HAL
open science

De la perception à l'action : l'asservissement visuel, de l'action à la perception : la vision active

François Chaumette

To cite this version:

François Chaumette. De la perception à l'action : l'asservissement visuel, de l'action à la perception : la vision active. Robotique [cs.RO]. Université Rennes 1, 1998. tel-00843890

HAL Id: tel-00843890

https://theses.hal.science/tel-00843890

Submitted on 12 Jul 2013

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

HABILITATION À DIRIGER DES RECHERCHES

présentée devant

L'Université de Rennes 1
Institut de Formation Supérieure
en Informatique et en Communication

par

François Chaumette

De la perception à l'action : l'asservissement visuel, de l'action à la perception : la vision active.

soutenue le 28 janvier 1998 devant le jury composé de

MM.	Jean-Pierre Banâtre	Président
	Raja Chatila	Rapporteur
	Bernard Espiau	Rapporteur
	Giulio Sandini	Rapporteur
	Patrick Bouthemy	Examinateur
	Patrick Rives	Examinateur
	Chris Brown	Examinateur

Ce document de synthèse est le fruit de dix ans de travail au sein du projet TEMIS de l'IRISA. Cet institut, grâce notamment à sa composante INRIA dont je dépend, offre un exceptionnel environnement de travail de par ses moyens, son organisation, sa gestion, et l'ambiance qui y règne. Il en a été de même du projet TEMIS, qui vient de donner naissance aux projets TEMICS et VISTA, et je voudrais remercier chaleureusement les deux créateurs qui lui sont restés fidèles jusqu'à sa récente disparition : Claude Labit, qui a dirigé efficacement l'équipe depuis sa création tout en nous laissant une autonomie et des responsabilités appréciables et appréciées, et Patrick Bouthemy, mon principal et indispensable interlocuteur sur des sujets aussi divers et variés que, par exemple, le couplage de l'analyse du mouvement et de la vision active, et les résultats footballistiques des équipes bretonnes, principalement celle de Nantes bien sûr.

Mes activités de recherche ont été initiées sous la direction conjointe de Bernard Espiau et Patrick Rives, partis sous d'autres cieux plus ensoleillés en 1988, et s'inscrivent dans la suite naturelle de leurs travaux en commande référencée capteurs et vision dynamique. Je les remercie filialement pour le tremplin sur lequel ils m'ont accompagné, puis laissé m'envoler de mes propres ailes.

Je remercie paternellement les étudiants en thèse que j'ai ou ai eu le plaisir d'encadrer : Samia Boukir, Maître de Conférences à l'Université de La Rochelle, Farabi Bensalah, qui a ensuite obtenu un poste d'ATER à l'Université de Caen, Eric Marchand, qui nous a rejoint au sein du projet VISTA après un séjour post-doctoral à l'Université de Yale, Christophe Collewet du CEMAGREF Rennes, et enfin Armel Crétual et Ezio Malis, qui sont bien engagés dans leur troisième et dernière année de doctorat. Ce document n'existerait pas sans leur travail, ni sans celui de V. Sundareswaran et Nathalie Giordana avec lesquels j'ai travaillé lors de leur séjour post-doctoraux. Je remercie également Fabien Spindler pour l'aide inestimable apportée depuis trois ans dans le développement et la maintenance des deux cellules de vision robotique de l'IRISA. J'ai aussi bénéficié des discussions enrichissantes que j'ai eues au sein ou en dehors de la communauté Orasis : certains membres de l'IRISA, Eric Rutten notamment, de l'INRIA Sophia-Antipolis, du LIFIA/INRIA Rhône-Alpes, du LASMEA, du LAAS, de l'IFREMER, de la société EDIXIA et de la DER-EDF avec lesquels j'ai collaboré se reconnaîtront sans peine.

Je remercie également Raja Chatila, Giulio Sandini et Chris Brown qui m'ont fait l'honneur et le plaisir de participer au jury de mon habilitation, ainsi, bien sûr, que Jean-Pierre Banâtre.

Je ne sais si je dois remercier Sophie, Anna et Awen pour m'avoir permis de consacrer quelques trop nombreuses soirées et week-ends au couplage de la perception et de l'action en robotique.

Finalement, il est clair, pour ceux qui me connaissent, que je ne serais pas à l'IRISA depuis dix ans si cet institut était situé hors de Bretagne. Je rends grâce à qui de droit que ce ne soit pas le cas !

Table des matières

Présentation **5**

1 Les techniques d'asservissement visuel **7**
1.1 L'asservissement visuel 3D 7
1.2 L'asservissement visuel 2D 10
1.2.1 Modélisation 11
1.2.2 Gestion de la redondance par l'approche fonction de tâche . . 13
1.2.3 Poursuite de cible 16
1.3 L'asservissement visuel 2D 1/2 17
1.4 L'asservissement visuel $d2D/dt$ 19

2 La vision active **23**
2.1 Introduction 23
2.2 Stratégies locales de reconstruction 25
2.3 Stratégies globales d'exploration 27
2.3.1 Reconstruction et modélisation incrémentale 28
2.3.2 Exploration globale 30
2.3.3 Spécification et mise en œuvre sécurisée 31

Conclusion et perspectives **33**

Bibliographie **37**

Présentation

Nos travaux se situent à l'intersection des domaines de la robotique et de la vision par ordinateur. Ces deux domaines font l'objet de recherches fructueuses depuis de très nombreuses années et sont particulièrement intéressants par le spectre scientifique très large qu'ils recouvrent. A l'intérieur de ce spectre, nous nous sommes intéressés à la génération de cycles de perception et d'action, d'un point de vue commande d'une part (c'est-à-dire de la perception vers l'action - on parle alors d'**asservissement visuel**), et d'un point de vue perception d'autre part (c'est-à-dire de l'action vers la perception - on parle alors de **vision active**). Nous avons donc tenté, comme l'ensemble des chercheurs travaillant dans les domaines de la commande référencée capteurs et de la perception active, d'éviter un cloisonnement entre les phases de perception et d'action. On peut en effet rappeler que la perception 3D a longtemps été considérée comme un module préalable, indispensable et indépendant à tout module de génération de mouvement d'un robot dans un environnement imparfaitement connu. Les aspects de perception et d'action sont au contraire étroitement liés, comme nous le verrons par la suite.

Plus précisément, nous décrivons dans le premier chapitre de ce document les différentes techniques d'asservissement visuel qu'il est possible d'employer pour positionner une caméra par rapport à son environnement. Le capteur est alors considéré comme local et permet de réaliser des actions de type réflexe. Le second chapitre porte sur l'élaboration de stratégies de perception permettant d'atteindre une connaissance beaucoup plus riche de l'environnement.

Chapitre 1

Les techniques d'asservissement visuel

Nous présentons tout d'abord un regard critique sur les différentes techniques possibles pour réaliser des tâches de positionnement à l'aide d'une caméra embarquée sur un robot de type manipulateur. De nombreux travaux existent sur l'utilisation de caméras déportées [Allen 93, Dornaika 95, Hager 95]. Nous ne les décrirons pas outre mesure car la plupart des points évoqués ci-dessous s'appliquent également à cette configuration. Par ailleurs, l'utilisation de l'asservissement visuel en robotique mobile connaît depuis quelques années un essor important, notamment pour les applications de conduite automatique de véhicules [Dickmanns 91, Pissard-Gibollet 93, Khadraoui 96b, Tsakiris 97]. La difficulté essentielle réside alors dans l'élaboration de lois de commande prenant en compte les contraintes non holonomes de ce type de robots

De manière très classique, on peut classer les différentes techniques d'asservissement visuel en fonction des mesures utilisées en entrée du module de commande. Nous décrirons donc les techniques d'asservissement visuel 3D et 2D, qui ont fait l'objet de travaux depuis de longues années [Weiss 87, Espiau 92, Hashimoto 93, Hutchinson 96], mais aussi 2D 1/2 et $d2\mathrm{D}/dt$, qui sont des approches que nous avons introduites très récemment.

Une autre classification, transversale à la première, peut être effectuée en fonction du mode d'asservissement employé (boucle ouverte ou boucle fermée). Nous ne retiendrons et décrirons que les techniques basées sur la seconde approche, en raison du manque total de robustesse de la première vis-à-vis des erreurs de modélisation et de calibration du capteur et du manipulateur.

1.1 L'asservissement visuel 3D

Comme son nom l'indique, l'asservissement visuel 3D (ou *position-based visual servoing*) utilise en entrée de la boucle de commande des informations tri-

dimensionnelles exprimées dans un repère euclidien. Plus précisément, si l'on considère une tâche de positionnement rigide contrôlant les six degrés de liberté de la caméra (voir Figure 1.1), la consigne et la mesure sont exprimées sous la forme d'un changement de repère décrivant la position et l'orientation d'un repère lié à l'objet d'intérêt par rapport au repère lié à la caméra. On notera par la suite r, élément de $\mathbb{R}^6$, une représentation d'un tel changement de repère et $\hat{r}$ le résultat d'une estimation de r. La boucle de commande consiste alors à estimer à chaque itération le changement de repère $\hat{r}(t)$ pour la position courante de la caméra et à amener progressivement la caméra à sa position désirée $\hat{r}^*$ par rapport à l'objet.

On peut bien sûr aussi ne considérer qu'un sous-ensemble des six paramètres de r ou fixer des contraintes entre ces paramètres afin de réaliser des tâches contraignant un nombre inférieur de degrés de liberté entre la caméra et la scène considérée.

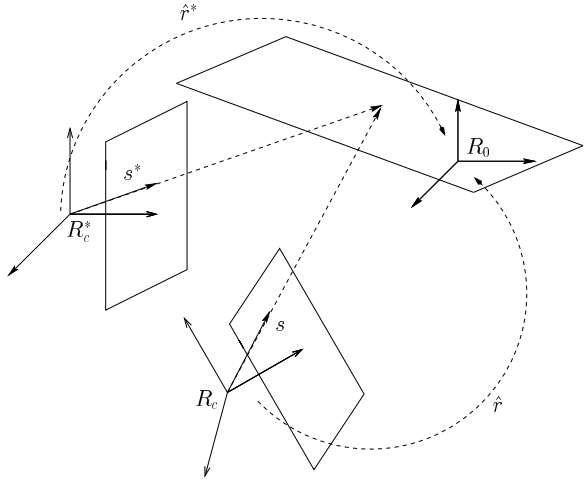

FIG. 1.1 – *Principe de l'asservissement visuel 3D*

En théorie, une seule itération est suffisante et une commande du robot en position permet d'atteindre la position désirée. En pratique évidemment, les diverses erreurs de modèle et de calibration nécessitent un rebouclage et un asservissement en vitesse pour assurer la stabilité du système [Wilson 96, Martinet 97].

De nombreuses méthodes existent pour estimer la position d'une caméra par rapport à un objet en utilisant une image de cet objet. Elles reposent très généralement sur la connaissance a priori d'un modèle 3D de l'objet et des paramètres intrinsèques de la caméra. Ces méthodes utilisent comme informations visuelles soit des points [Horaud 89, Haralick 89, Dementhon 95], soit des droites [Dhome 89], voire des ellipses résultant de la projection de cercles ou de sphères [Safaee-Rad 92], ou encore des objets cylindriques [Dhome 90]. Par contre, très peu de méthodes combinent des types différents d'informations visuelles (voir cependant [Phong 95] pour l'utilisation conjointe de points et de droites).

Des travaux existent par ailleurs, soit pour éviter d'utiliser un modèle 3D de l'objet (on retrouve là l'ensemble des travaux de reconstruction 3D par vision dy-

namique qui permettent d'estimer ce modèle et la localisation de la caméra à partir de la mesure du mouvement 2D et 3D), soit pour éviter d'utiliser des caméras calibrées [Morin 93], mais, à notre connaissance, ces diverses techniques n'ont jamais été appliquées en asservissement visuel 3D.

Concernant les aspects de calibration, il faut signaler qu'une calibration grossière est généralement suffisante pour assurer la convergence des lois de commande grâce à la boucle fermée employée. Deux cas de figure peuvent se présenter :

- soit la consigne $\hat{r}^*$ est calculée exactement de la même manière que les mesures $\hat{r}(t)$ (c'est-à-dire en utilisant l'image de l'objet acquise à partir d'une position d'apprentissage). On se retrouve alors dans la situation paradoxale où la consigne et les mesures sont notoirement fausses (plus exactement biaisées), mais où la position atteinte après convergence est correcte aux seules erreurs de mesure près.

- soit la consigne est fixée a priori. La position atteinte après convergence sera alors différente de celle désirée en raison du biais introduit dans l'estimation de $\hat{r}(t)$.

La situation est identique si le modèle 3D supposé de l'objet n'est pas parfaitement exact.

De plus, il est extrêmement difficile, voire même impossible, de déterminer des plages de variations de ces divers paramètres conservant la stabilité du système. En effet, une démonstration de la stabilité des techniques d'asservissement visuel 3D en présence d'erreurs de calibration ou de modèle semble, au jour d'aujourd'hui, hors de portée. Si l'on note s les informations visuelles utilisées pour estimer $\hat{r}$, il faut démontrer :

$$\left(\frac{\partial \hat{r}}{\partial s}\right) \cdot \left(\frac{\partial s}{\partial r}\right) > 0$$

Le deuxième terme de ce produit de matrices n'est rien d'autre que la matrice d'interaction associée à s sur lequel nous reviendrons en décrivant les techniques d'asservissement visuel 2D. En quelques mots, il représente la variation dans l'image des informations visuelles en fonction d'un mouvement de la caméra. Ce terme dépend des paramètres intrinsèques de la caméra. Le premier terme, lui, représente la variation de l'estimation de $\hat{r}$ en fonction d'une variation des informations visuelles. Ce terme, directement lié à la méthode d'estimation employée, dépend là encore des paramètres intrinsèques de la caméra, mais aussi du modèle 3D de l'objet. Dans le cas idéal, c'est-à-dire lorsque $\hat{r}$ correspond exactement à la position de la caméra par rapport à l'objet, le produit de matrices est égal à l'identité, ce qui assure la stabilité (et le découplage parfait) de la loi de commande. Malheureusement, en pratique, les erreurs de mesure dans l'image ajoutées aux erreurs de modèle et de calibration entraînent des perturbations qui peuvent parfois rendre le système instable. Une étude de la sensibilité aussi bien des méthodes d'estimation que des informations visuelles choisies (quant à leur type, mais aussi quant à leur disposition dans l'image et dans l'espace 3D) n'a encore jamais été entreprise, en raison principalement de

la complexité des calculs qu'elle impliquerait, ce qui fait que nous ne disposons d'aucun résultat théorique sur la stabilité des méthodes d'asservissement visuel 3D (nous verrons malheureusement que, même si la complexité des calculs est moindre en asservissement visuel 2D, elle reste suffisante pour n'avoir jusqu'à présent permis d'obtenir que très peu de résultats théoriques).

