d'asservissement 2D basée sur une boucle de commande visuelle utilisant une caméra.

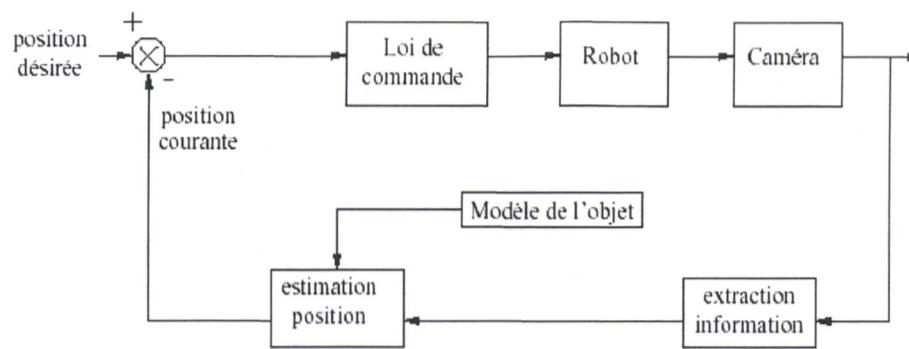
Depuis peu, de nouvelles approches ont été développées, nous citons comme exemple l'asservissement du type 2D 1/2 présenté par Malis dans [Mal.98], ou l'asservissement du type d2D/dt exposé dans [Cre.98]. Dans la suite de cette partie nous allons nous focaliser sur les deux approches de base de l'asservissement visuel.

### 2-1 Asservissement 3D

Cet asservissement est appelé aussi asservissement en position ou asservissement dans l'espace cartésien. Comme son nom l'indique, l'asservissement visuel 3D utilise à l'entrée de la boucle de commande des informations visuelles tridimensionnelles exprimées dans un repère euclidien [Dor.95].

Dans ce type de commande, les informations visuelles sont extraites afin de contrôler la position et l'orientation de la caméra par rapport à l'objet.

La Figure.4-1 représente un schéma bloc correspondant à un asservissement visuel 3D.



**Figure 4-1.** Asservissement visuel 3D.

L'asservissement visuel 3D présente l'avantage de fournir des mesures directement dans l'espace euclidien, ce qui permet l'emploi des lois de commande extrêmement simples pour aller itérativement d'une position à une autre. L'inconvénient qui en résulte est qu'aucun contrôle véritable dans l'image n'est effectué, ce qui implique que l'objet d'intérêt peut très bien sortir du champ de vue de la caméra pendant son déplacement, entraînant alors l'arrêt de la loi de commande.

Il existe deux manières de réaliser ce type d'asservissement suivant l'état statique ou dynamique de l'acquisition et du déplacement, ainsi nous avons :

- 1- **Static Look and Move** : La phase d'acquisition d'images avec déduction de la situation courante est enchaînée séquentiellement avec le déplacement du robot. Cette technique est surtout utilisée pour les objets fixes,
- 2- **Dynamic Look and Move** : Les deux étapes (acquisition et déplacement) sont effectuées d'une manière simultanée. Cette technique a l'avantage de pouvoir suivre des objets en mouvement puisque les informations visuelles sont prises en compte dès qu'elles sont opérationnelles.

## 2-2 Asservissement 2D

Appelé aussi Asservissement dans l'image ou Image-based Look-and-move. Contrairement à un asservissement 3D, les techniques d'asservissement visuel 2D ne nécessitent pas de reconstruction 3D [Chau.91]. Les raisonnements sont réalisés directement au niveau du capteur. Les lois de commande consistent alors à contrôler le mouvement de la caméra afin que les mesures dans l'image représentées par un vecteur " $s$ " atteignent une valeur désirée " $s^*$ ", voire suivent une trajectoire en fonction du temps " $s^*(t)$ ".

Cette technique repose essentiellement sur deux points fondamentaux :

- 1- Choix des informations visuelles,
- 2- Obtention d'une relation liant ces informations et les mouvements de la caméra.

Cette relation est en fait une matrice appelée suivant les auteurs, matrice d'interaction, jacobien de la tâche ou jacobien de l'image. Cette matrice joue un rôle important dans l'élaboration des diverses lois de commande possibles. L'obtention de la matrice d'interaction

pour les différentes primitives géométriques les plus courantes (points, droites, cercles, cylindres, sphères et ellipses) a permis d'effectuer une classification des différentes tâches robotiques réalisables par asservissement visuel 2D, en fonction des liaisons virtuelles souhaitées entre la caméra et l'objet considéré. La Figure 4-2 représente le schéma bloc correspondant à un asservissement visuel 2D.

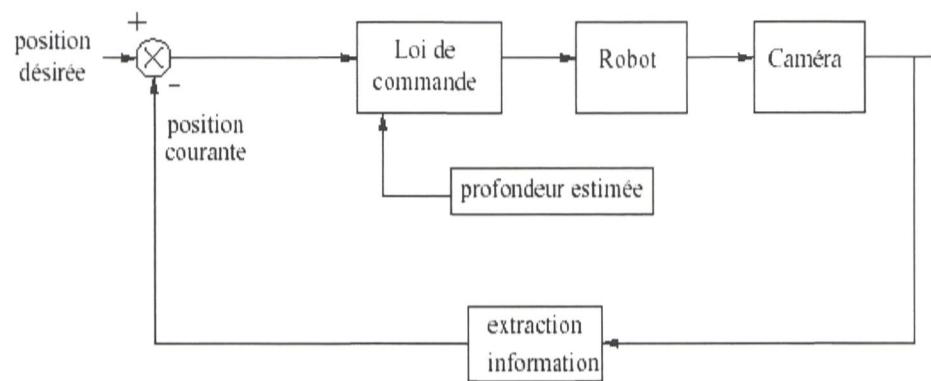


Figure 4-2. Asservissement visuel 2D.

Il faut noter que l'asservissement visuel 2D a apporté toute satisfaction dans un grand nombre de travaux. L'originalité de cette technique réside dans le fait que la tâche à réaliser est spécifiée directement dans l'espace image. Ainsi, un des principaux avantages de cette approche réside dans la précision et la facilité de mise en œuvre des tâches réalisées. On évite ici le calcul de pose à partir des mesures faites dans l'image, ce qui permet d'être moins sensible à d'éventuelles erreurs de calibration de la caméra.

### 2-3 Notre méthode d'asservissement

Pour notre travail de thèse nous allons nous situer dans le contexte de l'asservissement visuel 2D pour positionner la pince du bras robotique dans l'espace par rapport à l'objet cible. Dans la suite, nous allons voir quels sont les objets simples utilisés pour notre approche, comment calculer les paramètres de l'asservissement visuel 2D ainsi que les dimensions des objets à saisir. Après, nous allons exposer notre méthode globale d'asservissement visuel adaptée au bras Manus.

Pour obtenir le meilleur algorithme possible, trois approches utilisant la distance entre l'objet et le système de mesure ont été explorées. La première consiste à utiliser deux types d'objets simples de couleur unique. Dans la deuxième approche, nous utilisons la forme de l'objet et sa couleur dans des environnements de travail réels. Enfin, dans la troisième approche, nous utilisons plusieurs types d'objets avec des formes et des couleurs variées dans des scènes complexes.

#### 2-3-1 Système de mesure

Avant d'exposer les trois approches développées pour l'asservissement visuel de la pince du robot, nous allons présenter tout d'abord, les modes de commande du bras Manus et les types de mouvements qu'il peut exécuter. Puis, nous présentons le système expérimental d'acquisition installé sur la pince.

Grâce à ses six degrés de liberté et son extrémité dotée d'une pince à deux doigts comme l'illustre la Figure.4-3, le bras a la possibilité de saisir des objets d'un poids inférieur à 1,5 kg dans les différentes directions de l'espace [Eve.01].

Son installation sur une base mobile au lieu d'un fauteuil roulant va donner à la personne handicapée plus de maniabilité et plus d'autonomie dans les tâches quotidiennes.

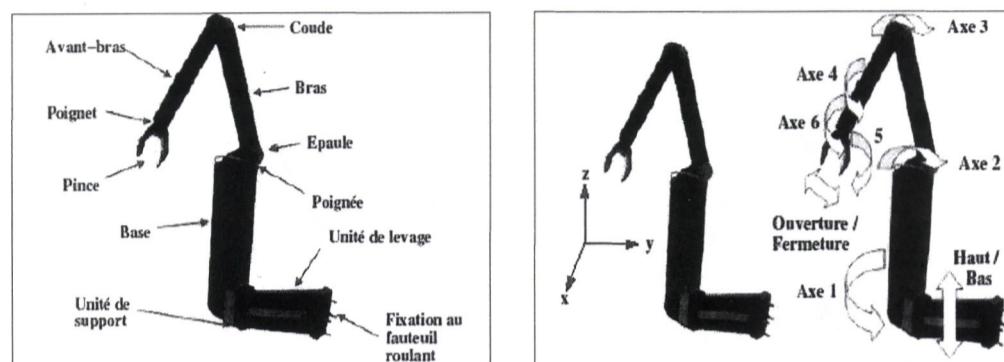


Figure 4-3. Différentes éléments du bras Manu.

D'après le constructeur, il existe deux modes de commande du bras Manus :

1- **Le mode manuel** : Il peut être commandé grâce à :

- Un clavier 16 touches spécial,
- Un joystick,
- Un PC en mode transparent (communication entre le robot et PC de commande à travers un bus CAN).

2- **Le mode automatique** : Représenté par deux manœuvres :

- Le Folding in (repliement automatique),
- Le Folding out (dépliement automatique).

Un indicateur digital à LED renseigne l'utilisateur sur l'état de fonctionnement. Le bras peut exécuter aussi deux types de mouvement dans l'espace :

- 1- **Des mouvements cartésiens** (mode XYZ) : Déplacer la pince du robot d'un point de l'espace à un autre en suivant des translations suivant les axes  $\vec{x}$ ,  $\vec{y}$  et  $\vec{z}$ ,
- 2- **Des mouvements articulaires** : Chaque articulation (ou degré de liberté) est contrôlée individuellement par le système de commande.

Le système expérimental utilisé dans l'acquisition des données vision et ultrasons est représenté sur la Figure.4-4. Il comporte deux éléments installés sur le préhenseur du bras manipulateur Manus.

Le système ultrasonore que nous utiliserons pour la reconnaissance de la forme de l'objet et le calcul de la distance pince / objet et une caméra couleur sans fil qui fournit les images à partir desquelles nous extrairons les paramètres de l'objet : contours, régions et couleurs.

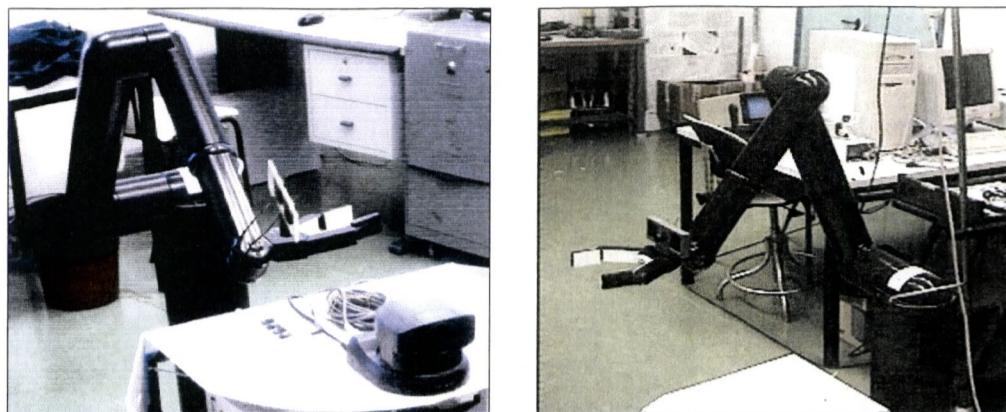


Figure 4-4. Système expérimental utilisé pour l'acquisition.

### 2-3-2 Calcul des corrections

Dans notre travail nous avons choisi d'utiliser un asservissement visuel 2D. Ce type d'asservissement génère un déplacement de la pince du robot vers l'objet qu'on souhaite saisir. Ce déplacement est composé de trois coordonnées (Figure 4-5), qui vont être injectées dans la commande cartésienne du bras Manus.

Ces coordonnées sont appelées paramètres de l'asservissement visuel :

- 1- Un déplacement horizontal de la pince que nous appelons correction horizontale  $C_x$ ,
- 2- Un déplacement vertical de la pince que nous appelons correction verticale  $C_y$ ,
- 3- Un déplacement en profondeur selon l'axe  $\vec{z}$ , qui est calculé à partir des informations fournies par le système ultrasonore.

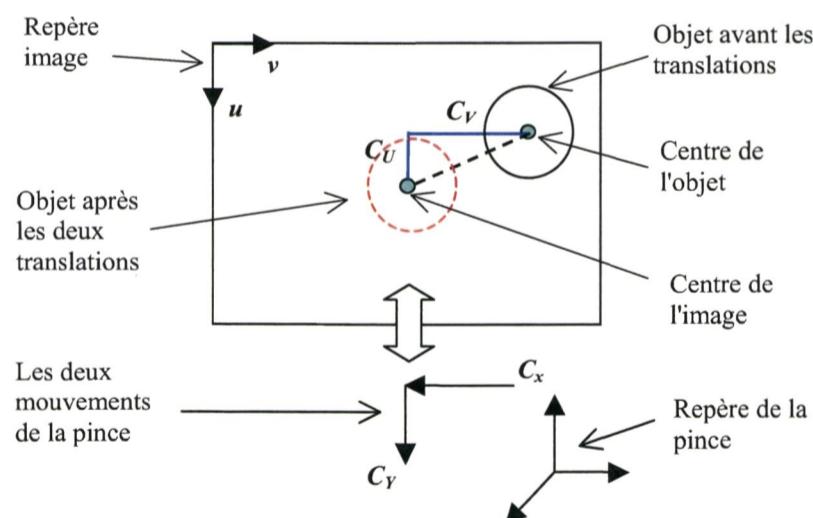


Figure 4-5. Les deux corrections de la pince  $C_x$  et  $C_y$ .

Les deux corrections  $C_x$  et  $C_y$  représentent la phase de centrage de l'image. Son objectif est de ramener le centre de l'objet vers le centre de l'image afin de les superposer. Nous utilisons dans cette partie un rectangle d'une couleur différente de celle de l'objet pour le localiser dans l'image. Les dimensions de ce rectangle vont suivre la taille de l'objet.

Le processus de centrage nécessite quatre étapes :

- 1- Calculer par traitement d'images les coordonnées en pixels du centre de l'objet ( $u_{co}, v_{co}$ ) dans le repère de l'image,
- 2- Transformer ces deux coordonnées du repère image vers le repère de la caméra grâce à son modèle interne. Les valeurs de ces nouvelles coordonnées sont  $X_{co}$  et  $Y_{co}$  données par l'équation (4-1), où  $\alpha_u$ ,  $\alpha_v$ ,  $u_0$  et  $v_0$  sont les paramètres intrinsèques de la caméra et  $z$  la distance de l'objet :

$$\begin{cases} X_{co} = z \cdot \frac{u_{co} - u_0}{\alpha_u} \\ Y_{co} = z \cdot \frac{v_{co} - v_0}{\alpha_v} \end{cases} \quad (4-1)$$

- 3- Transformer les coordonnées du centre de l'image ( $u_{ci}, v_{ci}$ ) du repère image vers le repère de la caméra par la même équation, ce qui donne :

$$\begin{cases} X_{ci} = z \cdot \frac{u_{ci} - u_0}{\alpha_u} \\ Y_{ci} = z \cdot \frac{v_{ci} - v_0}{\alpha_v} \end{cases} \quad (4-2)$$

- 4- Calculer les deux corrections  $C_x$  et  $C_y$  (en millimètres) qui correspondent aux mouvements de la pince du bras manipulateur dans les axes horizontaux et verticaux de la caméra par, et ainsi :

$$\begin{cases} C_x = \sqrt{(X_{co} - X_{ci})^2} \\ C_y = \sqrt{(Y_{co} - Y_{ci})^2} \end{cases} \quad (4-3)$$

Les quatre étapes de cette phase permettent d'apporter l'axe  $\vec{z}$  de la caméra vers le centre de l'objet. Une fois il y a superposition entre le centre de l'objet et de l'image, le bras pourra ainsi se déplacer vers l'objet d'une distance  $z$ .

Les trois déplacements calculés dans cette partie sont appelés les paramètres de l'asservissement visuel. En plus des ces paramètres nous calculons également la largeur de l'objet. Ceci est détaillé dans la partie suivante.

### 2-3-3 Calcul de la largeur de l'objet

La largeur d'un objet est la distance horizontale entre ses bords extérieurs comme il est montré dans la Figure.4-6. Si l'objet à saisir a une forme cylindrique ou sphérique, sa largeur est donnée par son diamètre. Par contre si l'objet a une forme atypique, sa largeur est donnée par la distance horizontale entre les deux droites parallèles passant par ses deux points extrêmes gauche et droite.

Pour calculer cette largeur nous suivons une stratégie similaire à celle que nous venons de décrire. La seule différence est qu'au lieu d'extraire le centre de l'objet nous allons déterminer deux points appartenant au contour de l'objet  $P_1 (u_1, v_1)$  et  $P_2 (u_2, v_2)$ .

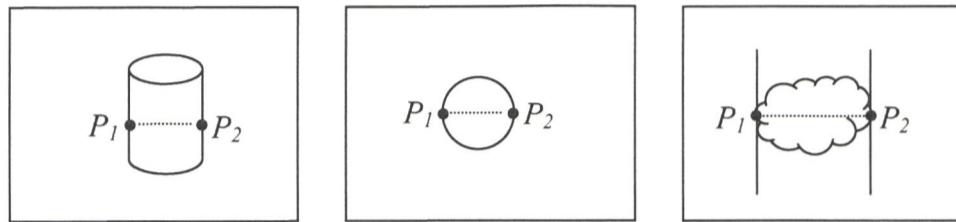


Figure 4-6. Largeur de trois types d'objets.

Puis, nous transformons les coordonnées de ces deux points du repère de l'image vers le repère de la caméra en utilisant l'équation précédente (4-1), ainsi, les nouvelles coordonnées du point  $P_1$  sont données par :

$$\begin{cases} X_1 = z \cdot \frac{u_1 - u_0}{\alpha_u} \\ Y_1 = z \cdot \frac{v_1 - v_0}{\alpha_v} \end{cases} \quad (4-4)$$

Et celles du point  $P_2$  sont données par :

$$\begin{cases} X_2 = z \cdot \frac{u_2 - u_0}{\alpha_u} \\ Y_2 = z \cdot \frac{v_2 - v_0}{\alpha_v} \end{cases} \quad (4-5)$$

La largeur de l'objet représente ainsi, la distance entre ces deux points. Elle est calculée par l'équation de distance suivante :

$$\text{Largeur} = \sqrt{(X_1 - X_2)^2 + (Y_1 - Y_2)^2} \quad (4-6)$$

Le calcul de la largeur de l'objet va nous permettre dans une étape prochaine de reconnaître un objet parmi d'autres objets de même forme mais de dimensions différentes. Comme par exemple, pouvoir distinguer une balle de ping-pong et une balle de tennis

#### 2-4 Stratégie d'approche

L'algorithme d'asservissement visuel développé pour approcher les objets utilise deux types de capteurs, une caméra couleur sans fil qui fournit des informations 2D et un système ultrasonore comme celui utilisé dans le chapitre précédent qui fournit une information de forme de l'objet et de sa profondeur par rapport à la pince.

