



UNIVERSITÉ DE
SHERBROOKE

Faculté de génie
Génie électrique et génie informatique

Interfaces graphiques tridimensionnelles de téléopération de plateformes robotiques mobiles

Mémoire de maîtrise ès sciences appliquées
Spécialité: génie électrique

François FERLAND

Jury: François Michaud (dir.)
Chon Tam Le Dinh (co-dir.)
Michel Lauria
Dominic Létourneau

Sherbrooke (Québec) Canada

Janvier 2009

IV-1951



Library and Archives
Canada

Published Heritage
Branch

395 Wellington Street
Ottawa ON K1A 0N4
Canada

Bibliothèque et
Archives Canada

Direction du
Patrimoine de l'édition

395, rue Wellington
Ottawa ON K1A 0N4
Canada

Your file Votre référence
ISBN: 978-0-494-53160-0
Our file Notre référence
ISBN: 978-0-494-53160-0

NOTICE:

The author has granted a non-exclusive license allowing Library and Archives Canada to reproduce, publish, archive, preserve, conserve, communicate to the public by telecommunication or on the Internet, loan, distribute and sell theses worldwide, for commercial or non-commercial purposes, in microform, paper, electronic and/or any other formats.

The author retains copyright ownership and moral rights in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

In compliance with the Canadian Privacy Act some supporting forms may have been removed from this thesis.

While these forms may be included in the document page count, their removal does not represent any loss of content from the thesis.

AVIS:

L'auteur a accordé une licence non exclusive permettant à la Bibliothèque et Archives Canada de reproduire, publier, archiver, sauvegarder, conserver, transmettre au public par télécommunication ou par l'Internet, prêter, distribuer et vendre des thèses partout dans le monde, à des fins commerciales ou autres, sur support microforme, papier, électronique et/ou autres formats.

L'auteur conserve la propriété du droit d'auteur et des droits moraux qui protègent cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

Conformément à la loi canadienne sur la protection de la vie privée, quelques formulaires secondaires ont été enlevés de cette thèse.

Bien que ces formulaires aient inclus dans la pagination, il n'y aura aucun contenu manquant.


Canada

RÉSUMÉ

Les besoins grandissant en santé rendent des technologies comme la téléprésence à domicile de plus en plus intéressantes. Cependant, dans le domaine des interfaces humains-machines, il est souvent noté que négliger la façon dont est présentée l'information provenant du robot peut nuire à l'opérateur dans sa compréhension de la situation, ce qui entraîne une efficacité réduite. C'est en considérant la façon dont est traitée l'information chez l'opérateur que nous arriverons à développer une interface permettant d'allouer le maximum des capacités cognitives de l'opérateur à la tâche. De plus, les développements récents de matériel à haute performance et à coûts réduits nous permettent de mettre en œuvre des techniques modernes de traitement d'images en temps réel.

Nous proposons donc de développer un système flexible pour étudier les différentes façons de présenter l'information pertinente à la navigation efficace d'une plateforme robotique mobile. Ce système est basé sur une reconstruction en trois dimensions de l'environnement parcouru à partir des lectures de capteurs retrouvés couramment sur ces plateformes. De plus, l'utilisation d'une caméra vidéo stéréoscopique permet de reproduire l'effet de perspective tel qu'une personne sur place le percevrait. La présence d'un flux vidéo est souvent appréciée par les opérateurs et nous croyons que d'ajouter la profondeur dans notre reproduction de celui-ci est un avantage. Finalement, la caméra virtuelle de l'interface peut être continuellement réorientée de façon à fournir une perspective soit égocentrique, soit exocentrique, selon les préférences de l'opérateur.

Nous validons l'utilisation de ce système en évaluant selon différentes métriques les performances d'opérateurs, autant néophytes qu'experts en robotique mobile, de façon à bien cibler les besoins fonctionnels de ce genre d'interfaces et leurs évaluations avec des populations-cibles. Nous croyons que la flexibilité quant au positionnement de la caméra virtuelle de l'interface demeure l'aspect le plus important du système. En effet, nous nous attendons à ce que cela permette à chaque opérateur d'adapter l'interface à ses préférences et les tâches en cours pour qu'il effectue son travail le plus efficacement possible. Bien que nous n'incluons pas de tâches spécifiques au domaine de la télésanté dans nos expérimentations, nous croyons que les observations de ce travail quant à la téléopération en général pourront s'appliquer éventuellement à ce domaine en particulier.

Mots clés Interfaces graphiques, Interaction humain-robot, Téléprésence, Compréhension de la situation, Facteurs humains, Vidéo stéréoscopique, Images de synthèse 3D, Réalité virtuelle et augmentée

REMERCIEMENTS

Tout d'abord, j'aimerais remercier mes deux directeurs de maîtrise, Chon-Tam Le Dinh et François Michaud, pour leur soutien constant.

Je désire remercier le Fonds de soutien en traitement d'images de l'Université de Sherbrooke et le Conseil de recherche en sciences naturelles et génie du Canada pour leurs supports financiers.

J'aimerais également remercier l'équipe complète de Laborius et particulièrement François Pomerleau pour leur aide précieuse dans le développement du système présenté dans ce mémoire.

Finalement, je dois souligner l'indispensable collaboration des 13 testeurs volontaires sans qui les résultats de l'expérimentation n'auraient pu exister.

TABLE DES MATIÈRES

1	Introduction	1
2	Interfaces de téléopération et visualisation 3D	3
2.1	Compréhension de la situation	3
2.1.1	Particularités en recherche et sauvetage urbain	5
2.2	Pertinence de l'utilisation de la vidéo	6
2.3	Fusion d'information	7
2.3.1	Fusion avec la vidéo	7
2.3.2	Représentation cartographique de l'environnement	8
2.4	Représentation et positionnement de la vue en 3D	8
2.5	Reconstruction d'environnements 3D	10
2.6	Pertinence de l'utilisation d'éléments translucides	11
2.7	Transmission de vidéo stéréoscopique	12
3	Conception d'une interface-opérateur 3D	13
3.1	Méthodologie retenue	17
4	Article	19
4.1	Introduction	20
4.2	Experimental Platform	21
4.3	Egocentric and Exocentric 3D Interface	22
4.3.1	Laser-based Projection	25
4.4	Experimental Results	29
4.4.1	Gaze	31
4.4.2	Viewpoint Positions	31
4.4.3	Control and Task Accuracy	32
4.4.4	LASSO Classification	34
4.4.5	Usefulness	34
4.5	Conclusion and Future Work	37
4.6	Acknowledgements	37
5	Conclusion	38
A	Calibration de la caméra	40
B	Supplément aux résultats des tests	41
B.1	Profils des utilisateurs	41
B.2	Point de vue ajustable	43
	BIBLIOGRAPHIE	44

LISTE DES FIGURES

2.1	Systèmes étudiés par Drury <i>et al.</i>	5
2.2	Interface développée par Ricks <i>et al.</i>	9
2.3	Effet de la semi-transparence d'un curseur 3D.	11
3.1	Maquette de l'interface à réaliser.	16
4.1	<i>Telerobot</i> , a teleoperated robot for assistive homecare.	21
4.2	Screenshots from our ego-/exocentric 3D interface.	24
4.3	Viewpoint configuration.	26
4.4	Screenshots of exocentric viewpoint presets, noted as (θ, ϕ, ρ)	27
4.5	Experimental course.	30
4.6	Operator's gaze on the screen distribution.	31
4.7	Virtual camera position distributions.	33
4.8	Stereoscopic projection screenshots.	35
B.1	Positions de la caméra virtuelle pour toutes les participations.	43

LISTE DES TABLEAUX

- 4.1 LASSO classification. 34
- 4.2 Usefulness average, from 1 to 5, according to operators after their trial run. 36

- B.1 Profil des opérateurs. 42

LEXIQUE

Graphe de scène Graphe permettant de représenter hiérarchiquement les différents éléments d'une scène à présenter et l'application des transformations communes à des groupes d'éléments. Par exemple, une rotation appliquée à un noeud parent se retrouve appliquée aux éléments enfants de ce noeud.

Vue égocentrique Vue telle que le robot la perçoit. Aussi appelée «à la première personne» et parfois *endocentric* dans la littérature.

Vue exocentrique Vue telle que la percevrait une tierce partie en présence du robot. Aussi appelée «à la troisième personne».

CHAPITRE 1

Introduction

La robotique mobile téléopérée est une technologie de plus en plus présente. Que ce soit pour désamorcer des bombes [Graves et Czarnecki, 1999], assister des personnes en perte d'autonomie [Michaud *et al.*, 2007] ou même explorer d'autres planètes [Stone, 1996], l'usage de robots transforme plusieurs corps de métiers et leur permet d'effectuer des tâches auparavant impossibles. Le robot devient alors une extension physique de l'humain dans un environnement distant, et s'avère efficace dans la mesure où l'opérateur arrive à bien s'y retrouver.

Or, les interfaces couramment utilisées pour représenter l'environnement distant à partir du robot et pour diriger ce dernier sont souvent abstraites et demandent une formation avancée à quiconque voulant s'en servir. L'information provenant des capteurs y est souvent présentée sous forme brute et disjointe, et une attention particulière à plusieurs éléments disposés arbitrairement sur un tableau de bord est nécessaire. De plus, lorsqu'elles tentent de présenter une vue égocentrique du robot, elles le font souvent par une fenêtre vidéo fixe de petite taille et de largeur de champ restreinte. Ces phénomènes s'expliquent par le fait que la disposition de l'interface d'opération est souvent une considération de dernière minute de la part des développeurs [Adams, 2002].

Le domaine connexe des jeux vidéos voit quant à lui un développement fulgurant des méthodes de manipulation d'environnement virtuels en 3D (trois dimensions). Aujourd'hui, le commun des mortels peut, après quelques minutes d'acclimatation, piloter une voiture et ce, sans utiliser autre chose qu'une télécommande rudimentaire et en ayant comme seule source d'information l'image d'une seule caméra virtuelle. Cette caméra propose à l'utilisateur une caractéristique cruciale : la possibilité d'être réorientée instantanément. Ainsi, l'utilisateur peut se placer dans l'univers virtuel à l'endroit qu'il juge le plus apte pour piloter son véhicule, sans entrer en collision avec son environnement.

Nous nous intéressons donc à mettre en place un système flexible permettant de présenter l'information requise selon différents modes (allant de la simple vue vidéo à une représentation 3D de l'environnement) pour la téléopération d'une plate-forme robotique. Une des difficultés à surmonter pour passer d'une représentation d'une vue vidéo normale à un environnement 3D est de fusionner en temps réel les informations provenant de plusieurs

capteurs (principalement la vue vidéo avec des capteurs de proximité) afin de créer cette représentation 3D de l'environnement dynamique dans lequel le robot évolue. À cette fin, nous avons développé une interface permettant de combiner les données d'une caméra stéréoscopique couleur avec un laser de proximité, permettant de créer une représentation 3D de l'environnement selon deux méthodes : 1) par la projection 3D d'une des deux vues vidéo sur les plans identifiées par le capteur laser ; 2) par la projection stéréoscopique des vues vidéo sur laquelle se superpose les données du capteur laser. Dans cette représentation 3D, l'opérateur peut sélectionner sa perspective à sa guise, soit pour adopter une vue égocentrique, soit pour dégager une vue exocentrique à partir d'un point de vue externe au robot. Puisqu'un des objectifs visés par ce système est qu'il soit utilisé dans des expérimentations de téléopération avec des opérateurs ciblés (par exemple par des cliniciens dans un contexte de téléassistance à domicile), nous avons validé son utilisation et l'évaluation de différentes métriques avec un groupe d'opérateurs, démontrant ainsi ses fonctionnalités ainsi que les améliorations possibles à apporter.

Le présent travail est organisé de la façon suivante. Le chapitre 2 présente une revue des interfaces de téléopération robotique. Le chapitre 3 donne les grandes lignes du travail réalisé et des hypothèses de travail formulées. Le chapitre 4 présente l'article de conférence rédigé sur les travaux. Cet article est accepté pour présentation à la *4th ACM/IEEE International Conference on Human-Robot Interaction*. Il est l'un des 23 articles acceptés parmi les 120 articles soumis (19% d'acceptation) à cette conférence, qui a eue lieu du 11 au 13 mars 2009 à San Diego en Californie. Enfin, le chapitre 5 vient conclure le travail.

CHAPITRE 2

Interfaces de téléopération pour plateformes mobiles robotiques et concepts entourant la visualisation en 3D

Cette section présente les approches couramment utilisées en téléopération pour représenter un robot dans son environnement, dans le but de maintenir une bonne compréhension de la situation¹ chez l'opérateur. De plus, quelques techniques pertinentes et déjà utilisées en imagerie et interfaces-utilisateurs graphiques en 3D sont étudiées.

2.1 Compréhension de la situation

Yanco et Drury [Yanco et Drury, 2004] ont adapté aux interactions humains-robots une définition populaire par Endsley [Endsley, 1988] de la compréhension de la situation dans la littérature anglaise. :

[La compréhension de la situation est] la perception de la position, de l'entourage et de l'état du robot, la compréhension de leur signification et la projection dans un avenir proche du comportement du robot.

De plus, Drury *et al.* [Drury *et al.*, 2003] avaient déjà présenté, un an auparavant, une réponse plus détaillée définissant cinq dimensions de la compréhension de la situation :

- *Humain-Robot* : La compréhension que les humains ont du positionnement, de l'identité, des activités, de l'état et de l'entourage des robots. Également, la compréhension de la certitude avec laquelle les humains connaissent cette information.
- *Humain-Humain* : La compréhension que les humains ont du positionnement, de l'identité et des activités de leurs collaborateurs humains.
- *Robot-Humain* : La connaissance que les robots ont du commandement des humains nécessaire pour suivre des activités, et n'importe quelles contraintes

¹*Situation Awareness*

définies clairement par l'humain qui pourraient nécessiter d'ignorer un commandement ou de modifier le cours normal des opérations.

- *Robot-Robot* : La connaissance qu'ont les robots des commandes qu'ils ont reçues par d'autres robots et la coordination robot à robot nécessaire pour redistribuer dynamiquement des tâches entre robots.
- *La conscience générale des humains de la mission* : La compréhension qu'ont les humains des objectifs généraux des activités communes aux robots et aux humains, et la mesure étape par étape du progrès par rapport à ces objectifs.

L'objectif d'une bonne interface de téléopération d'un robot mobile est donc nécessairement de maintenir un bon niveau de compréhension de la situation. Afin de bien évaluer ceci, Drury, Keyes et Yanco [Drury *et al.*, 2007] ont décrit un système intéressant pour des interfaces qui sont centrées soit sur une image vidéo, soit sur une carte de l'environnement parcouru. Nommé *LASSO* (*Location, Activities, Surroundings, Status, Overall mission*), ce système étudie cinq catégories de compréhension de la situation, l'objectif étant de maximiser cette dernière.

Pour évaluer la compréhension de la situation de l'opérateur, l'expérience consiste à l'encourager à penser tout haut pendant ses interactions avec le robot et à analyser ce qu'il dit sur le champ plutôt que de recueillir ses commentaires à la fin. Il est noté toutefois qu'il est souvent difficile, à partir de ce que dit l'utilisateur, de déterminer à quelle catégorie du système *LASSO* un commentaire appartient. C'est particulièrement vrai pour le positionnement et l'entourage du robot, qui sont dans certains cas difficiles à distinguer. Cependant, les chercheurs ont décidé de noter chaque commentaire comme positif, négatif ou neutre pour chacune des catégories. Le besoin de classifier les commentaires est ainsi pratiquement éliminé, ce qui aide grandement l'analyse.