Il est bien connu que, si on utilise trois points, il existe quatre solutions distinctes au problème de l'estimation de la position de la caméra (on parle alors du problème P3P pour *Persective from 3 Points*). A moins d'introduire une heuristique qui limite grandement l'espace de convergence (à savoir choisir la solution la plus proche de la position désirée, ce qui nécessite en outre de connaître cette position désirée), il faut donc utiliser au moins quatre points pour aboutir à une solution unique. Cependant, comme la plupart des problèmes inverses, le problème P4P est parfois instable. En effet, si l'on considère par exemple quatre points coplanaires, une très faible variation de la position des quatre points dans l'image peut entraîner une très forte variation dans l'estimation de $\hat{r}$ (ce qui implique que la matrice $\left(\frac{\partial \hat{r}}{\partial s}\right)$ est très mal conditionnée). Une instabilité de la commande peut donc apparaître en raison de simples erreurs de mesure dans l'image. Il semble que de tels problèmes n'apparaîssent pas en pratique si les points choisis ne sont pas coplanaires [Wilson 96]. Cela reste malheureusement, comme nous l'avons déjà dit, à démontrer.

Finalement, l'asservissement visuel 3D présente l'avantage de fournir des mesures directement dans l'espace euclidien, ce qui permet l'emploi de lois de commande extrêmement simples pour aller itérativement d'une position à une autre, voire une certaine planification pour éviter par exemple les butées et singularités du robot ou des obstacles connus de la scène. L'inconvénient qui en résulte est qu'aucun contrôle véritable dans l'image n'est effectué, ce qui implique que l'objet d'intérêt peut très bien sortir du champ de vue de la caméra pendant son déplacement, entraînant alors l'arrêt de la loi de commande. Un tel phénomène peut se produire pour des erreurs importantes introduites dans les paramètres intrinsèques de la caméra, mais aussi dans le modèle géométrique du robot ou le passage effecteur-caméra.

1.2 L'asservissement visuel 2D

Les techniques d'asservissement visuel 2D (*image-based visual servoing*) utilisent, elles, directement les informations visuelles s extraites de l'image, c'est-à-dire qu'elles ne nécessitent pas la phase d'estimation de $\hat{r}$ contrairement à l'asservissement visuel 3D. Les lois de commande consistent alors à contrôler le mouvement de la caméra afin que les mesures dans l'image $s(t)$ atteignent une valeur désirée s^*, voire suivent une trajectoire spécifiée $s^*(t)$ (voir Figure 1.2).

Notre contribution à ce domaine est importante. Elle porte tout d'abord sur l'élaboration d'une méthode de calcul explicite de la matrice d'interaction associée aux informations visuelles géométriques [Chaumette 90, Espiau 92]. Cette matrice, notée L, joue un rôle essentiel dans l'élaboration des diverses lois de commande possibles. Elle est également appelée Jacobien de l'image par analogie évidente avec

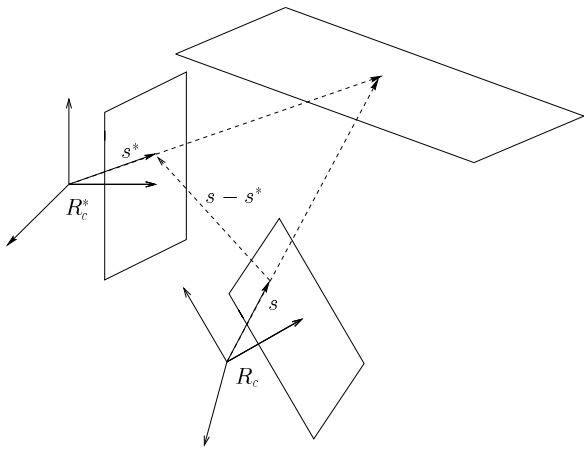

FIG. 1.2 – *Principe de l'asservissement visuel 2D*

le Jacobien d'un robot. Plus précisément, L est défini par $L = \left(\frac{\partial s}{\partial r}\right)$, soit encore :

$$\dot{s} = L(s, z)\ T$$

où $T = \frac{dr}{dt}$ est la vitesse relative entre la caméra et l'objet considéré. On peut démontrer que, si l'on considère des informations de type géométrique, L dépend de la valeur des informations visuelles s, mais aussi des paramètres intrinsèques de la caméra, et des paramètres z décrivant la configuration 3D de l'objet (en l'occurrence, la position de la surface des limbes dans le cas d'un objet volumétrique).

On peut signaler que la méthode de calcul de la matrice d'interaction que nous avons développée s'est généralisée simplement à l'emploi d'un capteur constitué d'une caméra couplée à des plans lasers [Motyl 92, Khadraoui 96a].

1.2.1 Modélisation

L'obtention de la matrice d'interaction pour les primitives géométriques les plus courantes a permis d'effectuer une classification des différentes tâches robotiques réalisables par asservissement visuel 2D en fonction des liaisons virtuelles souhaitées entre la caméra et les objets considérés [Chaumette 93a]. De plus, et contrairement à l'approche précédente, il est immédiat de considérer au sein d'une même tâche des informations visuelles de type différent (telles la surface d'une ellipse et l'orientation d'une droite, ou la distance d'un point à une droite), à partir du moment où la matrice d'interaction qui leur est associée est connue. Certains auteurs ont préféré utiliser des réseaux de neurones [Suh 93] ou une estimation en ligne [Hosoda 94, Jägersand 97] pour déterminer L. Cependant, comme nous allons le voir, il est alors impossible de démontrer la stabilité des lois de commande associées.

En effet, l'asservissement visuel 2D consiste schématiquement en la régulation de la fonction de tâche suivante :

$$e = \hat{L}^{+}(s - s^*)$$

où $\hat{L}^+$ est la pseudo-inverse d'un modèle ou d'une approximation de L. Pour assurer la stabilité et la convergence de e, on dispose de la condition suffisante suivante :

$$\hat{L}^+ . L > 0$$

c'est-à-dire que $\hat{L}$ doit être suffisamment correct et proche de L pour ne pas trop perturber le système (puisque si $\hat{L} = L$, le produit de matrice est égal à l'identité).

Concernant l'influence de la calibration, les expérimentations présentées dans [Espiau 93] indiquent une forte robustesse des lois de commande par rapport à des variations des paramètres intrinsèques. Par contre, et excepté pour des cas très simples (comme le contrôle de l'orientation de la caméra pour qu'un objet soit situé au centre de l'image), la complexité des calculs fait qu'il est quasiment impossible de démontrer la positivité du produit de matrices ci-dessus dans les cas les plus généraux.

Concernant la valeur choisie pour les informations visuelles et leur profondeur respective, deux choix sont couramment utilisés :

- $\hat{L} = L(s(t), \hat{z}(t))$, c'est-à-dire que l'on calcule à chaque itération la valeur courante de la matrice d'interaction. Une estimation des paramètres z doit alors être réalisée en ligne, soit en utilisant les techniques de reconstruction 3D par vision dynamique, soit par les méthodes d'estimation décrites précédemment si l'on dispose d'un modèle 3D de l'objet.

- $\hat{L} = L(s^*, z^*)$, c'est-à-dire que la matrice choisie est constante et correspond à la configuration désirée. Une valeur de profondeur à la position finale, même très approximative, est alors nécessaire.

Ces deux possibilités présentent chacune un certain nombre d'avantages, mais aussi d'inconvénients [Chaumette 98] :

Le choix $\hat{L} = L(s(t), \hat{z}(t))$ contraint fortement la trajectoire que doivent suivre dans l'image les informations visuelles pour atteindre leur but. Par exemple, si l'on choisit des points, ce choix implique que ceux-ci doivent aller simultanément en ligne droite vers leur position désirée (voir Figure 1.3.a). Cependant, ce comportement souhaitable dans l'image peut parfois entraîner des mouvements de la caméra inadéquats (correspondant à un comportement en singularité), voire même impossible à réaliser. La caméra peut ainsi atteindre un minimum local de l'erreur dans l'image à partir duquel il est impossible de rejoindre le minimum global correspondant à une erreur nulle (voir Figure 1.4).

Le choix $\hat{L} = L(s^*, z^*)$, lui, ne contraint quasiment pas la trajectoire dans l'image (voir Figure 1.3.b), le comportement exponentiel découplé et la positivité de $\hat{L}^+ . L$ n'étant assurés que dans un voisinage de la position désirée. Du coup, et même si nous ne pouvons pas l'expliquer, il s'avère en pratique que la commande est alors beaucoup moins sujette aux problèmes des mouvements irréalisables et des minima locaux, la trajectoire dans l'image calculée par la loi de commande permettant d'éviter ces configurations défavorables (voir Figure 1.5). Par contre, il est tout à fait possible que l'objet sorte du champ de vue de la caméra au cours du positionnement (pour y rentrer, mais trop tard, ensuite...).

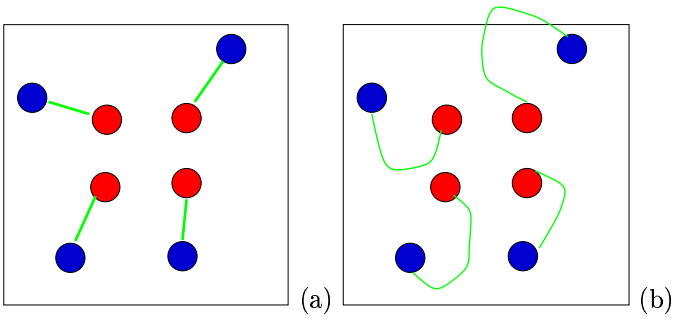

FIG. 1.3 – *a) trajectoire attendue dans l'image en choisissant* $\hat{L} = L(s(t), \hat{z}(t))$, *b) trajectoire possible dans l'image en choisissant* $\hat{L} = L(s^*, z^*)$

Finalement, dans les deux cas, il est possible de rencontrer une singularité de la matrice d'interaction [Michel 93, Chaumette 98] ou, pour le moins, des positions où son conditionnement est très faible, entraînant là encore soit une instabilité de la commande, soit un échec dans la convergence du système.

Ces différents problèmes théoriques, qui peuvent parfois avoir des répercussions pratiques, ne doivent pas faire oublier que l'asservissement visuel 2D apporte toute satisfaction dans grands nombres de cas. Nous avons notamment, dans le cadre d'un transfert technologique avec la société EDIXIA pour la DER-EDF Chatou, appliqué ces techniques à des tâches de préhension d'outils et de vissage qui ont été implémentées aussi bien sur le robot de l'IRISA (voir Figure 1.6) que sur plusieurs robots de la DER-EDF à des fins de tâches de manipulation en environnement nucléaire. Nous pensons que les problèmes évoqués ci-dessus sont principalement dûs à des problèmes de représentation, le choix de coordonnées de points ne paraîssant pas optimal pour représenter un objet. Il n'empêche que la recherche d'une représentation assurant une convergence à tout coup quelle que soit la position initiale choisie reste un problème largement ouvert.

1.2.2 Gestion de la redondance par l'approche fonction de tâche

L'intégration de l'asservissement visuel 2D dans l'approche générale de la fonction de tâche [Samson 91] permet de résoudre de manière efficace et élégante les problèmes de redondance que l'on rencontre lorsqu'une tâche visuelle ne contraint pas l'ensemble des degrés de liberté de la caméra [Chaumette 90, Espiau 92]. Il est alors possible d'exécuter parallèlement à la tâche visuelle des tâches secondaires. Encore faut-il combiner adéquatement ces deux tâches, ce qui n'a pas été fait dans [Nelson95] pour un problème d'évitement des butées et singularités. Dans cet article en effet, les deux tâches sont purement et simplement ajoutées, ce qui fait que la commande résultante perturbe la réalisation de la tâche visuelle.

Pour notre part, nous avons tout d'abord appliqué l'approche fonction de tâche,

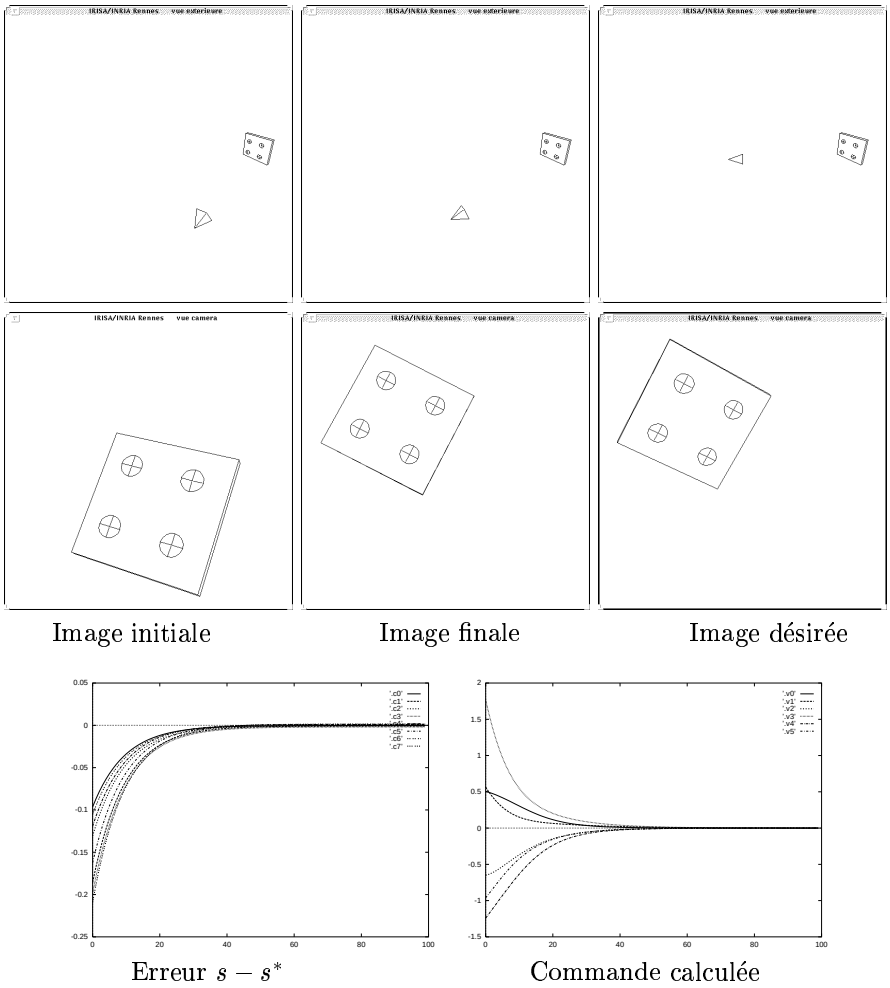

FIG. 1.4 – *Convergence vers un minimum local en utilisant* $\hat{L} = L(s(t), \hat{z}(t))$