Nous allons présenter dans la suite l'algorithme général qui gère l'opération d'approche, puis nous développons les trois stratégies développées.

#### 2-4-1 Organigramme de l'approche d'objets

Cet organigramme est composé de cinq blocs de base dont les trois premiers appartiennent à la fonction de perception et les deux autres blocs correspondent à la fonction commande (Figure.4-7).

Une fonction de choix de l'objet à saisir est nécessaire au début de la stratégie d'approche. Cette dernière permet à la personne handicapée de sélectionner un objet simple parmi d'autres objets stockés dans la mémoire du système de commande sous forme d'une table contenant les objets saisissables et leurs paramètres.

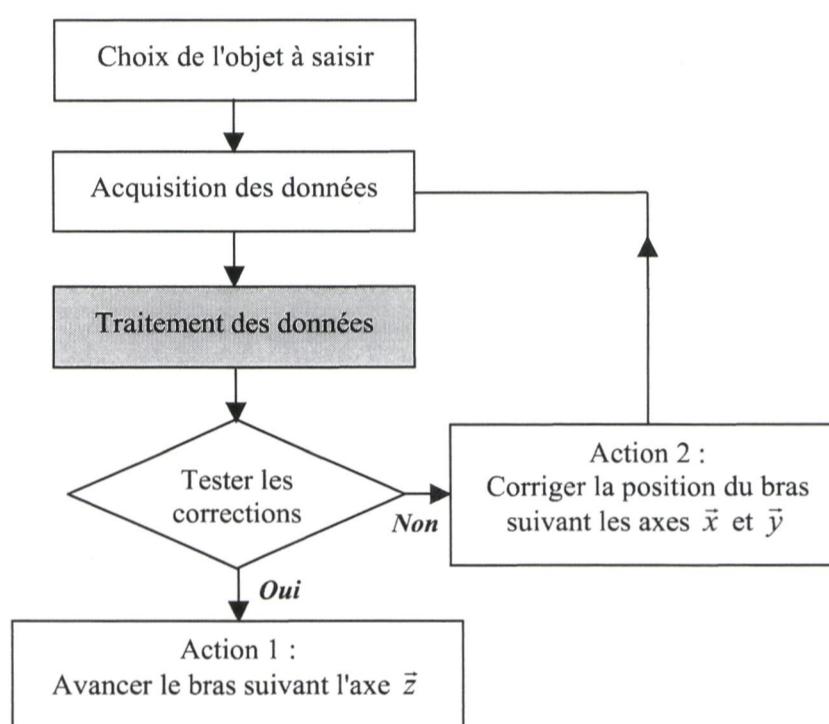


Figure 4-7. Schéma bloc de l'asservissement visuel.

Chaque objet simple est caractérisé dans cette table par un ensemble de paramètres comme sa forme, sa couleur et ses dimensions. Une interface homme / machine, va permettre plus tard, à la personne handicapée de faire le choix sur l'objet simple qu'il souhaite saisir.

Les cinq blocs de base qui constituent notre stratégie d'approche sont les suivants :

- 1- **L'acquisition** : Ce premier bloc permet de recevoir les caractéristiques de l'objet que nous voulons approcher comme ses dimensions, sa couleur, sa distance par rapport au système de mesure et sa forme. Deux types de capteurs sont utilisés dans notre algorithme d'approche, la vision et un les ultrasons,
- 2- **Le traitement** : L'image et l'écho reçus nécessitent plusieurs traitements pour extraire leurs paramètres comme la segmentation, le filtrage et le seuillage,

- 3- **Le test** : Les paramètres de l'asservissement visuel obtenus dans l'étape de traitement vont être comparés avec des valeurs de référence pour déplacer le bras du robot vers la position souhaitée,
- 4- **Action 1** : Ce bloc représente l'exécution d'un mouvement du bras suivant les consignes données auparavant. Il permet d'avancer la pince vers l'objet
- 5- **Action 2** : Ce dernier bloc permet de corriger la position de la pince dans l'espace avant d'avancer vers l'objet à saisir.

Il faut noter ici que chacun de ces cinq blocs est composé à son tour d'un ou de plusieurs éléments. La stratégie d'approche que nous avons développée est itérative, c'est à dire, elle fonctionne en boucle fermée.

A chaque étape de la stratégie, le système d'acquisition va prendre une image et un écho ultrasonore, puis, suivant les traitements obtenus et les tests effectués, le système de traitement va prendre une des deux décisions suivantes :

- Avancer le bras d'un pas de 10 cm vers l'objet suivant l'axe  $\vec{z}$  de la caméra, qui correspond à l'action 1,
- Effectuer une ou deux corrections de la position du bras suivant le plan déterminé par les axes  $\vec{x}$  et  $\vec{y}$  ce qui correspond à l'action 2.

Comme nous l'avons annoncé auparavant, le développement de notre stratégie d'approche a été fait d'une façon progressive, d'une situation simple utilisant un seul objet avec deux à trois attributs vers une situation plus complexe comportant plusieurs objets simples et utilisant différents paramètres.

C'est le bloc "traitement de données" présenté en gris sur l'algorithme précédent qui va décider de l'approche d'asservissement visuel à suivre. Dans la suite de ce chapitre, nous allons étudier les trois configurations développées et discuter les résultats obtenus avec chacune d'elles.

Tout d'abord, nous allons montrer comment extraire la distance séparant un objet simple et le système d'acquisition à partir d'un écho ultrasonore reçu. Ce paramètre de profondeur va être utilisé par les trois stratégies d'approche que nous allons exposer dans la suite.

#### 2-4-2 Calcul de la distance

Pour extraire la distance de l'objet que nous voulons approcher, nous utilisons l'écho ultrasonore de l'objet obtenu par l'intermédiaire du système d'acquisition. Un seuillage sur la valeur de l'amplitude de l'écho permet de mesurer la distance de l'objet en face du capteur avec une bonne précision de l'ordre de quelques centimètres.

La distance  $z$  entre le capteur ultrasonore et l'objet cible est proportionnelle au temps de vol (TOF) vu dans le chapitre II. L'équation qui relie ces deux paramètres est donnée par :

$$z = ct_0 / 2 \quad (4-7)$$

Où  $t_0$  est le temps de vol enregistré en seconde et  $c$  est la vitesse du son dans l'air, égale à 343.3 m/s à une température ambiante de 25°.

Le seuil d'amplitude utilisé dans ce travail pour extraire la distance de l'objet correspond à une valeur de 0.03 Volts comme le montre la Figure 4-8. Elle montre aussi la largeur de l'écho qui varie suivant la forme de l'objet.

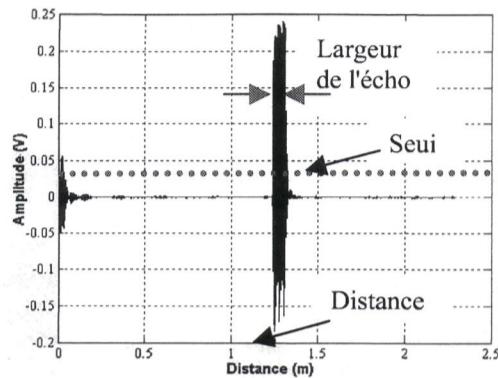


Figure 4-8. Mesure de distance avec l'écho de l'objet.

La valeur de la distance objet / capteur  $z$  obtenue par seuillage de l'écho ultrasonore, ainsi que les deux corrections  $C_x$  et  $C_y$  vues précédemment et obtenues grâce à la vision représentent les trois paramètres d'un asservissement visuel.

### 3- Première approche

Dans cette première approche nous allons calculer les paramètres de l'asservissement visuel en utilisant une configuration simple constituée d'un objet d'une seule couleur et d'un fond d'image de couleur uniforme. Deux types d'informations sont employés pour localiser l'objet dans l'image et la scène :

- 1- L'écho ultrasonore qui permet de déterminer la distance  $z$  de l'objet à la pince,
- 2- Les contours de l'objet qui sont utilisés pour calculer les corrections  $C_x$  et  $C_y$ .

Avant de donner l'algorithme qui gère cette approche et les résultats obtenus, nous exposons d'abord les objets utilisés et la méthode de segmentation par contours développée pour localiser l'objet dans l'image.

#### 3-1 Objets utilisés

Deux types d'objets sont utilisés, le cylindre est représenté par une canette de soda d'une couleur bleue et d'un diamètre de 64.3 mm. La sphère est représentée par une balle de ping-pong de couleur orange et d'un diamètre de 39.2 mm. Comme nous l'avons cité auparavant, la personne handicapée pourra utiliser une interface pour sélectionner à distance les objets qu'il souhaite saisir.

Cette interface peut être un écran tactile, un microphone pour recevoir des commandes vocales ou encore un ensemble composé d'un joystick et d'un écran pour le retour d'effort. Le problème posé actuellement n'est pas l'interface que la personne handicapée utilise pour faire ses choix, mais comment approcher les différents objets. La partie qui va suivre permet de localiser l'objet d'intérêt dans l'image par analyse des contours.

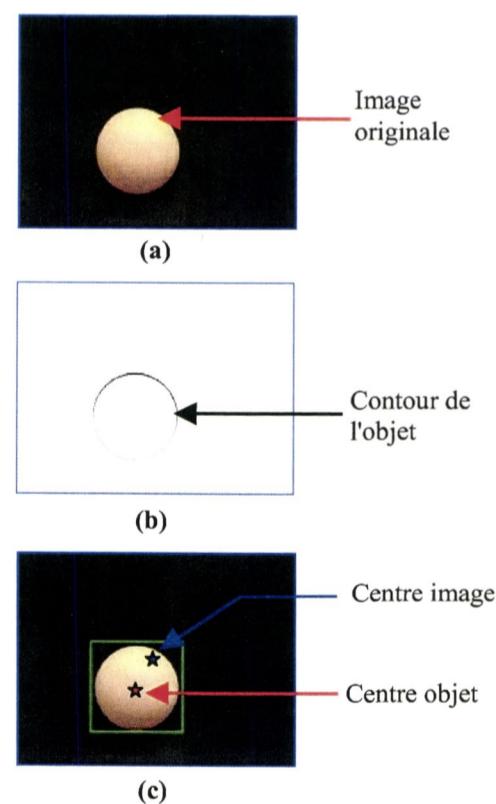
### 3-2 Segmentation par contours

Comme pour le chapitre précédent, nous avons utilisé un détecteur de contours optimal du type Deriche [Der.87] pour extraire les paramètres de l'objet que nous voulons approcher dans l'image. L'image du gradient obtenue à la sortie de ce filtre est transformée en image binaire grâce à un seuillage par hystérosis.

Après cette première étape, les pixels des points contours sont stockés dans deux accumulateurs, un pour l'axe  $\vec{u}$  et l'autre pour l'axe  $\vec{v}$ . Puisqu'il y a un seul objet dans l'image, la moyenne des pixels dans les deux accumulateurs correspond automatiquement aux coordonnées  $u_c$  et  $v_c$  du centre de l'objet dans l'image.

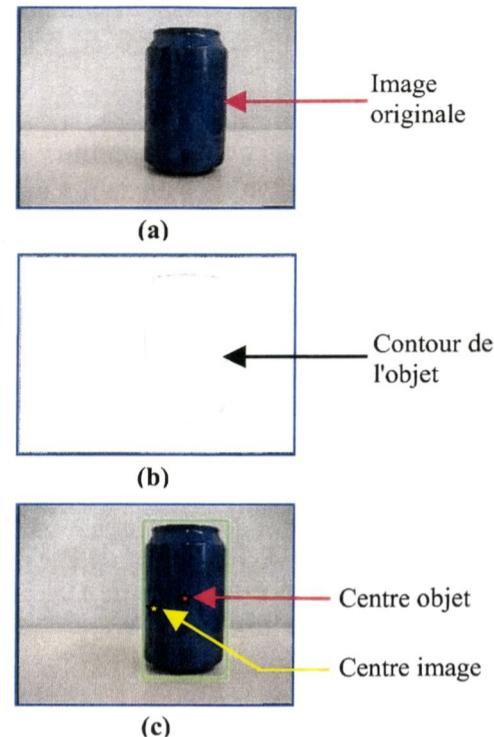
La connaissance des coordonnées de ce point, ainsi que les coordonnées du centre de l'image permettent de calculer les corrections, horizontale  $C_v$  et verticale  $C_u$  en pixels dans le repère image. Le modèle interne de la caméra permet à la fin, de mesurer les deux corrections  $C_x$  et  $C_y$  en centimètres, nécessaires pour le centrage de l'objet dans l'image comme nous l'avons expliqué au paragraphe 2-3-2.

La Figure 4-9 montre un exemple de localisation d'un objet dans l'image par la segmentation contours. Elle représente les contours extraits par un filtre de Deriche pour l'objet balle. Les coordonnées du centre de l'objet représenté par une étoile rouge sont calculées à partir de la moyenne sur les points contours. Le centre de l'image donné par l'étoile bleue.



**Figure 4-9.** Localisation de l'objet balle dans l'image, (a) : Objet balle, (b) : Contours, (c) : Centres objet et image.

La Figure.4-10 quant à elle, donne un deuxième exemple de localisation dans l'image par la segmentation contours d'un objet canette. Le centre de l'objet est représenté sur la figure par une étoile jaune.



**Figure 4-10.** Localisation de l'objet canette dans l'image, (a) : Objet canette, (b) : Contours, (c) : Centres objet et image.

La Table.4-1 donne les deux coordonnées réelles en pixels du centre des deux objets utilisés ( $U_{cr}, V_{cr}$ ), les coordonnées en pixels de ces centres calculées par la méthode de segmentation exposée auparavant ( $U_{cc}, V_{cc}$ ) et l'erreur en pixels sur les valeurs de ces coordonnées ( $E_u, E_v$ ) par rapport aux valeurs réelles.

Objets	Coordonnées réelles du centre (pixels)		Coordonnées calculées du centre (pixels)		Erreurs (pixels)	
	$U_{cr}$	$V_{cr}$	$U_{cc}$	$V_{cc}$	$E_u$	$E_v$
Balle	384.12	324.48	385.99	325.77	1.87	1.29
Canette	273.16	468.73	271.61	467.34	1.55	1.39

**Table 4-1.** Coordonnées réelles et calculées des centres d'objets.

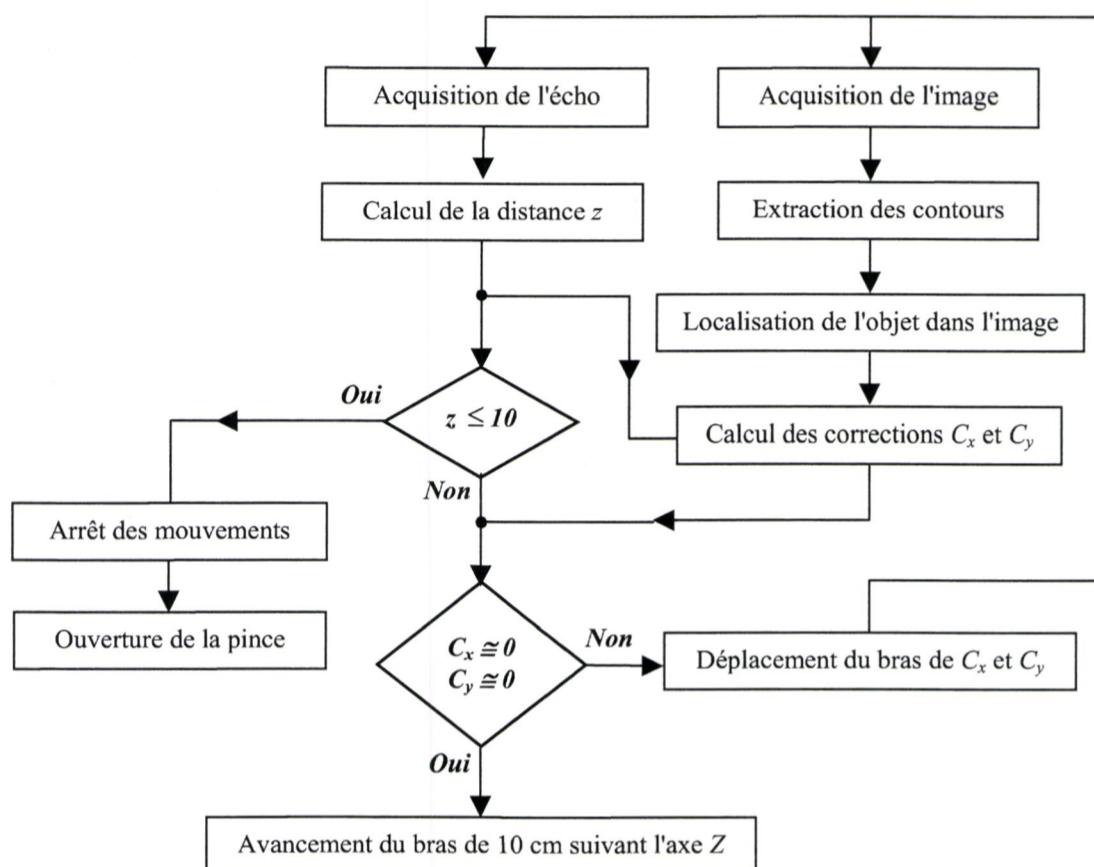
Les résultats obtenus avec cette méthode montrent une précision de l'ordre de quelques pixels entre les coordonnées calculées et les coordonnées réelles des centres d'objets utilisés dans cette première approche.

### 3-3 Algorithme

L'algorithme développé pour cette première approche est représenté par la Figure.4-11. Il s'appuie sur deux types d'informations, la distance  $z$  à l'objet donnée par le système

ultrasonore et les corrections  $C_x$  et  $C_y$  obtenues à partir de l'image de l'objet prise par une caméra HF. Si la distance entre la pince du robot et l'objet donnée par le sonar est supérieure à un seuil de 10 cm, l'algorithme va calculer les deux corrections  $C_x$  et  $C_y$ .

Les valeurs de ces deux corrections vont déterminer la commande qu'il faut appliquer au bras manipulateur. Soit l'avancer vers l'objet de 10 cm suivant l'axe  $Z$  de la caméra si ces deux distances sont proches de zéro (inférieur à 0.2 cm), soit, corriger la position de la pince de  $C_x$  suivant l'axe  $X$  et de  $C_y$  suivant l'axe  $Y$  et relancer l'algorithme.



**Figure 4-11.** Algorithme de la première approche d'asservissement.

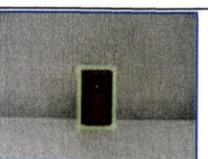
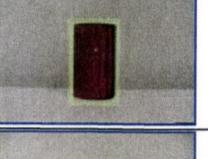
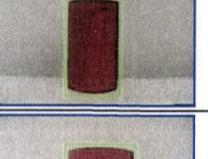
La partie commande du bras dans cette approche commence au niveau de la deuxième boucle de test. Il y a deux types de manœuvres de déplacement que l'algorithme peut ordonner, une translation suivant  $Z$  pour approcher l'objet ou un repositionnement de la pince suivant le plan  $XY$  pour centrer l'objet.

Le fonctionnement en boucle fermée permet de corriger à chaque étape de l'algorithme les coordonnées de la pince pour assurer une bonne stratégie et approcher l'objet que nous souhaitons saisir avec succès.