Deux interfaces ont été étudiées plus spécifiquement. L'interface A présente à l'opérateur une carte en 3D à laquelle une représentation virtuelle du robot et l'image vidéo sur un plan flottant devant ce dernier sont ajoutées. L'interface B présente plutôt deux vues vidéos, une vers l'avant et l'autre vers l'arrière. La vue arrière est plus petite que la première et est placée un peu à la façon d'un rétroviseur d'automobile, mais leurs contenus peuvent être interchangeables par l'opérateur. Un plan en 2D est placé à gauche de la vue vidéo principale. Cette section de l'interface peut également afficher les lectures d'un laser de proximité, mais ceci a pour effet de désactiver complètement la présentation du plan. Sous ces trois vues s'affichent les contrôles de direction du robot sur un plan en perspective, un peu

comme si ces contrôles étaient affichés au sol par rapport à la vue vidéo. Des captures d'écran de chacune de ces interfaces sont disponibles à la figure 2.1.

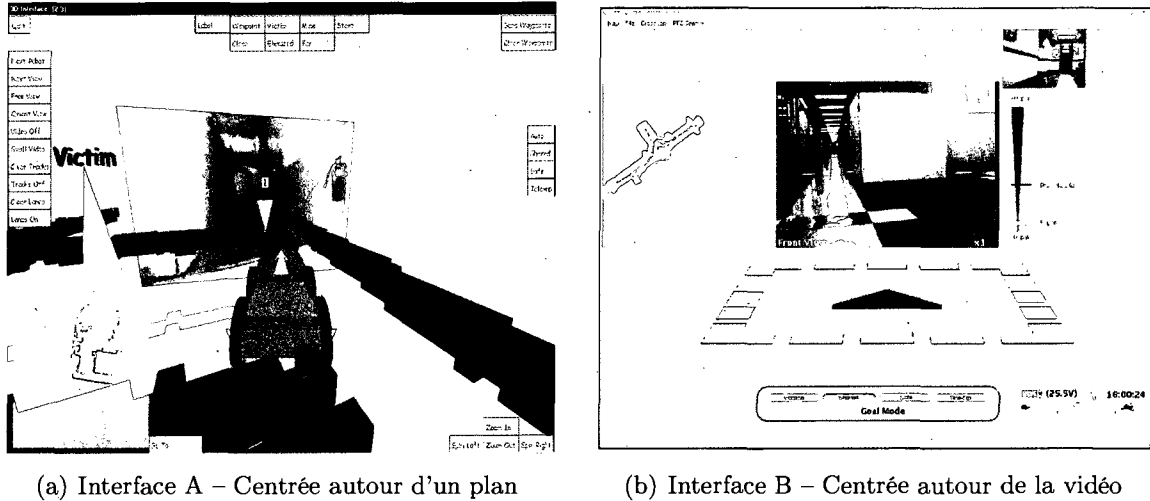


Figure 2.1 Systèmes étudiés par Drury *et al.* [Drury *et al.*, 2007]

Le résultat de cette étude est que l'interface centrée autour de la vidéo (interface B) était particulièrement efficace pour maintenir une bonne compréhension de l'entourage immédiat et des tâches du robot (le premier S et le A de *LASSO*), tandis que l'interface centrée autour d'un plan était meilleure pour ce qui est du positionnement global et de l'état du robot (le L et le deuxième S de *LASSO*). Une interface idéale devrait cependant maintenir une bonne compréhension des cinq catégories. Ils concluent que les travaux futurs porteront probablement sur une union de ces deux types d'interfaces et que le système *LASSO* devrait être comparé à d'autres systèmes d'évaluations subjectifs de la compréhension de la situation.

2.1.1 Particularités en recherche et sauvetage urbain

Drury *et al.* [Drury *et al.*, 2004] ont adapté aux applications de recherche et sauvetage urbain le guide en six questions de Scholtz [Scholtz, 2002] sur la compréhension de la situation. Deux des questions sont particulièrement pertinentes :

Est-ce qu'il y a suffisamment d'information sur l'état et le positionnement du robot pour que l'opérateur sache si le robot fonctionne correctement et qu'il évite les obstacles ?

Est-ce que les moyens fournis par l'interface pour interagir avec le robot sont efficaces à la fois pour l'humain et le robot (e.g., est-ce que des raccourcis sont présentés à l'humain) ?

Ces deux questions soulèvent l'importance qui doit être apportée à la vitesse d'exécution de la tâche en cours (e.g., retrouver le plus rapidement possible un blessé) et à la connaissance de l'état du robot (e.g., s'assurer que le robot est bel et bien en mesure de retrouver un blessé).

2.2 Pertinence de l'utilisation de la vidéo

Il est important de se questionner sur la pertinence d'utiliser un flux vidéo dans des tâches de navigation. En effet, il a été observé [Baker *et al.*, 2004; Drury *et al.*, 2007; Yanco *et al.*, 2006] qu'en général, un opérateur porte toute son attention sur la vidéo si elle est disponible, jusqu'à exclure tout le reste qui lui est présenté. Dans tous les cas, la navigation par la vidéo était préférée à une navigation qui pourrait être qualifiée de strictement «aux instruments.»

La vidéo est donc souhaitée, mais est-elle vraiment souhaitable ? Selon ce qu'elle affiche et la façon dont elle est présentée, la vidéo peut parfois nuire à la navigation. En effet, les travaux de Nielsen et Goodrich [Nielsen et Goodrich, 2006] mettant en œuvre différentes combinaisons d'une carte et d'un flux vidéo le montrent bien. Si la vidéo ne possède pas suffisamment d'indices permettant de se situer dans l'environnement, les temps de complétion d'un parcours et le nombre de collisions augmentent par rapport à une interface qui ne présente qu'un plan de l'environnement. Puisque l'opérateur concentre la majeure partie de son attention sur la vidéo, la carte se retrouve pratiquement ignorée. Cependant, lorsque la vidéo est superposée à la carte, la qualité de la navigation était quasiment la même que pour l'interface à carte seulement.

Un autre contre-argument à l'utilisation de la vidéo peut être fait à partir des travaux de Labonté *et al.* [Labonté *et al.*, 2006]. En comparant deux interfaces de téléopération, une basée sur la navigation par points sur une carte et l'autre basée sur la navigation par points directement visibles par un flux vidéo, il a été observé que des opérateurs novices obtiennent de meilleurs résultats avec la première qu'avec la deuxième. C'est l'inverse qui a été observé chez des opérateurs habitués, mais avec moins d'écart entre les résultats des deux interfaces. La pertinence de l'utilisation de la vidéo dépendrait donc aussi de l'opérateur et non seulement de l'environnement ou de sa présentation.

2.3 Fusion d'information

Il est critique d'atteindre un bon niveau de fusion d'information. En adaptant aux tâches de recherche et sauvetage urbain les méthodes d'évaluation des performances humain-système de Scholtz [Scholtz, 2002], Yanco *et al.* [Yanco *et al.*, 2004] posent une question particulièrement importante :

Est-ce que l'information provenant des robots est présentée de façon à minimiser l'usage de la mémoire de l'opérateur, incluant la quantité de fusion d'information devant être effectuée par le cerveau de l'opérateur ?

En effectuant une fusion d'information sensée, i.e., qui combine des informations liées à une même tâche ou un même contexte, le travail de compréhension effectué par l'opérateur, aussi qualifiée de charge cognitive, peut être minimisé. De cette façon, il peut mieux se concentrer à sa tâche de navigation. Nielsen *et al.* [Nielsen *et al.*, 2007] qualifient ce type d'interface d'intégrée, où une seule fenêtre sert à combiner toutes les sources d'information nécessaires.

2.3.1 Fusion avec la vidéo

Toujours dans l'analyse de Yanco *et al.* [Yanco *et al.*, 2004], il a été observé une fois de plus que de demander à l'opérateur de porter attention à différents endroits de l'interface nuisait à sa tâche. C'est à partir de cette observation que Hestand et Yanco [Hestand et Yanco, 2004] ont développé une méthode de superposition de l'information à une image vidéo dans le but d'assister des opérateurs de robots en situation de recherche et sauvetage urbain. En utilisant une technique simple de composition vidéo par incrustation-couleur, l'interface ajoute à un flux vidéo une prise de vue d'une caméra infrarouge. Il devient ainsi beaucoup plus facile d'associer des signatures thermiques à des éléments visuels de l'environnement. Bien qu'il s'agisse d'un travail strictement validé en laboratoire, Hestand et Yanco font des observations intéressantes pour rendre le système réellement utilisable :

- La composition doit absolument être effectuée en temps réel.
- Un flux vidéo devrait idéalement disposer de 30 images/seconde pour être utile, quoique 15 images/seconde sont suffisantes dans certains cas. Cependant, il serait probablement difficile d'effectuer le traitement complet par logiciel à ce rythme et l'usage de matériel spécialisé serait justifié.

- L'opérateur doit pouvoir ajouter ou retirer de couches d'informations de la composition. De cette façon, il peut ajuster la fusion d'information à la tâche qu'il effectue.

Il est à noter que la crainte de ne pas avoir suffisamment de puissance pour traiter la vidéo en temps réel s'estompe rapidement si le traitement est effectué sur un poste de travail possédant un processeur graphique moderne. *GPUCV* [Farrugia *et al.*, 2006], une adaptation de la bibliothèque *OpenCV*² à ce type de processeurs, rend beaucoup plus facile l'implémentation optimisée de compositions vidéos simples.

2.3.2 Représentation cartographique de l'environnement

Comme noté avec les travaux de Nielsen et Goodrich [Nielsen et Goodrich, 2006], ajouter une représentation cartographique de l'environnement à l'interface de l'opérateur peut être un atout si cette représentation est bien présentée. Dans d'autres travaux de Yanco et Drury [Yanco et Drury, 2004] sur la compréhension de la situation, il est noté que les opérateurs vont même jusqu'à qualifier d'inutile le plan de l'environnement lorsqu'il est placé trop loin de la vidéo ou qu'il est difficile de retrouver rapidement où se trouve le robot sur la carte. Un mauvais choix de couleur pour représenter le robot était la principale cause de cette difficulté. Il faut donc attacher une importance particulière à un choix qui peut sembler futile. De plus, la carte n'était pas réorientée selon la position du robot.

2.4 Représentation et positionnement de la vue en 3D

Les travaux récents exploitent souvent une représentation en 3D de l'environnement parcouru [Nielsen et Goodrich, 2006; Yanco *et al.*, 2006]. Dans les travaux de Ricks *et al.* [Ricks *et al.*, 2004], cette représentation provient en temps réel de senseurs de proximité. Les obstacles détectés sont représentés par des barils verts ou bleus selon le type de senseur (laser ou sonar) autour d'un baril rouge montrant l'emplacement du robot téléopéré. Ensuite, la vue d'une caméra vidéo posée sur le robot y est superposée. Une capture d'écran de cette interface est disponible à la figure 2.2.

La caméra virtuelle est placée légèrement en retrait de l'emplacement du robot et vise l'horizon. La caméra conserve cette position en suivant «en laisse» le modèle du robot, un peu à la manière d'un chien. Il est observé qu'une perspective exocentrique facilite

² *Open Source Computer Vision Library*. <http://www.intel.com/technology/computing/opencv/>

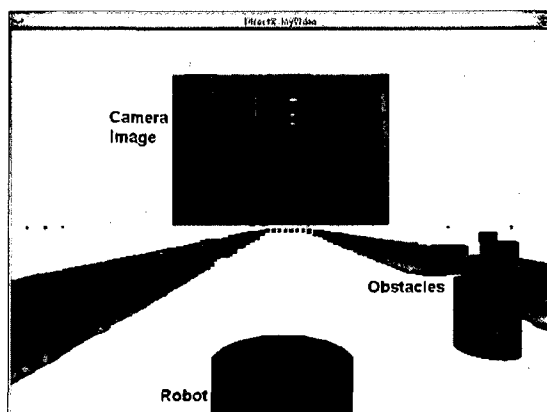


Figure 2.2 Interface développée par Ricks *et al.* [Ricks *et al.*, 2004]

l'intégration de l'information spatiale chez l'opérateur. Il est également noté que ce genre de perspective est préférable lorsqu'il est nécessaire de comprendre la structure spatiale de l'environnement, mais qu'une perspective égocentrique est préférable pour des tâches de navigation. Il propose donc une interface où l'opérateur peut constamment réorienter la caméra virtuelle selon ses préférences, puisque les interactions humain-robot comportent généralement ces deux types de tâches à la fois.

Par la suite, ces observations ont été détaillées par Nielsen *et al.* [Nielsen *et al.*, 2007]. Tout d'abord, une distinction entre perspective ajustable et adaptative est présentée : une perspective ajustable est un point de vue qui peut être modifié par l'opérateur, alors qu'une perspective adaptative est un point de vue qui s'ajuste automatiquement au contexte. Les jeux vidéos présentent souvent ce genre de perspective où la caméra virtuelle se déplace de façon à éviter les obstacles qui pourraient gêner la vue. Les perspectives ajustables permettraient d'aider les trois niveaux de la compréhension de la situation selon Endsley [Endsley, 1988] : la perception, car elles permettent de visualiser que l'information requise ; la compréhension, car elles permettent de supporter l'opérateur dans différentes tâches ; et la projection, car elles permettent de maintenir l'attention lors de changement de perspectives, puisque les transitions sont continues.

Des observations similaires ont été faites par Wang dans ses travaux sur les points de vues dynamiques dans des environnements virtuels [Wang, 2004]. Il avance que l'usage d'un point de vue dynamique suivant toujours «en laisse» le sujet à contrôler permet d'unir les avantages des vues purement égo ou exocentriques. Pour la laisse virtuelle, il propose un modèle masse-ressort-amortisseur à six degrés de liberté. Ses résultats expérimentaux suggèrent que les longues lasses favorisent la perception spatiale de l'environnement alors

que les lisses plus courtes favorisent la navigation locale. De plus, un modèle plutôt rigide de lisse est préféré à un plus souple en navigation locale.

Plus récemment, Cooper *et al.* [Cooper et Goodrich, 2008] ont expérimenté trois types de perspectives pour interfaces intégrées pour la téléopération simulée d'avion sans pilote :

- Une perspective de chasse (*chase*), où le point de vue est situé en retrait de l'appareil.
- Une perspective nord-vers-le haut (*north-up*), où le point de vue est situé directement au-dessus de l'appareil qui se déplacent sur un plan de l'environnement toujours orienté de la même façon.
- Une perspective direction-vers-le-haut (*track-up*) où, comme pour la perspective nord vers le haut, on observe l'appareil directement d'au-dessus, mais c'est plutôt ce dernier qui est toujours orienté vers le haut, et le plan est aligné en conséquence.