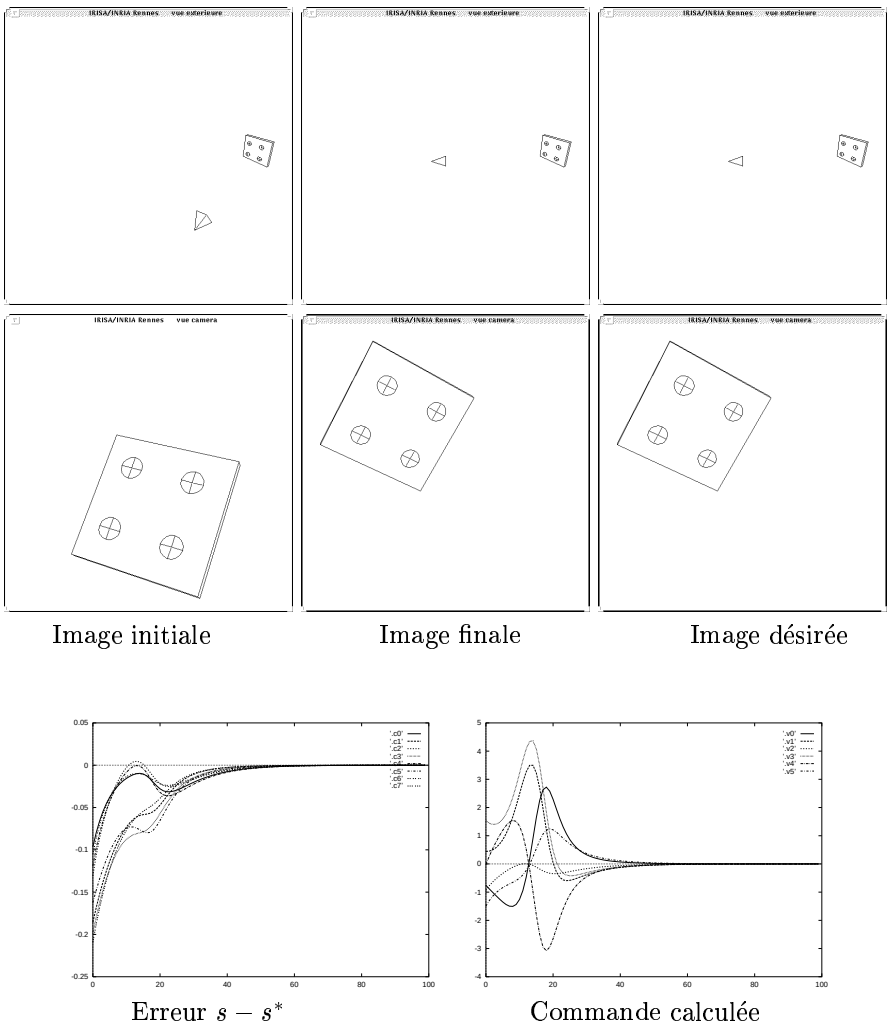

FIG. 1.5 – *Convergence vers le minimum global en utilisant* $\hat{L} = L(s^*, z^*)$

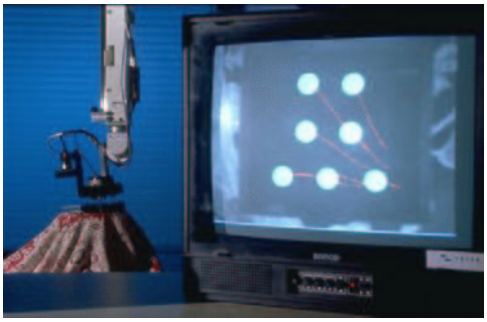

FIG. 1.6 – *Préhension d'un objet marqué par asservissement visuel 2D*

qui consiste à ne considérer dans la tâche secondaire que les combinaisons de degrés de liberté qui ne sont pas contraints par la tâche principale, à des opérations de suivi de trajectoire [Chaumette 94] (voir Figure 1.7). Celles-ci peuvent être utiles dans des applications d'inspection et de contrôle de conformité, mais aussi pour la reconstruction 3D des objets considérés. Nous reviendrons sur ce point dans le chapitre suivant. Nous avons également considéré le problème d'évitement des butées et singularités [Marchand 96c]. Il nous reste à déterminer des techniques de réglage automatique de gain similaires à celles employées dans [Chan 95] pour terminer notre étude sur ce sujet.

Nous avons également, en collaboration avec le LIFIA, considéré le problème du contrôle d'attention d'une tête stéréoscopique active. L'approche fonction de tâche s'est, ici aussi, révélée fructueuse pour assurer la symétrie de vergence de la tête [Crowley 95].

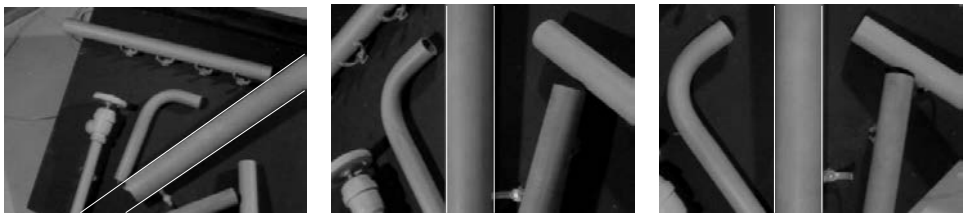

FIG. 1.7 – *Positionnement et rotation autour d'un cylindre.*

1.2.3 Poursuite de cible

Aussi surprenant que cela puisse paraître, une bonne partie des travaux en asservissement visuel traitant de la poursuite de cible considère le mouvement de l'objet comme une perturbation négligeable [Papanikolopoulos 93, Nelson95] ou suppose connue la trajectoire de l'objet [Allen 93, Hashimoto 95]. Nous faisons bien sûr abstraction ici de l'ensemble des travaux ayant trait au contrôle d'attention des têtes stéréoscopiques actives où la compensation du mouvement de la cible est

le problème essentiel. On pourra par exemple consulter à ce propos [Brown 90, Milios 93, Murray 95, Viéville 95, Brunnström 96]. Nous avons pour notre part utilisé tout d'abord un simple intégrateur dans la commande pour supprimer les erreurs de traînage en présence d'un objet à vitesse constante [Chaumette 91]. En raison de la difficulté de réglage du gain intégral et du manque de généricité de cette approche, nous avons ensuite proposé un estimateur de la vitesse de l'objet dans l'image, basé sur le mouvement global et la mesure du mouvement de la caméra. Cette estimation est ensuite filtrée par un filtre de Kalman à accélération constante et bruit corrélé avant d'être introduit dans la commande, aboutissant alors à un système de commande adaptatif indirect [Chaumette 93b]. Une approche similaire est décrite dans [Corke 93], basée sur un simple modèle d'état à vitesse constante. Plus récemment, dans le cadre de la thèse de Farabi Bensalah [Bensalah 95, Bensalah 96], nous avons utilisé l'algorithme dit du GLR (*Generalized Likelihood Ratio*) [Willsky 76, Basseville 83] afin de détecter, d'estimer et de compenser d'éventuels sauts en direction ou en amplitude de la vitesse de l'objet. Pour achever cette étude, nous pensons nous inspirer des techniques de reconnaissance de trajectoires, basées par exemple sur les filtres de Kalman multimodèles [Bar-Shalom 93] pour sélectionner au mieux le modèle d'état correspondant au mouvement effectif.

1.3 L'asservissement visuel 2D 1/2

Nous décrivons à présent succinctement une nouvelle approche d'asservissement visuel que nous avons développée dans le cadre de la thèse d'Ezio Malis en collaboration avec la DER-EDF Chatou [Malis 98]. Elle nous semble particulièrement originale et intéressante puisqu'elle ne présente pas la plupart des inconvénients respectifs des deux approches précédemment présentées. Nous avons qualifié cette approche d'asservissement visuel 2D 1/2 car les informations utilisées comme mesures et consignes sont, pour certaines d'entre elles, exprimées directement dans l'image et, pour les autres, exprimées dans le repère de la caméra.

Plus précisément (voir Figure 1.8), cette approche est basée sur l'estimation de l'homographie qui relie l'image de trois points [Faugeras 93, Couapel 94, Boufama 95]. Dans le cas d'un objet plan, un minimum de quatre points appariés sur les images courante et désirée permet une telle estimation par la résolution d'un simple système linéaire. Dans le cas d'un objet quelconque, l'appariement de huit points permet d'aboutir, là-encore par la résolution de systèmes linéaires. A partir de l'homographie obtenue, il est possible de calculer le déplacement en rotation que la caméra doit effectuer pour atteindre sa position spécifiée, ainsi que la direction de son déplacement en translation. L'homographie fournit également le rapport d/d^* entre les distances courante et désirée de la caméra à l'objet. Comme la translation à réaliser n'est connue qu'à un facteur d'échelle près, un asservissement visuel purement 3D est impossible. Cependant, la combinaison des informations 2D et 3D disponibles permet d'aboutir à une solution aux propriétés particulièrement intéressantes.

En effet, l'erreur en rotation estimée est controlée directement avec les trois degrés de liberté en rotation du robot. De plus, le mouvement le long de l'axe optique

est contraint de manière à ce que le rapport des profondeurs entre la caméra et l'objet tende vers 1. Finalement, les deux degrés de liberté de translation parallèlement au plan image sont utilisés pour garder la cible au centre de l'image à l'aide, par exemple, des coordonnées d'un point de l'objet (voir Figure 1.9).

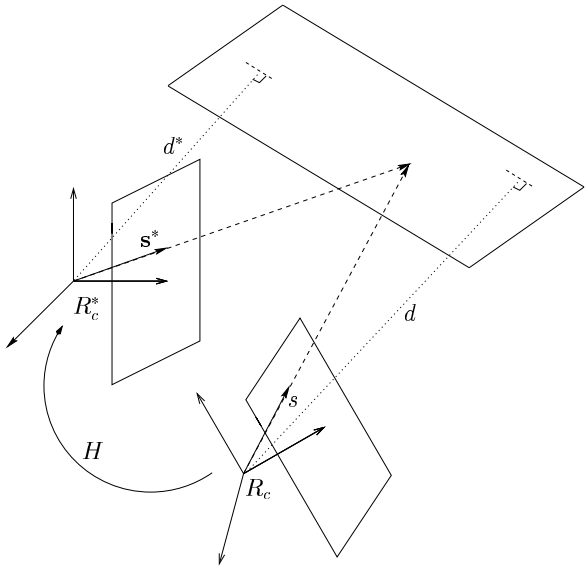

FIG. 1.8 – *Principe de l'asservissement visuel 2D 1/2*

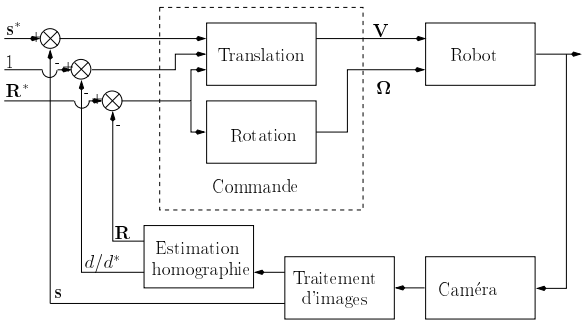

FIG. 1.9 – *Schéma bloc de l'asservissement visuel 2D 1/2*

On peut ainsi séparer les boucles d'asservissement en rotation et en translation de la caméra, ce qui permet :

- un fort découplage de la loi de commande ;
- un contrôle partiel dans l'image permettant de conserver l'objet en permanence dans le champ de vision de la caméra ;

– une étude de la stabilité et du domaine de convergence de la loi de commande. On a ainsi pu obtenir des conditions suffisantes de stabilité prenant en compte les erreurs de calibration et les erreurs d'estimation de l'homographie qui en découlent afin que le domaine de convergence soit constitué de l'ensemble de la demi-sphère surplombant l'objet.

Ainsi, pour des positions initiales de la caméra très éloignées de la position désirée (voir Figure 1.10), la convergence est obtenue là où l'asservissement visuel 2D échoue.

FIG. 1.10 – *Positionnement pas asservissement visuel 2D 1/2 : (a) image désirée, (b) image initiale, (c) trajectoire dans l'image*

L'avantage majeur de cette approche est que la connaissance d'un modèle géométrique 3D de l'objet n'est plus nécessaire, ce qui lui donne un domaine d'application très important à partir du moment où le motif à atteindre est préalablement déterminé, par exemple lors d'une phase d'apprentissage. La seule information 3D utilisée dans la commande est la profondeur désirée approximative d'un point de l'objet, soit un besoin d'informations a priori bien moindre qu'en asservissement visuel 3D et 2D. L'utilisation de droites comme informations visuelles pour estimer l'homographie est également possible dans le cas d'un objet planaire. Par contre, l'utilisation de primitives plus complexes (telles des ellipses par exemple) semble plus délicate.

1.4 L'asservissement visuel $d2D/dt$

L'ensemble des méthodes que nous avons décrit jusqu'à présent reposait sur l'utilisation d'informations visuelles géométriques (coordonnées de points, paramètres représentant l'image d'une droite, etc). Les contraintes imposées sont donc que ces primitives géométriques existent dans la scène, mais surtout qu'il est possible de les extraire et de les suivre par traitement d'image à cadence élevée afin de conserver la robustesse et la stabilité des lois de commande. Actuellement, peu de systèmes fiables existent et sont limités soit au suivi de segments [Crowley 88, Deriche 90, Boukir 98], soit au suivi de points caractéristiques ou de formes simples [Rives 93, Hager 98]. En pratique, on utilise le plus souvent des marqueurs disposés sur la scène observée afin d'extraire les positions de leur projection perspective dans l'image. Des travaux récents ont permis de supprimer ces contraintes fortes par l'emploi

d'informations visuelles dynamiques [Questa 95, Colombo 95, Sundareswaran 96, Santos-Victor 97]. Le principe de la commande consiste alors à contrôler les mouvements de la caméra de telle sorte que le mouvement 2D mesuré atteigne un champ de vitesse désiré, d'où l'appellation d'asservissement visuel $d2\mathrm{D}/dt$ (voir Figure 1.11).