### 3-4 Résultats

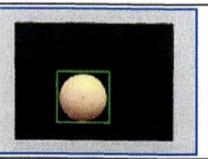
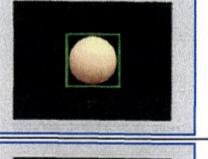
Nous avons testé cet algorithme sur deux objets différents, une canette de soda qui représente la forme cylindrique et une balle de ping-pong pour la forme sphérique. A chaque étape, nous fixons la distance entre le système de mesure et l'objet par le dispositif ultrasonore, puis, nous prenons une image de la scène.

La Table.4-2 montre l'objet canette encadré par un rectangle vert qui prouve que l'objet a été bien localisé dans l'image. Une fois cette étape réalisée, l'algorithme calcule les deux corrections  $C_x$  et  $C_y$  et donne la distance à l'objet  $z$ . Ces trois paramètres vont être transmis au système de commande pour pouvoir approcher l'objet à saisir.

Distance $z$	Canette	$C_x$ (cm)	$C_y$ (cm)
60 cm		1.22	0.71
50 cm		1.27	0.96
40 cm		1.28	0.17
30 cm		1.19	0.16
20 cm		1.16	-0.45

**Table 4-2.** Calcul des paramètres pour la canette rouge.

La Table.4-3 donne les trois paramètres de l'asservissement visuel pour l'objet balle de ping-pong encadré par un rectangle vert.

Distance $z$	Balle	$C_x$ (cm)	$C_y$ (cm)
40 cm		1.28	2.03
30 cm		0.06	0.04
20 cm		-0.33	0.72

**Table 4-3.** Calcul des paramètres pour la balle orange.

Sur ces deux exemples nous pouvons constater l'efficacité de cette première approche qui est simple, rapide et précise. Toutefois, son efficacité est liée à deux conditions qui définissent ses limites : un seul objet dans la scène et de couleur unie.

#### 4- Deuxième approche

La deuxième approche étudie toujours des objets de forme cylindrique et sphérique mais appartenant à des scènes complexes et comportant plusieurs objets de couleurs et de dimensions multiples. Nous utilisons aussi dans cette partie un réseau de neurones spécialisé attaché au système d'acquisition ultrasonore et qui permet de reconnaître la forme des objets situés face à lui.

Nous allons présenter brièvement les objets utilisés dans cette approche, puis, les méthodes de traitement d'images employées pour localiser l'objet d'intérêt dans l'image et calculer le centre de l'objet et les corrections  $C_x$  et  $C_y$ . Après, nous exposons le réseau de neurones utilisé pour reconnaître la forme des objets et nous donnons les résultats de classifications. Enfin, nous présentons l'algorithme développé et les résultats obtenus.

##### 4-1 Objets utilisés

Nous utilisons dans cette partie huit objets dans cette partie d'asservissement visuel, trois balles de ping-pong de couleurs différentes, une balle de tennis verte, une petite canette rouge et trois canettes de taille normale et de couleurs différentes. La Table.4-4 présente ces différents objets et leurs caractéristiques de forme et de couleur.

Numéro d'objet	Objet	Couleur	Forme
1	Balle	Bleue	Sphère
2	Balle	Orange	Sphère
3	Balle	Jaune	Sphère
4	Balle	Verte	Sphère
5	Canette	Rouge	Cylindre
6	Canette	Bleue	Cylindre
7	Canette	Orange	Cylindre
8	Canette	Jaune	Cylindre

**Table 4-4.** Objets utilisés dans la deuxième approche.

Les objets utilisés dans cette deuxième approche gardent toujours les deux formes de base qui sont le cylindre et la sphère. Nous pouvons ainsi regrouper ces objets en deux groupes suivant leur forme. C'est l'attribut couleur qui va différencier les éléments de ces deux groupes. Notons ici que nous utilisons dans nos travaux des couleurs vives comme le jaune, le vert ou l'orange.

##### 4-2 Traitement d'images

Pour localiser l'objet d'intérêt dans l'image et calculer les paramètres de l'asservissement visuel, nous allons combiner deux méthodes de traitements d'images. La première extrait les

régions de l'image et la deuxième choisit une des régions à partir de sa couleur. Ainsi nous allons détailler dans la suite l'algorithme utilisé pour la segmentation de l'image en régions, puis, la méthode de seuillage couleurs utilisée dans la sélection de l'objet d'intérêt parmi les autres objets.

#### 4-2-1 Segmentation par région

Pour extraire les régions de l'image nous utilisons un algorithme de morphologie mathématique appelé "Ligne de Partage des Eaux" (LPE) ou "Watershed" [Gur.05]. C'est un algorithme de segmentation d'images qui sert à décomposer une image en régions homogènes. Il utilise la description des images en termes morphologiques. Ainsi, une image peut être perçue comme un relief si l'on associe le niveau de gris de chaque point à une altitude. Il est alors possible de définir la ligne de partage des eaux comme étant la crête formant la limite entre deux bassins versants.

Notons enfin, que l'on n'applique pas la LPE à l'image originale mais à son gradient obtenu après le filtrage de Deriche développé précédemment. Cela crée ainsi des lignes de partage des eaux aux points de discontinuité du niveau de gris, ce qui est intéressant pour les applications de segmentation. Le paramètre le plus important qui détermine la qualité de la segmentation en régions est "la taille d'une région". Cet attribut se fixe au début de l'algorithme par rapport aux dimensions des objets étudiés. Dans notre travail nous avons choisi de garder les régions de l'image dont l'aire est supérieure ou égale à 1000 pixels.

#### 4-2-2 Seuillage par les couleurs

L'algorithme LPE permet d'avoir les différentes régions de l'image. Pour sélectionner la région formée par l'objet que nous voulons saisir, nous devons déterminer sa couleur au début de l'algorithme de segmentation. Dans notre travail, nous effectuons tout d'abord un étiquetage des régions, puis, nous choisissons la région qui nous intéresse à partir de l'analyse de sa couleur.

Dans l'espace RGB que nous utilisons, une couleur est formée de trois composantes, rouge, verte et bleue [Mon.05]. Une image couleur peut être décomposée en trois images suivant les trois espaces couleurs, le rouge, la verte et la bleue. La Figure 4-12 montre un exemple de cette décomposition d'une image RGB en trois images.

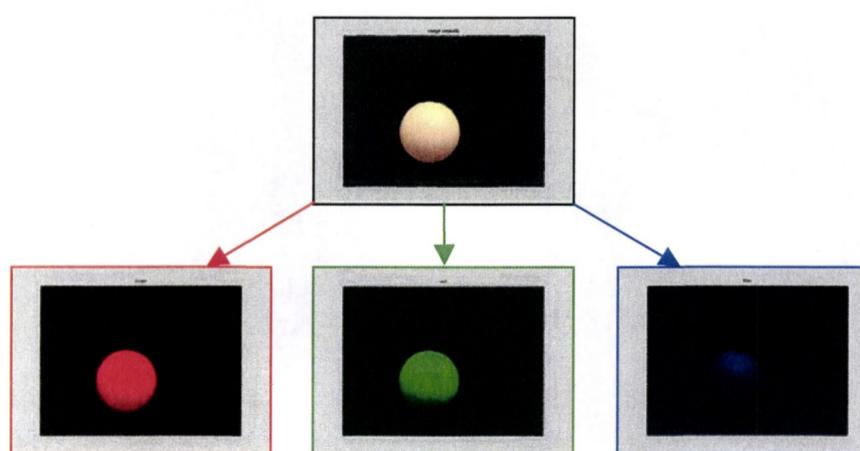


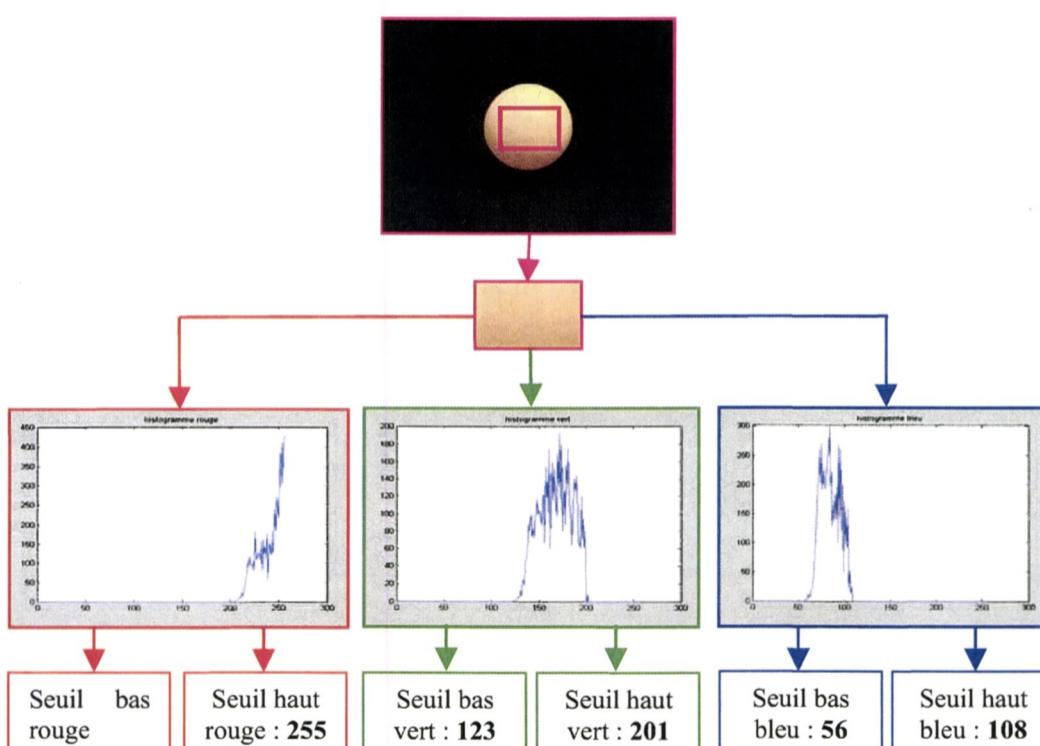
Figure 4-12. Séparation des trois couleurs d'une image RGB.

L'étude des différents systèmes de couleurs et des histogrammes de couleurs nous a permis de caractériser n'importe quelle couleur par un ensemble de six seuils, trois du niveau haut et trois du niveau bas [Tra.01]. Ainsi pour localiser un objet avec une couleur donnée dans une image il suffit de reconnaître le seuil haut et le seuil bas de chaque composante.

L'analyse des histogrammes couleurs nous permet de définir ces six seuils et de localiser une couleur donnée. Chaque histogramme couleur est représenté par un vecteur de taille 256 correspondant à 256 niveaux de gris.

Les amplitudes de cet histogramme vont se situer dans une plage définie à l'intérieur de ce vecteur. Ainsi, le rang Min de ce vecteur correspond au seuil bas d'une couleur et le rang Max correspond à son seuil haut.

Nous allons donner dans la Figure.4-13 un exemple des six seuils d'une couleur orange. Au début nous sélectionnons une partie de l'image avec la couleur de l'objet, puis, nous décomposons cette partie en trois composantes, rouge, verte et bleue et nous calculons l'histogramme de chacune. A partir de ces histogrammes nous extraireons les seuils haut et bas de chaque composante couleur.



**Figure 4-13.** Les trois histogrammes de la couleur de l'objet et les six seuils qui correspondent à cette couleur orange.

Dans la Table.4-5 nous donnons les 6 seuils correspondants aux couleurs des objets utilisés dans cette deuxième approche d'asservissement visuel. Les valeurs bas et haut des ces seuils se situent entre 0 et 255 (256 niveaux de gris).

Pour chaque objet nous possédons 6 seuils que nous calculons d'une manière automatique à partir des histogrammes couleurs. L'extraction de ces différents seuils se fait d'une manière hors-ligne, avant de lancer la procédure d'asservissement visuel.

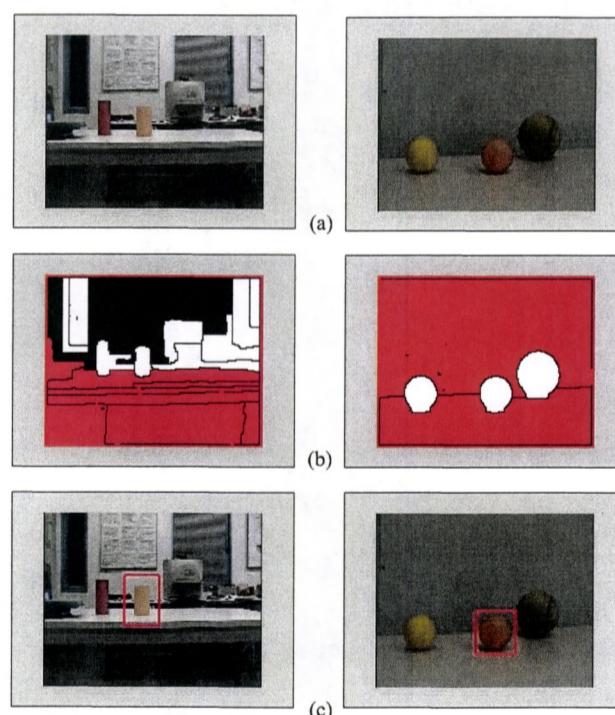
Couleur d'objet	Composante rouge		Composante verte		Composante bleue	
	Seuil bas	Seuil haut	Seuil bas	Seuil haut	Seuil bas	Seuil haut
Orange	202	255	123	201	56	108
Jaune	145	172	119	153	37	75
Rouge	146	172	47	69	34	63
Bleu	25	52	41	64	62	86
Vert	75	117	71	115	37	77

**Table 4-5.** Seuils bas et haut des couleurs utilisées.

Pour voir comment nous localisons un objet par traitement d'images, nous présentons les deux exemples suivants et la les coordonnées des centres d'objets calculées.

#### 4-2-3 Exemples de segmentation

Dans les Figure.4-14 nous avons deux exemples du traitement d'images utilisé dans cette deuxième approche d'asservissement visuel. Ces exemples montrent comment nous pouvons localiser puis sélectionner un objet (canette ou balle) de forme cylindrique ou sphérique et de couleur orange, dans une scène qui comporte d'autres objets de la même forme mais de couleurs différentes.



**Figure 4-14.** Exemple de sélection d'objets, (a) : Image d'origine, (b) : Algorithme LPE, (c) : Choix de l'objet à approcher à partir de sa couleur.

A la fin de ces traitements, l'objet d'intérêt que nous souhaitons saisir se trouve à l'intérieur d'un cadre rouge. Cette méthode d'analyse donne des résultats satisfaisants comme nous allons le voir dans la partie résultats.

Dans la Table.4-6 nous calculons les coordonnées du centre des deux objets avec la méthode de segmentation développée. Nous donnons aussi, les coordonnées réelles des centres et les erreurs sur les valeurs calculées. Ces erreurs ne dépassent pas les 2 pixels comme dans le cas de la première approche utilisant les contours.

Objets simples	Coordonnées réelles du centre (pixels)		Coordonnées calculées du centre (pixels)		Erreurs (pixels)	
	$U_{cr}$	$V_{cr}$	$U_{cc}$	$V_{cv}$	$E_u$	$E_v$
Balle orange	394.55	419.76	393.32	420.99	1.23	0.98
Canette orange	257.95	324.44	259.99	322.99	2.04	1.45

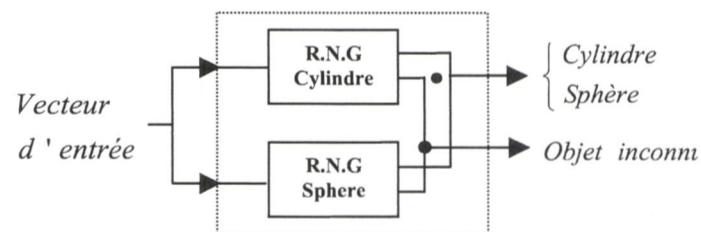
**Table 4-6.** Calcul des coordonnées du centre objet.

### 4-3 Classification des objets

En plus de la distance entre le capteur et l'objet, le système ultrasonore peut également reconnaître la forme de certains objets simples. Dans le chapitre précédent, nous avons utilisé un réseau spécialisé pour déterminer la forme de l'environnement (plan, coin ou arête). Dans cette partie, nous allons proposer un nouveau réseau de neurones qui va permettre de reconnaître les formes des objets simples utilisés dans cette deuxième approche d'asservissement visuel et cela par l'intermédiaire de l'écho ultrasonore.

#### 4-3-1 Réseau de neurones spécialisé

Ce réseau spécialisé est constitué de deux réseaux de neurones globaux du type PCM (Perceptron Multi-Couches), un pour reconnaître les formes cylindriques et l'autre pour les formes sphériques (Figure.4-15). La fonction d'activation comme dans le chapitre précédent, utilise le maximum de la probabilité à posteriori. Pour la construction de la base de données (apprentissage et généralisation), nous utilisons un balayage angulaire d'un pas de  $0,9^\circ$  et un balayage en distance d'un pas de 2 cm.



**Figure 4-15.** Structure du réseau de neurones spécialisé.

La base de données contient environ 4000 échantillons, 2000 pour la balle de ping-pong et les 2000 autres pour la canette de boisson. Le vecteur d'entrée du réseau de neurones spécialisé utilise 35 paramètres qui sont :

- Les 30 composantes de la FFT,
- La durée de l'écho et la différence de durée entre les deux récepteurs,
- L'amplitude de l'écho,
- La distance transducteur / cible pour chaque récepteur,
- L'énergie de l'écho.

#### 4-3-2 Taux de reconnaissance

Les résultats de la reconnaissance des deux objets, obtenus avec le réseau de neurones spécialisé développé, sont donnés dans la Table 4-7.

Forme de l'objet	Neurones en couche cachée	Paramètres en entrées	Taux de bonne classification	Taux de rejet
Cylindre	18	35	78 %	22 %
Sphère	18	35	75 %	25 %

Table 4-7. Taux de reconnaissance des deux formes d'objets.

Avec cette configuration (18 neurones en couche cachée et 35 paramètres pour le vecteur d'entrée), le taux de reconnaissance de l'objet cylindrique est de l'ordre de 78 % et celui de l'objet sphérique atteint les 75 %.

#### 4-4 Algorithme

Le réseau de neurones utilisé dans cette approche nous donne la forme de l'objet que souhaitons saisir en plus de la distance à cet objet. En même temps, le traitement d'images nous permet de localiser et de centrer l'objet (Figure 4-16).