L'objectif des opérateurs était de retrouver des marques au sol tout en déplaçant l'avion. Les marques étaient distribuées de trois façons : uniforme, gaussienne ou suivant un trajet rectangulaire. Pour éviter de noter des marques redondantes, les auteurs ont observé que la distribution de ses dernières avait un impact sur la performance. En effet, les perspectives de chasse et nord-vers-le-haut étaient plus efficaces pour des distributions uniformes, alors que celle de chasse l'était pour des distributions gaussienne et suivant un trajet. De plus, pour limiter les faux-positifs, la perspective nord-vers-le-haut était meilleure de manière significative pour la distribution gaussienne, mais la perspective de chasse se démarquait pour la distribution suivant un trajet. Ils concluent donc qu'une seule perspective n'est pas efficace pour toutes les situations.

2.5 Reconstruction d'environnements 3D

Dans certaines situations, un plan en 3D de l'environnement parcouru est disponible. Ce n'est évidemment pas le cas lorsque l'environnement est visité pour la première fois. Il est donc pertinent de s'intéresser à la reconstruction en 3D des lieux parcourus à partir de ce que le robot capte.

Surmann *et al.* [Surmann *et al.*, 2003] proposent une technique intéressante basée sur la fusion d'une image vidéo conventionnelle et les lectures d'un laser de proximité 3D. Pour combiner les lectures faites tout au long du parcours, l'étude met en œuvre un algorithme décrit au départ par Chen *et al.* [Chen et Medioni, 1992]. De plus, elle met en œuvre des optimisations, proposées par Simon *et al.* [Simon *et al.*, 1994], basées sur des arbres

de recherche à k dimensions, ce qui permet une estimation en temps réel des poses des lectures.

Thrun *et al.* [Thrun *et al.*, 2000] ont mis en œuvre un système similaire, mais en utilisant deux lasers de proximité 2D, un pointant vers l'avant du robot et l'autre vers le haut. L'étude présente également une technique permettant la cartographie d'un environnement à partir des lectures de plusieurs robots parcourant un même endroit simultanément.

Les techniques d'accumulation et d'estimation de poses des données présentées dans ces études sont particulièrement intéressantes puisqu'elles sont applicables à d'autres sources de points 3D, comme une caméra stéréoscopique qui fournirait à la fois la couleur des pixels captés et une estimation de la distance à laquelle ils se trouvent.

2.6 Pertinence de l'utilisation d'éléments translucides

Dans un environnement virtuel ou dans des applications de réalité augmentée, la transparence est souvent utilisée pour présenter des éléments qui n'appartiennent pas réellement à l'environnement. Ces éléments peuvent être par exemple des curseurs pointant des objets importants de l'environnement. Un curseur en semi-transparence avec l'arrière-plan permet à l'utilisateur de voir à la fois le curseur et l'objet qu'il pointe, peu importe de quel angle l'environnement est observé.

Ces éléments translucides apportent un avantage intéressant lorsqu'ils sont utilisés avec des systèmes d'affichage conventionnels à deux dimensions : il est possible de voir rapidement si un objet est devant ou derrière un élément translucide. C'est en effet quelque chose qui est perdu lorsque notre affichage ne permet pas de reproduire complètement les effets de perspectives.

De plus, une dimension additionnelle à cette information est ajoutée si des solides translucides, par exemple un cube, sont utilisés. En effet, il est alors possible de voir si un objet est derrière (la lumière émise par l'objet est atténuée par deux faces translucides), à l'intérieur (la lumière émise par l'objet est atténuée par une seule face translucide) ou devant (la lumière de l'objet n'est pas atténuée) le curseur.

Des travaux par Zhai *et al.* [Zhai *et al.*, 1996] montrent que l'usage de ce genre de curseurs dans des tâches de pointage d'objets en trois dimensions améliore grandement les résultats. Un exemple de l'effet de la transparence est disponible dans les captures d'écrans de la figure 2.3. Dans ces travaux, des utilisateurs arrivent à attraper des objets en mouvements

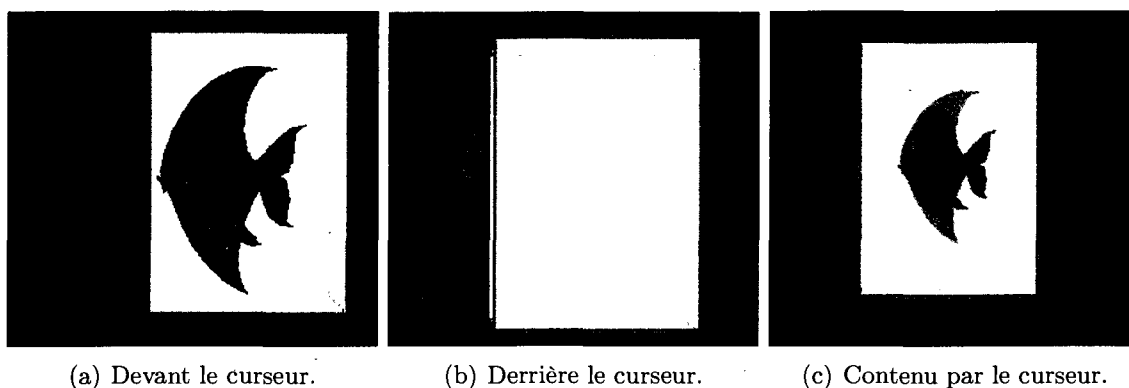


Figure 2.3 Effet de la semi- transparence d'un curseur 3D [Zhai *et al.*, 1996].

dans un environnement en trois dimensions en beaucoup moins de temps qu'avec un curseur n'ayant que des arêtes, même lorsque qu'un affichage stéréo reproduisant l'effet de perspective est utilisé.

2.7 Transmission de vidéo stéréoscopique

La transmission de vidéo stéréoscopique prend nécessairement plus de bande passante que la transmission de vidéo conventionnelle puisqu'elle utilise deux flux vidéos. Il ne faut cependant pas oublier l'objectif d'utiliser la vidéo stéréoscopique, c'est-à-dire la reconstruction de l'information de profondeur de la scène observée. L'ajout de cette information à un flux vidéo conventionnel peut être plus efficace que l'ajout d'un deuxième flux à la transmission. En effet, la profondeur peut être encodée sur un seul canal plutôt que trois pour une image couleur [Grewatsch et Miiller, 2004]. Il y a beaucoup de recherche dans le domaine de la télévision à trois dimensions pour trouver une façon efficace de la diffuser.

Une approche repose sur la similarité entre les vecteurs de mouvement de la séquence vidéo et de ceux de la séquence de la profondeur. En modifiant la procédure normale d'un codec comme le H.264/AVC, il est possible de récupérer les vecteurs de mouvements calculés pour la vidéo et les appliquer en partie à l'encodage de la profondeur, ce qui accélère l'encodage de la séquence complète. Des travaux de Pourazad *et al.* [Pourazad *et al.*, 2006] mettent en œuvre cette technique et montrent qu'il est possible de diminuer le calcul des vecteurs de mouvement de l'information de profondeur d'environ 60%. Ces travaux sont inspirés des travaux précédents de Grewatsch et Miiller [Grewatsch et Miiller, 2004] qui modifiaient plutôt le codec MPEG-2 et obtenait des résultats satisfaisants en limitant la bande passante de l'information de profondeur à 20% de celle de l'image conventionnelle.

CHAPITRE 3

Principes dans la conception d'une interface-opérateur 3D flexible

Après avoir étudié le domaine des interfaces humain-robots en navigation téléopérée, nous remarquons deux grandes tendances dans les travaux récents : la multiplication de vues juxtaposées nuit à un bon maintien de la compréhension de la situation, et l'usage de la vidéo, particulièrement superposée à un plan de l'environnement, aide à réduire le nombre de collisions et les temps de complétion des parcours. De plus, l'usage d'une représentation en 3D plutôt que strictement en 2D semble donner de meilleurs résultats.

Une bonne interface-opérateur doit maintenir une bonne compréhension de la situation chez un opérateur de plateforme robotique mobile dans deux types de conditions :

- *Navigation en environnement connu.* Par exemple, fournir de l'assistance, via un robot de téléprésence, à des personnes âgées dans un domicile dont un plan statique de l'environnement est disponible à l'avance. Dans ce cas, l'opérateur doit être en mesure d'identifier rapidement l'état d'éléments dynamiques (e.g., la personne à assister) par rapport à l'environnement statique parcouru (e.g., le domicile).
- *Navigation en environnement inconnu.* Par exemple, retrouver une personne enfouie sous des décombres dans une application de recherche et sauvetage. Ici, il est toujours important d'être capable de distinguer une personne du décor, mais il est particulièrement important d'être en mesure d'éviter des obstacles pour ne pas se retrouver coincé en pleine recherche. De plus, il faut s'assurer d'avoir parcouru l'environnement au complet.

Pour les deux types de conditions, nous formulons les hypothèses de travail suivantes :

- *La vidéo est importante.* Dans les cas présentés à la section 2.2, la présence de la vidéo était appréciée par les opérateurs. De plus, elle était particulièrement efficace pour aider à prendre conscience de l'entourage local du robot. Finalement, c'est à peu près le seul moyen d'identifier rapidement si une personne se trouve en face du robot. Il est possible d'y arriver à partir

- de sources sonores ou de lectures de capteurs de proximité, mais la vidéo demandera toujours moins de temps de réflexion. Finalement, nous avons remarqué que lorsqu'un flux vidéo est fusionné à un plan, il a tendance à en cacher une bonne partie, étant donné qu'il est souvent rendu de façon opaque. Nous supposons que, grâce à l'usage d'une caméra stéréoscopique, le fait de projeter en 3D l'image vidéo à même la représentation cartographique de l'environnement plutôt que sur un plan fixe pourrait diminuer l'importance de ce problème, surtout s'il est rendu en semi-transparence.
- *Une représentation cartographique de l'environnement est nécessaire.* Pour les tâches d'assistance en milieu connu, il est particulièrement important d'avoir un plan d'ensemble pour planifier le chemin le plus court menant à un endroit donné. Il est possible de supposer qu'un opérateur typique pourrait être associé à plusieurs patients demeurant à des domiciles différents. Il ne devrait donc pas être espéré qu'il connaisse par cœur la dispositions des pièces de tous les domiciles qu'il aurait à naviguer.
 - *Plusieurs vues juxtaposées peuvent nuire à la compréhension de la situation.* Il a été remarqué que le fait d'avoir dans une interface plusieurs endroits où se trouve l'information n'est pas nécessairement appréciée de l'opérateur. Celui-ci doit constamment déplacer son attention d'un endroit de l'interface à l'autre et peut aller jusqu'à en ignorer complètement.
 - *L'opérateur doit pouvoir adapter l'interface à ses préférences pendant qu'il opère le robot.* Certaines dispositions ou positions de caméra virtuelles sont plus adaptées à certaines tâches qu'à d'autres. Il est possible de supposer qu'un opérateur apprécierait d'être en contrôle de ces paramètres pendant la navigation et d'adapter l'interface selon la tâche en cours. Par exemple, sortir d'une pièce demande une plus grande compréhension de l'entourage immédiat du robot que de ce qui se trouve dans les autres pièces. C'est le contraire lorsqu'il réussit à quitter la pièce et que l'opérateur veut connaître par quelles autres pièces il est nécessaire de passer pour se rendre à la destination voulue.

Notre objectif est donc de permettre à l'opérateur de trouver la disposition idéale en 3D pour une interface de téléopération d'un robot mobile dans des tâches de navigation. Cette disposition permettrait de maintenir un maximum de compréhension de la part de l'opérateur de la plateforme téléopérée. La situation visée est principalement celle de la téléprésence en assistance à des personnes en perte d'autonomie dans leurs propres

domiciles. Ainsi, un bon maintien de cette compréhension permettrait à l'opérateur de rejoindre le plus rapidement possible une personne qui demanderait de l'aide en minimisant les mauvais trajets et les blocages de la plateforme dans des obstacles statiques ou dynamiques. C'est ce que nos tests devront mesurer.

Pour tenter de trouver la meilleure disposition d'interface possible, nous devons en développer une suffisamment flexible pour nous permettre de tester efficacement différentes configurations. Voici quelques critères que nous croyons importants de respecter :

- *L'interface doit présenter le plus d'information à l'échelle possible.* Nous croyons par exemple qu'il est plus facile d'évaluer les distances s'il est possible de comparer la taille d'un modèle en 3D du robot contrôlé (en admettant que l'opérateur a déjà vu le robot en personne) avec son environnement immédiat. Un simple symbole ne permet pas cela. Ce principe s'applique également pour la vidéo. Avec une caméra stéréoscopique, il est possible de projeter l'image vidéo à travers l'environnement reproduit et ainsi mieux apprécier la distance à laquelle se trouvent les éléments vus par le robot.
- *La reconfiguration doit être possible lors de l'exécution du système.* Si l'application doit être redémarrée ou pire, recompilée, l'objectif ne sera pas atteint. Les interactions de l'opérateur avec le robot doivent être interrompues le moins possible.
- *L'opérateur doit pouvoir activer ou désactiver rapidement l'affichage de certaines données.* Par exemple, si la télémétrie n'est plus suffisamment précise pour localiser le robot dans un environnement connu, l'opérateur devrait pouvoir désactiver l'affichage du plan pour qu'il ne nuise pas à la lecture des données locales comme la vidéo ou les capteurs de proximité.
- *L'application de contrôle doit pouvoir être compilée et exécutée sous plusieurs environnements.* Le code développé doit être suffisamment générique pour être compilé sous au moins trois plateformes : Linux, Windows et Mac OS X. Ainsi, l'application sera suffisamment mobile pour être testée à la fois sur des postes fixes au laboratoire, des postes fixes de maison ou différents ordinateurs portables.
- *L'interface de contrôle doit être compatible avec plusieurs plateformes robotiques.* Pour cela, l'utilisation d'un protocole commun comme *Player*¹ et

¹The Player Project. Free software tools for robot and sensor applications. <http://playerstage.sourceforge.net/>

la mise en œuvre dans le cadre de la plateforme MARIE² du laboratoire est fortement considérée. De plus, d'autres projets du laboratoire ont déjà manifesté de l'intérêt pour le développement d'une interface générique d'affichage en 3D. Le système doit donc être facilement réutilisable dans divers environnements:

Nous croyons donc qu'en développement une interface d'affichage d'informations propices à la navigation présentant une reproduction en 3D de l'espace parcouru pourrait aider aux développements des différents travaux en télérobotique du laboratoire. Finalement, le fait de projeter un flux vidéo stéréoscopique à travers cette représentation est une contribution novatrice et ouvrira potentiellement la porte à d'autres projets qui pourraient pousser encore plus l'usage de ce type d'imagerie.

En s'inspirant des travaux présentés à la section 2, nous présentons une maquette de ce que l'interface finale pourrait être à la figure 3.1. Dans cette maquette, une perspective exocentrique est utilisée pour intégrer un modèle du robot téléopéré (en rouge) se déplaçant dans un modèle 3D de l'environnement (en gris). Les lectures du laser de proximité (rideau translucide turquoise) et la projection de la vidéo stéréoscopique y sont superposées.