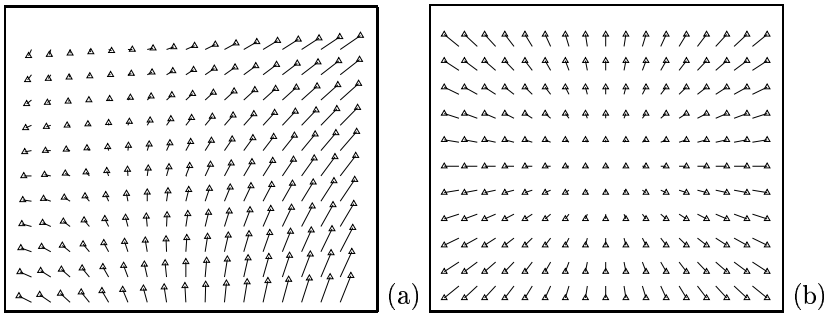

FIG. 1.11 – *Exemple synthétique de champs de vitesse mesuré (a) et désiré (b).*

Les algorithmes d'analyse du mouvement 2D, génériques quant au contenu de la séquence d'images, peuvent en effet être employés et fournir les paramètres d'un modèle choisi du mouvement 2D, à l'issue ou non d'une segmentation au sens du mouvement [Odobez 94]. Plus précisément, si l'on considère un modèle quadratique particulier de la vitesse d'un point de coordonnées (x, y) dans le plan image :

$$\left\{ \begin{array}{lcl} \dot{x} & = & a_1 + a_2 x + a_3 y + b_1 x^2 + b_2 xy \\ \dot{y} & = & a_4 + a_5 x + a_6 y + b_2 y^2 + b_1 xy \end{array} \right.$$

la relation entre cette vitesse 2D et le mouvement relatif $T = (V, \Omega)$ entre la caméra et le point considéré est donnée par [Subarrao 86, Negahdaripour 92] :

$$\left\{ \begin{array}{ll} a_1 = -V_x/Z_p - \Omega_y & a_4 = -V_y/Z_p + \Omega_x \\ a_2 = \gamma_1 V_x/Z_p + V_z/Z_p & a_5 = \gamma_1 V_y/Z_p - \Omega_z \\ a_3 = \gamma_2 V_x/Z_p + \Omega_z & a_6 = \gamma_2 V_y/Z_p + V_z/Z_p \\ b_1 = -\gamma_1 V_z/Z_p - \Omega_y & b_2 = -\gamma_2 V_z/Z_p + \Omega_x \end{array} \right.$$

où la surface locale au point considéré est décrite par $Z = Z_p + \gamma_1 X + \gamma_2 Y$. Le principe de la commande consiste alors à sélectionner les paramètres nécessaires à la réalisation de la tâche donnée et à établir une loi de commande afin que ces paramètres atteignent une valeur désirée.

Par ailleurs, l'emploi de ces nouvelles informations visuelles permet d'appréhender simplement des tâches qu'il était extrêmement délicat à modéliser en utilisant des informations visuelles géométriques. Ainsi, dans le cadre du séjour post-doctoral de V. Sundareswaran, l'alignement de l'axe optique d'une caméra, montée sur l'effecteur d'un robot, avec la direction de translation de ce robot a été réalisé en considérant comme informations visuelles les termes a_1 et a_4 [Sundareswaran 96]. On peut également utiliser les techniques classiques d'asservissement visuel 2D basé sur les coordonnées du foyer d'expansion (FOE) obtenu là encore par analyse du mouvement 2D.

Dans le cadre de la thèse d'Armel Crétual, nous avons poursuivi ces travaux pour réaliser une tâche de positionnement telle que le plan image de la caméra soit parallèle au plan d'un objet observé [Crétual 97]. La réalisation de cette tâche revient à annuler les termes b_1 et b_2 du développement quadratique du mouvement projeté dans l'image. Sous l'hypothèse d'une translation non nulle le long de l'axe optique nous garantissant d'obtenir des mesures significatives, on peut alors obtenir une relation linéaire entre la dérivée temporelle de ces paramètres et la rotation de la caméra orthogonalement à son axe optique. Une commande en boucle fermée sur ces informations peut alors être établie. L'utilisation des seuls degrés de liberté en rotation amenant le risque de perte de l'objet dans l'image, cette commande a été complétée par une tâche de fixation qui contraint les deux degrés de liberté de translation parallèle au plan image. Cette tâche supplémentaire permet de conserver toujours le même point au centre de l'image, sans mesurer explicitement la position de ce point dans l'image. Sur la Figure 1.12 sont représentées les images initiale et finale d'une expérimentation. On peut remarquer la difficulté qu'il aurait fallu surmonter pour extraire des informations géométriques pertinentes capables de réaliser cette tâche.

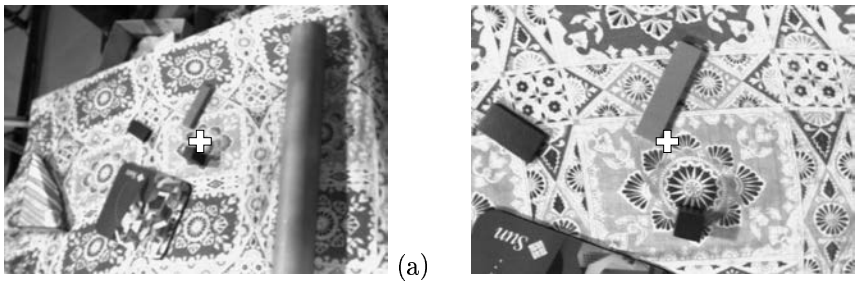

FIG. 1.12 – *Images initiale (a) et finale (b) pour une tâche de fixation et de positionnement de la caméra parallèlement à un plan*

Dans un cadre plus général, la relation linéaire entre les variations de l'information visuelle et la vitesse de la caméra, représentée par la matrice d'interaction pour des informations visuelles géométriques, n'est généralement plus valide, en particulier en raison de l'apparition de termes liés à l'accélération de la caméra et de termes quadratiques. Nous travaillons donc actuellement à l'élaboration de lois de commande en accélération, ou, si besoin, combinant accélération et vitesse.

Il faut finalement signaler que, même si des progrès énormes ont été faits récemment, les algorithmes de traitements d'images utilisés sont encore relativement coûteux en temps de calcul si l'on considère un modèle affine du mouvement (et encore davantage pour un modèle quadratique). Ils fournissent en outre des résultats beaucoup plus bruités que ceux employés pour extraire des informations visuelles géométriques. A des cadences de l'ordre du hertz, la validation expérimentale des techniques d'asservissement visuel $d2\mathrm{D}/dt$ est donc particulièrement délicate.

Par ailleurs, il est possible de remonter à des informations visuelles géométriques

par simple intégration des paramètres de mouvement estimés (si une localisation initiale est disponible). On peut alors utiliser directement les lois de commande qui ont été développées en asservissement visuel 2D. Cette approche a été validée sur une tâche de poursuite d'un objet quelconque (tel un piéton par exemple, voir Figure 1.13) en contrôlant l'orientation de la caméra [Crétual 98]. La position du centre de gravité de l'objet peut être estimée à partir des seuls paramètres constants du modèle du mouvement 2D, ce qui a permis une implémentation d'un algorithme d'estimation robuste et multi-résolation [Odobez 94] à la cadence video sur station de travail (en l'occurrence une Ultra Sparc). On peut signaler que nous n'avons pas rencontré les problèmes de dérive qui surviennent généralement par intégration. Cela indique que l'estimation fournie n'est pas biaisée et adéquate pour réaliser la tâche choisie.

FIG. 1.13 – *Suivi d'un piéton par intégration de son mouvement 2D estimé : visualisation d'une image sur dix (cadence de traitement : 25 Hz)*

Chapitre 2

La vision active

Nous présentons dans ce chapitre notre contribution au domaine de la vision active. Elle porte sur l'élaboration de stratégies de perception pour la recontruction optimale et l'exploration complète de scènes statiques simples à l'aide d'une caméra embarquée sur un robot manipulateur. L'originalité de notre travail porte, d'une part, dans l'utilisation des techniques d'asservissement visuel 2D pour générer les mouvements de la caméra dans les principales phases de la reconstruction et de l'exploration et, d'autre part, dans la réalisation d'un système complet reliant étroitement des stratégies dites de bas-niveau (la reconstruction d'une primitive en tant que telle) à des stratégies de haut niveau (calcul de points de vue, prédiction/vérification d'hypothèses pour la gestion des occlusions et l'obtention d'une représentation de type facettes). Nous évoquons également à la fin de ce chapitre les aspects essentiels et transversaux de spécification et de mise en œuvre sécurisée des systèmes de vision robotique.

2.1 Introduction

Le concept de la vision active date d'une dizaine d'années seulement [Bajcsy 88, Aloimonos 87, Aloimonos 90, Ballard 91, Brown 92]. Il consiste à élaborer des stratégies de perception et d'action afin de résoudre les problèmes soulevés lors de la conception de systèmes de vision par ordinateur, à savoir principalement leur sensibilité au bruit, leur faible précision et surtout leur manque de réactivité. La vision active a en effet pour but d'améliorer, en contrôlant les paramètres du capteur, la qualité de la perception par rapport à l'approche passive classique, où l'on se restreint à observer, mesurer et interpréter les données issues du capteur. D'un point de vue anthropomorphique, la vision active peut être vue comme une tentative de simulation du système visuel animal et humain visant à approcher leurs facultés d'adaptation.

On peut dire que la vision active est née d'un constat d'échec dans les tentatives passées de réaliser des solutions universelles aux problèmes de perception. Par principe, les travaux réalisés dans le domaine de la vision active sont moins ambitieux,

puisque dédiés à une application ou un but précis (la vision est alors intentionnelle) dans un cadre déterminé. Il semble évident que les stratégies à élaborer sont différentes d'une application à l'autre. Ainsi, si l'on cherche à reconstruire un objet pour une application de préhension ou d'évitement d'obstacles, la différence de précision des résultats souhaités entrainera l'utilisation de méthodes plus ou moins élaborées et contraignant plus ou moins les mouvements du capteur, pour atteindre, et atteindre seulement, la précision demandée. Les stratégies dépendent également très fortement des ressources disponibles. Par exemple, si l'on utilise un robot manipulateur, ou au contraire un robot mobile, comme moyen de déplacement du capteur, les stratégies pour résoudre le même problème seront généralement très différentes en raison des problèmes d'odométrie que l'on rencontre en robotique mobile et qui sont quasiment inexistants pour la plupart des robots manipulateurs.

L'inconvénient majeur de la vision active est donc l'absence de généricité des travaux qui découlent de ce concept pourtant générique. Cependant, des classes de méthodes, comme par exemple les techniques d'asservissement visuel ou de prédiction et vérification d'hypothèses, semblent particulièrement bien adaptées à la vision active. C'est pourquoi nous avons tenté de suivre une méthodologie qui devrait permettre de considérer une vaste classe de problèmes liés à la réalisation de tâches robotiques en environnement statique ou dynamique à l'aide d'informations visuelles. Cette méthodologie repose sur les trois relations suivantes :

- *la relation entre le global et le local.* Généralement, une tâche est définie de manière globale par son but alors que les informations disponibles pour y parvenir sont locales. La relation entre cette modélisation globale et cet ensemble de sous-modèles locaux doit donc être étudiée afin de réaliser la tâche spécifiée.

  Dans le cadre de la reconstruction et de l'exploration d'une scène, nous avons géré cette dualité par le développement de stratégies locales pour la reconstruction et globales pour l'exploration. Nous avons également considéré le passage d'une modélisation locale de la scène (en termes de primitives simples et éparses) à une modélisation globale (en termes de facettes) à l'aide de réseaux Bayesiens. Finalement, le calcul de points de vue apportant un gain d'information permet d'aboutir à une reconstruction complète de la scène. Réciproquement, en considérant l'information disponible dans son ensemble, il est possible de résoudre des problèmes locaux telles les occultations ou la focalisation sur des zones d'intérêt.

- *la relation entre le continu et le discret.* Cet aspect du problème est très fortement relié au précédent. Si la mise en œuvre des tâches élémentaires s'appuie le plus souvent sur des méthodes continues, comme les lois de commandes qui contrôlent les mouvements du capteur, l'enchaînement des différentes tâches menant à la réalisation de la tâche nominale repose sur la manipulation d'informations logiques, temporelles, etc.

  Cette relation a été étudiée tant du point de vue algorithmique que du point de vue génie logiciel. Dans notre cas, les mouvements de la caméra sont gérés soit de manière continue, en utilisant l'asservissement visuel 2D, soit de ma-

nière discrète, en utilisant un repositionnement à une position précédemment occupée ou les techniques de calcul de points de vue.

- *la relation entre la perception et l'action.* Le point fondamental de la vision active est la relation existant entre le mouvement du capteur et les informations perçues pendant ce mouvement. L'information permet de guider le capteur dans son déplacement lorsque le déplacement sert à acquérir l'information. Cette boucle de rétroaction se retrouve à tous les niveaux du système que nous avons proposé, tant aux niveaux local que global, continu que discret.

2.2 Stratégies locales de reconstruction

De nombreux travaux menés en vision par ordinateur se sont fixés pour objectif la réalisation de systèmes capables d'accéder à la géométrie spatiale d'une scène à partir de son observation par une caméra mobile. Ces systèmes doivent fournir une description 3D géométrique claire et complète de la scène à partir d'une séquence d'images 2D.

L'observabilité du mouvement de la caméra qui est souvent essentielle en reconstruction 3D caractérise un domaine de recherches appelé *vision dynamique*. Deux principales approches ont été mises en œuvre dans la littérature pour traiter ce problème :

- l'approche discrète, dans le sens où le mouvement de la caméra est exprimé sous la forme d'un déplacement et le mouvement dans l'image sous la forme de disparité. La plupart des travaux porte sur la reconstruction de points ou de droites mis en correspondance [Chien 89, Crowley 92, Weng 92, Faugeras 93, Zhang 95, Christy 96] en raison de la complexité des calculs nécessaires à la reconstruction d'objets plus complexes [Safaee-Rad 92, Zhao 94].

- l'approche continue, dans le sens où le mouvement de la caméra est alors exprimé sous forme de vitesse, de même que le mouvement perçu dans l'image. Les images doivent alors être acquises à une cadence proche de la cadence vidéo. On peut alors obtenir une reconstruction dense, mais généralement très bruitée sur les zones de fortes variations de la profondeur, à partir des dérivées spatio-temporelles de l'intensité lumineuse [Negahdaripour 87, Huang 91] ou d'une estimation du mouvement 2D [Waxman 87, Adiv 89, Matthies 89, Vernon 90, Sandini 90]. On peut également remonter à une information locale de la surface à partir du suivi des contours d'occultation [Cipolla 92, Kutulakos 94a, Cipolla 95, Boyer 96]. Par ailleurs, on peut obtenir une reconstruction éparse si l'on ne s'intéresse qu'à des primitives particulières. Dans ces deux derniers cas, la détermination du flot optique n'est pas à réaliser et, comme les approches continues impliquent de faibles mouvements dans l'image, la mise en correspondance est fortement simplifiée et se ramène au suivi 2D de contours ou de primitives (*token trackers*).

Nos travaux sur ce thème, effectués dans le cadre de la thèse de Samia Boukir [Boukir 93, Chaumette 96], se situent dans la suite logique de ceux déjà effectués

à l'IRISA sur la reconstruction de points et de segments [Xie 89]. Nous avons en effet généralisé l'approche continue employée aux primitives géométriques paramétrables les plus classiques. En se plaçant dans un contexte de vision active, nous avons pu en outre déterminer les mouvements de la caméra assurant une reconstruction optimale des primitives 3D en termes de suppression des erreurs de discrétisation et de minimisation des erreurs de mesure.