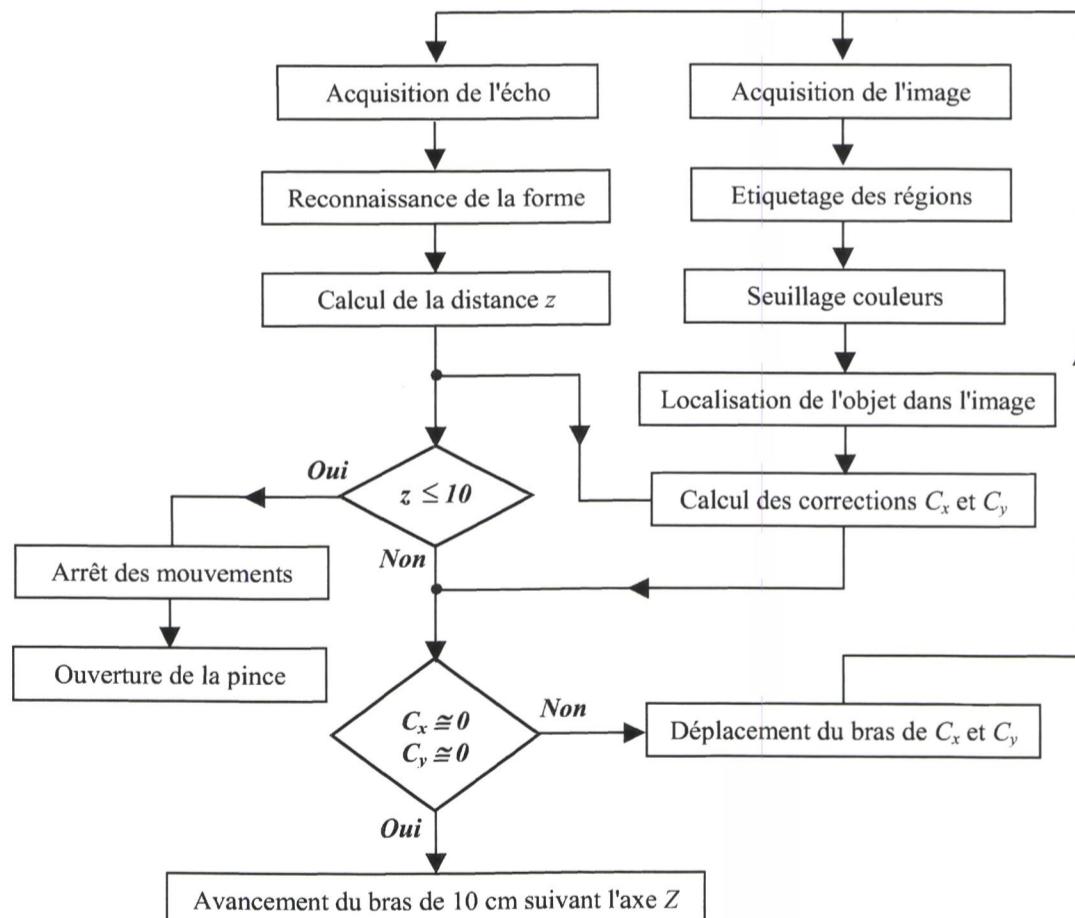


Figure 4-16. Algorithme de la deuxième approche d'asservissement.

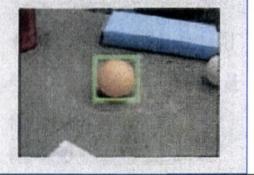
Tant que la distance  $z$  de l'objet reste supérieure à 10 cm, nous calculons chaque fois les deux corrections  $C_x$  et  $C_y$  et nous adaptons la position du bras par rapport à ces deux paramètres comme dans le cas de la première approche.

Cette deuxième approche permet une meilleure connaissance de l'objet que nous voulons saisir à partir de sa forme et sa couleur. Nous pouvons ainsi saisir un objet d'une couleur donnée et définie, se trouvant dans une scène complexe.

#### 4-5 Résultats

Dans la Table.4-8 nous présentons un exemple qui permet la reconnaissance de la forme d'un objet et le calcul de ces paramètres d'asservissement visuel qui sont, la distance à l'objet  $z$ , la correction verticale  $C_x$  et la correction horizontale  $C_y$ .

L'objet utilisé dans cette première manipulation est une balle de ping-pong de couleur orange. Elle se trouve dans un environnement réel contenant d'autres objets de la même forme ou de forme différente comme le cylindre ou le parallélépipède. La couleur aussi, elle est différente d'un objet à l'autre dans la scène.

Distance $z$	Balle orange	$C_x$ (cm)	$C_y$ (cm)
200 cm		2.28	1.29
150 cm		2.07	1.44
100 cm		1.19	0.85
80 cm		0.80	0.61
60 cm		-0.63	-0.51
40 cm		-0.43	-0.47
30 cm		0.25	0.26
20 cm		-0.25	0.18

**Table 4-8.** Calcul des paramètres de l'asservissement pour une balle orange.

La Table 4-9 donne les paramètres de l'asservissement visuel extraits pour une canette de couleur rouge pour des distances allant de 200 cm à 20 cm. La scène utilisée est plus complexe que dans la première approche, elle contient d'autres objets de formes et de couleurs différentes.

Distance $z$	Canette rouge	$C_x$ (cm)	$C_y$ (cm)
200 cm		-2.79	1.45
150 cm		2.32	1.85
100 cm		-1.46	6.21
80 cm		-1.25	0.50
60 cm		-0.86	0.43
40 cm		-0.51	0.22
30 cm		0.25	0.26
20 cm		0.53	0.80

**Table 4-9.** Calcul des paramètres de l'asservissement pour une canette rouge.

Grâce à ce deuxième algorithme, nous arrivons à localiser un objet par rapport à d'autres ayant la même forme ou la même couleur. Nous obtenons les trois paramètres de l'asservissement qui vont être transmis au fur et à mesure au système de commande pour approcher l'objet.

Cette deuxième approche d'asservissement est plus intéressante que la première car elle apporte une variété d'objets respectant les formes de base mais avec des couleurs différentes.

## 5- Troisième approche : Objets réels

La dernière approche regroupe les attributs utilisés dans les deux premières comme le calcul de la distance, la reconnaissance de la forme de l'objet et l'utilisation d'une segmentation en régions et d'un seuillage couleurs. Elle utilise également la largeur de l'objet qui permet de distinguer deux objets de la même couleur mais avec des formes et des dimensions différentes.

Nous allons donner dans la suite, les objets utilisés avec cette approche, la méthode de sélection d'un objet à partir de sa largeur, l'algorithme d'approche et enfin, les résultats obtenus sur de multiples objets.

### 5-1 Objets utilisés

En plus des objets classiques utilisés dans les deux approches précédentes (canette ou balle de couleur unique), nous utilisons dans cette partie des peluches, des canettes et des balles de dimensions et de couleurs différentes. La Table.4-10 présente ces objets et leurs paramètres.

Numéro	Objet	Couleur	Forme	Largeur (mm)
1	Balle	Bleue	Sphère	39.20
2	Balle	Orange	Sphère	39.20
3	Balle	Jaune	Sphère	39.20
4	Balle	Verte	Sphère	63.50
5	Canette	Rouge	Cylindre	64.30
6	Canette	Bleue	Cylindre	64.30
7	Canette	Orange	Cylindre	64.30
8	Canette	Jaune	Cylindre	64.30
9	Canette	Rouge	Cylindre	54.00
10	Canette	Orange	Cylindre	54.00
11	Canette	Verte / Rouge	Cylindre	64.30
12	Canette	Orange / Bleue	Cylindre	64.30
13	Peluche	Orange	Inconnue	50.00
14	Peluche	Noire	Inconnue	100.0

**Table 4-10.** Objets utilisés dans la troisième approche.

Dans cette table nous trouvons les huit objets de l'approche précédente et six autres objets dans le but de généraliser cette approche. Les six nouveaux objets sont :

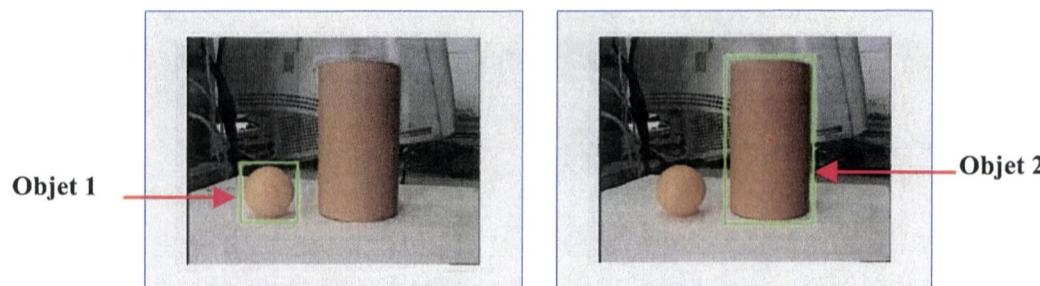
- Deux petites canettes d'une largeur de 54 mm,
- Deux canettes de largeur normale mais avec aux moins deux couleurs,
- Deux peluches de forme inconnue et de couleur et largeur définies.

### 5-2 Choix de l'objet à partir de la largeur

Nous utilisons dans cette partie l'équation (4-6) qui permet de calculer la largeur d'un objet localisé dans l'image. Après l'application de l'algorithme LPE, chaque région de l'image ayant une aire supérieure à 1000 pixels sera étiquetée par un numéro. Ainsi, nous calculons la largeur de chaque zone dans l'image et nous comparons ces largeurs avec les largesses réelles données dans la table précédente pour identifier l'objet.

Le calcul de la largeur d'un objet dans l'image permet de reconnaître deux objets ayant la même couleur mais de formes différentes. Cette largeur permet aussi la reconnaissance de deux objets de la même forme et de la même couleur, mais de dimensions différentes.

La Figure.4-17 montre la sélection de deux objets dans une image, qui ont la même couleur et une forme différente.



**Figure 4-17.** Sélection d'un objet orange à partir de sa largeur.

Dans la Table.4-11 nous caractérisons chaque objet de l'image précédente par différents paramètres comme la forme, la largeur ou la couleur.

En connaissant d'avance la largeur réelle de chaque objet utilisé dans la saisie, nous pouvons déterminer quel est l'objet qui nous intéresse parmi les autres objets de la scène qui ont la même couleur mais une largeur différente.

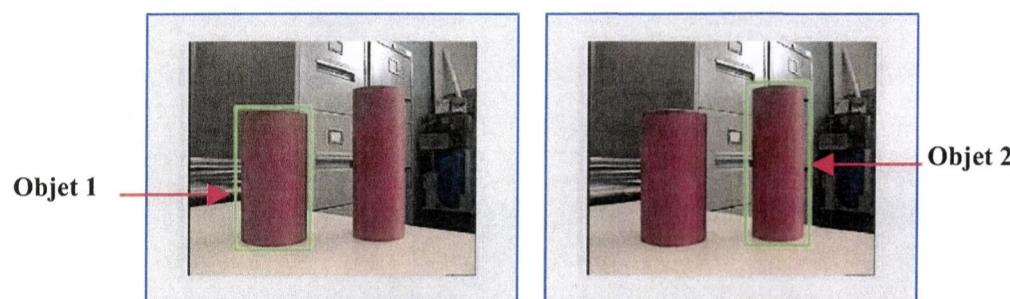
Nous calculons à la fin de cette table l'erreur sur la largeur calculée par rapport à la largeur réelle donnée précédemment.

Numéro d'objet	"1"	"2"
Distance (cm)	20	20
Forme de l'objet	Sphère	Cylindre
Couleur de l'objet	Orange	Orange
Correction $C_x$ (cm)	1.24	-0.64
Correction $C_y$ (cm)	3.61	-3.57
Largeur (cm)	3,82	6,32
Erreur (cm)	<b>0.10</b>	<b>0.11</b>

**Table 4-11.** Caractéristiques des deux objets orange de la scène.

La Figure.4-18 montre aussi comment sélectionner un des deux objets de couleurs et de formes identiques ayant une largeur différente. Le rectangle vert dans la première image encadre la première canette rouge, ce qui correspond à la sélection du premier objet.

Dans la deuxième image, c'est le deuxième objet qui est la petite canette rouge, qui est sélectionné par le rectangle vert.



**Figure 4-18.** Sélection d'un objet rouge à partir de sa largeur.

Ces deux objets sont caractérisés sur la Table.4-12 où nous trouvons la distance de l'objet, sa forme, sa couleur, les corrections, la largeur et enfin, l'erreur sur la largeur calculée par rapport à la largeur réelle.

Numéro d'objet	"1"	"2"
Distance (cm)	30	30
Forme de l'objet	Cylindre	Cylindre
Couleur de l'objet	Rouge	Rouge
Correction $C_x$ (cm)	0.71	-0.56
Correction $C_y$ (cm)	2.88	-3.25
Largeur (cm)	6.50	5.32
Erreur (cm)	<b>0.07</b>	<b>0.08</b>

**Table 4-12.** Caractéristiques des deux objets rouges de la scène.

Ce travail qui concerne le choix de l'objet à partir de sa largeur va être intégré dans l'algorithme d'asservissement visuel, non pas pour sélectionner un objet mais pour vérifier que nous avons choisi le bon objet.

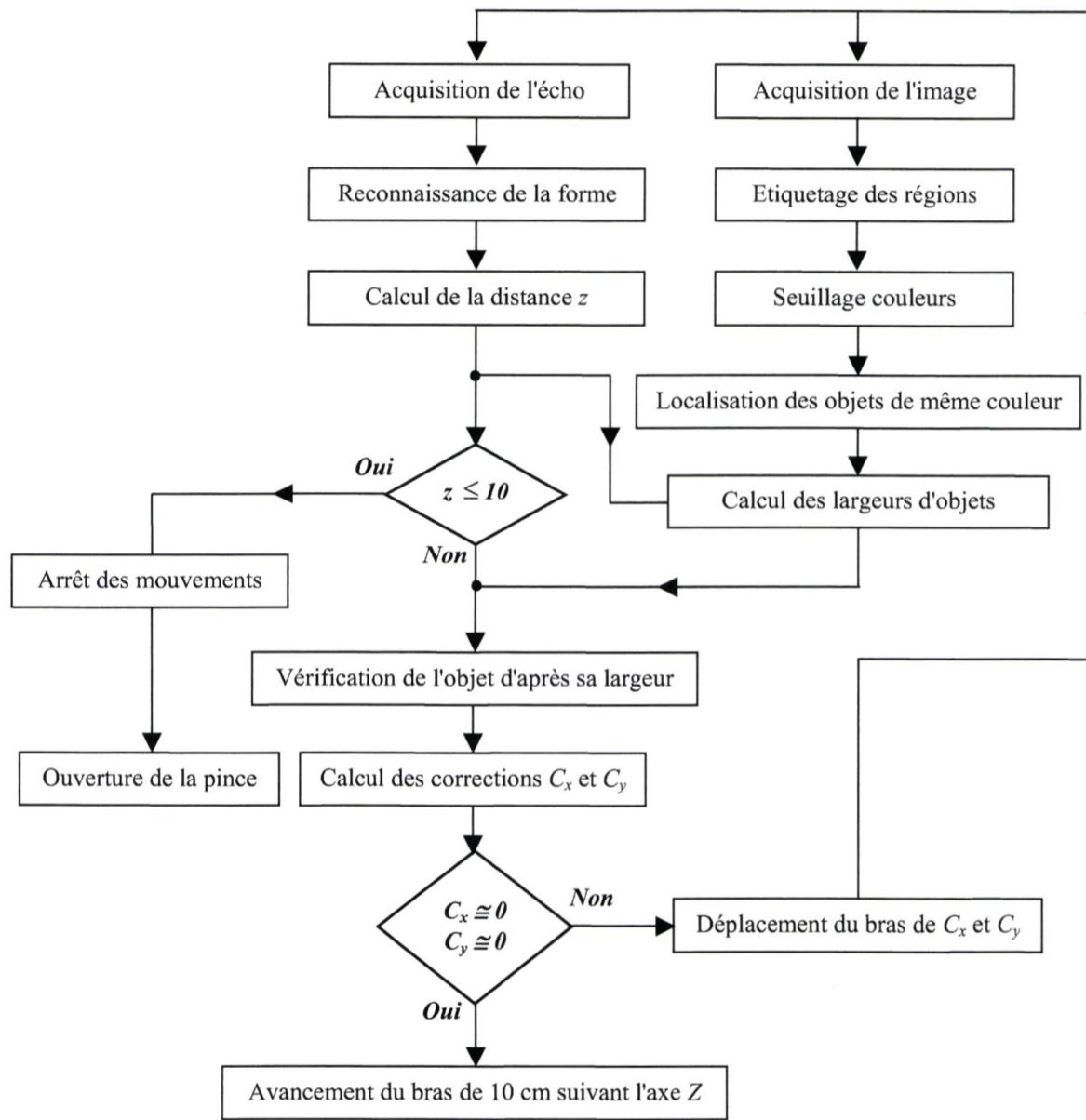
### 5-3 Algorithme

Après la localisation de l'objet sur l'image à partir de sa couleur, nous calculons sa largeur en pixels d'après l'équation (4-6), puis nous transformons cette valeur en valeur métrique grâce au modèle interne de la caméra.

Si la distance de l'objet par rapport à la pince est supérieure à 10 cm, nous comparons cette valeur avec les différentes largeurs d'objets enregistrés dans la Table.4-9 et nous choisissons un de ces objets qui représente l'élément le plus proche d'après sa largeur.

Une fois l'objet que nous désirons saisir correspond à un objet de la table, il sera sélectionné, puis, nous calculons les deux corrections  $C_x$  et  $C_y$  et nous déplaçons la pince suivant ces valeurs pour l'approcher (Figure 4-19).

Si les valeurs de  $C_x$  et de  $C_y$  sont proches de zéro (des valeurs inférieures à 0.2 cm), le bras avance vers l'objet de 10 cm. Sinon, la position du bras sera corrigée dans l'espace suivant les valeurs des deux corrections et la boucle de mesure recommence.



**Figure 4-19.** Algorithme de la troisième approche d'asservissement.

Cette stratégie est conçue pour approcher plusieurs types d'objets dans des environnements complexes dans lesquels d'autres objets de même forme et/ou de même couleur peuvent être présents. Elle représente la situation générale que nous pouvons rencontrer, c'est à dire, des scènes comportant une variété d'objets.

#### 5-4 Résultats

Dans cette partie, nous allons calculer les trois paramètres de l'asservissement visuel à des distances différentes et pour plusieurs objets (les peluches ou les canettes de boisson) en se basant sur la troisième approche. Ces objets ont des formes et des couleurs différentes comme nous l'avons indiqué dans la table des objets. Nous allons mesurer aussi la largeur des objets que nous souhaitons approcher.

Nous notons ici que les peluches ont des formes atypiques, ainsi le réseau de neurones développé n'arrive pas à les reconnaître. Dans cette situation, les ultrasons nous fournissent uniquement la distance de la peluche dans l'environnement de travail. Si la peluche comporte une seule couleur, nous pouvons la localiser facilement dans l'image, par contre si elle comporte plusieurs couleurs, sa localisation devient très difficile car, nous ne connaissons pas également sa forme.

##### a) Peluche orange

Distance $z$	Peluche orange	$C_x$ (cm)	$C_y$ (cm)	Largeur (cm)
140 cm		0.71	-0.43	5.03
120 cm		0.40	-0.69	5.01
100 cm		0.75	0.30	5.0
80 cm		1.53	-1.71	4.90
60 cm		-1.34	-2.43	4.85

Table 4-13. Paramètres de l'asservissement pour la peluche orange.

**b) Peluche noire**

Distance $z$	Peluche noire	$C_x$ (cm)	$C_y$ (cm)	Largeur (cm)
150 cm		1.47	1.05	10.12
130 cm		1.68	1.93	10.10
110 cm		1.64	2.77	10.08
90 cm		0.54	3.12	9.96

**Table 4-14.** Paramètres de l'asservissement pour la peluche noire.**c) Canette orange / bleue**

Nous calculons dans cette partie les paramètres de l'asservissement visuel pour approcher une canette comportant deux de couleurs, orange et bleue, et de forme cylindrique. Nous mesurons aussi la largeur de cette canette pour que nous puissions l'identifier sur la table d'objets.