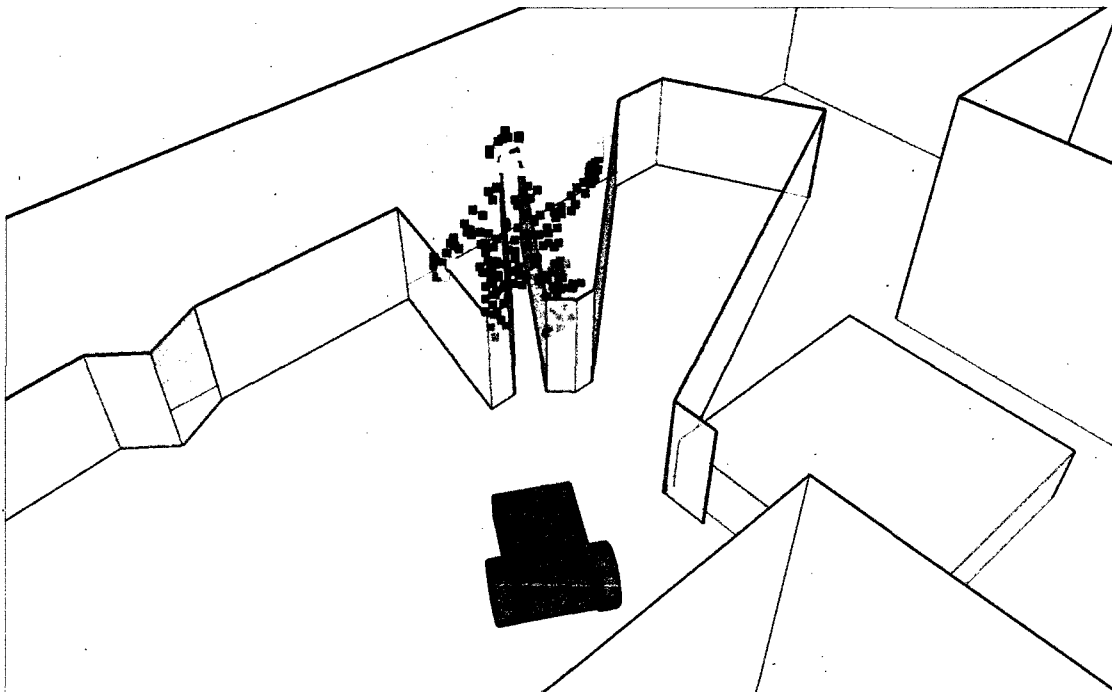


Figure 3.1 Maquette de l'interface à réaliser.

²Mobile Autonomous Robot Integrated Environment (MARIE). <http://marie.sourceforge.net>

3.1 Méthodologie retenue

Pour évaluer la qualité de l'interface développée, nous recueillons les commentaires à chaud des opérateurs et les analysons selon le système *LASSO* [Drury *et al.*, 2007]. Nous avons imaginé deux types de parcours :

- Un parcours dont le plan d'ensemble et l'endroit à atteindre sont connus à l'avance.
- Un parcours inconnu dans lequel un élément facile à distinguer visuellement doit être retrouvé.

Nous nous intéressons principalement au deuxième type de parcours impliquant plusieurs tâches, comme suivre des indications ou identifier des éléments distinctifs. Le parcours est assemblé à même les locaux disponibles dans l'entourage du laboratoire, afin de limiter les coûts et les risques du transport de la plateforme robotique. L'environnement reproduit respecte cependant le plus possible les dimensions d'un domicile conventionnel.

Nous pouvons classer en quatre catégories les opérateurs observés :

- *Type A.* Cet opérateur n'a jamais travaillé avec un robot mobile auparavant. Il nous permet d'évaluer le niveau de formation requis chez un néophyte.
- *Type B.* Cet opérateur n'a jamais travaillé avec un robot mobile auparavant, mais connaît le domaine d'application, e.g., un professionnel de la santé. Cet opérateur est donc en mesure de nous renseigner sur les attentes dans le domaine.
- *Type C.* Cet opérateur n'a pas nécessairement déjà travaillé avec un robot mobile, mais occupe une profession qui lui demande d'opérer un système complexe en mouvement, souvent à partir de lectures d'instruments, e.g., un pilote d'avion ou un opérateur de machinerie lourde. En plus d'avoir déjà prouvé avoir les capacités motrices nécessaires, cet opérateur devrait nous permettre de comparer les problèmes communs des différentes facettes des interactions humains-machines.
- *Type D.* Cet opérateur a déjà travaillé avec des robots mobiles ou en a déjà développé. Cet opérateur peut nous aider à comparer ce qui se fait déjà dans le domaine.

Idéalement, nous devrions faire appel à plus d'une personne par catégorie. Cela pourrait être plus difficile pour les types B et C, qui ne sont pas nécessairement disponibles dans

l'entourage immédiat du laboratoire. Le minimum est donc d'au moins une personne par catégorie. Il faut toutefois spécifier que l'objectif de nos travaux n'est pas d'arriver à formuler des observations par rapport à ces catégories d'opérateurs, mais bien d'évaluer les fonctionnalités du système flexible d'interfaces de téléopération et son utilisation dans un contexte expérimental réaliste afin de bien préparer son utilisation dans des études ultérieures.

Pour les deux types de parcours, nous évaluons le nombre d'obstacles avec lesquels l'opérateur a été en contact et sa performance dans l'achèvement des différentes tâches sélectionnées. En disposant de suffisamment d'utilisateurs, nous pouvons établir la moyenne de chaque groupe dans les deux métriques. En comparant les moyennes des différentes configurations de l'interface, nous pouvons donc cibler quelles valeurs des paramètres, i.e., perspective égocentrique vs. exocentrique, améliorent ou diminuent les performances de chaque type d'opérateur. Les commentaires pris à chaud permettent également de tirer des conclusions intéressantes à cet égard.

Finalement, les positions occupées par la caméra virtuelle s'avèrent aussi être une métrique intéressante. Puisque l'opérateur peut réorienter la caméra à sa guise, il est intéressant de mesurer à quelle fréquence chaque type d'opérateur change la position de la caméra et dans quel type de perspective, égocentrique ou exocentrique, il reste le plus longtemps.

Les tests s'effectuent avec le *Télérobot*³, un projet en développement au sein du laboratoire. Certains tests préliminaires pendant le développement, i.e., sans les opérateurs décrits plus haut, ont été faits sur d'autres plateformes puisqu'elles permettaient l'installation d'au moins une caméra stéréoscopique et d'un laser de proximité. Ces tests ont permis notamment de confirmer que notre interface est réutilisable dans d'autres contextes.

³*Télérobot* - Robot mobile de téléprésence à domicile. <http://www.gel.usherbrooke.ca/laborius/projects/Telerobot/index.html>

CHAPITRE 4

Article

Cet article, intitulé *Egocentric and Exocentric Teleoperation Interface using Real-time, 3D Video Projection* a été soumis à la *4th ACM/IEEE International Conference on Human-Robot Interaction* qui aura lieu du 11 au 13 mars 2009 à San Diego en Californie. Il expose le développement du système et les tests qui ont été entrepris pour étudier ses possibilités. Les auteurs sont François Ferland, François Pomerleau, Chon Tam Le Dinh et François Michaud.

Résumé - Interface de téléopération égocentrique et exocentrique utilisant une projection vidéo 3D en temps réel

L'interface opérateur est l'élément central d'un système de téléprésence robotique. Les modalités de visualisation utilisées affectent grandement la compréhension de la situation de l'opérateur et ainsi ses performances dans la complétion de ses tâches. Selon la tâche en cours, passer d'un point de vue égocentrique à un autre exocentrique et améliorer la représentation de la profondeur peuvent fournir une meilleure représentation de l'environnement d'opération. Le système développé, qui combine une reconstruction 3D des lectures d'un capteur laser de proximité et deux techniques de projection vidéo, permet à l'opérateur de passer facilement d'un point de vue à un autre. Cet article présente le développement de l'interface et démontre ses possibilités en ayant 13 utilisateurs téléopérant un robot mobile dans un contexte de navigation.

Abstract

The user interface is the central element of a telepresence robotic system and its visualization modalities greatly affect the operator's situation awareness, and thus its performance. Depending on the task at hand and the operator's preferences, going from ego- and exocentric viewpoints and improving the depth representation can provide better perspectives of the operation environment. Our system, which combines a 3D reconstruction of the

environment using laser range finder readings with two video projection methods, allows the operator to easily switch from ego- to exocentric viewpoints. This paper presents the interface developed and demonstrates its capabilities by having 13 operators teleoperate a mobile robot in a navigation task.

4.1 Introduction

The user interface is the central element of a telepresence robotic system. It is through this interface that users can interact remotely with the intended environment, via the robot. Such interfaces must be designed in accordance with user's skills and application purposes, and not from the designer's standpoint, which is too often the case [Adams, 2002]. Information display, layout and visualization capabilities directly affect teleoperation performance. For instance, displaying the robot's reference frame and status, combining readings from multiple sensors, providing the ability to inspect it and automatically presenting contextually-appropriate information (e.g., showing a rear-facing camera when moving backward) all need to be considered in teleoperation interfaces for urban search and rescue (USAR) applications [Scholtz *et al.*, 2004].

Because it provides a general and multi-dimensional (2 dimensional with color, texture) view of the remote environment, a video display is found in most interfaces. It can however be enhanced by adding other information (e.g., laser range finder, virtual reality) to improve situation awareness [Drury *et al.*, 2003; Yanco et Drury, 2004] while minimizing the operator's cognitive load [Yanco *et al.*, 2004]. Such fusion must be done in appropriate ways, and simply using distinct windows results in operators focusing mostly on the video display [Baker *et al.*, 2004; Drury *et al.*, 2007; Yanco *et al.*, 2006]. Combining information modalities is also influenced by the visualization viewpoint, which can be egocentric (seeing the world from the robot's perspective) or exocentric (observing the world from an external perspective). An egocentric viewpoint is usually better for navigating and avoiding obstacles, while an exocentric viewpoint gives the operator a better understanding of the environment's structure [Ricks *et al.*, 2004]. Egocentric multi-modal displays can superimpose graphical information to a 2D video feed (e.g., distance markers and translucent 2D maps [Yanco et Drury, 2007], infrared camera superposition for heat sources detection [Hestand et Yanco, 2004]) or superimpose a 2D video feed with a virtual reality model of the world [Labonté *et al.*, 2008]. Proportion and localization issues are then key challenges to address. In interfaces following the ecological paradigm [Nielsen *et al.*, 2007], where the teleoperated robot is displayed along with its environment, exo-

centric multi-modal displays can be made by constructing 3D structures from exclusively 2D data [Ricks *et al.*, 2004]. However, displaying video on a fixed 2D plane in front of the robot introduces a serious occlusion problem, because the 2D plane can completely hide the 3D extrusion from the range finder readings. Finally, including the representation of the robot's chassis in either ego- or exocentric views contributes significantly to situation awareness [Keyes *et al.*, 2006].

As a logical extension to such observations, this paper demonstrates that it is possible to use 3D data representations in real-time in a new interface that 1) combines laser range finder data with 3D video projections from a stereo camera (to handle occlusion problems), and 2) provides the operator with the ability to change viewpoints as desired. After presenting the hardware and software components used in our experimental settings to situate the capabilities and constraints of our development, the paper describes our ego-/exocentric 3D interface and the two video projection methods implemented. It also reports observations made from having 13 operators use the interface in a navigation task in terms of operator's gaze, viewpoints, control and task accuracy, situation awareness and usefulness. This pilot study was meant to validate the possibility of evaluating precise metrics of our system in preparation of a performance-based comparison of different display modalities.

4.2 Experimental Platform

Our experimental setup makes use of Telerobot [Labonté *et al.*, 2008; Michaud *et al.*, 2007], a differential drive robot platform shown in Figure 4.1 and designed to provide health care assistance in home environments. The robot's onboard computer is an Intel Pentium® M processor at 1.7 GHz, 512 MB of RAM and using a 802.11B wireless LAN card. The laser range finder is a Hokuyo URG-04LX placed approximately at 0.4 m from the ground. This particular model has a 270 degrees field of view over a range of 4 meters. A Videre Design STOC (*STereo On a Chip*) stereo camera (with two 320×240 video feeds) is placed between a webcam (used for other purposes) and the laser range finder. This position provides a good view of the ground in front of the platform and does not significantly hinder the recognition of higher-placed objects or individuals. This Videre camera model processes disparity information on a onboard FPGA, avoiding taking up processing time on the robot's onboard computer. A Logitech® wireless gamepad, which features two analog joysticks and ten digital buttons, is used to teleoperate the platform.

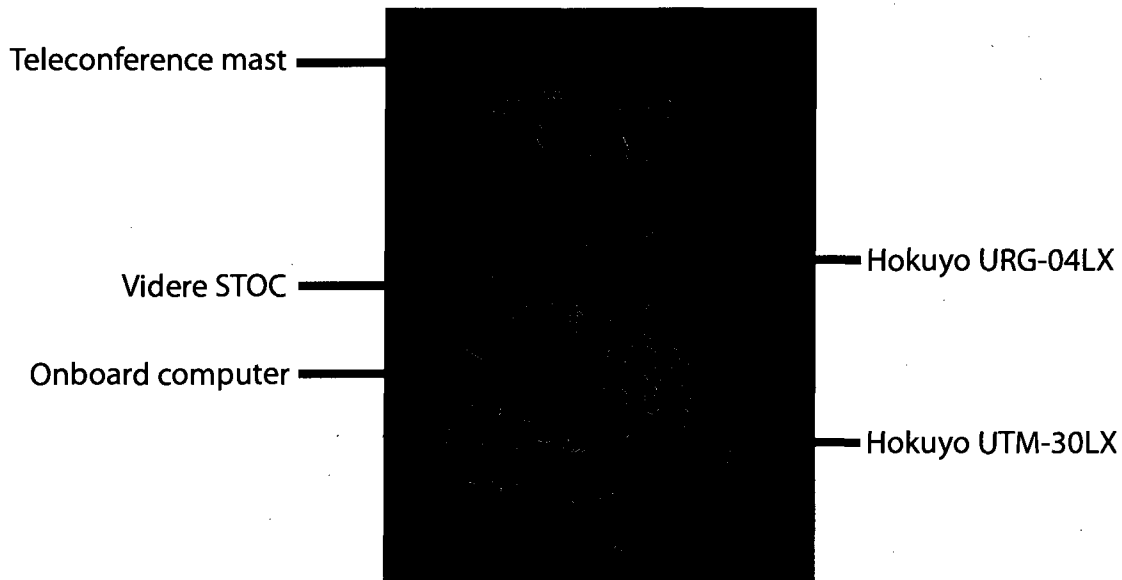


Figure 4.1 *Telerobot*, a teleoperated robot for assistive homecare.

The teleoperation interface has been mostly developed and entirely tested on an laptop computer featuring an Intel Core[®] Duo clocked at 1.83GHz with 2GB of RAM and an ATI[®] X1600 GPU with 128 MB of dedicated video memory. The developed software is distributed on both the robot's onboard computer and the operator's laptop. The developed server reads various sensors, calculates odometry, sends data at 10 Hz over the network, receives asynchronous command from the client's software and applies them to the motor drives. A set of libraries were developed to manage user inputs and communicate with robot servers. Except for stopping automatically the motors when the network connection is lost, there are no security modalities applied to the operator's commands. Video acquisition and transmission is done concurrently on a distinct TCP/IP connection. Once acquired on the robot, the left image and the disparity information are encoded in a single, double-width image (640 × 240) and transmitted at 10 Hz using the FFmpeg H.263+ codec implementation to compress data at 500 kbps. For the 3D display, we developed a simple library to manage the rendering of a asynchronously-updated scene graph. Our library requires OpenGL 2.1 and currently runs on Mac OS X and Linux. Asynchronous incoming data, represented as events, immediately updates internal structures. These changes are picked up by various synchronized threads, notably our rendering thread which needs to be closely coupled with the underlying operating system.