La méthode que nous avons développée repose sur l'expression analytique de la matrice d'interaction associée à la primitive considérée. Plus précisément, comme la matrice d'interaction dépend des paramètres P_L représentant la configuration de la surface des limbes, il est possible d'estimer itérativement leur valeur. On obtient :

$$\widehat{P_L} = \widehat{P_L}(p, \dot{p}, T_c)$$

où p est l'ensemble des paramètres décrivant la position de la primitive dans l'image, et où $\dot{p}$ est la vitesse de p induite par la vitesse T_c de la caméra. Il est ensuite possible de remonter aux paramètres P qui décrivent la configuration de la primitive à partir de sa position dans l'image :

$$\widehat{P} = \widehat{P}(P_L, p)$$

Cette approche a été utilisée pour la reconstruction des primitives les plus représentatives, à savoir le point, la droite, le cercle, la sphère et le cylindre [Boukir 93]. Il faut noter que, pour l'ensemble de ces objets, cette méthode ne nécessite que la résolution de deux systèmes linéaires. D'un point de vue géométrique (voir Figure 2.1), elle revient à calculer l'intersection entre la surface des limbes et un cône généralisé (de centre le centre optique de la caméra et de génératrice l'image de la primitive). Les méthodes discrètes, équivalentes à la triangulation en stéréovision, reposent elles sur le calcul de l'intersection de deux cônes et semblent donc d'une complexité supérieure.

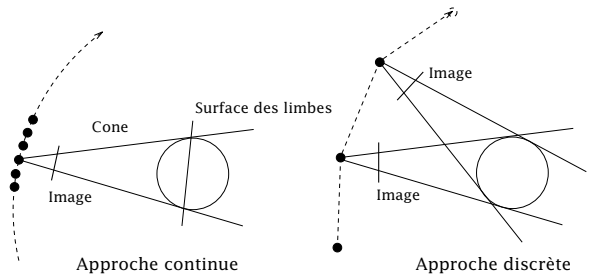

FIG. 2.1 – *Différence entre l'approche continue et l'approche discrète*

Quand les mouvements de la caméra sont quelconques, d'importantes erreurs sur l'estimation des paramètres de la primitive apparaissent. La qualité de l'estimation est en effet très sensible à la nature des mouvements effectués. Une méthode reposant sur le concept de la vision active a donc été développée afin d'optimiser la qualité de l'estimation en générant les mouvements de la caméra permettant de supprimer les effets des erreurs de discrétisation et de minimiser les effets des erreurs de

mesures. Deux tâches (dites de fixation et de focalisation) permettent d'obtenir une estimation optimale, à savoir robuste et non biaisée. Pour les cas d'un cylindre ou d'un segment par exemple, ceux-ci doivent apparaître vertical (ou horizontal) centré dans la séquence d'images (voir Figure 2.2) et rester à une distance constante de la caméra. On retrouve donc ici des résultats similaires à ceux décrits dans [Sandini 90] (sur l'intérêt d'avoir un point de fixation immobile, résultat que l'on retrouve également dans les travaux sur les têtes stéréoscopiques actives) et dans [Cipolla 95] (sur l'intérêt de conserver la même distance entre la caméra et l'objet). Une stratégie différente est par contre utilisée dans [Kutulakos 94a] : les mouvements appropriés sont dans ce cas situés dans le plan tangent à la surface pour observer en permanence le même point de cette surface. Cette différence s'explique par le fait que nous nous sommes intéressés à la reconstruction d'une primitive dans son ensemble (ce qui suppose que cette primitive existe) alors que les travaux de Kutulakos concernent plus généralement la reconstruction locale de surfaces convexes.

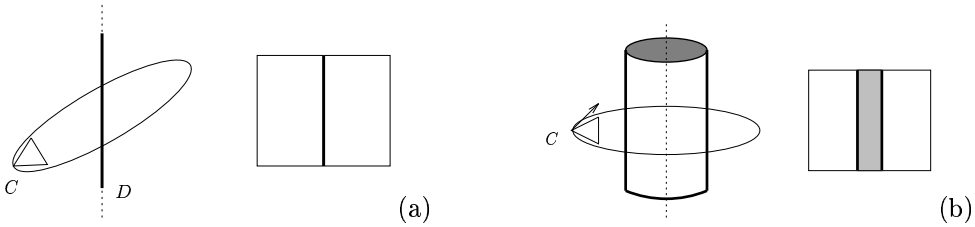

FIG. 2.2 – *Mouvements optimaux de la caméra pour la reconstruction d'une droite (a) et d'un cylindre (b)*

Dans notre cas, les mouvements nécessaires à la réalisation des tâches de fixation et de focalisation sont automatiquement générés par asservissement visuel. Comme la matrice d'interaction, élément essentiel de la loi de commande, dépend de paramètres qui sont itérativement estimés, l'algorithme d'estimation améliore le comportement de la loi de commande, qui lui-même améliore le comportement de l'algorithme d'estimation, aboutissant ainsi à un cycle complet de perception-action. Les résultats obtenus par cette approche sont beaucoup plus fiables, robustes et précis que ceux obtenus sans contrôle explicite du mouvement de la caméra. Dans de bonnes conditions, on peut par exemple atteindre une précision de l'ordre du dixième de millimètre dans l'estimation du rayon d'un cylindre.

2.3 Stratégies globales d'exploration

La stratégie de reconstruction présentée dans le paragraphe précédent ne permet de reconstruire qu'une seule primitive à la fois puisque les mouvements de la caméra sont contraints pour estimer de manière optimale la primitive sélectionnée. La suite de ce travail, effectué dans le cadre de la thèse d'Eric Marchand [Marchand 96b], a donc consisté à définir des stratégies de perception capables d'aboutir à une représentation complète et globale de la scène à reconstruire. Pour limiter ce problème

très complexe, nous n'avons à ce jour considéré que des scènes relativement simples, c'est-à-dire constituées uniquement d'objets polyédriques et de cylindres. De manière schématique (voir Figure 2.3), la méthode développée consiste à enchaîner :

- des phases d'exploration afin de détecter la présence d'objets non encore reconstruits ;
- des phases de reconstruction incrémentale et de modélisation pour traiter les objets découverts ;
- des phases de reconstruction optimale des primitives composant ces objets.

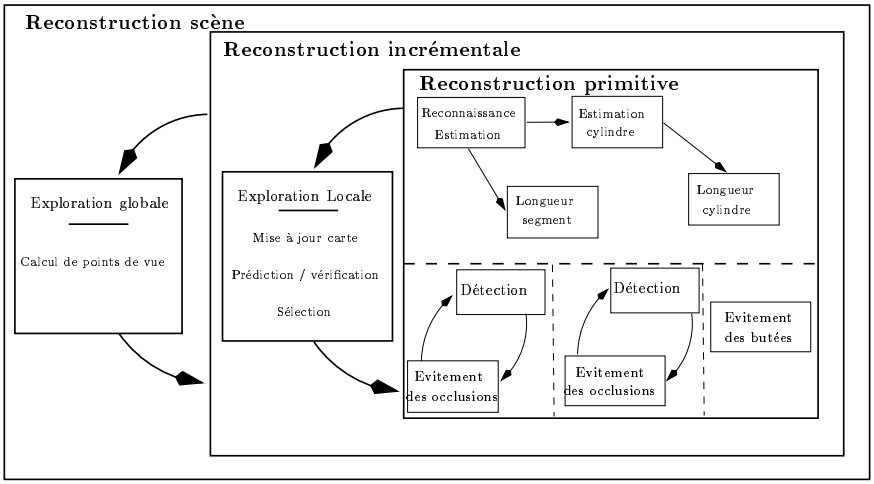

FIG. 2.3 – *Automate hiérarchique parallèle assurant la reconstruction globale de la scène*

2.3.1 Reconstruction et modélisation incrémentale

En raison du type d'objets que nous avons considéré, leur projection dans l'image est simplement constituée d'un ensemble de segments. Il a donc été très simple de proposer une stratégie capable de sélectionner successivement les différents segments observés par la caméra à l'aide d'un module de mise en correspondance entre les segments observés dans une image et une rétro-projection du modèle courant de la scène dans cette image (voir Figure 2.4). Le segment choisi sert de base à la reconnaissance et à la reconstruction de la primitive 3D associée. On obtient ainsi une modélisation 3D de la partie observée de la scène. Cette modélisation est cependant de bas niveau et généralement incomplète. Afin de remédier à ces inconvénients, nous avons développé une technique de prédiction et vérification d'hypothèses gérées à l'aide de réseaux Bayesiens [Marchand 98a]. Ceux-ci se prêtent en effet très bien au raisonnement et à la prise de décision en présence d'incertitudes [Pearl 88, Rimey 94, Djian 97]. Dans notre cas, ils permettent d'émettre des

hypothèses sur l'existence et la localisation de nouveaux objets, de proposer l'exécution d'une action conduisant à vérifier ou à infirmer cette hypothèse, et enfin, en fonction du résultat de l'étape de vérification, de compléter le modèle 3D de la scène (voir Figure 2.4). La génération des hypothèses repose sur la configuration des segments déjà reconstruits et sur un certain nombre d'a priori très généraux sur les caractéristiques d'une scène (du style : deux segments proches et parallèles ont une forte probabilité d'appartenir à un même polygone). La vérification des hypothèses s'appuie, elle, sur les observations 2D déjà réalisées, mais aussi, si besoin, sur une acquisition d'informations supplémentaires nécessitant un déplacement du capteur (déplacement qui est alors automatiquement réalisé par asservissement visuel, voir Figure 2.5). On peut remarquer que l'on retrouve ici un cycle complet de perception-action puisqu'un résultat d'une perception locale génère une action dédiée à l'amélioration de la perception globale. Finalement, la modélisation de la scène et la création de nouveaux objets (jonctions, polygones, ...) reposent sur les informations disponibles initialement et sur celles introduites par les hypothèses validées.

Cette approche permet de gérer les problèmes locaux d'occultation et de complétude (voir Figure 2.6), tout en fournissant une modélisation des objets de haut niveau sur laquelle il semble envisageable de greffer un module de reconnaissance.

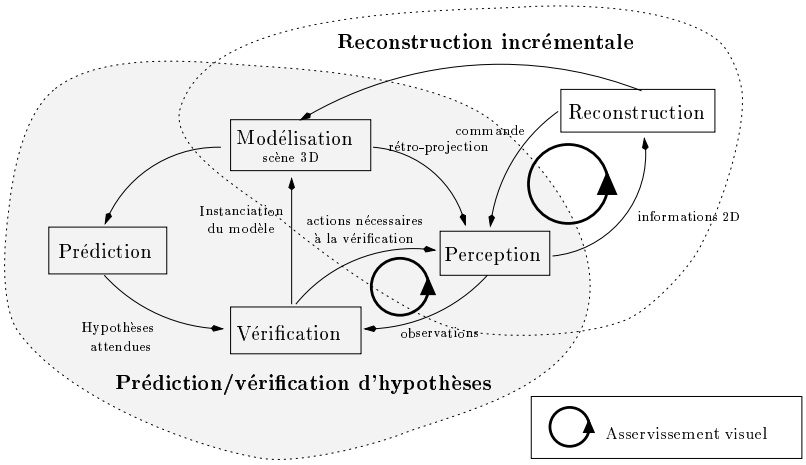

FIG. 2.4 – *Principe de la reconstruction et de la modélisation incrémentale*

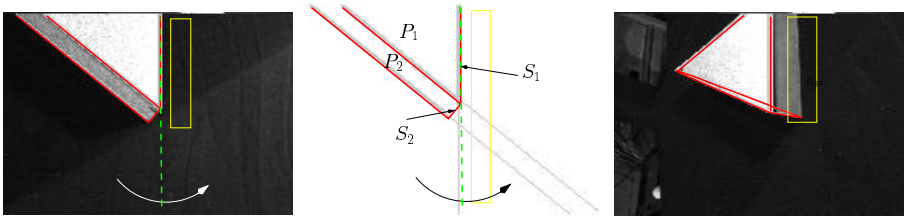

FIG. 2.5 – *Vérification de la présence d'un segment prédit par asservissement visuel)*

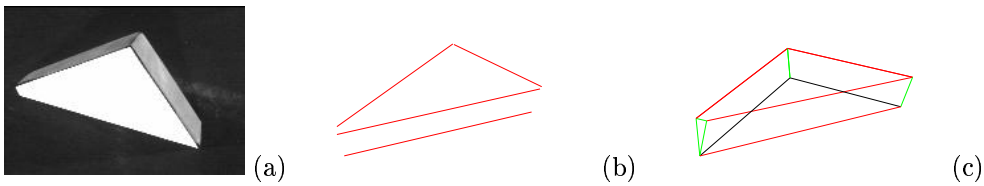

FIG. 2.6 – *(a) Image initiale de la scène, (b) modèle 3D acquis en se servant uniquement du module de reconstruction incrémentale, (c) modèle reconstruit en se servant du module de prédiction/vérification d'hypothèses*

2.3.2 Exploration globale

La phase précédente de reconstruction incrémentale permet de traiter l'ensemble des objets observés par la caméra. Afin d'assurer la complétude de la reconstruction, des mouvements sont ensuite nécessaires pour explorer les parties cachées et les parties non encore observées de la scène. Cette exploration passe par le calcul d'un certain nombre de points de vue qui permettront l'observation de nouveaux objets ou, au contraire, qui permettront d'assurer que telle ou telle partie de l'espace est vide. La méthode que nous avons développée s'inspire des travaux décrits dans [Connoly 85, Wixson 94, Triggs 95]. On peut également signaler les travaux sur l'exploration complète d'un seul objet [Kutulakos 94b], la découverte des zones occultées [Maver 93], la recherche d'un objet donné dans un environnement [Rimey 94, Ye 96], l'exploration autonome d'une scène basée sur la fusion de données imprécises [Moutarlier 91, Chatila 93, Whaite 94, Bulata 96], ainsi que les travaux sur l'observation optimale d'un objet (qui reposent sur une connaissance a priori de la scène) [Cowan 88, Tarabanis 95].

Dans notre cas, la stratégie de calcul de points de vue repose sur la modélisation et l'optimisation d'une fonction de coût [Marchand 96a]. Quatre critères complémentaires ont été intégrés dans cette fonction. En premier lieu, il convient de maximiser le gain d'observation apporté par une nouvelle position. Ce gain détermine le volume potentiel découvert calculé en utilisant les techniques de lancers de rayons. Afin de minimiser la distance totale parcourue par la caméra, le coût du déplacement d'un point de vue au suivant est également pris en compte. De plus, les contraintes mécaniques du robot conduisent à l'introduction d'un critère éloignant le robot de ses butées articulaires. Enfin, les connaissances déjà acquises sur la scène permettent de définir un critère binaire représentant l'accessibilité d'un point de vue. La minimisation de la fonction obtenue est effectuée à l'aide d'un algorithme ICM multi-échelle (préférable à une méthode stochastique de type recuit simulé pour des raisons de temps de calcul) qui détermine la prochaine position de la caméra à l'intérieur d'une demi-sphère englobant la scène. Un algorithme de focalisation sur les zones inconnues résiduelles a également été proposé : il permet de traiter de manière adéquate les parties occultées par des objets. Ajoutons finalement que nous avons proposé des techniques issues de la programmation dynamique afin de minimiser (sous certaines hypothèses) le nombre de points de vue nécessaires à l'exploration complète de la scène [Marchand 97b].

L'exploration de la scène s'achève quand, quel que soit le point de vue choisi parmi tous les points de vue accessibles, il n'y a plus d'apport supplémentaire d'informations. Ceci signifie que la reconstruction est alors aussi complète que possible compte tenu des contraintes imposées par le manipulateur et par la scène.