Distance $z$	Balle orange	$C_x$ (cm)	$C_y$ (cm)	Largeur (cm)
100 cm		0.92	0.65	6.44
70 cm		-0.59	-0.58	6.38
40 cm		0.23	-0.37	6.35

**Table 4-15.** Paramètres de l'asservissement pour la canette orange / bleue.

#### d) Canette verte / rouge

Nous calculons les mêmes paramètres ( $C_x$ ,  $C_y$ ,  $z$ ) que dans l'exemple précédent pour une canette de forme cylindrique avec deux couleurs, verte et rouge.

Distance $z$	Balle orange	$C_x$ (cm)	$C_y$ (cm)	Largeur (cm)
110 cm		-0.89	-0.54	6.48
80 cm		0.04	-1.54	6.41
50 cm		0.35	-0.42	6.35

**Table 4-16.** Paramètres de l'asservissement pour la canette verte / rouge.

Dans toutes ces tables, l'approche développée arrive à calculer avec précision la largeur des objets utilisés. Nous arrivons aussi à localiser les différents objets que nous souhaitons approcher dans l'image comme les peluches, ce qui permet d'avancer vers ces objets avec sûreté.

L'utilisation d'objets avec deux couleurs différentes entre dans le processus de généralisation de la saisie sur des objets réels.

Cette dernière approche d'asservissement visuel représente aussi certaines limites que nous pouvons les regrouper dans les points suivants :

- Utilisation des ultrasons pour reconnaître deux formes d'objets simples qui sont le cylindre et la sphère, car il existe une infinité de formes complexes dans la réalité comme le parallélépipède ou le cube,
- Utilisation des objets avec des couleurs vives, car ils sont facilement reconnaissables par l'analyse des histogrammes couleurs,
- Limitation à des objets comportant deux couleurs au maximum. Une multiplicité des couleurs rend la localisation difficile et lente.

## 6- Conclusion

Dans ce dernier chapitre nous avons décidé d'ajouter une nouvelle tâche dans le cadre de la conception et du développement du robot mobile du projet ARPH. Cette fonctionnalité est l'asservissement visuel de la pince du bras manipulateur, qui va permettre d'approcher et de saisir des objets simples se trouvant dans l'environnement du robot.

Après avoir cité les différentes méthodes d'asservissement visuel existantes, nous avons proposé notre méthode d'approche utilisant une boucle visuelle. Cette méthode se décompose en trois stratégies d'asservissement, basée chacune sur la combinaison de deux types d'informations issues de la vision et des ultrasons.

La première stratégie utilise deux types d'objets simples de couleur unique. Ainsi, le système ultrasonore fournit la distance à l'objet et les contours extraits de l'image permettent de calculer la position de son centre. Ces deux informations combinées permettent de calculer les paramètres de l'asservissement visuel et d'approcher ces objets.

La deuxième approche utilise la forme et la couleur des objets. Ainsi, d'un coté, un classifieur basé sur les réseaux de neurones fournit la forme de l'objet à saisir. D'un autre coté, deux algorithmes de traitement d'images permettent d'extraire les régions de l'image et d'identifier la région d'intérêt correspondante à l'objet. L'algorithme utilisant ces informations a pu donner les paramètres de l'asservissement visuel pour des objets de couleurs différentes se trouvant dans des scènes simples ou complexes.

La troisième stratégie regroupe les informations traitées dans les deux précédentes approches, en plus une information sur la largeur de l'objet. Ainsi, avec cette donnée, nous pouvons reconnaître deux objets du même type avec deux couleurs différentes ou encore, deux objets de type différent ayant la même couleur. Cette dernière approche représente le cas général, car elle regroupe l'ensemble de paramètres utilisés dans les autres approches et elle s'applique à des environnements complexes.

La forme parallélépipède est facilement reconnaissable par les ultrasons et elle représente plusieurs types d'objets de la vie quotidienne comme les livres, les cassettes VHS et les boîtiers des CD/DVD. Ainsi, l'intégration de cette forme dans classifieur neuronal va permettre de reconnaître plus d'objets. C'est une des perspectives de ce travail.

Une deuxième perspective concerne un type d'objets très important, c'est le livre. Ainsi l'idée est de mettre certains nombres de livres intéressant dans des boîtiers à couleurs différentes. La reconnaissance de la couleur permet de localiser l'objet et l'approcher facilement en se basant sur l'analyse des histogrammes couleurs vues auparavant.

Conclusion  
et perspectives

Les travaux présentés dans cette thèse entrent dans le cadre du développement d'un robot mobile d'aide aux personnes handicapées moteurs capable d'apporter de l'assistance dans les différentes tâches de la vie journalière. Ils se composent de deux parties principales correspondant à deux objectifs majeurs, se localiser dans un environnement d'intérieur et approcher les objets à saisir avec précision.

La partie bibliographique présentée dans le premier chapitre de la thèse nous a permis de mettre en lumière les différents travaux de recherche effectués pour résoudre ces deux objectifs. Pour commencer, nous avons cité quelques exemples de projets utilisant un robot mobile équipé de plusieurs types de capteurs. Puis, nous nous sommes intéressés au problème de la localisation d'intérieur avec les différentes approches développées. Enfin, les différentes configurations et stratégies de saisie qui existent.

Le premier objectif consistait à rendre plus robuste et plus rapide une méthode de localisation dans un environnement d'intérieur. Cette méthode est basée sur une mise en correspondance entre des segments 2D extraits de l'image et des segments 3D représentant le modèle filaire de l'environnement. Elle utilisait également deux contraintes géométriques, une transformation basée sur l'approche de Phong-Horaud et une fonction de choix utilisant la pose initiale donnée par le système odométrique.

L'introduction de la contrainte ultrasonore dans l'algorithme d'appariement 2D/3D et dans le modèle 3D de l'environnement, ainsi que l'optimisation des différentes parties de la méthode de localisation existante nous ont permis de développer un algorithme plus fiable, plus rapide et surtout général. Parmi les améliorations apportées, nous pouvons citer les quelques points suivants :

- Les ultrasons et les réseaux de neurones nous ont permis de proposer un nouveau type de modèle de l'environnement. Ce nouveau modèle que nous avons appelé "modèle enrichi" donne en plus des segments de l'environnement une information sur leur contexte topologique.
- L'utilisation de ce modèle dans le processus d'appariement 2D/3D ainsi que la contrainte ultrasonore ont permis de proposer deux nouvelles méthodes d'appariement, la pré-combinaison et la post-combinaison.
- La transformation entre le repère caméra et le repère image dans la méthode initiale représentait le point noir de ce travail car le temps de calcul pouvait atteindre plusieurs minutes. Pour résoudre ce problème de temps de traitement, nous avons remplacé cette première fonction par une transformation basée sur l'approche de Lowe. La solution linéaire de cette approche répond exactement à la contrainte "temps réel" nécessaire à la procédure de localisation.
- L'utilisation d'une fonction de coût basée sur les distances entre segments dans le choix de la meilleure hypothèse de localisation rend cette dernière indépendante de la valeur de l'initialisation odométrique.

Les résultats obtenus dans les différentes étapes de l'algorithme sont très significants. La combinaison des meilleurs éléments permet d'avoir une localisation précise, pour l'application visée, avec une erreur moyenne de 5 cm en distance et de 1° en rotation, dans un temps inférieur à 2 secondes. L'utilisation de la contrainte ultrasonore avec le modèle de

l'environnement enrichi a permis de réduire le nombre d'hypothèses de localisation générées dans la phase d'appariement d'environ deux tiers. Cette réduction dans le nombre d'hypothèses permet aussi de réduire le temps de traitement nécessaire à la localisation. La généralisation de cette approche sur plusieurs environnements de travail confirme la robustesse et la qualité de la méthode de localisation développée.

Pour compléter ce travail, il est maintenant nécessaire d'implanter sur le robot cette technique de localisation et d'évaluer performances et robustesse en situation réelle.

Malgré les nombreuses améliorations apportées, nous devons toutefois mettre en évidence une limite de cette méthode, c'est sa non-efficacité dans un environnement trop peu éclairé. Il est bien possible de prévoir une source lumineuse liée au robot mais cela ne résoudra pas toutes les situations. Ceci ouvre la voie à des perspectives de travail à moyen terme. Il faudrait développer des modes de fonctionnement en cas de défaillance de la vision. Ceci suppose de renforcer le rôle des ultrasons dans le travail de localisation.

Le deuxième objectif de cette thèse était d'apporter une contribution à la saisie d'objets simples, effectuée par l'intermédiaire de la pince du bras manipulateur Manus. Après avoir cité quelques méthodes d'asservissement visuel existantes comme l'asservissement 2D ou 3D, nous avons présenté notre méthode pour approcher des objets simples de formes et de couleurs différentes. Cette méthode s'articule sur une boucle de commande visuelle, qui apporte ces coordonnées de la vision et des ultrasons.

A partir de l'image 2D de l'objet, nous localisons l'objet de couleur définie dans l'image, puis nous calculons les corrections horizontale et verticale à apporter à la position de la caméra de manière à obtenir l'objet au centre de l'image avant de déplacement en  $z$  de la pince du robot.

Le système ultrasonore quant à lui, fournit la distance objet / pince et donne sa forme (cylindrique ou sphérique). Le développement de cette méthode a donné lieu à la proposition de trois stratégies d'asservissement visuel, basée chacune sur la combinaison des deux types d'informations issues de la vision et des ultrasons.

- La première stratégie utilise uniquement deux types d'objets simples et de couleur unique. Le système ultrasonore fournit la distance à l'objet et à partir des contours obtenus par le filtrage de Deriche sur l'image, on connaît la position du centre de l'objet. Ces deux informations combinées permettent de calculer les paramètres de l'asservissement visuel et d'approcher ces objets.
- La deuxième approche utilise la forme et la couleur des objets. Un classifieur basée sur les réseaux de neurones fournit la forme de l'objet situé dans le champ des capteurs. Deux algorithmes de traitement d'images permettent d'extraire les régions de l'image et d'identifier la région d'intérêt qui correspond à l'objet.
- La troisième stratégie utilise les informations des deux précédentes approches ainsi que le diamètre de l'objet. Cette dernière donnée nous permet de reconnaître deux objets du même type ayant la même couleur mais des dimensions différentes. Cette approche représente le cas général, car elle regroupe l'ensemble de paramètres utilisés dans les autres approches et elle s'applique à des environnements complexes.

Les résultats donnés par ces trois approches d'asservissement visuel et notamment les valeurs des corrections  $C_x$  et  $C_y$  et de la distance  $z$  nous permettent d'avoir l'objet au centre de l'ouverture de la pince. Les coordonnées du centre d'un objet sont calculées avec une erreur ne dépassant pas 3 pixels d'image. Le système ultrasonore fournit des distances précises avec des erreurs de l'ordre de quelques centimètres, ce qui est suffisant pour l'application envisagée.

Pour compléter ce travail et le rendre opérationnel il reste à développer une stratégie de saisie des objets. C'est à dire, déterminer les commandes à appliquer à la pince pour réaliser la dernière étape de la boucle visuelle. Il faut également concevoir une interface homme machine (IHM) conviviale permettant à l'handicapé moteur de sélectionner l'objet. Enfin, il faut planter l'ensemble de ce travail sur le robot mobile et valider ses performances en situation réelle.

Il faut noter que ces approches d'asservissement visuel sont confrontées à deux limites majeures. La première est l'utilisation d'un nombre limité d'objets simples. La deuxième est la nécessité que ces objets soient de couleurs vives. Ainsi, parmi les perspectives à plus long terme, nous pouvons citer :

- Généraliser l'utilisation de l'attribut couleur à tous les spectres de couleurs existants. La méthode du seuillage par les couleurs que nous avons proposée utilise l'espace RGB (Rouge, vert et Bleu). L'existence d'autres types d'espaces couleurs comme le HSV (Teinte, Saturation, et Valeur) par exemple, laisse la porte ouverte à d'autres méthodes de seuillage permettant de discriminer un plus grand nombre de couleurs entre elles.
- La capacité des réseaux de neurones spécialisés à reconnaître les formes de l'environnement nous laisse penser à étendre la base de données utilisée dans cette partie aux objets de formes parallélépipédiques comme les livres, les cassettes VHS et les boîtiers des CD / DVD. Cet élargissement permet à la personne handicapée d'atteindre un maximum d'objets.

Pour finir, les deux parties de cette thèse ont été validées par la communauté scientifique grâce à cinq conférences internationales et une européenne. Une revue portant sur la localisation du robot mobile est en soumission et une autre sur la partie asservissement visuel est en préparation.

## Annexe A

### Les contraintes géométriques

### A-1 Première contrainte géométrique : contrainte unaire

La première contrainte géométrique utilisée dans nos travaux de localisation est appelée contrainte unaire.

Soit  $L_i$  un segment du modèle de vecteur directeur  $V_i$  et de vecteur position  $P_i$ . Soit  $l_i$  un segment de la cène dont  $N_i$  est le vecteur normal au plan d'interprétation (Figure A-1).

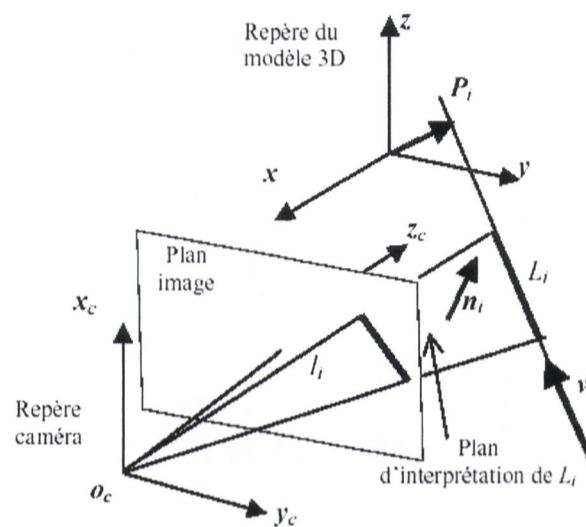


Figure A-1. Projection d'un segment 3D.

En reprenant les deux contraintes de l'équation (2-17), nous pouvons écrire :

$$\begin{cases} a_{ij} \cos \phi + b_{ij} \sin \phi + c_{ij} = 0 \\ A_{ij} x + B_{ij} y + C_{ij} = 0 \end{cases} \quad (A-1)$$

avec les termes  $a_{ij}$ ,  $b_{ij}$  et  $c_{ij}$  donnés par :

$$\begin{cases} a_{ij} = (n_x \cos \theta - n_z \sin \theta)V_x + n_y V_y \\ b_{ij} = n_y V_x - (n_x \cos \theta - n_z \sin \theta)V_y \\ c_{ij} = (n_x \sin \theta + n_z \cos \theta) V_z \end{cases} \quad (A-2)$$

et les termes  $A_{ij}$ ,  $B_{ij}$  et  $C_{ij}$  donnés par :

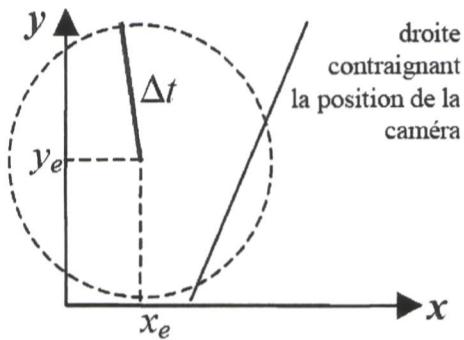
$$\begin{cases} A_{ij} = -(n_x \cos \theta - n_z \sin \theta) \cos \phi - n_y \sin \phi \\ B_{ij} = -n_y \cos \phi + (n_x \cos \theta - n_z \sin \theta) \sin \phi \\ C_{ij} = [(n_x \cos \theta - n_z \sin \theta)x_i + n_y v_i] \cos \phi + [n_x - (n_x \cos \theta - n_z \sin \theta)y_i] \sin \phi + [n_x \sin \theta + n_z \cos \theta](z_i - z) \end{cases} \quad (A-3)$$

Ceci signifie qu'avec un seul appariement entre segments modèle et image, nous pouvons calculer l'orientation  $\phi_{ij}$  de la caméra grâce à la première équation du système (A-1).

Une fois cette orientation calculée, elle peut servir à contraindre la position du robot à appartenir à une droite  $D_{ij}$  puisque la deuxième équation du système devient une équation de droite d'inconnues  $x$  et  $y$ .

Si l'appariement considéré est bon alors l'orientation doit être incluse dans l'intervalle d'orientation défini par une marge  $\Delta\phi$  autour de la valeur de l'orientation estimée  $\phi_e$ .

De plus, la droite définie par la deuxième équation du système (A-1) doit couper la zone du plan horizontal de déplacement définie par les marges  $\Delta x_e$  et  $\Delta y_e$  autour de la position estimée  $(x_e, y_e)$  comme le montre la Figure A-2.



**Figure A-2.** La contrainte unaire.

**Enoncé de la contrainte :**

Un appariement " $L_i - l_j$ " d'orientation  $\phi_{ij}$  et de droite de contrainte  $D_{ij}$  est cohérent si et seulement si :

- $|\phi - \phi_e| < \Delta\phi$
- $D_{ij} \cap C_{ij} \neq \emptyset$

où  $C_{ij}$  représente un cercle de centre  $(x_e, y_e)$  et de rayon  $\Delta t = \sqrt{x_e^2 + y_e^2}$

**A-2 Deuxième contrainte géométrique : contrainte binaire**

Lorsque le segment  $L_i$  du modèle est vertical, les paramètres  $a_{ij}$  et  $b_{ij}$  obtenus dans le cas de la contrainte unaire s'annulent. Le calcul de l'orientation ne peut pas se faire de la même façon dans cette situation particulière et la contrainte unaire n'est pas utilisable pour les segments verticaux du modèle.

Nous allons définir dans la suite une deuxième contrainte géométrique pour résoudre ce cas de figure. Elle est appelée contrainte binaire.

Soit un couple de segments verticaux du modèle  $(L_i, L_j)$  et un couple de segments image  $(l_k, l_l)$  supposé correspondre au couple modèle. Si nous supposons que la caméra est dans une position horizontale, ceci correspond à un angle  $\theta = \pi / 2$ . Dans cette configuration, l'angle est nul et les deux segments  $(l_k, l_l)$  de la scène sont verticaux dans le plan image. Ils peuvent être définis par les équations  $v=v_k$  et  $v=v_l$ .

A partir de la donnée  $v_k$  et connaissant les paramètres du modèle interne de la caméra, nous pouvons calculer les angles de vue de chaque segment,  $w_k$  et  $w_l$ . C'est l'angle entre le rayon optique de la caméra et la droite horizontale passant par la focale et appartenant au plan d'interprétation de l'un des deux segments (Figure A-3).