4.3 Egocentric and Exocentric 3D Interface

Figure 4.2 features screenshots of our ego-/exocentric 3D interface. The top-left window (parts (b) and (c)) shows the left video feed coming from the stereoscopic camera. The text on the bottom-left of the main window displays status information about the robot (e.g., the batteries' voltage level, odometry). The main window displays either the egocentric or the exocentric viewpoints, constructed using laser-based video projection. In the exocentric view, laser range finder data are extruded in red, and the yellow parts are zones that are considered empty. The displayed square grid has $1\text{ m} \times 1\text{ m}$ cells and is fixed at the world's origin, and the circular grid, separated by 1 m , follows the robot.

Internal representation of the environment is based on a sparse 3D cell array with a granularity of 0.04 m . Incoming spatial data is continuously accumulated to provide the operator with a global map of the environment. For laser range finder data, our interface first maps the data in a 3D discretized point cloud, to then create a flattened 2D representation from the ground plane to create the local map. This local map's cells are then extruded at the same height as the laser range finder to create virtual, translucent walls. Odometry readings are used to position the robot and local maps. After having moved from more than either 0.25 m in translation or 10 degrees in rotation, local maps are compared and registered using an Iterative Closest-Point (ICP) variant named Kd-ICP [Pomerleau *et al.*, 2009]. When the corrected position is obtained, the robot's model position in the interface is changed, a new local map is created and the 2D global map is updated. Because the dimensions of the global map are unknown at the initialization phase, the global map is managed as a set of sparsely located tiles of a fixed size. The global map resolution is independent of the original local maps and is defined as an amount of cells in a single square size covering 1 m in size. The tile set is indexed in a hash map. To place the incoming local map, our interface uses a single matrix translating and rotating its bitmap, built from the corrected position. This process, implemented using a set of OpenGL Shading Language (GLSL) vertex and fragment programs, provides an efficient way to display a large bitmap on a plane in 3D space. Global maps frequently reach 1280×1280 cells.

The robot's model is displayed using multi-element triangular meshes. Surface normals are defined for each triangle vertice. Each element can have a different material defined by its diffuse, specular and ambient color, and specular shininess factor. Rendering is done using a single directional light source with the Blinn-Phong [Blinn, 1977] shading model. To create the robot's model, the complete SolidWorks[®] model is exported to a Virtual Reality Markup Language (VRML) file, rotated using Blender (an open-source

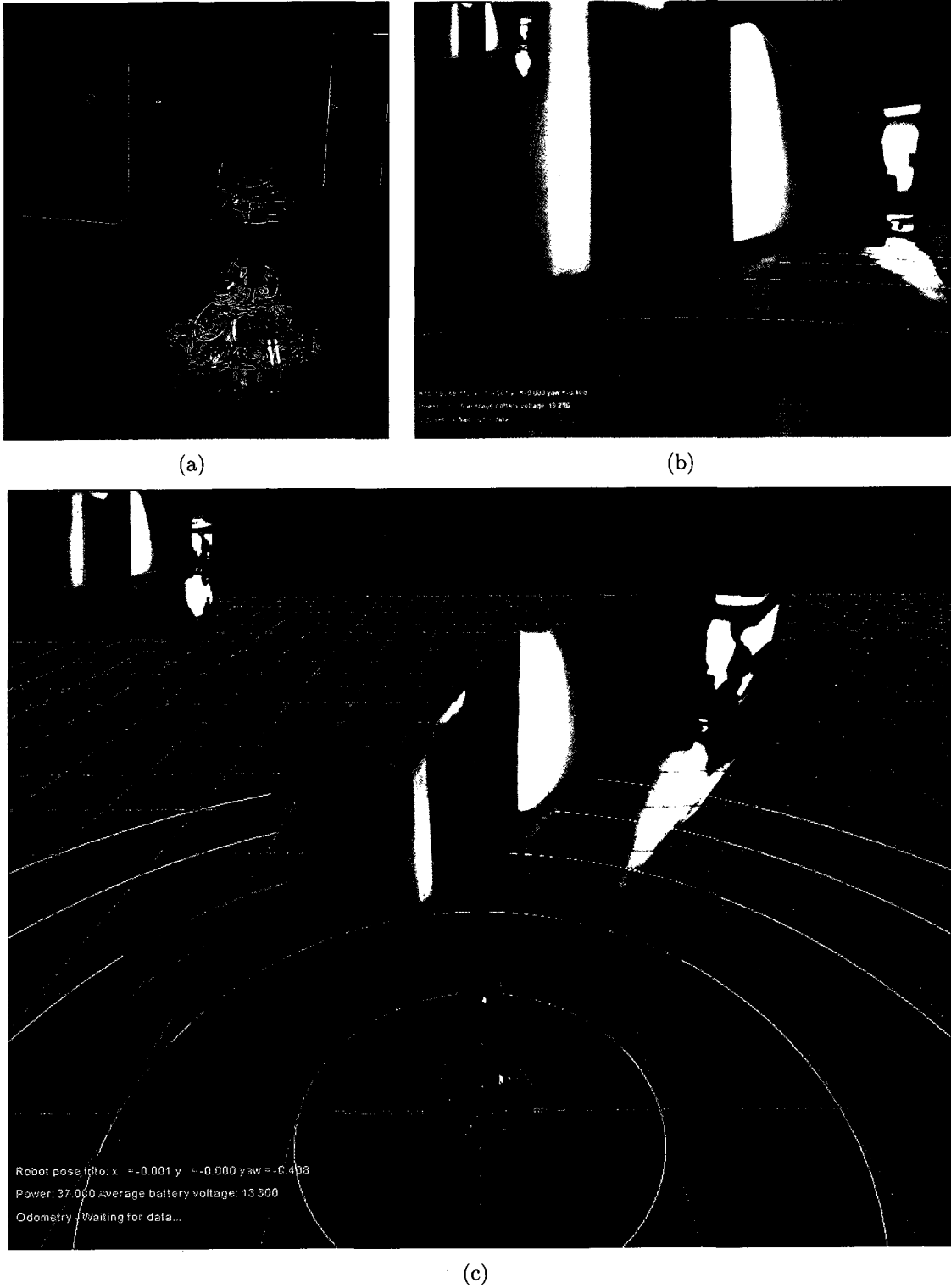


Figure 4.2 Screenshots from our ego-/exocentric 3D interface : (a) the scene ; (b) egocentric viewpoint ; (c) Exocentric viewpoint.

application) to make it fit with our coordinate system, and then exported to a Collada file, a XML-based format popular with 3D modeling animation packages. The Telerobot model features more than 400,000 triangles and does not tax significantly the system's rendering performance. Being able to provide a true-to-life representation of the operated robot is important for proportion representation in the 3D interface. Creating such models quickly and effortlessly facilitates the use of other robot platforms with our interface.

A virtual camera, acting as the operator's viewpoint and displaying the reconstructed 3D scene in the main window, is always following the robot by targeting a point located at 1 m over the ground and 1 m in front of it. This target point places the robot's model in the lower third portion the main window, with the virtual camera at its default position of 3 m ($\rho = 3$) behind the target, looking straight ahead ($\theta = 0$) at 30 degrees tilt angle from the ground ($\phi = \frac{\pi}{6}$). Figure 4.3 illustrates this configuration using a cylinder to represent the robot's base, an arrow for its heading and an eye for the operator's viewpoint. By modifying ϕ and keeping $\theta = 0$, the interface can switch from chase and track-up perspectives [Cooper et Goodrich, 2008]. Adjustable perspectives have been noted as a way to reduce the operator's cognitive workload [Nielsen *et al.*, 2007]. To enable this, the operator can continuously rotate and zoom around this target point using one of the analog joysticks and two buttons. In addition to these controls, the operator can cycle through five presets positions using a single button. These presets include a purely egocentric point of view and a top-down and track-up map-like one. Figure 4.4 shows screenshots of the three first exocentric presets, with the video projection and extrusion disabled.

According to Milgram's taxonomy of Mixed Reality (MR) of display integration [Milgram et Colquhoun, 1999], our system could be classified as variably-congruent. Its centricity is continuously variable, since the operator can move freely between ego- and exocentric viewpoints. Even though our control scheme is always ego-referenced, it closely matches the display's reference frame when the viewpoint is situated right behind our robot. However, the Control/Display (C/D) offset increases when the selected θ angle is farther from zero. Finally, our control order is 1 because our joystick is mapped to the speed of the robot.

To display the stereo camera data, our interface uses one of two video projection methods. The first simply projects color images from the video feed to the virtual representation of the world as set by laser range finder data, and the second projects stereoscopic data.

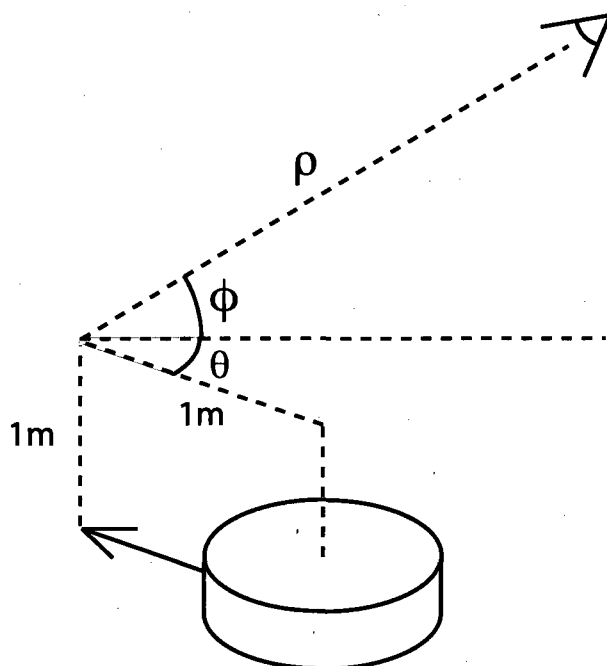


Figure 4.3 Viewpoint configuration.

4.3.1 Laser-based Projection

The laser-based projection method creates a 3D mesh covering the floor and elevating virtual walls based on the laser range finder data. Disparity information from the stereoscopic camera is not considered. This mesh covers the field-of-view of the video feed and serves as a virtual projection screen. Each of the mesh's vertices is passed through conventional model, view and projection matrices describing the transformation from the mesh's geometry to 2D coordinates on the left video image plane. These coordinates represent texture coordinates used to map the video image on the mesh. As expressed by Equation 4.1 [Shreiner *et al.*, 2008], the vector $(mv_{14}, mv_{24}, mv_{34})^T$ represents the translation from the stereo camera's left viewpoint to the laser range finder's origin. In our case, the camera and the laser both have the same orientation, but a rotation matrix could be applied to the *ModelView* matrix if the laser range finder and/or stereo camera were placed on a pan-tilt head, for instance.

$$ModelView = \begin{bmatrix} 1 & 0 & 0 & mv_{14} \\ 0 & 1 & 0 & mv_{24} \\ 0 & 0 & 1 & mv_{34} \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (4.1)$$

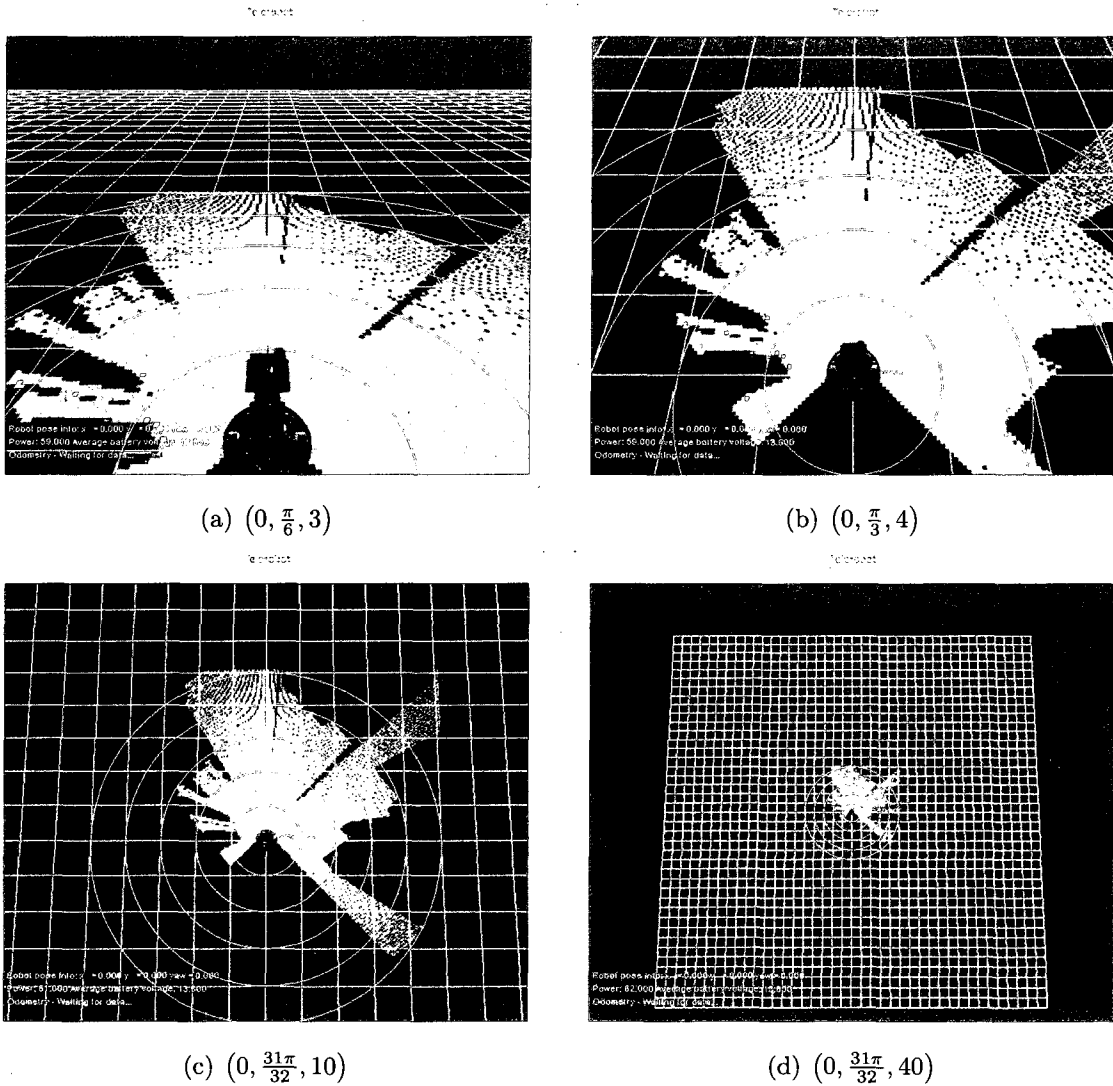


Figure 4.4 Screenshots of exocentric viewpoint presets, noted as (θ, ϕ, ρ) .