En conclusion, le système que nous avons développé est totalement autonome et permet d'obtenir avec une excellente précision la cartographie complète d'une scène composée de cylindres et d'objets polyédriques, sans a priori initial sur le nombre, la position et les dimensions de ces objets (voir Figure 2.7). Les objets traités sont encore relativement simples. Le passage à la reconstruction d'objets plus complexes constitue l'une des perspectives naturelles de ces travaux.

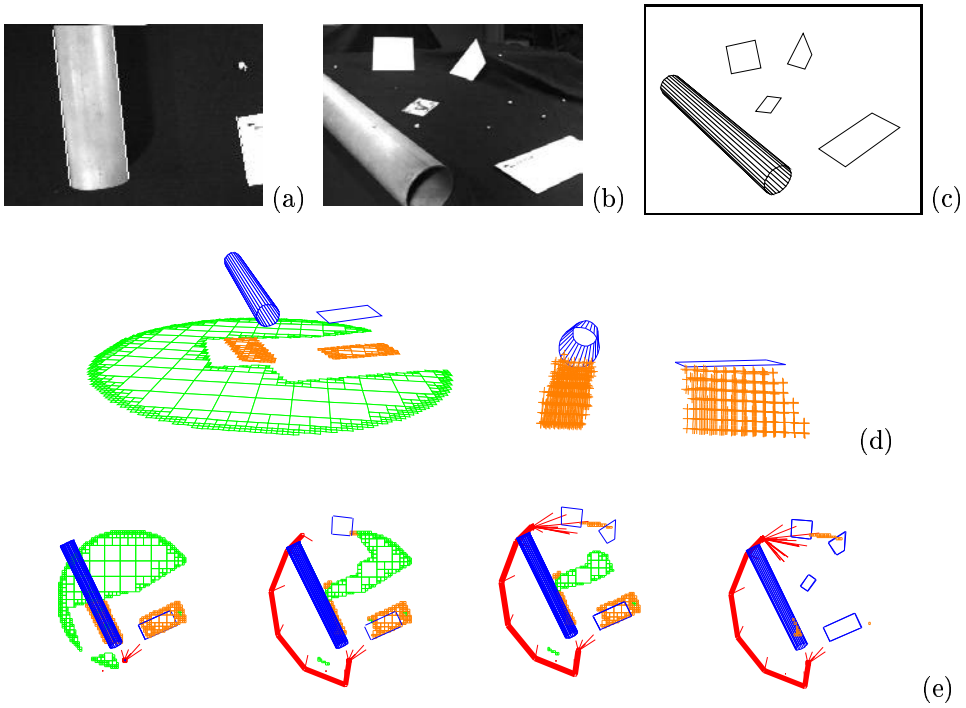

FIG. 2.7 – *Résultat de l'exploration : (a) image initiale acquise par la caméra, (b) image extérieure à comparer au résultat final (c) de la reconstruction, (d) reconstruction et zones inconnues après la première reconstruction incrémentale, (e) quelques étapes de l'exploration*

2.3.3 Spécification et mise en œuvre sécurisée

Nous voudrions très brièvement insister pour terminer ce chapitre sur les aspects importants de spécification et de mise en œuvre des stratégies de perception et d'action. Les principales difficultés résident, d'une part, dans la façon de gérer

les aspects continus et événementiels relatifs aux stratégies locales et globales qu'il faut définir dans toute application significative, et, d'autre part, dans la façon de gérer le parallélisme entre tâches. Le travail que nous avons effectué dans ce domaine, mené en collaboration avec Eric Rutten dans le cadre de la thèse d'Eric Marchand, repose sur l'emploi du langage synchrone SIGNAL développé au sein du projet EP-ATR de l'IRISA [Le Guernic 91]. Les intérêts de SIGNAL se sont révélés multiples [Marchand 98b]. D'une part, son caractère flot de données est parfaitement adapté aux tâches d'asservissement visuel et aux algorithmes d'estimation ou de filtrage. D'autre part, la notion d'intervalle de temps introduite dans SIGNAL*GTi* offre la possibilité de gérer la hiérarchisation, le séquencement, le parallélisme ainsi que la préemption de tâches. On peut ainsi considérer de manière unifiée les différents aspects d'une application temps réel à base de systèmes réactifs, d'autant plus que certaines parties du programme (tels les calculs matriciels ou la gestion des listes de données) peuvent être écrites dans un langage plus adapté sans pour autant sortir du cadre synchrone. De plus, il est possible de vérifier certaines propriétés statiques ou dynamiques des programmes produits (comme par exemple l'absence de *dead-locks* ou l'attractivité), ce qui permet de sécuriser fortement le fonctionnement attendu du système. De nombreux points restent à étudier dans ce domaine, comme nous le verrons dans la dernière et prochaine partie de ce document.

Conclusion et perspectives

Nous récapitulons en guise de conclusion notre contribution en asservissement visuel et perception active. Nous présentons également les perspectives que nous souhaitons mener dans ces deux axes étroitement liés.

Asservissement visuel

L'asservissement visuel consiste à utiliser les informations fournies par une caméra afin d'en contrôler le mouvement. L'approche que nous avons retenue repose sur la modélisation de fonctions de tâches appropriées et consiste à spécifier le problème en termes de régulation dans l'image. Elle permet de compenser les imprécisions des modèles (erreurs de calibration), aussi bien du capteur que du porteur de la caméra, par des lois de commande robustes en boucle fermée sur les informations visuelles extraites de l'image.

Nous pensons avoir apporté une contribution importante à ce domaine, aussi bien pour les tâches de positionnement réalisant diverses liaisons virtuelles, combinées à l'évitement des butées et singularités du manipulateur, que pour les tâches de poursuite d'objet mobile. Le savoir-faire obtenu a notamment débouché sur la réalisation de tâches de manipulation (préhension, vissage) pour des opérations de maintenance dans les bols de générateur de vapeur des centrales nucléaires.

Plus récemment, nous avons entrepris une étude sur l'utilisation d'informations visuelles dynamiques (et non plus seulement géométriques) dans les boucles de commande par asservissement visuel. Cette approche, qualifiée d'asservissement visuel $d2\mathrm{D}/dt$, a été validée par une tâche d'alignement du capteur dans la direction du déplacement de celui-ci, et par une tâche de positionnement telle que le plan image de la caméra soit parallèle à un objet plan quelconque. Cette thématique originale nous semble élargir grandement le champ d'application de la commande référencée vision par la réalisation de nouvelles tâches, la suppression de connaissances géométriques fortes sur la scène et son applicabilité à des scènes complexes grâce à des traitements de séquences d'images fondés sur l'analyse du mouvement.

Finalement, nous développons depuis peu une nouvelle approche, dite d'asservissement visuel 2D 1/2, qui permet, contrairement à l'asservissement visuel 2D, de converger vers la position désirée quelle que soit la position initiale du capteur. L'avantage de cette nouvelle technique par rapport aux techniques classiques d'asservissement visuel 3D est de ne pas nécessiter de modèle 3D de l'objet considéré.

Nos perspectives en asservissement visuel concernent de manière générale la prise en compte d'objets complexes et la réduction des connaissances nécessaires sur ces objets afin d'accroître les domaines d'application et la facilité de mise en œuvre des techniques d'asservissement visuel. Du point de vue de la commande, les approches $d2\mathrm{D}/dt$ (de par la complexité des images que l'on peut appréhender) et 2D 1/2 (de par l'absence de besoin d'un modèle 3D des objets considérés) permettent d'atteindre en grande partie cet objectif.

Concernant l'asservissement visuel $d2\mathrm{D}/dt$, il convient de généraliser les résultats que nous avons obtenus pour la réalisation de tâches particulières. En effet, la relation linéaire entre les variations de l'information visuelle et la vitesse de la caméra, représentée par la matrice d'interaction pour des informations visuelles géométriques, n'est dans l'absolu plus valide, en particulier en raison de l'apparition de termes liés à l'accélération de la caméra et de termes quadratiques. Nous souhaitons donc déterminer des classes de tâches pour lesquelles l'élaboration de lois de commande en accélération, en vitesse, ou combinant accélération et vitesse est nécessaire.

En asservissement visuel 2D 1/2, nous pensons, en raison de la forme très intéressante de la matrice d'interaction (triangulaire), qu'il est possible d'élaborer une loi de commande, basée sur une détermination automatique de gains, capable d'assurer la convergence quelle que soit la position initiale de la caméra, et ce sous simple respect de conditions nécessaires et suffisantes aisément réalisables en pratique. Actuellement, certaines parties de l'objet d'intérêt peuvent éventuellement sortir du champ de vision de la caméra si des paramètres de calibration très grossiers sont employés. Il s'agit donc de jouer sur la loi de commande afin d'éviter de tels phénomènes. Des techniques d'autocalibration pourront également être employées afin d'améliorer la dynamique du système obtenu.

Des études méritent également d'être menées en asservissement visuel 2D. En effet, les quelques problèmes évoqués (minima locaux, singularité) sont principalement dûs à des problèmes de représentation, le choix usuel de coordonnées de points comme informations visuelles ne paraîssant pas optimal pour représenter un objet. Nous espérons donc déterminer, si elles existent, des informations visuelles génériques pour tout type d'objets. Il s'agit de trouver une représentation adéquate, calculable à partir d'éléments caractéristiques extraits des images, offrant un découplage maximal par rapport aux degrés de liberté de la caméra (afin de les sélectionner aisément en fonction de la tâche à accomplir), et permettant en outre d'éviter les problèmes de divergence parfois rencontrés pour des positions initiales très éloignées de leur position désirée. Les six premiers moments d'inertie de la projection d'un objet semblent intuitivement être de bons candidats. Il reste à

démontrer leurs véritables propriétés et, en cas d'échec, à déterminer d'autres jeux d'informations visuelles.

La réalisation de tâches définies de manière qualitative fait également partie de nos perspectives importantes. Cela permettrait d'élargir la gamme des tâches que l'on peut considérer et simplifierait grandement le calcul des consignes à atteindre, actuellement déterminées soit par apprentissage, soit à partir de connaissances géométriques. Par exemple, une tâche aussi simple que "se positionner parallèlement à un segment" est en fait extrêmement délicate à modéliser (et donc à réaliser) si l'on ne dispose d'aucune information 3D sur le segment. La difficulté réside dans le choix des informations visuelles à utiliser afin d'aboutir à des conditions d'observabilité, et ensuite dans la modélisation de l'interaction associée.

Un autre aspect, plus applicatif, est également envisagé. Il concerne la modélisation d'enchaînements adéquats de tâches de bas niveau afin de réaliser des tâches complexes, comme par exemple l'approche puis la saisie d'un objet dans un environnement encombré, ou la prise en compte de changements possibles d'aspect de l'objet au cours du mouvement de la caméra. Ce thème est bien évidemment étroitement lié à celui de la perception active que nous décrivons dans le prochain paragraphe. Cependant, de telles décompositions peuvent également apporter des solutions aux problèmes de divergence. Une tâche de positionnement peut en effet se décomposer en sous-tâches de centrage, contrôle de la surface, de l'orientation, etc. Il doit être possible d'établir des conditions nécessaires et suffisantes de stabilité de chacune de ces sous-tâches. Il reste alors à déterminer, à partir des conditions obtenues, l'enchaînement à effectuer pour aboutir à la réalisation de la tâche nominale. Nous souhaitons donc repousser autant que possible les limites de l'aspect local des techniques d'asservissement visuel 2D.

Finalement, à plus long terme, et comme nous le verrons par la suite, nous pensons engager des études sur l'utilisation de plusieurs capteurs visuels embarqués et/ou déportés.

Parallèlement, les aspects de traitement d'images pour extraire à une cadence proche de la cadence video les informations visuelles liées à la paramétrisation de la tâche ne sont évidemment pas à négliger. Cette phase nécessite une segmentation initiale de la forme d'intérêt, puis de suivi dans l'image. Par le passé, dans le cadre de la thèse de Samia Boukir, une méthode performante de suivi de segments 2D, basée sur le suivi d'éléments de contours en mouvement [Bouthemy 89], a été développée [Boukir 98]. Ces travaux sont actuellement poursuivis, en collaboration avec Patrick Bouthemy dans le cadre d'une convention avec la DER-EDF Chatou (séjour post-doctoral de Nathalie Giordana [Giordana 98]), pour traiter des objets complexes dans un environnement non maîtrisé. On utilise pour cela des algorithmes d'estimation robuste du modèle affine du mouvement 2D [Odobez 94] et de suivi de structures déformables 2D [Kervrann 95] (voir Figure 2.8). De nombreux travaux restent cependant à accomplir afin de gérer les connaissances disponibles (modèle CAO) et les changements d'aspects au cours du mouvement. Il s'agit également de diminuer autant que possible les temps de calcul obtenus (cadence actuelle de 1 Hz sur station de travail de type Ultra Sparc).

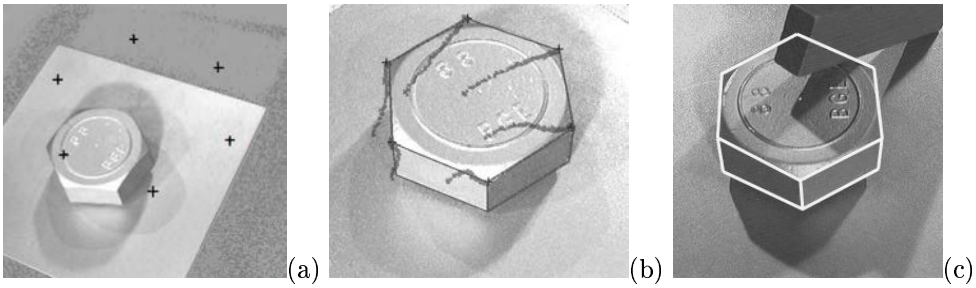

FIG. 2.8 – *Positionnement par rapport à une vis : (a) position initiale, (b) position finale, (c) modèle suivi (robuste aux occultations partielles)*

Perception active

Dans le domaine de la vision active, nous nous sommes exclusivement intéressés au contrôle de la position, de la vitesse, ou de la trajectoire d'une caméra embarquée. Deux types de stratégies ont été considérées et combinées : d'une part, des stratégies locales, réagissant en continu au flot des informations perçues et fournissant une perception de bas niveau, et d'autre part, des stratégies globales, activées sur des événements particuliers et remontant à une perception de haut niveau.

Notre première contribution a porté sur la reconstruction 3D optimale de primitives géométriques paramétrables. Le principe consiste à réaliser de manière continue un cycle de perception-action en contraignant et en générant automatiquement par asservissement visuel les mouvements de la caméra afin d'obtenir une reconstruction aussi précise que possible (à savoir non biaisée et robuste aux erreurs de mesure). Ce travail nous ouvre de très nombreuses perspectives en reconstruction 3D d'objets complexes (c'est-à-dire non paramétrables), et en trajectographie 3D par vision active. Pour la reconstruction 3D d'objets complexes, nous pensons pouvoir généraliser l'approche basée sur la matrice d'interaction que nous avons développée à la reconstruction locale de surfaces gauches. Des stratégies de haut niveau seront ensuite à développer pour obtenir une reconstruction complète et globale des objets. Concernant les aspects de trajectographie 3D [Aggarwal 88], des résultats préliminaires ont été obtenus récemment dans le cas d'un objet ponctuel, permettant notamment de lever l'ambiguïté entre la vitesse de translation et la profondeur de l'objet [Bensalah 96]. Une généralisation de ces résultats à des objets plus complexes s'avère indispensable, tout d'abord dans le cadre de la vision dynamique, puis dans le cadre de la vision active.