Pour le vecteur  $v=v_k$ ,  $w_k$  est donné par :

$$\omega_k = \arctan \left( \frac{v_k - v_0}{\alpha_v} \right) \quad (A-4)$$

Pour le vecteur  $v=v_l$ ,  $w_l$  est donné par :

$$\omega_l = \arctan \left( \frac{v_l - v_0}{\alpha_v} \right) \quad (A-5)$$

l'angle de vue entre les deux segments est calculé par :

$$\omega_{kl} = |\omega_k - \omega_l| \quad (A-6)$$

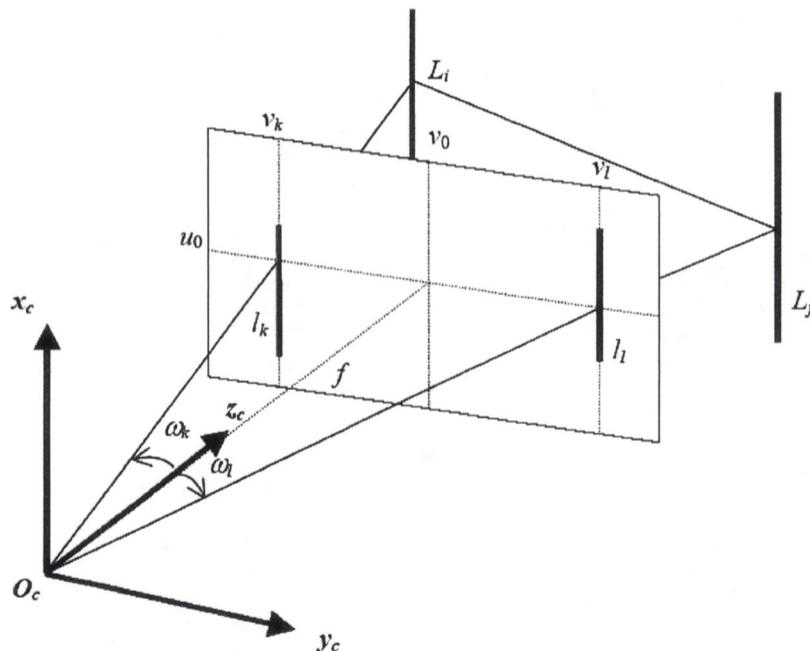


Figure A-3. Les angles de vue.

En appliquant la règle du cosinus sur le triangle formé par le centre focal  $O_c$  et les deux points d'intersection du plan horizontal passant par ce centre et les deux segments  $L_i$  et  $L_j$ , nous obtenons l'équation :

$$d_{ij}^2 = d_i^2 + d_j^2 - 2d_i d_j \cos(w_{kl}) \quad (A-7)$$

où  $d_{ij}$  est la distance horizontale connue entre les deux segments parallèles  $L_i$  et  $L_j$  du modèle et  $d_i$  et  $d_j$  sont respectivement les distances horizontales des segments  $L_i$  et  $L_j$  au centre focal de la caméra, données par :

$$d_i = \sqrt{(x_i - x)^2 + (y_i - y)^2} \quad (A-8)$$

$$d_j = \sqrt{(x_j - x)^2 + (y_j - y)^2} \quad (A-9)$$

Cette équation (A-7) est en fait celle d'un cercle par rapport aux coordonnées x et y du centre focale de la caméra. la position de la caméra et ainsi du robot est donc contrainte à appartenir à un arc de cercle défini par l'équation (A-7).

De la même façon que pour la contrainte unaire, si le couple d'appariement  $(L_i-l_k, L_j-l_l)$  est correct, le cercle correspondant  $C_{ij}'$  coupe la zone définie précédemment (le cercle  $C_{ij}$  de la contrainte unaire). La Figure A-4 montre cette intersection.

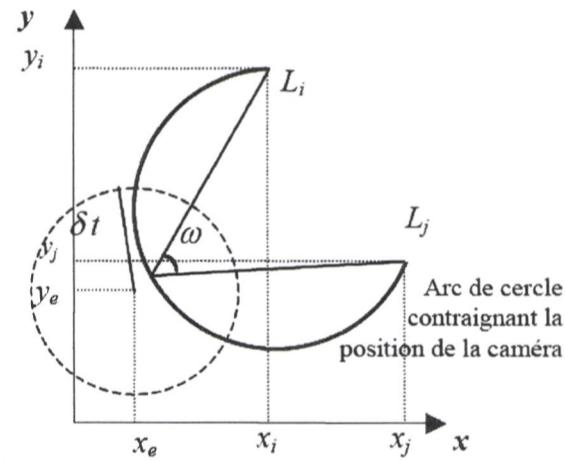


Figure A-4. La contrainte binaire.

**Enoncé de la contrainte :**

Un couple de correspondances  $(L_i-l_k, L_j-l_l)$  de cercle de contrainte  $C_{ij}'$  est cohérent si est seulement si :

- $|\phi - \phi_e| < \Delta\phi$
- $C_{ij} \cap C_{ij}' \neq \emptyset$

## Annexe B

### Les quaternions

### B-1 Les quaternions

Les quaternions peuvent être vus comme des vecteurs de dimension 4 ou encore comme des nombres complexes à trois parties imaginaires :

$$q = q_0 + iq_x + jq_y + kq_z \quad (B-1)$$

avec :

$$i^2 = j^2 = k^2 = ijk = -1 \quad (B-2)$$

### B-2 Propriétés des quaternions

Nous pouvons remarquer qu'on peut obtenir également :

$$ij = -ji = k \quad (B-3)$$

$$jk = -kj = i \quad (B-4)$$

$$ki = -ik = j \quad (B-5)$$

Grâce à ces formules, nous pouvons facilement calculer le produit de deux quaternions, noté par "\*" :

$$r * q = (r_0 + ir_x + jr_y + kr_z)(q_0 + iq_x + jq_y + kq_z) \quad (B-6)$$

Le produit de deux quaternions n'est pas commutatif. Comme le produit vectoriel, le produit de deux quaternions peut s'écrire sous forme matricielle :

$$r * q = Q(r) q \quad (B-7)$$

avec  $Q(r)$  donnée par :

$$Q(r) = \begin{pmatrix} r_0 & -r_x & -r_y & -r_z \\ r_x & r_0 & -r_z & r_y \\ r_y & r_z & r_0 & -r_x \\ r_z & -r_y & r_x & r_0 \end{pmatrix} \quad (B-8)$$

Nous pouvons également écrire le produit sous la forme suivante :

$$q * r = W(r) q \quad (B-9)$$

avec  $W(r)$  défini par :

$$W(r) = \begin{pmatrix} r_0 & -r_x & -r_y & -r_z \\ r_x & r_0 & r_z & -r_y \\ r_y & -r_z & r_0 & r_x \\ r_z & r_y & -r_x & r_0 \end{pmatrix} \quad (\text{B-10})$$

Nous pouvons facilement vérifier les propriétés suivantes quant à ces matrices :

$$Q(r)^t Q(r) = Q(r) Q(r)^t = r^t r I \quad (\text{B-11})$$

$$W(r)^t W(r) = W(r) W(r)^t = r^t r I \quad (\text{B-12})$$

$$Q(r)q = W(q)r \quad (\text{B-13})$$

$$Q(r)^t r = W(r)^t r = r^t r e \quad (\text{B-14})$$

$$Q(r)Q(q) = Q(Q(r)q) \quad (\text{B-15})$$

$$W(r)W(q) = W(W(r)q) \quad (\text{B-16})$$

$$Q(r)W(q)^t = W(q)^t Q(r) \quad (\text{B-17})$$

$e$  étant le quaternion unité :  $e = (1000)^t$ .

Le produit scalaire de deux quaternions est donné par :

$$r \cdot q = r_0 q_0 + r_x q_x + r_y q_y + r_z q_z \quad (\text{B-18})$$

Nous avons bien évidemment :

$$q \cdot q = \|q\|^2 \quad (\text{B-19})$$

Le quaternion conjugué de  $q$  est  $\bar{q}$  :

$$\bar{q} = q_0 - iq_x - jq_y - kq_z \quad (\text{B-20})$$

Notons que si  $Q(r)$  et  $W(r)$  sont les matrices associées à  $q$ ,  $Q(q)^t$  et  $W(q)^t$  sont les matrices associées au quaternion conjugué  $\bar{q}$ , nous avons :

$$q * \bar{q} = q \cdot q = \|q\|^2 \quad (\text{B-21})$$

et nous obtenons ainsi le quaternion inverse de  $q$  :

$$q^{-1} = \frac{1}{\|q\|^2} \bar{q} \quad (\text{B-22})$$

Pour un quaternion unitaire (de norme égale à 1) l'inverse est égal à son conjugué. Nous allons maintenant établir quelques propriétés du produit qui nous seront utiles par la suite. Ces propriétés sont facilement vérifiables si nous utiliserons la notation matricielle :

$$(q * p).(q * r) = (q \cdot q)(p \cdot r) \quad (\text{B-23})$$

Nous déduisons facilement les propriétés suivantes :

$$(q * p).(q * p) = (p \cdot p)(q \cdot q) \quad (\text{B-24})$$

$$\|p * q\|^2 = \|p\|^2 \|q\|^2 \quad (\text{B-25})$$

$$(p * q).r = p.(r * \bar{q}) \quad (\text{B-26})$$

Un vecteur de dimension 3 peut s'écrire comme un quaternion purement imaginaire :

$$q = 0 + iq_x + jq_y + kq_z \quad (\text{B-27})$$

Les matrices associées à un quaternion purement imaginaire sont des matrices antisymétriques. Ainsi :

$$Q(r) = \begin{pmatrix} 0 & -r_x & -r_y & -r_z \\ r_x & 0 & -r_z & r_y \\ r_y & r_z & 0 & -r_x \\ r_z & -r_y & r_x & 0 \end{pmatrix} \quad (\text{B-28})$$

et :

$$W(r) = \begin{pmatrix} 0 & -r_x & -r_y & -r_z \\ r_x & 0 & r_z & -r_y \\ r_y & -r_z & 0 & r_x \\ r_z & r_y & -r_x & 0 \end{pmatrix} \quad (\text{B-29})$$

avec les propriétés évidentes :

$$Q(r)^t = -Q(r) \quad (\text{B-30})$$

$$W(r)^t = -W(r) \quad (\text{B-31})$$

## Annexe C

### Le filtre de Canny / Deriche

### C-1 Le filtre de Canny / Deriche

Ce filtre permet de détecter les contours d'une image bruitée et floue en tenant compte de paramètres liés au rapport signal sur bruit et à la localisation des contours. Ce filtre fonctionne à partir d'un modèle de contour prédéfini donné dans la Figure C-1. Il a pour équation :

$$C(x) = \begin{cases} 1 - \frac{e^{-sx}}{2} & x \geq 0 \\ \frac{e^{sx}}{2} & x \leq 0 \end{cases} \quad (C-1)$$

avec une valeur de  $s$  positif.

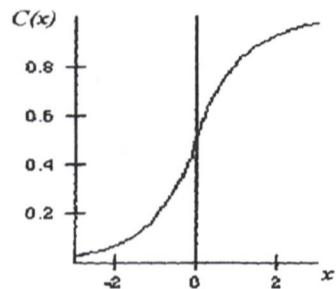


Figure C-1. Contour sous forme exponentielle avec  $s = 0,5$ .

Nous cherchons à maximiser les mêmes trois critères que ceux de Canny :

1- Le rapport signal sur bruit RSB :

$$RSB = \frac{\int_{-\infty}^{+\infty} C(-x) f(x) dx}{\eta_0 \sqrt{\int_{-\infty}^{+\infty} f^2(x) dx}} \quad (C-2)$$

2- La localisation :

$$L = \frac{\int_{-\infty}^{+\infty} C'(-x) f'(x) dx}{\eta_0 \sqrt{\int_{-\infty}^{+\infty} f'^2(x) dx}} \quad (C-3)$$

3- La suppression des réponses multiples :

$$MRC = \sqrt{\frac{\int_{-\infty}^{+\infty} f'^2(x) dx}{\int_{-\infty}^{+\infty} f''^2(x) dx}} \quad (C-4)$$

$f(x)$  étant la fonction du filtre,  $C(x)$  le contour prédéfini et  $\eta_0$  la variance du bruit.

En maximisant le produit RSB x L sous la contrainte MRC, nous trouvons la forme générale de  $f(x)$  :

$$f(x) = ke^{\alpha x} \sin(wx) + e^{\alpha w} \cos(wx) - e^{sx} \quad \text{avec } x \leq 0 \quad (\text{C-5})$$

Nous pouvons séparer cette fonction en deux fonctions causale et anti-causale :

$$f^+(x) = ke^{-\alpha x} \sin(wx) + e^{-\alpha w} \cos(wx) - e^{-sx} \quad \text{avec } x > 0 \quad (\text{C-6})$$

$$f^-(x) = ke^{\alpha x} \sin(wx) + e^{\alpha w} \cos(wx) - e^{sx} \quad \text{avec } x \leq 0 \quad (\text{C-7})$$

Pour obtenir une implantation numérique de ce filtre sous forme récursive il suffit de calculer la transformée en Z de  $f^+(x)$  et  $f^-(x)$  :

$$y^+(i) = a_1 x(i-1) + a_2 x(i-2) + a_3 y^+(i-1) - a_4 y^+(i-2) + a_5 y^+(i-3) \quad (\text{C-8})$$

$$y^-(i) = -a_1 x(i+1) - a_2 x(i+2) + a_3 y^-(i+1) - a_4 y^-(i+2) + a_5 y^-(i+3) \quad (\text{C-9})$$

avec :

$$\begin{cases} a_1 = ke^{-\alpha} \sin(w) + e^{-s} - e^{\alpha} \cos(w) \\ a_2 = -ke^{-(s+\alpha)} \sin(w) - e^{-(s+\alpha)} \cos(w) + e^{-2\alpha} \\ a_3 = 2e^{-\alpha} \cos(w) + e^{-s} \\ a_4 = e^{-2\alpha} + 2e^{-(s+\alpha)} \cos(w) \\ a_5 = e^{-(s+\alpha)} \end{cases} \quad (\text{C-10})$$

Cet opérateur correspond à un filtre récursif d'ordre 3 et ses performances peuvent être facilement exprimées en fonction de  $m$  et  $w$  ( $m = \alpha \times s$  détermine la largeur du filtre  $f(x)$  et  $w$  le nombre de passages par zéro du filtre  $f(x)$  au voisinage de l'origine).

Ce filtre nécessite 4 lectures de l'image : de gauche à droite, de droite à gauche, de haut en bas et de bas en haut. A chaque lecture une fonction du filtre est calculée. C'est la somme de ces 4 fonctions qui donne le contour final.

Il faut encore faire subir à l'image du contour un processus identique mais de lissage, qui se calcule de la même façon que le gradient, pour une détection optimale des contours d'images bruitées.

Ce filtre peut s'implanter sous la forme récursive d'ordre 3 suivante :

$$y(n) = a_1 x(n-1) + a_2 x(n-2) + a_3 y(n-1) - a_4 y(n-2) + a_5 y(n-3) \quad (\text{C-11})$$

Les  $x(n-1)$  et  $x(n-2)$  représentent les 2 pixels qui précèdent le pixel  $x(n)$  pour lequel nous calculons le gradient  $y(n)$ .

## Annexe D

### La calibration de la caméra

### D-1 Introduction

Dans cette partie, nous allons décrire le modèle géométrique associé au processus de saisie d'image par une caméra. Ce modèle est caractérisé par un nombre de paramètres qui vont être estimés par le processus de calibration.

Les paramètres d'une caméra liant le repère caméra et le repère image sont appelés paramètres intrinsèques, ils seront estimés en même temps que les paramètres extrinsèques qui lient un repère associé à la caméra au repère associé à l'objet étalon. Un modèle géométrique de la caméra est caractérisé souvent à l'aide de deux transformations [Hor.95] :

- 1- Une projection perspective qui transforme un point de l'espace 3D en un point de l'image 2D,
- 2- Une transformation permettant de passer d'un repère métrique lié à l'image à un repère lié à la caméra.

Nous allons voir dans les paragraphes suivants les différentes propriétés de ces deux transformations, les paramètres extrinsèques et intrinsèques d'une caméra et le mode de calibration utilisé. Pour formaliser ces deux transformations, trois systèmes de coordonnées sont définis :

- 1- Un repère 3D ( $XYZ$ ) appelé repère monde (repère modèle ou repère objet), dans lequel sont exprimée les points 3D des objets de l'environnement constituant la scène,
- 2- Un repère 3D ( $xyz$ ) lié à la caméra, de telle façon que l'origine de ce repère soit le centre optique et l'axe  $z$  soit confondu avec l'axe optique de la caméra,
- 3- Un repère 2D ( $uv$ ) lié au plan image situé à une distance focale  $f$  de l'origine du repère caméra.

### D-2 Projection perspective

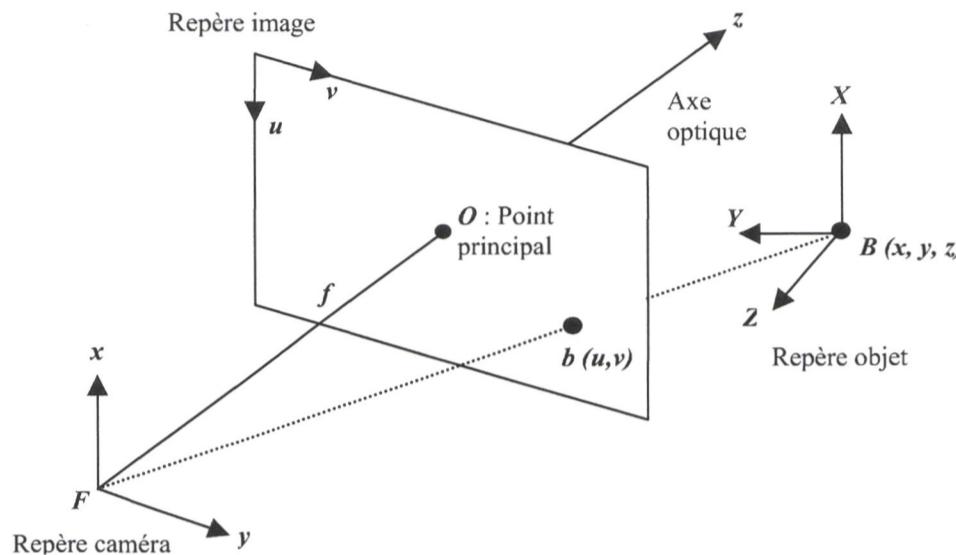
La projection perspective permet de passer des coordonnées caméra d'un point de l'espace 3D vers les coordonnées caméra de sa projection dans le plan image. Soient les notations suivantes :

- $O$  : Point principal,
- $F$  : Centre de projection (origine du repère caméra),
- $f$  : Distance focale,
- $B$  : Un point de l'espace de coordonnées  $(x, y, z)$  dans le repère caméra,
- $b$  : La projection perspective du point  $B$  sur le plan image

Les coordonnées  $(x', y', z')$  du point  $b$  de l'image dans le repère de la caméra sont données par le système (3-2) :

$$\begin{cases} x' = f \cdot x / z \\ y' = f \cdot y / z \\ z' = f \end{cases} \quad (D-1)$$

La Figure.D-1 montre la configuration d'un modèle géométrique d'une caméra composé généralement de trois repères. Elle montre aussi le point  $B$  de l'espace et sa projection sur l'image donnée par le point  $b$ .



**Figure D-1.** Modèle géométrique d'une caméra.

On peut représenter cette première transformation entre le point du modèle  $B$  et sa projection sur l'image  $b$  sous une forme matricielle de la manière suivante :

$$P = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 1/f & 0 \end{pmatrix} \quad (\text{D-2})$$

Les coordonnées homogènes de  $B$  sont donc  $(x, y, z, 1)$  et celles de  $b$  sa projection sont données par :

$$\begin{pmatrix} sx' \\ sy' \\ sz' \\ s \end{pmatrix} = P \cdot \begin{pmatrix} x \\ y \\ z \\ 1 \end{pmatrix} \quad (\text{D-3})$$

La matrice  $P$  est appelée matrice de projection perspective. Le terme  $s$  représente le facteur d'échelle, il vaut souvent l'unité.