The projection matrix is expressed by Equation 4.2 [Shreiner *et al.*, 2008], where (t, b, l, r, n, f) , for top, bottom, left, right, near and far, are the frustum parameters, which can be calculated from the lens' horizontal field of view in radians (θ) , the aspect ratio of the image (a) and a fixed near and far plane :

$$\begin{aligned}
 a &= w/h \\
 n &= 1.0 & f &= 1000.0 \\
 w_h &= n \sin(\theta/2) & h_h &= \frac{w_h}{a} \\
 l &= -w_h & r &= w_h \\
 b &= -h_h & t &= h_h
 \end{aligned}$$

$$Projection = \begin{bmatrix} \frac{2n}{r-l} & 0 & \frac{r+l}{r-l} & 0 \\ 0 & \frac{2n}{t-b} & \frac{t+b}{t-b} & 0 \\ 0 & 0 & \frac{-(f+n)}{f-n} & \frac{-2fn}{f-n} \\ 0 & 0 & -1 & 0 \end{bmatrix} \quad (4.2)$$

So, for a mesh vertex v , its location v' on the video image plane can be calculated using Equation 4.3.

$$v' = (v'_x, v'_y, v'_z, v'_w)^T = Projection \times (ModelView \times v) \quad (4.3)$$

After a conventional perspective division using v'_w , the normalized (from -1.0 to 1.0 in each dimension) 2D coordinate on the image frame v'' is derived from Equation 4.4 :

$$v'' = \frac{(v'_x, v'_y)^T}{v'_w} \quad (4.4)$$

With the full set of v'' , it is possible to apply standard texture mapping using OpenGL. In our setup, these texture coordinates are generated in a GLSL vertex shader used in a single rendering pass for the whole projection.

Stereoscopic Projection This method exploits depth information that can be derived from the camera's stereoscopic disparity data. 2D image vectors (defined as $v_s = (x, y, d, 1)$) from such a camera can be projected in 3D space using the projection matrix defined by Equation 4.5 [Videre Design, 2007], where a is the image's aspect ratio, f_s the lens' focal length and b_s the stereo camera baseline, i.e., the space between our head's optical axes.

$$StereoProjection = \begin{bmatrix} a & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & f_s \\ 0 & 0 & -\frac{1}{b_s} & 0 \end{bmatrix} \quad (4.5)$$

To efficiently transform entire images into a 3D representation at 10 Hz, we developed a set of GLSL fragment shaders to generate and render a resolution-independent triangular mesh that can be resized according to a system's processing abilities. Our first shader

takes the single-channel disparity map as its input and applies the matrix of Equation 4.5 to vertices created from the normalized texture coordinates (between 0.0 and 1.0) for the x and y components and the sampled disparity at these same coordinates. The result is stored in a 4 channel, single precision float texture that is the size of the desired mesh. This texture is copied into a vertex buffer using a pixel buffer. The vertex buffer is then rendered using a constant index buffer connecting all coordinates into a square grid. The result is a deformed surface texture-mapped with the original left-camera image. We are currently using a 256×256 mesh, which is a little bit smaller than the image's resolution while still giving us acceptable results, since this mesh usually only covers a portion of the main window.

When sampling the disparity map, our method applies a local variance filter to reject pixels that are situated between zones that are far away from each other. By calculating the variance in a 3×3 window around each disparity sample, a mask of pixels going over a certain threshold is created. This mask acts as a second rendering target in the projection fragment shader and is used during the rendering stage to leave these fragments as transparent. This disparity threshold can be adjusted at runtime.

4.4 Experimental Results

The objectives of the experiments conducted are to evaluate our ego-/exocentric 3D interface modalities in conditions that challenge both the operator's cognitive load and situation awareness when navigating in an unknown course and while accomplishing observational tasks. By letting the operator choose the visualization viewpoint, we also want to observe what preferences emerge for the tasks. The experimental scenario consists of teleoperating Telerobot in the area represented in Figure 4.5. The operators have not and cannot directly see the course and have two tasks to accomplish : 1) start at the black dot, locate table A, turn around it and count the number of visible orange cones ; 2) follow a path identified by arrows on walls and furniture (at approximately 0.5 m from the ground). The course features different lighting conditions (grey : natural illumination ; white : fluorescent lighting). A trial starts with an explanation of the system's control and display, followed with a short familiarization period. When ready, the operator starts executing the tasks. During a trial, at the locations identified by black triangles, the operator gets interrupted by the experimenter to estimate the distance between two colored tubes that are clearly visible from the video feed but not registered with the laser range finder. At the end of the course, the operators are instructed to go back to the starting position and

place the robot at 2 m from the wall, directly aligned with a specific target placed at the height of the camera. One trial lasts from 15 and 45 minutes, and the maximum robot's velocity allowed is set to 0.3 m/s.

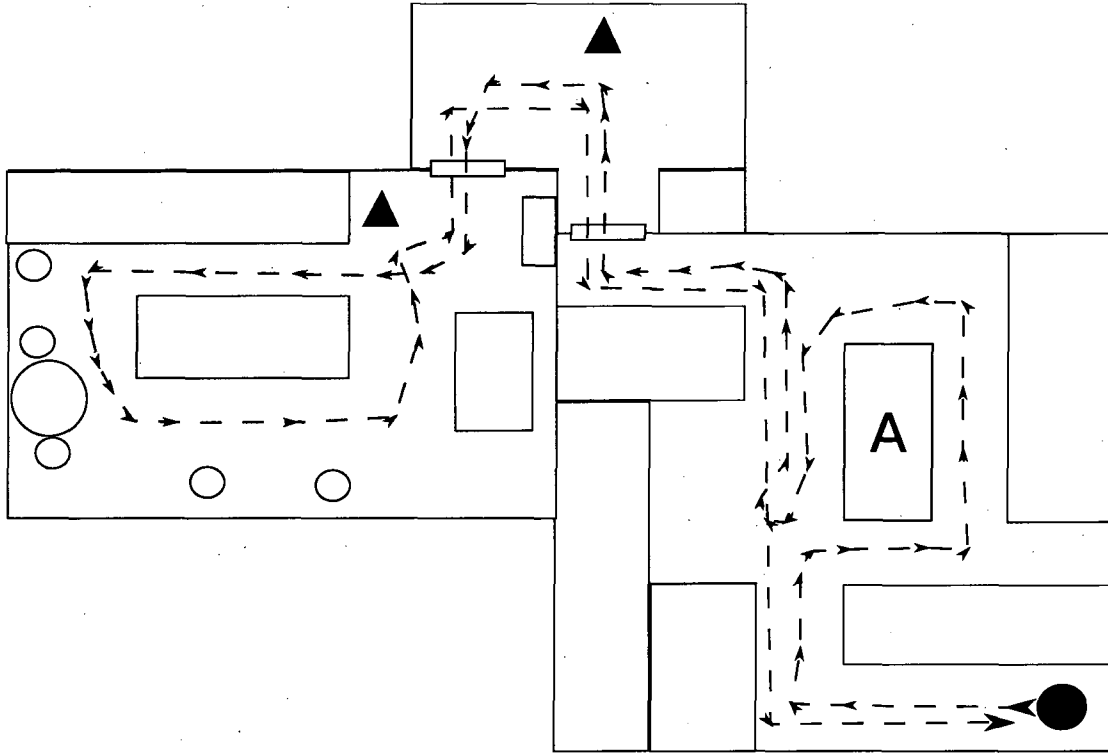


Figure 4.5 Experimental course.

Using a faceLAB[®] system (developed by Seeing Machines Inc.), we measured the operator's eye gaze directed toward the interface and the reference sheet placed on the table below the screen. We also logged the position of the interface's virtual camera, collisions with obstacles, commands given to the robot and the time taken to accomplish the tasks. We also encouraged operators to think aloud during the trial and tallied their comments using the LASSO technique [Drury *et al.*, 2007]. Once the tasks accomplished, we asked operators to rank, from one to five, their impression of the usefulness of different features of our interface, i.e., the 2D Video Window (top-left window), Laser Mapping and Extrusion (red-yellow lines), the Adjustable Viewpoint feature, Laser-based Projection and Stereoscopic Projection.

The test population was made of a convenience sample of 13 voluntary male participants, aged from 19 to 58. Four had previous experience in teleoperating a mobile robot, seven mentioned playing videogames at least one hour per week, and two never worked on or used a robot of any form.

4.4.1 Gaze

Figure 4.6 represents the total distribution of the operators' gaze on the screen. A gaze point or glance represents a single sample recorded by the faceLAB[®] system running at 60 Hz. Glances at the 2D video window represent 4.8% of total trial samples. The gaze is mostly concentrated in the top two thirds of the screen, which corresponds, for exocentric viewpoints, to what lies ahead of the robot. Glances at our text display are extremely rare and coincide with the lack of utterances relative to the information there. The small size of the text might have played an important role in these results.

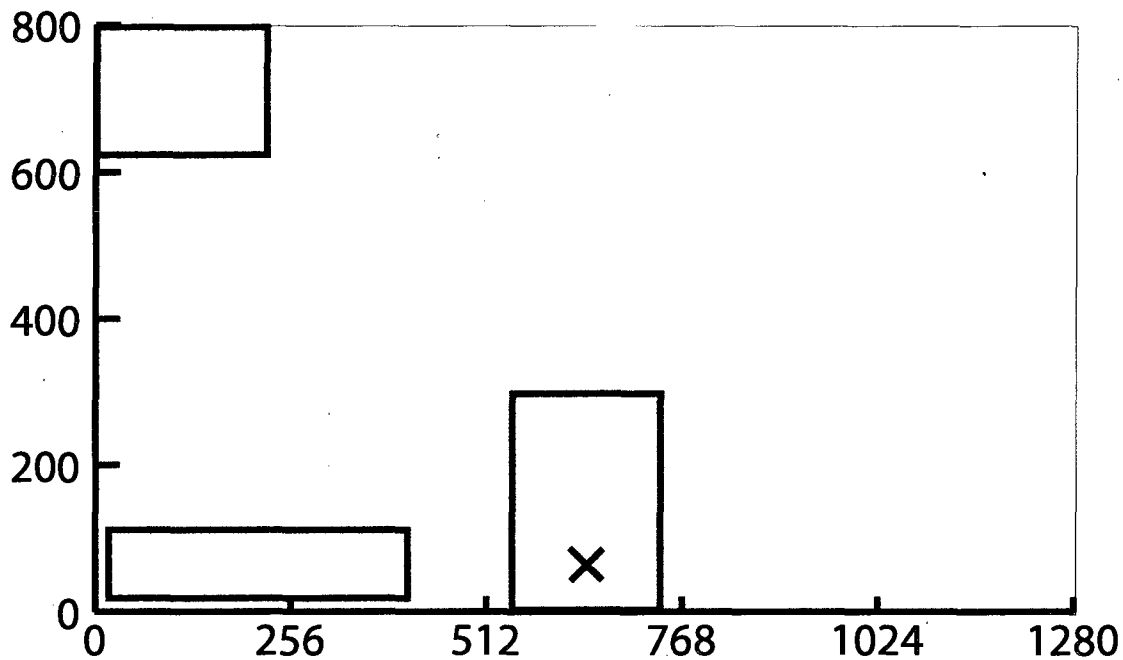


Figure 4.6 Operator's gaze on the screen distribution, in pixel coordinates. The top-left rectangle represents the 2D video window, and the bottom-left one the text display. As a reference, the cross and the center rectangle represent the base and model display location of the robot in the default initial viewpoint configuration.

4.4.2 Viewpoint Positions

From the virtual camera's positions for all the trials (500 000 positions), we filtered out the initial default position, and each component of the viewpoint (θ, ϕ, ρ) were considered independently to produce the chosen viewpoint dataset, shown in Figure 4.7. Changes in θ , ϕ and ρ account for 35.9%, 68.7% and 27.2%, respectively. The θ distribution shows that the operators positioned the virtual viewpoint to look to the left ($\theta > 0$) more often than

to the right, possibly because the tasks were mostly accomplished following a counter-clockwise path. The ϕ distribution suggests that a lower, straight-ahead tilt angle was preferred, especially for tasks requiring identifying objects in the scene, like directional arrows or orange cones. More vertical, top-down positions were selected for shorter period of time, usually in tight navigation situations like going through doorways, situations that happened less frequently than other tasks. Finally, the ρ distribution indicates that ego- or closely located exocentric positions ($\rho < 5$ m) were preferred. In a few occurrences, operators positioned the virtual camera at more than 30 m to get a full top-down map view of the traveled path.

Changes in either θ , ϕ or ρ occurs 78.2% of the time. Our initial hypothesis was that operators familiar with video-games might adjust the viewpoint more often, since it is a feature available in most 3D games. However, this was not the case : the differences between the two groups are too small to be significant. Operators playing videogames at least one hour per week spent 76.6% of their time using the same viewpoint. This behavior was observed 80.1% of the time with the other operators.

4.4.3 Control and Task Accuracy

The average time spent while the robot moves (translation or rotational motions) is 34.2%. The average and median translational velocities when moving are 64.9% and 71.0% of the maximum velocity. Furthermore, five operators had their median linear velocities at 100%, meaning they spent at least half of the time in movement at full velocity. This coincides mostly with operators not often exposed to videogames and having difficulties using the analog joysticks for accurate movements. While the robot was in translation, rotational commands were given 54.9% of the time. This implies that operators were confident enough to change the robot's direction without pausing it first. The cases when rotational commands were given without translational motions represent 22.4% of the operational time.

Out of 13 participants, seven managed to complete the tasks without encountering any obstacles, four hit directly or grazed one or two objects, and two hit more than five obstacles. Three operators missed a cone and one counted the same cone twice. Estimating distances with the interface remains difficult, however. On average, operators overestimated distances by 32%. Overestimations occurred 77% of the time and were not consistent between operators. Only one underestimated both distances during its course. Many display factors such as the field of view have been shown to influence distance estimations

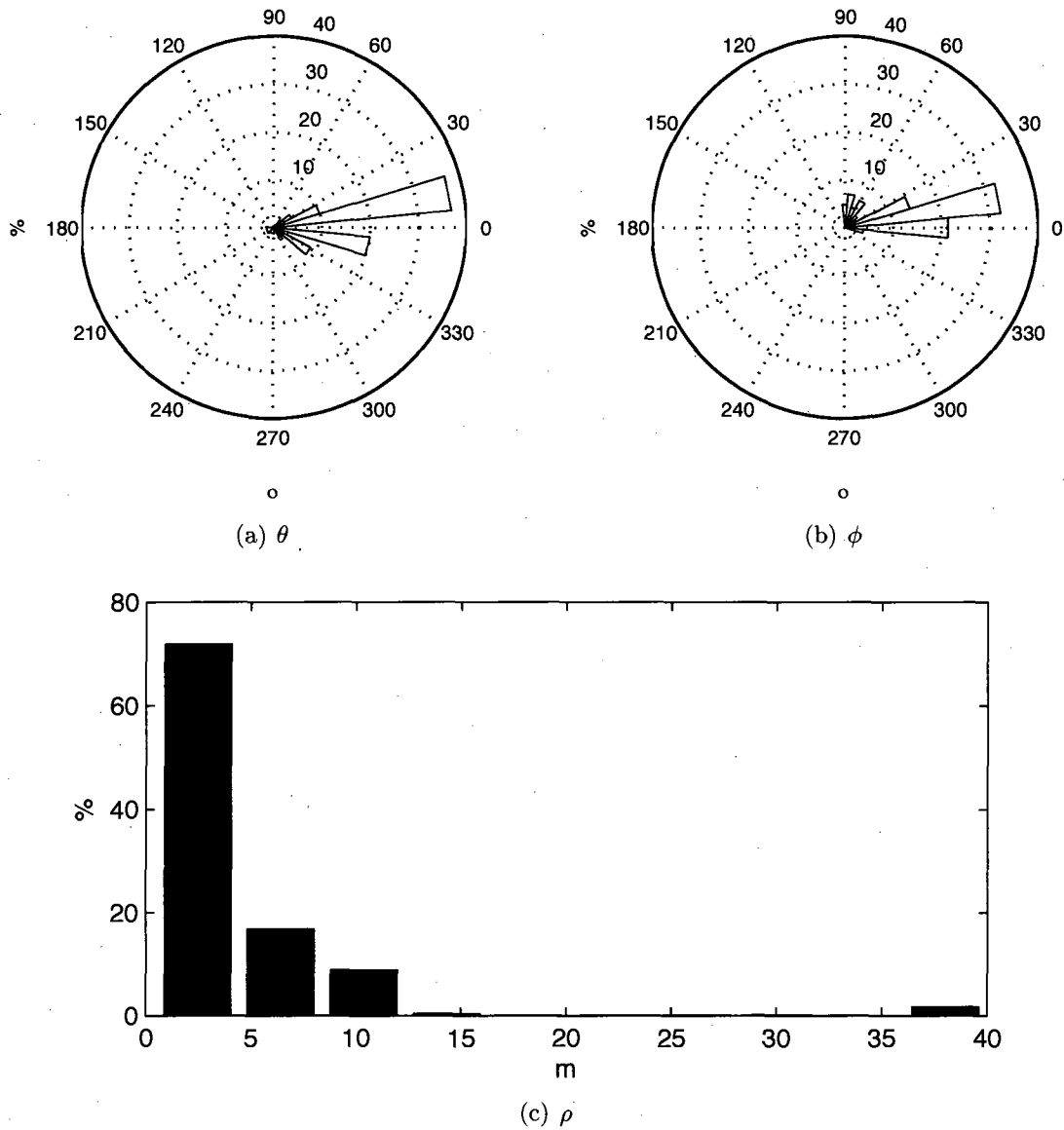


Figure 4.7 Virtual camera position distributions in terms of percentage of time spent in distinct positions. In (a) and (b), robot's heading is at 180° . The optical axis always cross the center of both graphs.

in virtual environments [Witmer et Kline, 1998] and would need to be fully assessed in further studies before formulating any conclusions on the distance estimation performance of our system.