Dans les deux cas (reconstruction d'objets non paramétrables et trajectographie 3D), il s'agit d'étudier ce que peuvent apporter à ces problèmes classiques de vision par ordinateur des contraintes sur le mouvement du capteur, contraintes qui seront générés par asservissement visuel. Les apports attendus concernent avant tout l'optimalité de l'estimation, mais aussi l'observabilité complète ainsi que son automatisation.

En reconstruction 3D, les applications potentielles à ces recherches se trouvent dans les domaines de la métrologie et du contrôle de qualité (vérification de la conformité d'objets manufacturés par exemple).

Pour aborder le vaste problème d'élaboration de stratégies de perception et d'action de haut niveau, nous avons choisi pour cadre applicatif l'exploration complète de scènes statiques constituées d'objets simples (cylindres et polyèdres). Ce travail nous a permis d'acquérir un savoir-faire important dans les domaines de calcul de point de vue, de prédiction et vérification d'hypothèses, et d'utilisation des réseaux Bayésiens. Les perspectives dans ce domaine concernent l'emploi de ces techniques pour l'exploration de scènes comportant des objets complexes, mais aussi pour des applications non reconstructionnistes, le 3D n'étant généralement pas une fin en soi. De telles applications sont par exemple la détection de changements à l'intérieur de scènes dont le modèle est connu, ou la recherche d'informations manquantes et cependant nécessaires à la réalisation d'une tâche donnée. De même, nous pensons nous intéresser au problème du couplage entre une vision globale (fournie par un capteur déporté) et une vision locale (fournie, elle, par un capteur embarqué capable de se focaliser sur les différents objets constituant la scène). Ce type de couplage paraît inévitable pour des environnements complexes, encombrées et comportant des objets mobiles. Les difficultés principales résident dans l'exploitation et la fusion adéquates de données complémentaires, provenant de points de vue distincts, et acquises avec des résolutions très différentes. De plus, il faudra prendre en compte le fait que les positions respectives des caméras ne pourront être connues précisément. Les stratégies de coopération seront également à définir en fonction des tâches à accomplir (exploration, assemblage, etc).

La robotique d'intervention, principalement dans les milieux hostiles (nucléaire, spatial, sous-marin, etc.) constitue le domaine d'applications naturel de ces recherches.

Enfin, concernant les aspects de spécification et de mise en œuvre sécurisée que nous avons traités à l'aide du langage synchrone SIGNAL, de nombreux points restent à étudier, de la gestion d'événements inattendus (comme par exemple l'échec d'un traitement d'image ou d'une mise en correspondance) aux aspects de communication et de synchronisation entre plusieurs capteurs et robots dans le cadre des visions globale et locale évoquées ci-dessus.

Bibliographie

[Adiv 89] G. Adiv : Inherent ambiguities in recovering 3D motion and structure from a noisy flow field. *IEEE Trans. on PAMI*, 11(5):477-489, Mai 1989.

[Aggarwal 88] J. Aggarwal, N. Nandhakumar : On the computation of motion from sequences of images. a review, *Proc. of the IEEE*, 76(8):917-935, Août 1988.

[Allen 93] P.K. Allen, B. Yoshimi, A. Timcenko, P. Michelman : Automated tracking and grasping of a moving object with a robotic hand-eye system. *IEEE Trans. on Robotics and Automation*, 9(2):152-165, Avril 1993.

[Aloimonos 87] Y. Aloimonos, I. Weiss, A. Bandopadhay : Active vision, *Int. Journal of Computer Vision*, 1(4):333-356, Janvier 1987.

[Aloimonos 90] Y. Aloimonos : Purposive and qualitative active vision, *IAPR Int. Conf. on Pattern Recognition, ICPR'90*, 1:346-360, Atlantic City, New Jersey, Juin 1990.

[Bajcsy 88] R. Bajcsy : Active perception, *Proc. of the IEEE*, 76(8):996-1005, Août 1988.

[Ballard 91] D.H. Ballard : Animate vision, *Artificial Intelligence*, 48(1):57-86, Février 1991.

[Bar-Shalom 93] Y. Bar-Shalom, X Li : *Estimation and tracking : principles, techniques and software.* Artech House, Boston, 1993.

[Basseville 83] M. Basseville, A. Benveniste : Design and comparative study of some sequential jump detection algorithms for digital signals. *IEEE Trans. on Acoustics, Speech, and Signal Processing*, 31(3):521-535, Juin 1983.

[Bensalah 95] F. Bensalah, F. Chaumette : Compensation of abrupt motion changes in target tracking by visual servoing, *IEEE/RSJ Int. Conf. on Intelligent Robots and Systems, IROS'95*, 1:181-187, Pittsburgh, Pennsylvannie, Août 1995.

[Bensalah 96] F. Bensalah : *Estimation du mouvement par vision active*, Thèse de l'Université de Rennes 1, IRISA, Juillet 1996.

[Boufama 95] B. Boufama, R. Mohr : Epipole and fundamental matrix estimation using the virtual parallax property. *IEEE Int. Conf. on Computer Vision, ICCV'95*, pp. 1030-1036, Cambridge, Massachusets, Juin 1995.

[Boukir 93] S. Boukir : *Reconstruction 3D d'un environnement statique à l'aide d'une caméra mobile commandable*, Thèse de l'Université de Rennes 1, IRISA, Octobre 1993.

[Boukir 98] S. Boukir, P. Bouthemy, F. Chaumette, D. Juvin : A local method for contour matching and its parallel implementation, *Machine Vision and Applications*, 10(5/6):321-330, Avril 1998.

[Bouthemy 89] P. Bouthemy : A maximum likelihood framework for determining moving edges. *IEEE Trans. on PAMI*, 11,(5):499-511, Mai 1989.

[Boyer 96] E. Boyer : *Reconstruction de surfaces d'objets courbes en vision par ordinateur*, Thèse de l'Institut National Polytechnique de Lorraine, INRIA Lorraine, Décembre 1996.

[Brown 90] C.M. Brown : Gaze control cooperating through prediction. *Image and Vision Computing*, 8(1):10-17, Février 1990.

[Brown 92] C. Brown : Issues in selective perception. *IAPR Int. Conf. on Pattern Recognition, ICPR'92*, 1:21-24, La Haye, Pays-Bas, Août 1992.

[Brunnström 96] K. Brunnström, J.O. Eklundh, T. Uhlin : Active fixation for scene exploration. *Int. Journal of Computer Vision*, 17(2):137-162, Février 1996.

[Bulata 96] H. Bulata : Modélisation d'un environnement structuré et localisation sur amers pour la navigation d'un robot mobile autonome. Thèse de l'Université Paul Sabatier, LAAS, Toulouse, Mai 1996.

[Cameron 90] A. Cameron, H. Durrant-Whyte : A bayesian approach to optimal sensor placement. *Int. Journal of Robotics Research*, 9(5):70-88, Octobre 1990.

[Chan 95] T.F. Chan, R.V. Dubey : A weighted least-norm solution based scheme for avoiding joints limits for redundant manipulators. *IEEE Trans. on Robotics and Automation*, 11(2):286-292, Avril 1995.

[Chatila 93] R. Chatila, S. Fleury, M. Herrb, S. Lacroix, C. Proust : Autonomous navigation in natural environment. *Third Int. Symposium on Experimental Robotics, ISER'93*, Kyoto, Japon, Octobre 1993.

[Chien 89] C. Chien, J.K. Aggarwal : Model construction and shape recognition from occluding contours. *IEEE Trans. on PAMI*, 11(4):372-389, Avril 1989.

[Chaumette 90] F. Chaumette : La relation vision-commande : théorie et applications à des tâches robotiques. *Thèse de l'Université de Rennes I, IRISA*, Juillet 1990.

[Chaumette 91] F. Chaumette, P. Rives, B. Espiau : Positioning of a robot with respect to an object, tracking it and estimating its velocity by visual servoing, *IEEE Int. Conf. on Robotics and Automation, ICRA'91*, 3:2248-2253, Sacramento, Californie, Avril 91.

[Chaumette 93a] F. Chaumette, P. Rives, B. Espiau : Classification and realization of the different vision-based tasks, *Visual servoing- Automatic Control of Mechanical Systems with Visual Sensors*, World Scientific Series in Robotics and Automated Systems, K. Hashimoto editor, Vol. 7:199-228, World Scientific Publishing Co Pte Ltd, London, 1993.

[Chaumette 93b] F. Chaumette, A. Santos : Tracking a moving object by visual servoing, *12th World Congress IFAC*, 3:643-648, Sydney, Australie, Juillet 1993.

[Chaumette 94] F. Chaumette : Visual servoing using image features defined upon geometrical primitives, 33rd *IEEE Conf. on Decision and Control, CDC'94*, 4:3782-3787, Orlando, Floride, Décembre 1994.

[Chaumette 96] F. Chaumette, S. Boukir, P. Bouthemy, D. Juvin : Structure from controlled motion, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18(5):492-504, Mai 1996.

[Chaumette 98] F. Chaumette : Potential problems of stability and convergence in image-based and position-based visual servoing, *The Confluence of Vision and Control* G. Hager, D. Kriegman, A. Morse (eds), LNCIS Series, Springer Verlag, 1998.

[Christy 96] S. Christy, R. Horaud : Euclidean reconstruction: from paraperspective to perspective. *European Conf. on Computer Vision, ECCV'96*, Cambridge, Angleterre, LNCS 1065, pp. 129-140, Avril 1996.

[Cipolla 92] R. Cipolla, A. Blake : Surface shape from the deformation of apparent contours. *Int. Journal of Computer Vision*, 9(2):83-112, Novembre 1992.

[Cipolla 95] R. Cipolla : *Active Visual Inference of Surface Shape*, LNCS 1016, Springer Verlag, 1995.

[Colombo 95] C. Colombo, B. Allota, P. Dario : Affine visual servoing: a framework for relative positioning with a robot. *IEEE Int. Conf. on Robotics and Automation, ICRA'95*, 1:464-471, Nagoya, Japon, Avril 1995.

[Connoly 85] C. Connoly : The determination of next best views. *IEEE Int. Conf. on Robotics and Automation, ICRA'85*, pp. 432-435, St Louis, Missouri, Mars 1985.

[Couapel 94] B. Couapel : Stéréovision par ordinateur : géométrie et expérimentation. Thèse de l'Université de Rennes 1, IRISA, Mars 1994.

[Corke 93] P.I. Corke, M.C. Goods : Controller design for high performance visual servoing *12th World Congress IFAC'93*, 9:395-398, Sydney, Australie, Juillet 1993.

[Crétual 97] A. Crétual, F. Chaumette : Positioning a camera parallel to a plane using dynamic visual servoing, *IEEE/RSJ Int. Conf. on Intelligent Robots and Systems, IROS'97*, 1:43-48, Grenoble, Septembre 1997.

[Crétual 98] A. Crétual, F. Chaumette : Image-based visual servoing by integration of dynamic measurements, *IEEE Int. Conf. on Robotics and Automation, ICRA'98*, Louvain, Belgique, Mai 1998.

[Cowan 88] C.K. Cowan, P.D. Kovesi : Automatic sensor placement from vision task requirements. *IEEE Trans. on PAMI*, 10(3):407-416, Mai 1988.

[Crowley 88] J.L Crowley, P. Stelmaszyk, S. Discours : Measuring image flow by tracking edge lines. *IEEE Int. Conf. on Computer Vision, ICCV'88*, pp. 658-664, Tampa, Floride, 1988.

[Crowley 92] J.L Crowley, P. Stelmaszyk, P. Puget : Measurement and integration of 3D structures by tracking edge lines. *Int. Journal of Computer Vision*, 8(1):29-52, Juillet 1992.

[Crowley 95] J. Crowley, M. Mesrabi, F. Chaumette : Comparison of kinematic and visual servoing for fixation, *IEEE/RSJ Int. Conf. on Intelligent Robots and Systems, IROS'95*, 1:335-341, Pittsburgh, Pennsylvannie, Août 1995.

[Dementhon 95] D. Dementhon, L. S. Davis : Model-based object pose in 25 lines of code. *Int. Journal of Computer Vision*, 15(1/2):123-141, Juin 1995.

[Deriche 90] R. Deriche, O. Faugeras : Tracking line segments. *Image and Vision Computing*, 8(4):261-270, Novembre 1990.

[Dhome 89] M. Dhome, M. Richetin, J.T. Lapreste, G. Rives : Determination of the attitude of 3D objects from a single perspective view. *IEEE Trans. on PAMI*, 11(12):1265-1278, Décembre 1989.

[Dhome 90] M. Dhome, J.T. Lapreste, G. Rives, M. Richetin : Spatial localization of modelled objects of revolution in monocular perspective vision. *European Conf. on Computer Vision, ECCV'90*, Antibes, LNCS 427, pp. 475-485, Avril 1990.

[Dickmanns 91] E. Dickmanns, T. Chritians : Relative 3D state estimation for autonomous visual guidance of road vehicles. *Robotics and Autonomous Systems*, 7(2/3):85-98, Août 1991,

[Djian 97] D. Djian : Contribution à l'analyse de scènes par vision active : utilisation de réseaux bayesiens, Thèse de l'Ecole des Mines de Paris, INRIA Sophia-Antipolis, Février 1997.

[Dornaika 95] F. Dornaika : *Contributions à l'intégration vision/robotique : calibrage, localisation, asservissement.* Thèse de l'Institut National Polytechnique de Grenoble, INRIA Rhône-Alpes, Septembre 1995.

[Espiau 92] B. Espiau, F. Chaumette, P. Rives : A new approach to visual servoing in robotics, *IEEE Transactions on Robotics and Automation*, 8(6):313-326, Juin 1992.

[Espiau 93] B. Espiau : Effect of camera calibration errors on visual servoing in robotics. *Third Int. Symposium on Experimental Robotics, ISER'93*, Kyoto, Japon, Octobre 1993.

[Faugeras 93] O. Faugeras : *Three-Dimensionnal Computer Vision: a Geometric Viewpoint*, MIT Press, Cambridge, Massachusetts, 1993.

[Giordana 98] N. Giordana, P. Bouthemy, F. Chaumette, F. Spindler, J.-C. Bordas, V. Just : 2D model-based tracking of complex shapes for visual servoing tasks, *Workshop on Robust Vision for Vision-based Control of Motion, ICRA'98*, Louvain, Belgique, Mai 1998.

[Hager 95] G. Hager, W. Chang, A.S. Morse : Robot feedback control based on stereo vision: Towards calibration-free hand-eye coordination. *IEEE Control Systems Magazine*, 15(1):30-39, Février 1995.