### D-3 Transformation caméra / image

Les points image sont mesurés en pixels dans le repère 2D associé à l'image. L'expression de la matrice de transformation du repère caméra au repère image nécessite l'introduction des paramètres suivants :

- 1-  $u_0, v_0$  sont les coordonnées du point  $F$  dans le repère 2D image (mesurées en pixels),
- 2-  $k_u$  est le facteur d'échelle vertical (mesuré en pixels/mm),
- 3-  $k_v$  est le facteur d'échelle horizontal (également mesuré en pixels/mm).

La transformation du repère caméra au repère image du point  $b$  s'écrit donc sous la forme matricielle suivante :

$$\begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} k_u & 0 & 0 \\ 0 & k_v & 0 \end{pmatrix} \begin{pmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{pmatrix} \begin{pmatrix} x' \\ y' \\ z' \end{pmatrix} + \begin{pmatrix} u_0 \\ v_0 \end{pmatrix} \quad (D-4)$$

C'est une transformation affine représentant un changement d'échelle, une rotation et une translation. Cette transformation donnée par une matrice  $K$  représente une application linéaire de l'espace projectif vers le plan projectif. Elle est donnée par :

$$K = \begin{pmatrix} -k_u & 0 & 0 & u_0 \\ 0 & k_v & 0 & v_0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \quad (D-5)$$

On obtient ainsi les coordonnées homogènes de la projection de  $B$  sur l'image :

$$\begin{pmatrix} su \\ sv \\ s \end{pmatrix} = K \cdot \begin{pmatrix} x' \\ y' \\ z' \\ 1 \end{pmatrix} \quad (D-6)$$

La matrice  $K$  est appelée matrice de transformation caméra / image.

#### D-4 Paramètres intrinsèques

Le produit des deux matrices  $K$  et  $P$  (projection perspective suivie d'une transformation affine caméra / image), nous permet d'écrire les équations du modèle géométrique de la caméra, c'est à dire la relation entre les coordonnées caméra  $(x, y, z)$  du point  $B$  et les coordonnées image du point  $b (u, v)$  :

$$\begin{cases} u = -k_u f \frac{x}{z} + u_0 \\ v = k_v f \frac{y}{z} + v_0 \end{cases} \quad (D-7)$$

Le produit matriciel  $K.P$  est donné par la matrice :

$$KP = \begin{pmatrix} -k_u & 0 & u_0/f & 0 \\ 0 & k_v & v_0/f & 0 \\ 0 & 0 & 1/f & 0 \end{pmatrix} \quad (D-8)$$

La multiplication des coefficients par la distance focale  $f$  nous donne :

$$I_c = \begin{pmatrix} \alpha_u & 0 & u_0 & 0 \\ 0 & \alpha_v & v_0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \quad (D-9)$$

avec  $\alpha_u = -k_u f$  et  $\alpha_v = k_v f$ . La matrice  $I_c$  est une application linéaire de l'espace projectif vers le plan projectif exprimant la transformation perspective :

$$\begin{pmatrix} su \\ sv \\ s \end{pmatrix} = I_c \cdot \begin{pmatrix} x \\ y \\ z \\ 1 \end{pmatrix} \quad (D-10)$$

La matrice  $I_c$  est appelée matrice interne ou matrice des paramètres intrinsèques. Elle comporte quatre paramètres principaux :  $\alpha_u$ ,  $\alpha_v$ ,  $u_0$  et  $v_0$ . Ces derniers sont les paramètres intrinsèques de la caméra, estimés par le processus de calibration.

#### D-5 Paramètres extrinsèques

La détermination des paramètres du modèle géométrique de la caméra est effectuée à l'aide d'une mire de calibration (ou autre objet étalon). Celle-ci est constituée d'un ensemble de points dont les coordonnées sont parfaitement connues dans son repère local appelé repère objet (Figure.D-2).

Chaque point de la mire se projette dans l'image et ses nouvelles coordonnées image sont mesurées en pixels dans le repère lié à image. La transformation "mire / image" se décompose en deux transformations : "mire / caméra" et "caméra / image". La transformation "mire / caméra" consiste en une rotation et une translation comme l'exprime l'équation (D-11) :

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{pmatrix} \begin{pmatrix} X \\ Y \\ Z \end{pmatrix} + \begin{pmatrix} t_x \\ t_y \\ t_z \end{pmatrix} \quad (D-11)$$

La transformation rigide composée d'une rotation et une translation, peut s'écrire sous la forme d'une transformation homogène  $A$  suivante :

$$A = \begin{pmatrix} r_{11} & r_{12} & r_{13} & t_x \\ r_{21} & r_{22} & r_{23} & t_y \\ r_{31} & r_{32} & r_{33} & t_z \\ 0 & 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} R & t \\ 0 & 1 \end{pmatrix} \quad (D-12)$$

Cette matrice  $A$  représente une transformation rigide composée de trois rotations et de trois translations, et appelée matrice des paramètres extrinsèques. La droite passant par le centre de

projection  $F$  et le point de l'image  $b$  ( $u, v$ ) s'appelle droite de vue associée à un point de l'image de coordonnées  $u$  et  $v$ .

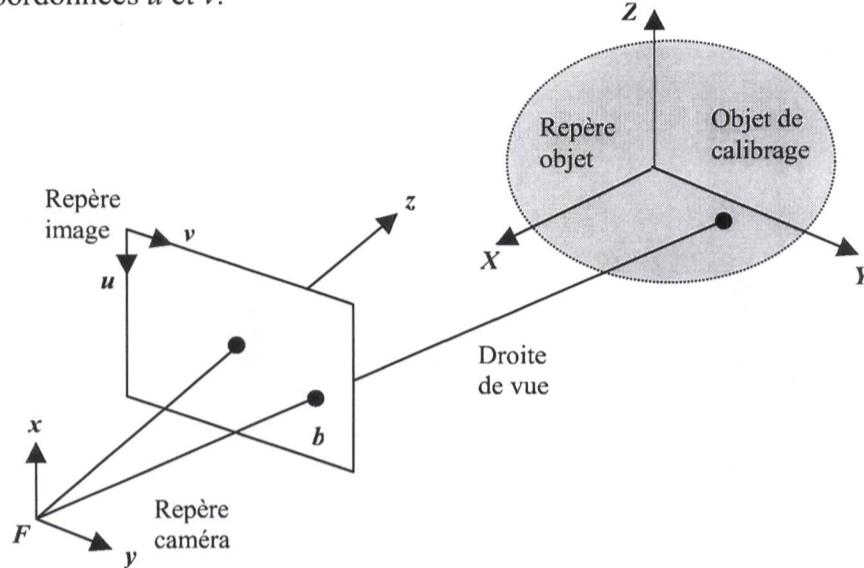


Figure D-2. Principe de calibrage d'une caméra.

#### D-6 Transformation mire / image

La deuxième transformation qui relie les repères mire et image peut être exprimée sous la forme d'une matrice  $3 \times 4$  appelée matrice de projection perspective  $M$ .

Cette dernière est le produit de la matrice des paramètres intrinsèques  $I_c$  et la matrice des paramètres extrinsèques  $A$ . Nous obtenons alors :

$$M = \begin{pmatrix} \alpha_u & 0 & u_0 & 0 \\ 0 & \alpha_v & v_0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} r_1 & t_x \\ r_2 & t_y \\ r_3 & t_z \\ 0 & 1 \end{pmatrix} \quad (D-13)$$

Après les simplifications,  $M$  s'écrit sous la forme suivant :

$$M = \begin{pmatrix} \alpha_u r_1 + u_0 r_3 & \alpha_u t_x + u_0 t_z \\ \alpha_v r_2 + v_0 r_3 & \alpha_v t_y + v_0 t_z \\ r_3 & t_z \end{pmatrix} \quad (D-14)$$

La matrice  $A$  a été exprimée sous une forme plus compacte en utilisant la notation  $r_1 = (r_{11} \ r_{12} \ r_{13})$ . La matrice de projection perspective  $M$  s'écrit en général sous la forme suivante :

$$\begin{pmatrix} su \\ sv \\ s \end{pmatrix} = \begin{pmatrix} m_{11} & m_{12} & m_{13} & m_{14} \\ m_{21} & m_{22} & m_{23} & m_{24} \\ m_{31} & m_{32} & m_{33} & m_{34} \end{pmatrix} \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix} \quad (D-15)$$

$X$ ,  $Y$ , et  $Z$  sont les coordonnées d'un point  $B$  de la mire dans le repère mire. Cette matrice peut également se présenter sous une forme plus simple comme :

$$M = \begin{pmatrix} m_1 & m_{14} \\ m_2 & m_{24} \\ m_3 & m_{34} \end{pmatrix} \quad (D-16)$$

avec  $m_i = (m_{i1} \ m_{i2} \ m_{i3})$  pour  $i = 1, 2$  et  $3$ . En identifiant  $M$  de l'équation (D-15) avec  $M$  obtenu dans l'équation (D-14), nous obtenons un ensemble d'équations qui permettent de calculer les paramètres intrinsèques et extrinsèques de la caméra. Nous avons alors :

$$\begin{cases} u_0 = m_1 \cdot m_2 \\ v_0 = m_2 \cdot m_3 \\ \alpha_u = -\|m_1 \wedge m_3\| \\ \alpha_v = \|m_2 \wedge m_3\| \\ r_1 = \frac{1}{\alpha_u} (m_1 - u_0 m_3) \\ r_2 = \frac{1}{\alpha_v} (m_2 - v_0 m_3) \\ r_3 = m_3 \\ t_x = \frac{1}{\alpha_u} (m_{14} - u_0 m_{34}) \\ t_y = \frac{1}{\alpha_v} (m_{24} - v_0 m_{34}) \\ t_z = m_{34} \end{cases} \quad (D-17)$$

Afin de trouver les paramètres intrinsèques et extrinsèques d'une caméra, nous devons passer par les deux étapes suivantes :

- 1- Estimer les coefficients de la matrice de projection  $M$  par calibration,
- 2- Extraire les paramètres de la caméra à partir de ces coefficients grâce à l'identification donnée dans l'équation (D-17).

#### D-7 Calibration de la caméra

L'équation (3-17) permet d'écrire les coordonnées dans le repère image d'un point du modèle en utilisant la matrice  $M$ , on obtient ainsi le système :

$$\begin{cases} u = \frac{m_{11}X + m_{12}Y + m_{13}Z + m_{14}}{m_{31}X + m_{32}Y + m_{33}Z + m_{34}} \\ v = \frac{m_{21}X + m_{22}Y + m_{23}Z + m_{24}}{m_{31}X + m_{32}Y + m_{33}Z + m_{34}} \end{cases} \quad (D-18)$$

Pour évaluer les 12 coefficients de la matrice  $M$ , on applique le système d'équations (D-18) sur tous les points de la mire, en calculant aussi toutes les projections de ces points dans l'image. Ainsi, chaque point  $(X_i, Y_i, Z_i)$  se projetant en  $(u_i, v_i)$  fournit deux équations linéaires

par rapport aux coefficients de la matrice  $M$ . Les deux équations du système (D-18) peuvent se réécrire comme une combinaison linéaire des éléments  $m_{ij}$ , on obtient :

$$X_i m_{11} + Y_i m_{12} + Z_i m_{13} + m_{14} - u_i X_i m_{31} - u_i Y_i m_{32} - u_i Z_i m_{33} = u_i m_{34} \quad (\text{D-19})$$

et :

$$X_i m_{21} + Y_i m_{22} + Z_i m_{23} + m_{24} - v_i X_i m_{31} - v_i Y_i m_{32} - v_i Z_i m_{33} = v_i m_{34} \quad (\text{D-20})$$

Se qui représente  $2n$  équations pour  $n$  points. La formulation générale est donnée alors par l'équation (D-21) :

$$\begin{pmatrix} X_i & Y_i & Z_i & 1 & 0 & 0 & 0 & \vdots & -u_i X_i & -u_i Y_i & -u_i Z_i \\ 0 & 0 & 0 & 0 & X_i & Y_i & Z_i & 1 & -v_i X_i & -v_i Y_i & -v_i Z_i \\ \vdots & & & & \vdots & & & & & & \end{pmatrix} \begin{pmatrix} m_{11} \\ m_{12} \\ m_{13} \\ m_{14} \\ m_{21} \\ m_{22} \\ m_{23} \\ m_{24} \\ m_{31} \\ m_{32} \\ m_{33} \end{pmatrix} = \begin{pmatrix} \vdots \\ u_i m_{34} \\ v_i m_{34} \\ \vdots \end{pmatrix} \quad (\text{D-21})$$

Pour effectuer la calibration plusieurs méthodes sont proposées comme celle de Zhang utilisés dans le travail initial de localisation. Dans notre travail, nous avons choisi la méthode basée sur les travaux de Faugeras et Toscani exposés dans [Fau.86], qui utilise la contrainte  $\|m_3\| = 1$  qui vient du fait que la somme des rotations  $r_{31}^2 + r_{32}^2 + r_{33}^2 = 1$  où encore  $m_{31}^2 + m_{32}^2 + m_{33}^2 = 1$ . Ainsi, l'équation (D-21) précédente peut s'écrire :

$$Bx_1 + Cx_2 = 0 \quad (\text{D-22})$$

avec l'élément  $x_1$  donné par :

$$x_1 = (m_1 \quad m_{14} \quad m_2 \quad m_{24} \quad m_{34})^t \quad (\text{D-23})$$

et  $x_2$  donné par :

$$x_2 = (m_3)^t \quad (\text{D-24})$$

La matrice  $B$  donnée par :

$$B = \begin{pmatrix} X_i & Y_i & Z_i & 1 & 0 & 0 & 0 & 0 & -u_i \\ 0 & 0 & 0 & 0 & X_i & Y_i & Z_i & 1 & -v_i \\ & & & & \vdots & & & & \end{pmatrix} \quad (D-25)$$

et la matrice  $C$  donnée par :

$$C = \begin{pmatrix} & & \vdots & \\ -u_i X_i & -u_i Y_i & -u_i Z_i & \\ -v_i X_i & -v_i Y_i & -v_i Z_i & \\ & & \vdots & \end{pmatrix} \quad (S-26)$$

Le critère à minimiser est alors  $Q = \|Bx_1 + Cx_2\|^2$  avec la contrainte  $\|x_2\|=1$ . Le critère peut donc s'exprimer comme suit :

$$Q = \|Bx_1 + Cx_2\|^2 + \lambda (1 - \|x_2\|^2) \quad (D-27)$$

avec  $\lambda$  est la valeur propre. Le développement de (D-27) donne :

$$Q = x_1^t B^t B x_1 + x_2^t C^t C x_2 + x_1^t B^t C x_2 + x_2^t C^t B x_1 + \lambda(1 - x_2^t x_2) \quad (D-28)$$

Faugeras et Toscani imposent des dérivées partielles de  $Q$  par rapport à  $x_1$  et  $x_2$  nulles et en substituant dans l'expression de  $Q$ , nous trouvons :

$$Q = x_2^t D x_2 \quad (D-29)$$

avec :

$$D = C^t C - C^t B (B^t B)^{-1} B^t C \quad (D-30)$$

$D$  est une matrice  $3 \times 3$  symétrique et positive. Elle a donc des valeurs propres réelles et positives. Le terme  $x_2$  est un vecteur propre de  $D$  associé à la valeur propre  $\lambda$ . Il est donné par l'équation :

$$D x_2 = \lambda x_2 \quad (D-31)$$

La minimisation du critère nécessite le calcul des valeurs propres de la matrice  $D$  et choisir la plus petite valeur propre qui le minimise. Ensuite, calculer le vecteur propre qui lui est associé  $x_2$ , le normaliser et finalement calculer  $x_1$  par :

$$x_1 = - (B^t B)^{-1} B^t C x_2 \quad (D-32)$$

Enfin, les coefficients de la matrice  $M$  sont fournis par les deux termes  $x_1$  et  $x_2$ .

Références  
bibliographiques

- [Aid.02] Ait Aider Omar, "Localisation référencée modèle d'un robot mobile d'intérieur par vision monoculaire," Thèse de doctorat de l'université d'Evry Val d'Essonne, Décembre 2002.
- [Ald.94] Aldon M.J., L. Le Bris, "Mobile Robot Localization Using a Light Stripe Sensor," Proc. of the Intelligent Vehicles Symposium, 1994, pp. 255-259.
- [All.91] Allen, P.K., Timchenko, A., Yoshimi, B., Michelman, P., "Real-time Visual Servoing," IEEE Int. Conf. on Robotics and Automation, vol. 4 pp. 851, 1991.
- [And.99] Andreff Nicolas, "Asservissement visuel à partir de droites et Auto-étalonage pince caméra," Thèse de doctorat de l'INP-Grenoble, Novembre 1999.
- [Ati.93] Atiya S. and G.D. Hager, "Real-Time Vision-Based Robot Localization," IEEE Trans. Robotics and Automation, vol. 9, n° 6, pp. 785-800, Dec. 1993.
- [Aya.86] Ayache N. and O. Faugeras and O.D. Hyper, "A New Approach for the Recognition and Positioning of Two-Dimensional Objects," IEEE Trans. on Pattern Analysis and Machine Intelligence, 8(1), 1986, pp. 44-54.
- [Bar.95] Barshan B. and H.F. Durrant-White, "Inertial navigation system for a mobile robot," IEEE Trans on Robotics and Automation, vol. 12, n° 5, pp. 328-342, 1995.
- [Bor.85] Borenstein, J. and Koren, Y., "A Mobile Platform For Nursing Robots," IEEE Transactions on Industrial Electronics, May 1985, pp.158-165,
- [Bor.89] Borenstein J. and Y. Koren, "Real-Time Obstacle Avoidance for Fast Mobile Robots," IEEE Trans. Systems, Man, and Cybernetics, vol. 19, n° 5, pp. 1179-1187, 1989.
- [Bor.91] Borenstein J. and Y. Koren, "The Vector Field Histogram-Fast Obstacle Avoidance for Mobile Robots," IEEE Trans. Robotics and Automation, vol. 7, n° 3, pp. 278-288, June 1991.
- [Bor.96] Borenstein J., H.R. Everett, and L. Feng, eds. Wellesley, "Navigating Mobile Robots: Systems and Techniques," Mass.: AK Peters, 1996. Roadways Using Range Data, Proc. SPIE-The Int'l Soc. Optica.
- [Bor.97] Borenstein J., H.R. Everett, L. Feng and D. Wehe, "Mobile Robot Positionning: Sensors and Techniques," Journal of Robotic Systems, 14(4), 1997, pp. 231-249.
- [Can.86] Canny J.F., "A computational approach to edge detection", IEEE Trans. on Pattern Analysis and Machine Intelligence, vol. PAMI-8, n° 6, pp. 679-698, Nov. 1986.
- [Cep.04] Cepolina F., F. Matteucci, R. C. Michelini, "Self Adaptable Clamping Tools for Multiple-Seizure," IEEE International Conf. on Intelligent Manipulation and Grasping, IMG'04, Genova, Italy, 1-2 July 2004, ISBN 88 900 426-1-3.