4.4.4 LASSO Classification

The amount of utterances varies greatly from operators to operators, as in [Drury *et al.*, 2007]. In our case, it could be as low as 7 or high as 61 per operator, for a total of 311 utterances. Table 4.4.4 shows that most of the negative utterances were tallied in the Surroundings category and were related to the state of the stereoscopic projection. Operators would complain that they could not figure out what was going on in front of them, until they switched off this projection method and went back to laser-based projection. Doing so sometimes generated a positive utterance in the same category. Figure 4.8 shows a scene with minimal distortion (top), visible as lines extending to infinity, in reconstructing the cabinets' doors. The bottom screenshot is more distorted : while most of the door stays flat, superfluous displacement is visible in other portions of the scene. Overall mission awareness was usually high, and most positive remarks were related to updating the status of a task, e.g., "I now have seen 3 cones".

TABLEAU 4.1 LASSO classification.

Awareness Type	% Positive	% Negative	% Neutral
Location	22.19	2.89	74.92
Activities	6.75	4.82	88.42
Surroundings	59.16	14.47	26.37
Status	15.76	8.36	75.88
Overall Mission	41.48	5.47	53.05

4.4.5 Usefulness

Table 4.2 suggests that the most useful feature in our interface is laser mapping and extrusion. The operators considered this modality as the only way to make sure that the robot was not colliding with obstacles. The display of the robot's model in conjunction with the laser range data was identified by many as essential. Another essential feature was the 2D video window. Many operators mentioned that they used it to confirm the nature of objects observed using either of the projection methods. This concurs with the data from Figure 4.6. Some operators however complained of the small size of this window, and would have like to have a quick way to temporarily display the unprocessed video stream in the main window.

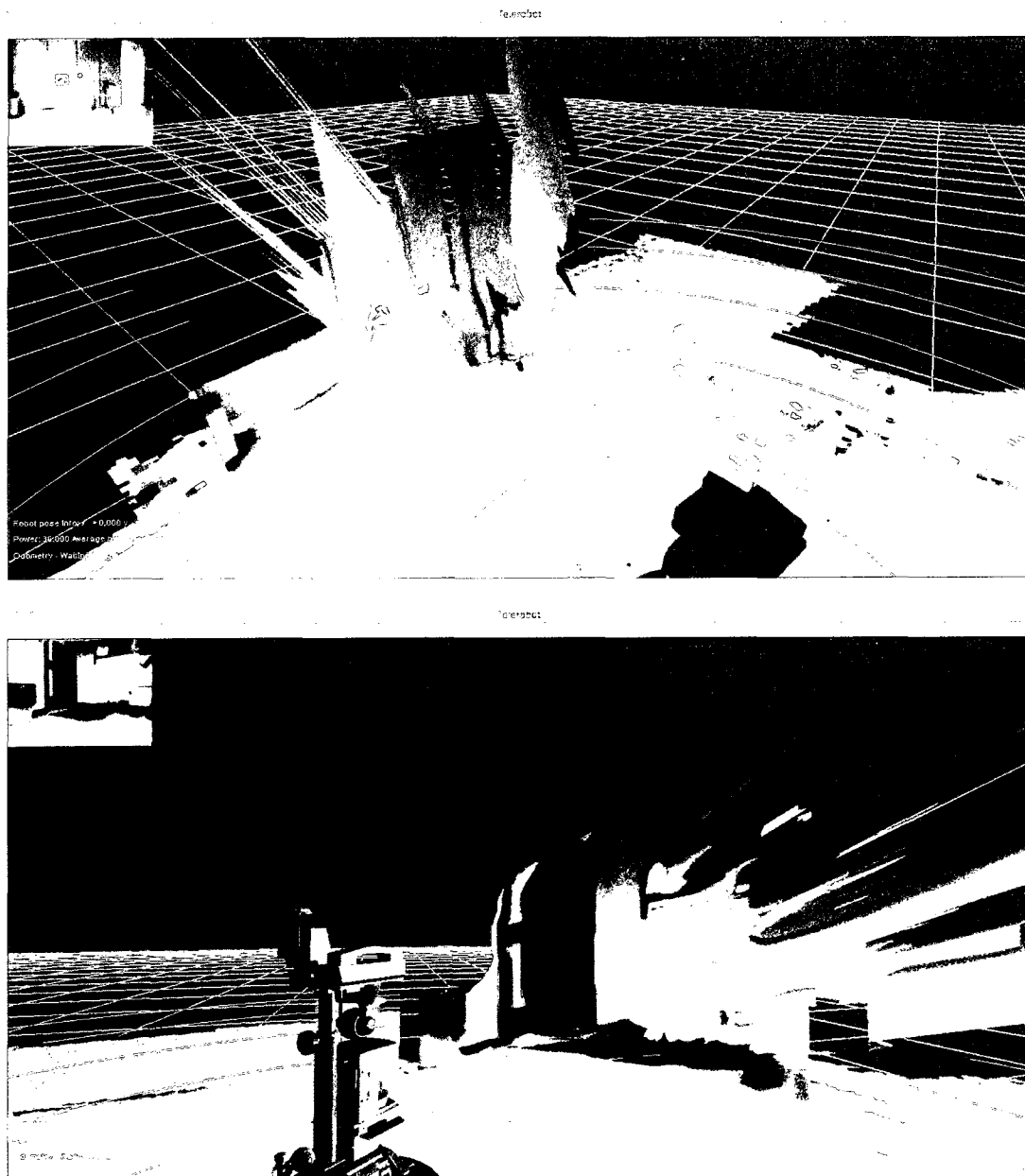


Figure 4.8 Stereoscopic projection screenshots.

TABLEAU 4.2 Usefulness average, from 1 to 5, according to operators after their trial run.

Laser Mapping and Extrusion	4.7
2D Video Window	4.2
Adjustable Viewpoint	4.0
Laser-based Projection	4.0
3D Stereoscopic Projection	1.9

The adjustable viewpoint feature was also appreciated. Being able to quickly tilt the viewpoint to obtain a close, top-down view of the robot's surroundings as perceived by the laser range finder, was considered to be essential for obstacle avoidance, especially while crossing doorways. However, one operator would have liked to be able to reconfigure the gamepad's layout by reassigning buttons or inverting the direction of the effect of the analog pad on the viewpoint's angle. This would be straightforward to implement in a manner similar to videogames and should be considered in future work.

Laser-based projection was also perceived as useful and, understandably, more than the stereoscopic projection. However, the stereoscopic projection was still turned on 49.1% of the time. Operators' ratings were often accompanied by a comment on how stereoscopic projection could certainly be helpful in figuring out the general structure of a room and would be really useful if the projection was more stable and less distorted. Both of our projection techniques provided good results when disparity data were sufficiently complete. In practice however, this proved to rarely be the case, which affected mostly stereoscopic projections. For images presenting little texture information, like a plain-colored wall, the disparity estimation algorithm implemented in Videre's FPGA simply could not provide enough useful information. It however performed fairly well for high-contrast zones like door frames over the same light-colored walls.

Finally, an unexpected remark was made concerning auditory feedback from the teleoperated robot. Even though the operators could not see the course, they could hear the robot moving for the first half of the trial. Some operators commented that they would have liked to have auditory feedback confirming that the robot is moving in areas where the robot could not be heard. Voice feedback communicating navigation decisions of a supervised, semi-autonomous robot has been noted as being effective and appreciated in similar situations [Perrin *et al.*, 2008]. Auditory feedback modalities would therefore be a nice feature that could be easily integrated in our interface.

4.5 Conclusion and Future Work

This paper presents a complete proof-of-concept demonstration of a new ego-/exocentric 3D interface that combines laser range finder data with 3D video information. We have shown that it is possible to efficiently combine in real-time laser range finder data with projection of 3D video data. Our interface's adjustable viewpoint feature gives flexibility to operators in selecting what reveals to work better in specific situations.

In future work, our system will be extended by adding an improved stereo disparity method, adding user-specifiable viewpoint presets and auditory feedback of the robot while in motion. The revised interface will then be a versatile tool to conduct rigorous evaluations of display modalities and visualization features with a larger number of operators and in conditions involving moving obstacles. It will also be used to conduct usability studies with novice operators in real telehealth homecare assistance tasks. Furthermore, we are currently extending our Kd-ICP mapping technique to reconstruct the navigated environment in 3D as in [Thrun *et al.*, 2004].

4.6 Acknowledgements

The authors gratefully acknowledge the contribution of the Natural Sciences and Engineering Research Council of Canada (NSERC), the Canada Research Chairs (CRC) and the Fondation of the Université de Sherbrooke for their financial support.

CHAPITRE 5

Conclusion

Les explications et les résultats présentés dans cet ouvrage démontrent bien que nous avons réussi à mettre en place un système flexible pour la téléopération d'une plateforme robotique. Celui-ci combine une reconstruction 3D des lectures d'un capteur laser de proximité et deux techniques de projection vidéo. De plus, il permet à l'opérateur de passer facilement d'un point de vue à un autre selon la tâche à effectuer. Après la réalisation du projet, nous croyons avoir atteint un résultat satisfaisant nos objectifs originaux. Le système développé est suffisamment flexible pour être étendu dans l'avenir. En pointant des métriques intéressantes, l'analyse des résultats de nos tests, malgré un petit nombre d'utilisateurs, est un bon point de départ pour l'ébauche d'une série d'expérimentations plus rigoureuses. Il est également possible d'envisager une comparaison des performances entre l'interface intégrée présentée dans ce document et d'autres semblables, intégrées ou non.

Une extension importante à considérer pour des expériences plus rigoureuses et de nouvelles métriques serait de synchroniser les données des systèmes externes comme *faceLAB* avec les commandes de l'opérateur. Avec les données actuelles, il est trop fastidieux de tenter d'associer un regard avec une commande précise, ce qui explique l'absence de telles analyses dans ce document. De plus, il faudrait envisager d'enregistrer l'affichage de l'interface pendant son utilisation. Une telle approche avait été envisagée avec la caméra de scène fournie avec le système *faceLAB*, mais la taille de l'écran, sa position relative et l'éclairage présent ne permettaient pas de bien distinguer ce qui était présenté. En enregistrant l'affichage de façon synchronisée, qui pourrait être à une fréquence ou résolution inférieure de façon à limiter la quantité de données, nous pourrions associer les stratégies des utilisateurs pour la configuration de leur point de vue et de leur regard sur l'écran en fonction des tâches qu'ils ont à accomplir.

Malgré les performances plutôt modestes de la projection stéréoscopique, nous croyons toujours qu'elle possède un bon potentiel. Plusieurs solutions à son amélioration sont imaginables, mais deux se démarquent. Premièrement, l'implémentation d'un meilleur algorithme d'évaluation de la disparité offrirait une bien meilleure reconstruction de la profondeur. Par contre, les meilleurs algorithmes disponibles peuvent être difficiles à utili-

ser en temps réel. Il faudra donc trouver un compromis. Deuxièmement, la reconstruction stéréoscopique pourrait être limitée aux éléments dynamiques ou distinctifs de la scène. Par exemple, en isolant les personnes de l'arrière plan et en appliquant l'analyse de disparité et la projection seulement sur ces parties de l'image, nous éviterions les situations où la projection stéréoscopique et celle basée sur le laser entrent en conflit.

Enfin, la construction en trois dimensions de la carte donnerait un avantage énorme à notre interface. Les algorithmes présentement mis en œuvre sont déjà extensibles en trois dimensions. La plus grande difficulté reste dans l'accumulation et le traitement des données, le système actuel n'étant pas suffisamment performant pour en accepter une beaucoup plus grande quantité. Par contre, le système actuel nous a permis, par sa flexibilité, de raffiner la technique de construction de carte en deux dimensions avec le Kd-ICP. Avec une simplification des données 3D par plans de l'environnement tel que présenté par Thrun *et al.* [Thrun *et al.*, 2004], qui n'est pas incompatible avec notre algorithme de cartographie, la technique de projection basée sur le laser pourrait facilement être étendue à tout l'environnement parcouru et accumulé par notre interface tout en conservant une performance d'affichage raisonnable en temps réel.

ANNEXE A

Calibration de la caméra

Pour obtenir une projection vidéo géométriquement correcte, les deux techniques doivent répliquer les caractéristiques de la lentille utilisée pour capturer l'image vidéo. À ce point de vue, les matrices de projection aux équations 4.2 et 4.5 dépendent toutes deux de θ , l'angle horizontal du champ de vision. Cet angle peut être calculé ainsi :

$$\theta = 2 \tan^{-1} \frac{y}{2l_f} \quad (\text{A.1})$$

où y est la taille horizontale de l'image formée derrière la lentille, donc du capteur, et l_f la longueur focale de la lentille, soit la distance entre celle-ci et l'image formée. Dans notre cas, nous possédons une lentille Rainbow de 2.8mm prévue pour des capteurs de taille diagonale de $\frac{1}{3}$ " , ce qui correspond à 4.8mm pour un ratio largeur/hauteur conventionnel de $\frac{4}{3}$. Nous devrions donc obtenir un angle de 1.42 radians. Or, il ne s'agit que d'un point de départ puisque la longueur focale est mécaniquement ajustable et que l'image ne se forme pas nécessairement sur la totalité du capteur. Pour finaliser les ajustement de la projection, il est possible, lorsque l'interface est en marche, d'ajuster à la fois l'angle du champ de vision et le ratio largeur sur hauteur de l'image par incrément de 0.01. Cela vaut également pour les vecteurs décrivant les distances qui séparent le laser et la caméra du centre du robot.