[Hager 98] G. Hager, K. Toyama : X-Vision: a prtable substrate for real-time vision applications. *Computer Vision and Image Understanding*, 69(1):23-37, Janvier 1998. .

[Haralick 89] R. Haralick, H. Joo, C.-N. Lee, X. Zhuang, V. Vaidya, M. Kim : Pose estimation from corresponding point data. *IEEE Trans. on Systems, Man and Cybernetics*, 19(6):1426-1445, Novembre 1989.

[Hashimoto 93] K. Hashimoto, editor : *Visual Servoing : Real Time Control of Robot manipulators based on visual sensory feedback.* Vol. 7 of *World Scientific Series in Robotics and Automated Systems.* World Scientific Press. Singapour, 1993.

[Hashimoto 95] K. Hashimoto, H. Kimura : Visual servoing with non linear observer. *IEEE Int. Conf. on Robotics and Automation, ICRA'95*, 1:484-489, Nagoya, Japon, Avril 1995.

[Horaud 89] R. Horaud, B. Conio, O. Leboulleux, B. Lacolle : An analytic solution for the perspective 4-points problem. *Computer Vision, Graphics, and Image Processing*, 47(1):33-44, Juillet 1989.

[Hosoda 94] K. Hosoda, M. Asada : Versatile visual servoing without knowledge of true jacobian. *IEEE/RSJ Int. Conf. on Intelligent Robots and Systems, IROS'94*, pp. 186-193, Munich, Allemagne, Septembre 1994.

[Huang 91] L. Huang, J. Aloimonos : Relative depth from motion using normal flow: an active and purposive solution. *IEEE Workshop on Visual Motion*, pp. 196-204, Princeton, New Jersey, Octobre 1991.

[Hutchinson 96] S. Hutchinson, G. Hager, P Corke : A tutorial on visual servo control. *IEEE Trans. on Robotics and Automation*, 12(5):651-670, Octobre 1996.

[Jägersand 97] M. Jägersand, O. Fuentes, R. Nelson : Experimental evaluation of uncalibrated visual servoing for precision manipulation, *IEEE Int. Conf. on Robotics and Automation, ICRA'97*, 3:2874-2880, Albuquerque, Nouveau Mexique, Avril 1997.

[Kervrann 95] C. Kervrann : Modèles statistiques pour la segmentation et le suivi de structures déformables bidimensionnelles dans une séquence d'images. Thèse de l'Université de Rennes 1, IRISA, Juin 1995.

[Khadraoui 96a] D. Khadraoui, G. Motyl, P. Martinet, J. Gallice, F. Chaumette : Visual servoing in robotics scheme using a camera/laser-stripe sensor, *IEEE Transactions on Robotics and Automation*, 12(5):743-750, Octobre 1996.

[Khadraoui 96b] D. Khadraoui : La commande référencée vision pour le guidage automatique de véhicules. Thèse de l'Université Blaise Pascal, LASMEA, Clermont-Ferrand, Décembre 1996.

[Kutulakos 94a] K. Kutulakos, C. Dyer : Recovering shape by purposive viewpoint adjustment. *Int. Journal of Computer Vision*, 12(2):113-136, Février 1994.

[Kutulakos 94b] K. Kutulakos, C. Dyer : Global surface reconstruction by purposive control of observer motion. *IEEE Int. Conf. on Computer Vision and Pattern Recognition, CVPR'94*, pp. 331-338, Seattle, Washington, Juin 1994.

[Le Guernic 91] P. Le Guernic, M. Le Borgne, T. Gautier, C. Le Maire : Programming real time application with SIGNAL. *Proc. of the IEEE*, 79(9):1321-1336, Septembre 1991.

[Malis 98] E. Malis, F. Chaumette, S. Boudet : Positioning a coarse calibrated camera with respect to an unknown object by 2D 1/2 visual servoing, *IEEE Int. Conf. on Robotics and Automation, ICRA'98*, Louvain, Belgique, Mai 1998 *(version longue disponible sous ftp://ftp.irisa.fr/techreports/1998/PI-1166.ps.gz)*.

[Marchand 96a] E. Marchand, F. Chaumette : Controlled camera motions for scene reconstruction and exploration, *IEEE Int. Conf. on Computer Vision and Pattern Recognition, CVPR'96*, pp. 169-176, San Fransisco, Californie, Juin 1996.

[Marchand 96b] E. Marchand : *Stratégies de perception par vision active pour la reconstruction et l'exploration de scènes statiques*, Thèse de l'Université de Rennes 1, IRISA, Juin 1996.

[Marchand 96c] E. Marchand, F. Chaumette, A. Rizzo : Using the task function approach to avoid robot joint limits and kinematic singularities in visual servoing, *IEEE/RSJ Int. Conf. on Intelligent Robots and Systems, IROS'96*, 3:1083-1090, Osaka, Japon, Novembre 1996.

[Marchand 97b] E. Marchand, F. Chaumette : Active sensor placement for complete scene reconstruction and exploration, *IEEE Int. Conf. on Robotics and Automation, ICRA'97*, 1:743-750, Albuquerque, Nouveau Mexique, Avril 1997.

[Marchand 98a] E. Marchand, F. Chaumette : Stratégies de perception par vision active : une approche par réseaux Bayesiens, *Revue d'Intelligence Artificielle*, 1998.

[Marchand 98b] E. Marchand, E. Rutten, H. Marchand, F. Chaumette : Specifying and verifying active vision-based robotics systems with the Signal environment, *Int. Journal of Robotics Research*, 17(4), Avril 1998.

[Martinet 97] P. Martinet, N. Daucher, J. Gallice, M. Dhome : Robot control using monocular pose estimation, *Workshop on New Trends in Image-based Robot Servoing, IROS'97*, pp. 1-12, Grenoble, Septembre 1997.

[Matthies 89] L. Matthies,T. Kanade and R. Szelisli : Kalman filter-based algorithms for estimating depth from image sequences, *Int. Journal of Computer Vision*, 3(3):209-238, Septembre 1989.

[Maver 93] J. Maver, R. Bajcsy : Occlusions as a guide for planning the next view. *IEEE Trans. on PAMI*, 15(5):417-433, Mai 1993.

[Michel 93] H. Michel, P. Rives : Singularities in the Determination of the Situation of a Robot Effector from the Perspective View of Three Points. *Rapport de recherche INRIA n. 1850*, Sophia-Antipolis, Février 1993.

[Milios 93] E. Milios, M. Jenkin, J. Tsosos : Design and performance of trish, a binocular robot head with torsional eye movements. *Int. Journal of Pattern Recognition and Artificial Intelligence*, 7(1):51-68, Février 1993.

[Morin 93] L. Morin : *Quelques contributions des invariants projectifs à la vision par ordinateur*, Thèse de l'Institut National Polytechnique de Grenoble, LIFIA-IMAG, Janvier 1993.

[Motyl 92] G. Motyl : Couplage d'une caméra et d'un faisceau laser en commande référencée vision, Thèse de l'Université Blaise Pascal, LASMEA, Clermont-Ferrand, Septembre 1992.

[Moutarlier 91] P. Moutarlier : Modélisation autonome de l'environnement par un robot mobile. Thèse de l'Université Paul Sabatier, LAAS, Toulouse, Octobre 1991.

[Murray 95] D. Murray, K. Bradshaw, P. Mc Lauchlan, P. Sharkey : Driving saccade to pursuit using image motion. *Int. Journal of Computer Vision*, 16(3):205-228, Mars 1995.

[Negahdaripour 87] S. Negahdaripour, B.K.P. Horn : Direct passive navigation. *IEEE Trans. on PAMI*, 9(1):168-176, Janvier 1987.

[Negahdaripour 92] S. Negahdaripour, S. Lee : Motion recovery from image sequences using only first order optical flow information. *Int. Journal of Computer Vision*, 9(3):163-184, Décembre 1992

[Nelson95] B. Nelson, P.K. Khosla : Strategies for increasing the tracking region of an eye-in-hand system by singularity and joint limits avoidance. *Int. Journal of Robotics Research*, 14(3):255-269, Juin 1995.

[Odobez 94] J.-M. Odobez : Estimation, détection et segmentation du mouvement : une approche robuste et markovienne. Thèse de l'Université de Rennes 1, IRISA, Décembre 1994.

[Papanikolopoulos 93] N. Papanikolopoulos, P. Khosla, T. Kanade : Visual tracking of a moving target by a camera mounted on a robot: A combination of control and vision. *IEEE Trans. on Robotics and Automation*, 9(1):14-35, Février 1993.

[Pearl 88] J. Pearl : *Probabilistic reasoning in intelligent systems : Networks of plausible inference.* – Morgan Kaufmann Publisher Inc., San Mateo, Californie, 1988.

[Phong 95] T.Q. Phong, R. Horaud, A. Yassine, P.D. Tao : Object pose from 2D to 3D point and line correspondences. *Int. Journal of Computer Vision*, 15(3):225-243, Juillet 1995.

[Pissard-Gibollet 93] R. Pissard-Gibollet : Conception et commande par asservissement visuel d'un robot mobile. Thèse de l'Ecole des Mines de Paris, INRIA Sophia-Antipolis, Décembre 1993.

[Questa 95] P. Questa, E. Grossmann, and G. Sandini : Camera self orientation and docking maneuver using normal flow. *SPIE AeroSense'95*, Orlando, Floride, Avril 1995.

[Rimey 94] R.D. Rimey, C. Brown : Control of selective perception using Bayes nets and decision theory, *Int. Journal of Computer Vision* 12(2/3):173-207, Avril 1994.

[Rives 93] P. Rives, J.-J. Borrelly, J. Gallice, P. Martinet : A versatile parallel architecture for vision-based applications, *Workshop on Computer Architecture for Machine Perception, CAMP'93*, Décembre 1993.

[Safaee-Rad 92] R. Safaee-Rad, I. Tchoukanov, K.C. Smith, B. Benhabib : Three-Dimensional Location Estimation of Circular Features for Machine Vision, *IEEE Trans. on Robotics and Automation*, 8(5):624-639, Octobre 1992.

[Samson 91] C. Samson, M. Le Borgne, B. Espiau : *Robot Control: the Task Function Approach.* Clarendon Press, Oxford, Angleterre, 1991.

[Sandini 90] G. Sandini and M. Tistarelli : Active tracking strategy for monocular depth inference over multiple frames, *IEEE Trans. on PAMI*, 12(1):13-27, Janvier 1990.

[Santos-Victor 97] J. Santos-Victor, G. Sandini : Visual behaviors for docking. *Computer Vision and Image Understanding*, 67(3):223-238, Septembre 1997.

[Subarrao 86] M. Subarrao, A. Waxman : Closed- form solutions to image equations for planar surface in motion. *Computer Vision, Graphics, and Image Processings*, 36(2/3):208-228, Novembre 1986.

[Suh 93] I.H. Suh : Visual servoing of robot manipulators by fuzzy membership function based neural networks. In *Visual Servoing*, K. Hashimoto ed., World Scientific Series in Robotics and Automated Systems, 7:285-315, Singapour, 1993.

[Sundareswaran 96] V. Sundareswaran, P. Bouthemy, F. Chaumette : Exploiting image motion for active vision in a visual servoing framework, *Int. Journal of Robotics Research*, 15(6):629-645, Décembre 1996.

[Swain 93] M.J. Swain, M.A. Stricker : Promising direction in active vision. *Int. Journal of Computer Vision*, 11(2):109-127, Octobre 1993.

[Tarabanis 95] K. Tarabanis, R. Tsai, P.K. Allen : The MVP sensor planning system for robotic vision tasks. *IEEE Trans. on Robotics and Automation*, 11(1):72-85, Février 1995.

[Triggs 95] B. Triggs, C. Laugier : Automatic camera placement for robot vision. *IEEE Int. Conf. on Robotics and Automation, ICRA'95*, 2:1732-1738, Nagoya, Japon, Mai 1995.

[Tsakiris 97] D. Tsakiris, P. Rives, C. Samson : Applying visual servoing techniques to control nonholonomic mobile robots. *Workshop on New Trends in Image-based Robot Servoing, IROS'97*, pp. 21-32, Grenoble, Septembre 1997.

[Vernon 90] D. Vernon, M. Tistarelli : Using camera motion to estimate range for robotic part manipulation. *IEEE Trans. on Robotics and Automation*, 6(5):509-521, Octobre 1990.

[Viéville 95] T. Viéville, E. Clergue, R. Enciso, H. Mathieu : Experimenting with 3D vision on a robotic head. *Robotics and Autonomous Systems*, 14(1):1-27, Juin 1995.

[Waxman 87] A.M. Waxman, B.K. Parsi, M. Subbarao : Closed-form solutions to image flow equations for 3D structure and motion. *Int. Journal of Computer Vision*, 1(3):239-258, Octobre 1987.

[Weiss 87] L. Weiss, A.C. Sanderson, C. Neuman : Dynamic sensor-based control of robots with visual feedback. *IEEE Journal of Robotics and Automation*, 3(5):404-417, Octobre 1987.

[Weng 92] J. Weng, T. Huang and N. Ahuja: Motion and structure from line correspondences: closed-form solution, uniqueness, and optimization, *IEEE Trans. on PAMI*, 14(3):318-336, Mars 1992.

[Whaite 94] P. Whaite, F. Ferrie : Autonomous exploration: Driven by uncertainty. – *IEEE Int. Conf. on Computer Vision and Pattern Recognition, CVPR'94*, pp. 339-346, Seattle, Washington, Juin 1994.

[Willsky 76] A. Willsky, H. Jones : A generalized likelihood ratio approach to the detection and estimation of jumps in linear systems. *IEEE Trans. on Automatic Control*, 21(1):108-112, Février 1976.

[Wilson 96] W. Wilson, C. Hulls and G. Bell : Relative end-effector control using cartesian position-based visual servoing. *IEEE Trans. on Robotics and Automation*, 12(5):684-696, Octobre 1996.

[Wixson 94] L.E. Wixson : Viewpoint selection for visual search. – *IEEE Int. Conf. on Computer Vision and Pattern Recognition, CVPR'94*, pp. 800-805, Seattle, Washington, Juin 1994.

[Xie 89] M. Xie : *Contribution à la vision dynamique : reconstruction 3D d'objets polyédriques par une caméra mobile*, Thèse de l'Université de Rennes 1, IRISA, Juin 1989.

[Ye 96] Y. Ye, J. Tsotsos : Sensor planning in 3D object search. *Int. Symposium on Intelligent Robotic Systems, IRS'96*, Lisbonne, Portugal, 1996.

[Zhao 94] C. Zhao, R. Mohr : Relative 3D regularized B-Splines surface reconstruction through image sequences. *European Conf. on Computer Vision, ECCV'94*, Stockhölm, Suède, LNCS 802, pp. 417-426, Mai 1994.

[Zhang 95] Z. Zhang : Estimating motion and structure from correspondences of line segments between two perspective images. *IEEE Trans. on PAMI*, 17(12):1129-1139, Décembre 1995.