- [Cha.03] Chatila R., "Modélisation et localisation simultanées en Robotique mobile," Quatrièmes journées nationales de recherche en robotiques, JNRR'03, 8-9 Octobre 2003, Clermont-ferrand, France.
- [Cha.85] Chatila R. and J. P. Laumond, "Position Referencing and Consistent World Modeling for Mobile Robots," Proc. IEEE Int'l Conf. Robotics and Automation, pp. 138-145, Mar. 1985.
- [Chau.91] Chaumette F., P. Rives, B. Espiau, "Positioning of a robot with respect to an object, tracking it and estimating its velocity by visual servoing," In IEEE Int. Conf. On Robotics and Automation, ICRA'91, vol. 3, Pages 2248-2253, Sacramento, Californie, Avril 1991.
- [Chau.90] Chaumette F., "La relation vision-commande: théorie et application à des tâches robotiques," Thèse de Doctorat, Université de Rennes I, 1990.
- [Che.95] Chenavier F., I. Lecoeur, and J. L. Crowley, "Estimation de la position d'un robot par odométrie et vision binoculaire," Traitement du Signal, 12(3) : 255-268, 1995.
- [Cou.83] Coulon P.Y. and M. Nougaret, "Use of a TV camera system in closed-loop position control of mechanisms," International Trends in Manufacturing Technology, A. Pugh Editor, IFS Publications, pages 117-127, 1983.
- [Cre.98] Cretual A. and F. Chaumette, "Image based visual servoing by integration of dynamic measurements," IEEE International Conference on Robotic and Automation, 1998.
- [Dal.92] Dallaway, J.L. and Jackson, R.D, "RAID - A Vocational Robotic Workstation," In Proceedings of the third international Conference on Rehabilitation Robotics (ICORR92), Keele University, September 15-16, 1992, UK.
- [Der.87] Deriche R., "Optimal edge detection using recursive filtering", 1st Int. Conf. on Computer Vision, pp. 501-505, London, 8-11 Juin 1987.
- [Der.93] Deriche R., "Using Canny's criteria to derive an optimal edge detector recursively implemented," The International Journal of Computer Vision, pp. 221-251, 1993.
- [Des.02] DeSouza G. N. and A. C. Kak, "Vision for Mobile Robot Navigation: A Survey," IEEE Trans. On Pattern Analysis and Machine Intelligence, vol. 24, n° 2, February 2002.
- [Dom.88] Dombre E., W. Khalil, "Modélisation et commande des robots," Editions Hermès, 1988.
- [Dor.95] Dornaika Fadi, "Contributions à l'Intégration Vision / Robotique : Calibrage, Localisation et Asservissement," Thèse préparée au sein du laboratoire Lifia/ Imag/Inria Rhône-Alpes, Septembre 1995.

- [Esp.96] Espiau B., F. Chaumette, and P. Rives, "A new approach to visual servoing in robotics," IEEE Transactions on Robotics and Automation vol. 12, n° 5 (1996), pp. 671-684.
- [Eve.01] Evers H.G., E. Beugels et G. Peters, "MANUS towards a new decade," Proc. ICORR 2001, vol. 9, pp. 155-161, 2001.
- [Fab.00] Fabrizi E. and A. Saffiotti, "Extracting Topology-Based Maps from Gridmaps," Proc. of the IEEE Int. Conf. on Robotics and Automation (ICRA), pp. 2972-2978. San Francisco, CA, 2000.
- [Fau.86] Faugeras O.D., G. Toscani, "The calibration problem for stereo," Proc. Second Proc. IEEE Conf. Computer Vision and Pattern Recognition, pp. 321-334, 1986.
- [Fed.89] Feddema. J, Mitchell. O, "Vision-guided Servoing with feature-based trajectory generation," IEEE Transaction on Robot and Automation, vol. 5, pp: 691-700, 1989.
- [Gan.84] Ganapahjy S., "Decomposition of transformation matrices for robot vision," in Proc. Int. Conf. Robotics and Automation, pp. 130-139, December 1984.
- [Gha.04] Gharbi R., N. Rezzoug, P. Gorce, P. Hoppenot & E. Colle, "Assistance à la saisie en milieu non structuré," Handicap 2004, Paris, 17-18 juin 2004.
- [Gir.79] Giralt G., R. Sobek, and R. Chatila, "A Multi-Level Planning and Navigation System for a Mobile Robot; A First Approach to Hilare," Proc. Sixth Int'l Joint Conf. Artificial Intelligence, vol. 1, pp. 335-337, 1979.
- [Gon.02] Gonzalez R. C. and R. E. Woods, "Digital Image Processing," Second Edition, Upper Saddle River: Prentice-Hall, 2002.
- [Gri.84] Grimson W.E.L. and Lozano-Perez T., "Model-based recognition and localization from sparse range or tactile data", International Journal of Robotics Research, 5(3), pp.3-34, 1984.
- [Gri.90] Grimson W.E.L., "Object Recognition: The Role of geometric constraints," MIT Press, 1990.
- [Gur.05] Guerfi S., J-P Gambotto, S. Lelandais, "Implementation of the watershed method in the HSI color space for the face extraction," IEEE Conference on Advanced Video and Signal Based Surveillance, AVSS'05, Como, Italy, pp. 282-286, 2005.
- [Ham.98] Hamadane Hafid, "Modélisation de dispositifs à ultrasons pour la perception de l'environnement," Thèse de Doctorat de l'Université d'Evry Val d'Essonne, Septembre, 1998.
- [Han.97] Hansen K., J.D. Andersen, "Understanding the Hough Transform: Hough Cell Support and its Utilization," Image and Vision Computing, vol. 15, pp. 205-218, Elsevier Science Publishers B.V, 1997.

- [Has.93] Hashimoto K., T. Bine, and H. Kimura, "Visual servoing with Head-Eye manipulator optimal control approach," *IEEE Transactions on Robotics and Automation* 12 (1993), n° 5, pp. 766-774.
- [Has.97] Hashima M., F. Hasegawa, S. Kanda, T. Maruyama, and T. Uchiyama, "Localization and Obstacle Detection for a Robot for Carrying Food Trays," *Proc. IEEE Int'l Conf. Intelligent Robots and Systems*, pp. 345-351, Sept. 1997.
- [Hil.79] Hill J. and W.T. Park, "Real time control of a robot with a mobile camera," *9th Int. Symp. On Industrial Robots* (1979), pp. 233-246.
- [Hop.01] Hoppenot P., E. Colle, O. Ait-Aider and Y. Rybaczuk, "ARPH- Assistant Robot for handicapped People - A pluridisciplinary Project," *IEEE Roman*, 2001, Bordeaux and Paris, 18-21 September 2001, pp. 624-629.
- [Horn.00] Horn O., A. Courcelle, M. Kreutner, "Le projet VAHM (Véhicule Autonome pour Handicapés Moteur) : la localisation," *Revue Traitement du signal*, Année 2000, vol. 17, num. 3, spécial 2000, pp. 221-232.
- [Hor.95] Horaud Radu, Olivier Monga, "Vision par ordinateur, Outils fondamentaux," Edition Hermès, Paris, 1995
- [Hor.98] Horaud R., F. Dornaika, and B. Espiau, "Visually guided object grasping," *IEEE Transactions on Robotics and Automation*, vol. 14, n° 4, pp. 525-532, August 1998.
- [Kim.94] Kim D. and R. Nevatia, "Representation and Computation of the Spatial Environment for Indoor Navigation," *Proc. Int'l Conf. Computer Vision and Pattern Recognition*, pp. 475-482, 1994.
- [Kos.92] Kosaka A. and A.C. Kak, "Fast Vision-Guided Mobile Robot Navigation Using Model-Based Reasoning and Prediction of Uncertainties," *Computer Vision, Graphics, and Image Processing-Image Understanding*, vol. 56, no. 3, pp. 271-329, 1992.
- [Kri.89] Kriegman D.J., E. Triendl, and T.O. Binford, "Stereo Vision and Navigation in Buildings for Mobile Robots," *IEEE Trans. Robotics and Automation*, vol. 5, no. 6, pp. 792-803, 1989.
- [Kro.89] Krotkov E., "Mobile Robot Localization Using a Single Image," *Proc. IEEE Int'l Conf. Robotics and Automation*, pp. 978-983, 1989.
- [Kum.89] Kumar R. and A.R. Hanson, "Robust estimation of camera location and orientation from noisy data having outliers," In *Proc. Workshop on Interpretation of 3-D Scenes*, pp. 52-60, Austin, Texas, USA, November 1989.
- [Lam.88] Lamdan Y. and Wolfson H.J., "Geometric hashing: A general and efficient model-based recognition scheme," *Second IEEE Int. Conf. Computer Vision*, Tampa, FL, pp. 238-24, December 1988.

- [Liu.90] Liu Y., T.S. Huang, O.D. Faugeras, "Determination of camera location from 2D to 3D line and point correspondences," IEEE Trans. On Pattern Analysis and Machine Intelligence, vol. 12, n°1, pp. 28-37, January 1990.
- [Low.87] Lowe D.G., "Three dimensional object recognition from single two-dimensional images," Artificial Intelligence, Elsevier Science Publishers B.V.(North-Holland), pp. 355-395, 1987.
- [Mall.96] Mallem M., M. Shaheen, X. Dourille and F. Chavand, "Method of matching object 3D model and its image using geometric constraints," IEEE / IMACS Int. Conf. CESA'96, Robotics and Cybernetics, pp. 565-569, Lille, 10-13 Juillet 1996.
- [Mal.98] Malis E., F. Chaumette, and S. Boudet, "Positionning a coarse calibrated camera with respect to an unknown object by 2d 1/2 visual servoing," IEEE International Conference on robotic and automation (1998).
- [Mar.01] Marchand E., F. Chaumette, F. Spindler, M. Perrier, "Controlling the manipulator of an underwater ROV using a coarse calibrated pan tilt camera," IEEE int. Conf. On Robotics and Automation, ICRA'01, vol. 3, South Korea, pp. 2773-2778, 2001.
- [Mat.03] Matsikis A., F. Schulte, F. Broicher & K. F. Kraiss, "A Behaviour Coordination Manager For A Mobile Manipulator," IEEE/RSJ, International Conference On Intelligent Robots And Systems, Intelligent Robots And Systems For Human Security, Health And Prosperity, IROS'03, Vol. 1, pp. 174-181, Las Vegas, USA, October 2003.
- [Mat.87] Matthies L. and S.A. Shafer, "Error Modeling in Stereo Vision," IEEE J. Robotics and Automation, vol. 3, no. 3, pp. 239-248, 1987.
- [Mei.99] Meier R., T.W. Fong, C. Thorpe, and C. Baur, "Sensor Fusion Based User Interface for Vehicle Teleoperation," International Conference on Field and Service Robotics (FSR'99), August, 1999.
- [Men.93] Meng M. and A.C. Kak, "NEURO-NAV: A Neural Network Based Architecture for Vision-Guided Mobile Robot Navigation Using Non-Metrical Models of the Environment," Proc. IEEE Int'l Conf. Robotics and Automation, vol. 2, pp. 750-757, 1993.
- [Mon.05] Montagne Christophe, "Apports de la couleur en vision : Application à la localisation globale et locale d'une base mobile d'aide aux handicapés," thèse de Doctorat de l'Université d'Evry Val d'Essonne, 13 Décembre 2005.
- [Mor.81] Moravec H.P., "Obstacle Avoidance and Navigation in the Real World by a Seeing Robot Rover," PhD thesis, Stanford Univ., Sept. 1980. Published as Robot Rover Visual Navigation. Ann Arbor, MI: UMI Research Press, 1981.
- [Mor.85] Moravec H.P. and A. Elfes, "High Resolution Maps from Wide Angle Sonar," Proc. IEEE Int'l Conf. Robotics and Automation, pp. 116-121, 1985.

- [Nil.84] Nilsson N.J., "Shakey the Robot," Technical Report 323, SRI Int'l, Apr. 1984.
- [Ori.95] Oriolo G., G. Ulivi, and M. Vendittelli, "On-Line Map Building and Navigation for Autonomous Mobile Robots," Proc. IEEE Int'l Conf. Robotics and Automation, pp. 2900-2906, May 1995.
- [Ouf.02] Aitoufroukh Naima, "Perception intelligente à partir de capteurs frustes pour la robotique de service," Thèse de Doctorat de l'Université d'Evry Val d'Essonne, Décembre, 2002.
- [Pam.88] Pampagnin L.H. and Devy M., "Identification d'un objet 3D par mise en correspondance entre une seule image et un modèle," Rapport interne, LAAS-Toulouse, France, 1988.
- [Pan.98] Pan J., D.J. Pack, A. Kosaka, and A.C. Kak, "FUZZY-NAV: A Vision-Based Robot Navigation Architecture Using Fuzzy Inference for Uncertainty-Reasoning," Proc. IEEE World Congress Neural Networks, vol. 2, pp. 602-607, July 1995. Feb. 1998.
- [Pho.95] Phong T. Q., R. Horaud, P. D. Tao, "Object pose from 2-D to 3-D point and line correspondences," Int. J. of Computer Vision, vol. 15, pp. 225-243, 1995.
- [San.80] Sanderson A.C. and L.E. Weiss, "Image based visual servo control using relational graph error signals," International Conference on Cybernetics and Society (1980), pp. 1074-1077.
- [Sha.99] Shaheen Mudar, "Reconnaissance d'objets Polyédriques à partir d'une image vidéo pour la téléopération," Thèse de Doctorat de l'Université d'Evry Val d'Essonne, Mars, 1999.
- [Shi.73] Shirai Y. and H. Inoue, "Guiding a robot by visual feedback in assembling tasks," Pattern Recognition 5 (1973), n°. 3, pp. 99-108.
- [Shi.73] Shirai Y. and H. Inoue, "Guiding a robot by visual feedback in assembling tasks," Pattern Recognition, vol. 5, pp. 99-108, 1973.
- [Smi.85] Smith, R. C. and P. Cheeseman (1985): "On the Estimation and Presentation of Spatial Uncertainty," Technical Report TR 4760 & 7239, SRI.
- [Sug.88] Sugihara K., "Some Location Problems for Robot Navigation Using a Single Camera," Computer Vision, Graphics, and Image Processing, vol. 42, pp. 112-129, 1988.
- [Thr.02] Thrun, S. (2002): "Robotic Mapping: A Survey," Report published by School of Computer Science, Carnegie Mellon University, USA.

- [Top.95] Topping M. J. "The Development of Handy 1 a Robotic Aid to Independence for the Severely Disabled," Proceedings of the IEE Colloquium "Mechatronic Aids for the Disabled", University of Dundee, May 17, 1995, pp.2/1-2/6, Digest No: 1995/107.
- [Tra.01] **Trabelsi Mohamed**, Rapport de stage, DEA Système Electronique et Traitement de l'Information, "Localisation et reconnaissance de plaques d'immatriculations," Laboratoire IEF, Université Paris-Sud, Orsay, Sep. 2002.
- [Tra.a.05] **Trabelsi Mohamed**, Naima Aitoufroukh et Sylvie Lelandais, "Saisie d'objets à l'aide d'une caméra et des ultrasons pour la robotique de service," SETIT'2005, 27-31 mars 2005, Sousse, Tunisie.
- [Tra.b.05] **Trabelsi Mohamed**, Sylvie Lelandais, Naima Aitoufroukh, "Development of an object grabbing strategy for mobile robotic by image processing and sonar distance computing," IMACS'2005, July 11-15, 2005, Paris, France.
- [Tra.c.05] **Trabelsi Mohamed**, Naima Aitoufroukh and Sylvie Lelandais, "New Object Seizure Method in Mobile Robotic Using a Visual Servoing and Neural Network Classification," IEEE / CIRA'2005, June 27- 30, 2005, Helsinki, Finland.
- [Tra.d.05] **Trabelsi Mohamed**, Naima Aitoufroukh and Sylvie Lelandais, "Fusion of Data Vision and Ultrasounds for Mobile Robot Localisation," AAATE'2005, September 6-9, 2005, Lille, France.
- [Tra.a.06] **Trabelsi Mohamed**, N. Aitoufroukh and S. Lelandais, " Localization and Object Grabbing in Robotic Using Color Image and Neural Network Classification," IEEE / IWSSIP'2006, Budapest, Hongrie, September 2006.
- [Tra.b.06] **Trabelsi Mohamed**, Naima Aitoufroukh and Sylvie Lelandais, " Improvements of Object Grabbing Method by Using Color Images and Neural Networks Classification," IEEE / IECON'2006, Paris, France, November 2006.
- [Vag.93] Vaganay J., "Conception d'un système de multisensoriel de localisation dynamique 3D pour robot mobile," Thèse de doctorat, Université Montpellier II, Sciences et Techniques de Languedoc, 1993.
- [Wei.84] Weiss Lee E., "Dynamic Visual Servo Control of Robots: An Adaptive Image-Based Approach," PhD thesis, Robotics Institute, School of Computer Science, Carnegie Mellon University, Pittsburgh, 1984.
- [Yua.89] Yuan J.S.C., "A general photogrammetric method for determining object position and orientation," IEEE Transactions on Robotics and Automation, vol. 5, n°2, pp. 129-142, April 1989.
- [Zha.98] Zhang Z., "A flexible new technique for camera calibration," IEEE Trans. On Pattern Analysis and Machine Intelligence, vol. 22, n°11, pp. 1330-1334, 1998.

## Résumé

Mon travail de recherche entre dans le cadre du projet « Assistance Robotisée aux Personnes Handicapées (ARPH). L'objectif de ce projet est d'apporter une assistance aux personnes handicapées moteur dans les différentes tâches de la vie quotidienne, à l'aide d'un robot mobile et un bras manipulateur.

La première partie de ma thèse est consacrée au développement d'un système de localisation d'intérieur pour le robot mobile. Nous utilisons un appariement 2D/3D entre le modèle filaire de l'environnement enrichi par l'information ultrasonore et les attributs de l'image. La transformation qui permet de passer des coordonnées des segments 3D du modèle aux coordonnées de la caméra se base sur le principe de Lowe linéaire. Le choix de la meilleure pose est obtenu par une fonction de coût basée sur les distances entre les groupes de segments dans l'espace.

La deuxième partie de ma thèse a été consacrée au développement d'une stratégie de saisie pour des objets de forme simple (cylindre ou sphère). Une caméra et un capteur ultrasonore équipent la pince du bras manipulateur Manus. La combinaison des informations reçues de ces deux capteurs permet de reconnaître l'objet à saisir en utilisant un réseau neuronal. Cette objet est centré dans le champ visuel de la caméra par traitement d'images et à l'aide de toutes de ces informations, la pince approche l'objet jusqu'à la saisie. Globalement, cette méthode est basée sur plusieurs éléments qui participent à l'élaboration de cette stratégie de type asservissement visuel.

**Mots clés :** robotique mobile, localisation, appariement 2D/3D, réseaux de neurones, ultrasons, combinaison d'informations, asservissement visuel.

## Abstract

My research belongs at the ARPH project. The aim of this project is to bring an assistance to disabled people in the various tasks of life, using a mobile robot and an arm manipulator.

The first part of my thesis is devoted to the development of a localization system for the mobile robot. We use a 2D/3D matching between the 3D environment model enriched with ultrasonic information and 2D image segments. Function which transforms the 3D coordinates of the model segments to the camera coordinates is based on the Lowe linear principle. The choice of the best position is obtained by measuring distances between the two groups of segments in space.

The second part of my thesis was devoted to the development of a seizure strategy for simple objects (cylinder or sphere). A camera and a sonar followed by a neural networks are installed at the robot gripper. The combination of information received from these two sensors allows the grabbing of the object. This object is centered in the visual field of the camera by image processing. The gripper approaches the object until the seizure. This method is based on several elements which take part in the development of a visual servoing strategy.

**Keywords:** mobile robot, localization, 2D/3D matching, neural networks, ultrasound, combination of information, visual servoing.