La technique d'ajustement est plutôt simple : il faut viser, avec le robot, une zone où des éléments sont franchement discernables avec le laser. Ces éléments doivent idéalement être distribués à diverses distances du robot, tout en restant dans la portée du laser. Les éléments les plus efficaces sont les coins de murs, particulièrement lorsque leur point de rencontre avec le sol et visible. Ensuite, il suffit d'ajuster les différents paramètres jusqu'à ce que l'image soit parfaitement alignée avec le support de projection généré par les lectures du laser. Il s'agit évidemment d'une approche essais-erreurs, mais la procédure demeurent très courte puisque nous disposons généralement d'un point de départ suffisamment près de la réalité.

ANNEXE B

Supplément aux résultats des tests

Certains résultats ont été omis de la rédaction de l'article et sont présentés dans cette section.

B.1 Profils des utilisateurs

La distribution détaillée des opérateurs qui ont participé aux tests est disponible au tableau B.1.

TABLEAU B.1 Profil des opérateurs.

No	Âge	Sexe	Occupation	Jeux vidéos par semaine, heures	Expérience en robotique
1	58	M	Consultant	0	Développement et Téléopération
2	23	M	Étudiant	0	Développement
3	25	M	Étudiant	5	Développement
4	19	M	Étudiant	1	Aucune
5	23	M	Étudiant	2 à 3	Développement
6	24	M	Étudiant	10 à 15	Développement
7	27	M	Étudiant	0	Développement et Téléopération
8	29	M	Professionnel de recherche, ing. méc.	0	Développement
9	34	M	Professeur	0	Développement et Téléopération
10	24	M	Étudiant	Environ 10	Développement et Téléopération
11	25	M	Étudiant	0	Développement et Téléopération
12	36	M	Technicien	5 à 15	Développement et Téléopération
13	22	M	Pilote d'avion, Instructeur	5	Aucune

B.2 Point de vue ajustable

La répartition 3D de toutes les données sur la position occupée par la caméra virtuelle de l'interface a été tracée à la figure B.1.

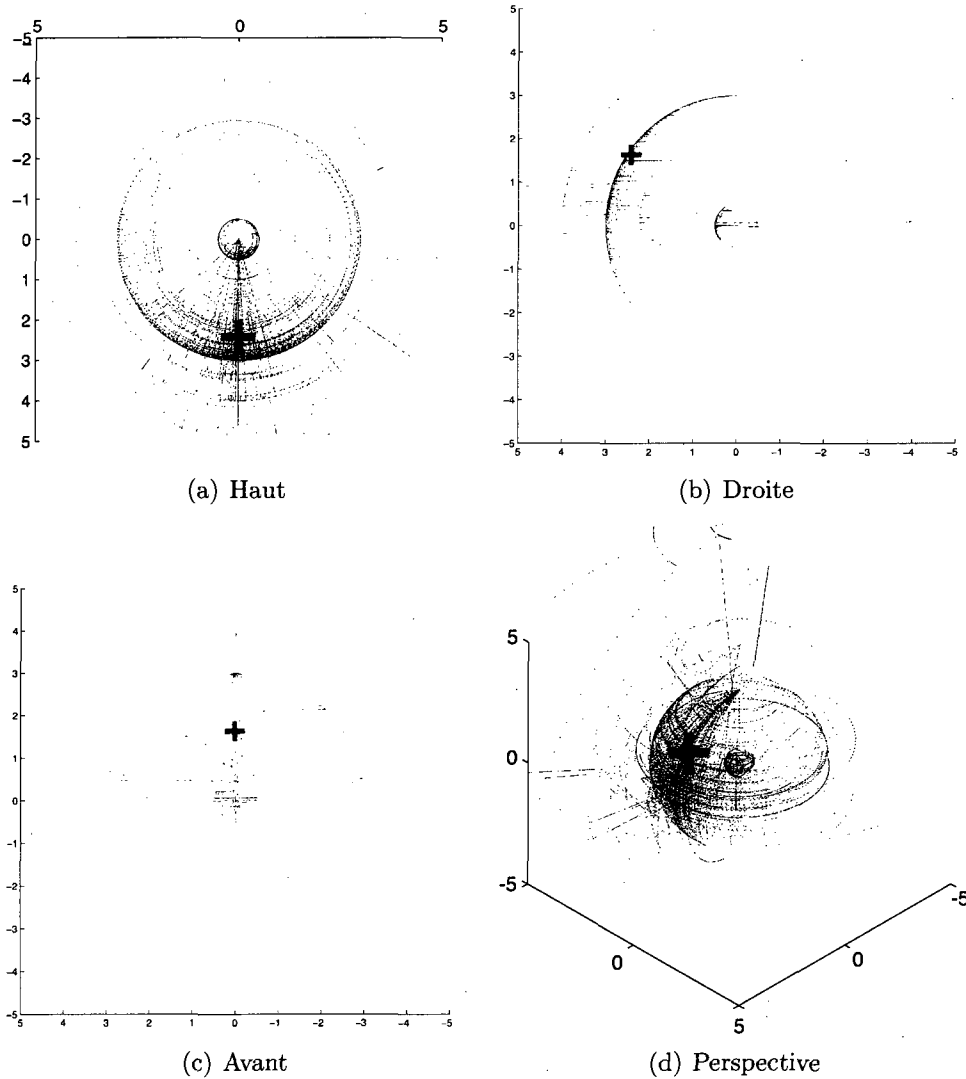


Figure B.1 Positions de la caméra virtuelle pour toutes les participations aux tests. La sphère intérieure représente les positions égocentriques. La croix est la position moyenne, qui est très proche de la position de départ par défaut. Le robot est positionné au centre à $(0,0,0)$ et est dirigé a) vers en haut à droite à la figure, b) vers le haut dans la figure, et c) vers la droite dans la figure.

BIBLIOGRAPHIE

- ADAMS, J. A. (2002). Critical considerations for human-robot interface development. *In AAAI Fall Symposium : Human Robot Interaction Technical Report FS-02-03*, pages 1–8.
- BAKER, M., CASEY, R., KEYES, B. et YANCO, H. A. (2004). Improved interfaces for human-robot interaction in urban search and rescue. *In Proceedings of the IEEE Conference on Systems, Man and Cybernetics*, volume 3, pages 2960–2965.
- BLINN, J. F. (1977). Models of light reflection for computer synthesized pictures. *In Proceedings of the 4th annual conference on Computer graphics and interactive techniques*, pages 192–198, New York, NY, USA.
- CHEN, Y. et MEDIONI, G. (1992). Object modelling by registration of multiple range images. *Image and Vision Computing*, 10(3):145–155.
- COOPER, J. et GOODRICH, M. A. (2008). Towards combining UAV and sensor operator roles in UAV-enabled visual search. *In Proceedings of the 3rd ACM/IEEE international conference on Human-Robot Interaction*, pages 351–358, New York, NY, USA.
- DRURY, J. L., KEYES, B. et YANCO, H. A. (2007). LASSOing HRI : Analyzing situation awareness in map-centric and video-centric interfaces. *In Proceedings of the ACM/IEEE International Conference on Human-Robot Interaction*, pages 279–286, New York, NY, USA.
- DRURY, J. L., SCHOLTZ, J. et YANCO, H. A. (2003). Awareness in human-robot interaction. *In Proceedings of the IEEE Conference on Systems, Man and Cybernetics*, volume 1, pages 912–918.
- DRURY, J. L., SCHOLTZ, J. et YANCO, H. A. (2004). Applying CSCW and HCI techniques to human-robot interaction. *In Proceedings of the CHI Workshop on Shaping Human-Robot Interaction.*, pages 13–16, Vienne.
- ENDSLEY, M. R. (1988). Design and evaluation for situation awareness enhancement. *In Proceedings of the Human Factors Society 32nd Annual Meeting*, pages 97–101, Santa Monica, CA.
- FARRUGIA, J. P., HORAIN, P., GUEHENNEUX, E. et ALLUSSE, Y. (2006). GPUCV : A framework for image processing acceleration with graphics processors. *In Proceedings of the International Conference on Multimedia and Expo*, pages 585–588.
- GRAVES, A. R. et CZARNECKI, C. (1999). A generic control architecture for telerobotics. *In Proceedings of Towards Intelligent Mobile Robots*, Bristol, UK.
- GREWATSCH, S. et MILLER, E. (2004). Sharing of motion vectors in 3D video coding. *In Proceedings of the International Conference on Image Processing*, volume 5, pages 3271–3274.

- HESTAND, D. et YANCO, H. A. (2004). Layered sensor modalities for improved feature detection. *In Proceedings of the IEEE Conference on Systems, Man and Cybernetics*, volume 3, pages 2966–2970.
- KEYES, B., CASEY, R., YANCO, H. A., MAXWELL, B. A. et GEORGIEV, Y. (2006). Camera placement and multi-camera fusion for remote robot operation. *In Proceedings of the IEEE International Workshop on Safety, Security and Rescue Robotics*.
- LABONTÉ, D., MICHAUD, F., BOISSY, P., CORRIVEAU, H., CLOUTIER, R. et ROUX, M. A. (2006). A pilot study on teleoperated mobile robots in home environments. *In Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 4466–4471.
- LABONTÉ, D., MICHAUD, F., CLOUTIER, R., ROUX, M.-A., LÉTOURNEAU, D. et BOISSY, P. (2008). Telerehabilitation via a mobile robot. *In Proceedings of Virtual Rehabilitation*.
- MICHAUD, F., BOISSY, P., LABONTÉ, D., CORRIVEAU, H., GRANT, A., LAURIA, M., CLOUTIER, R., ROUX, M.-A., IANNUZZI, D. et ROYER, M.-P. (2007). Telepresence robot for home care assistance. *In Proceedings of the American Association for Artificial Intelligence Spring Symposium on Multidisciplinary Collaboration for Socially Assistive Robotics*.
- MILGRAM, P. et COLQUHOUN, H. (1999). A taxonomy of real and virtual world display integration. *In Mixed Reality : Merging Real and Virtual Worlds*, ch. 1. Y. Ohta and H. Tamura (Ed.), Springer.
- NIELSEN, C., GOODRICH, M. et RICKS, R. (2007). Ecological interfaces for improving mobile robot teleoperation. *IEEE Transactions on Robotics*, 23(5):927–941.
- NIELSEN, C. W. et GOODRICH, M. A. (2006). Comparing the usefulness of video and map information in navigation tasks. *In Proceedings of the 1st ACM SIGCHI/SIGART Conference on Human-Robot Interaction*, pages 95–101.
- PERRIN, X., CHAVARRIAGA, R., RAY, C., SIEGWART, R. et del R. MILLÁN, J. (2008). A comparative psychophysical and eeg study of different feedback modalities for HRI. *In Proceedings of the 3rd ACM/IEEE International Conference on Human-Robot Interaction*, pages 41–48, New York, NY, USA.
- POMERLEAU, F., FERLAND, F. et MICHAUD, F. (2009). Kd-ICP for fast and robust map registration. Soumis à l'*IEEE International Conference on Robotics and Automation*.
- POURAZAD, M. T., NASIOPOULOS, P. et WARD, R. K. (2006). An H.264 encoding scheme for 3D TV. *In Proceedings of the European Signal Processing Conference*.
- RICKS, B., NIELSEN, C. W. et GOODRICH, M. A. (2004). Ecological displays for robot interaction : A new perspective. *In Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems*, volume 3, pages 2855–2860.
- SCHOLTZ, J. (2002). Evaluation methods for human-system performance of intelligent systems. *NIST Special Publication*, 990:161–166.

- SCHOLTZ, J., YOUNG, J., DRURY, J. et YANCO, H. (2004). Evaluation of human-robot interaction awareness in search and rescue. *In Proceedings of the IEEE International Conference on Robotics and Automation*, volume 3, pages 2327–2332.
- SHREINER, D., WOO, M., NEIDER, J. et DAVIS, T. (2008). *OpenGL(R) Programming Guide : The Official Guide to Learning OpenGL(R), Version 2.1 (6th Edition)*, chapitre F, pages 776–777. Addison-Wesley Professional.
- SIMON, D. A., HEBERT, M. et KANADE, T. (1994). Real-time 3-D pose estimation using a high-speed range sensor. *In Proceedings of the IEEE International Conference on Robotics and Automation*, volume 3, pages 2235–2241.
- STONE, H. W. (1996). Mars pathfinder microrover - A small, low-cost, low-power spacecraft. *In Proceedings of the AIAA Forum on Advanced Developments in Space Robotics*.
- SURMANN, H., NUCHTER, A. et HERTZBERG, J. (2003). An autonomous mobile robot with a 3D laser range finder for 3D exploration and digitalization of indoor environments. *Robotics and Autonomous Systems*, 45:181–198.
- THRUN, S., BURGARD, W. et FOX, D. (2000). A real-time algorithm for mobile robot mapping with applications to multi-robot and 3D mapping. *In Proceedings of the IEEE International Conference on Robotics and Automation*, volume 1, pages 321–328.
- THRUN, S., MARTIN, C., LIU, Y., HAHNEL, D., EMERY-MONTEMERLO, R., CHAKRABARTI, D. et BURGARD, W. (2004). A real-time expectation-maximization algorithm for acquiring multiplanar maps of indoor environments with mobile robots. *IEEE Transactions on Robotics and Automation*, 20(3):433–443.
- VIDERE DESIGN (2007). *SRI Small Vision System User's Manual version 4.4d*, chapitre 3, pages 45–47.
- WANG, W. (2004). *Human navigation performance using 6 DOF dynamic viewpoint tethering in virtual environments*. Thèse de doctorat, University of Toronto.
- WITMER, B. G. et KLINE, P. B. (1998). Judging perceived and traversed distance in virtual environments. *Presence : Teleoperators and Virtual Environments*, 7(2):144–167.
- YANCO, H. A., BAKER, M., CASEY, R., KEYES, B., THOREN, P., DRURY, J. L., FEW, D., NIELSEN, C. et BRUEMMER, D. (2006). Analysis of human-robot interaction for urban search and rescue. *In Proceedings of the IEEE International Workshop on Safety, Security and Rescue Robotics*.
- YANCO, H. A. et DRURY, J. L. (2004). «Where Am I?» Acquiring situation awareness using a remote robot platform. *In Proceedings of the IEEE Conference on Systems, Man and Cybernetics*, volume 3, pages 2835–2840.
- YANCO, H. A. et DRURY, J. L. (2007). Rescuing interfaces : A multi-year study of human-robot interaction at the AAAI robot rescue competition. *Autonomous Robots*, 22(4):333–352.

YANCO, H. A., DRURY, J. L. et SCHOLTZ, J. (2004). Beyond usability evaluation : Analysis of human-robot interaction at a major robotics competition. *Journal of Human-Computer Interaction*, 19(1-2):117-149.

ZHAI, S., BUXTON, W. et MILGRAM, P. (1996). The partial-occlusion effect : Utilizing semitransparency in 3D human-computer interaction. *ACM Transactions on Computer-Human Interaction*, 3(3):254-284.