

HABILITATION À DIRIGER DES RECHERCHES

présentée devant

**L'Université de Rennes 1**  
**Spécialité : Informatique**

par

Christophe Collewet

Vers des asservissements visuels sur des scènes inconnues

soutenue le 30 Septembre 2008 devant le jury composé de

M.	Michel	Dhome	Président
MM.	Seth	Hutchinson	Rapporteurs
	El Mustapha	Mouaddib	
	Patrick	Rives	
MM.	Patrick	Bouthemey	Examineurs
	François	Chaumette	
	Alain	Trémeau	



---

## Table des matières

---

<b>Remerciements</b>	<b>5</b>
<b>Avant-propos</b>	<b>7</b>
<b>Notations utilisées</b>	<b>9</b>
<b>1 Introduction</b>	<b>11</b>
<b>2 L'asservissement visuel</b>	<b>19</b>
2.1 Contrôler les 6 degrés de liberté d'un porteur . . . . .	20
2.1.1 Modélisation . . . . .	20
2.1.2 Extraction, mise en correspondance et suivi d'informations visuelles	21
2.1.3 Lois de commande . . . . .	22
2.1.4 Stabilité . . . . .	24
2.1.5 Choix des informations visuelles . . . . .	25
2.2 Contrôler $m$ degré(s) de liberté d'un porteur ( $m < 6$ ) . . . . .	27
2.3 Conclusion . . . . .	28
<b>3 Vers une indépendance vis-à-vis de l'objet observé</b>	<b>29</b>
3.1 Indépendance vis-à-vis de la texture de l'objet . . . . .	30
3.1.1 Utilisation d'un éclairage structuré . . . . .	31
3.1.1.1 Utilisation d'un éclairage déporté et codé . . . . .	33
3.1.1.2 Utilisation d'un éclairage embarqué . . . . .	35
3.1.2 Utilisation de la luminance . . . . .	40
3.2 Indépendance vis-à-vis de la texture de l'objet et de sa géométrie . . . . .	43

<b>4</b>	<b>Vers une indépendance vis-à-vis de la scène observée</b>	<b>51</b>
4.1	Modélisation de l'interaction lumière / matière . . . . .	52
4.1.1	Modélisation de la réflexion . . . . .	52
4.1.2	Modélisation de la luminance . . . . .	54
4.1.3	Modélisation de la couleur . . . . .	54
4.2	Application à la mise en correspondance de motifs texturés . . . . .	55
4.2.1	Utilisation de la luminance . . . . .	56
4.2.2	Utilisation de la couleur . . . . .	60
4.3	Application à l'asservissement visuel basé sur la luminance . . . . .	62
<b>5</b>	<b>Conclusion et perspectives</b>	<b>67</b>
	<b>Bibliographie</b>	<b>73</b>

---

## Remerciements

---

Écrire un document d'habilitation est une opportunité de se plonger dans le passé. Pour un ingénieur de formation que je suis, il faut bien avouer que le domaine de la recherche était un domaine bien mystérieux à mes yeux. C'est ainsi que je remercie tout d'abord ceux qui m'ont fait découvrir ce monde et qui m'ont donné l'envie de m'y investir. Il s'agit en premier lieu de mon professeur d'automatique de DEA, Yves Lecluse, qui m'a accueilli également en stage, et de François Chaumette qui dirigea par la suite mes travaux de thèse. Je tiens à remercier également Marc Richetin de m'avoir décidé, certes sans trop de difficultés, à m'inscrire au concours de chargé de recherche lors de la titularisation des agents du Cemagref.

François Chaumette doit être remercié à plus d'un titre. Il fut en effet mon interlocuteur privilégié pendant toutes ces années. Je le remercie ici tout aussi bien pour ses qualités humaines que scientifiques. Je le remercie également de m'avoir accueilli en détachement pendant ces trois dernières années au sein de l'équipe projet Lagadic dont il est le responsable.

Je remercie bien évidemment mes anciens doctorants Ali Alhaj, Michèle Gouiffès et Jordi Pagès sans qui ce travail ne serait pas. Je remercie également Chafik Kermad, que j'ai eu le plaisir de diriger lors de son séjour post-doctoral au Cemagref et tous les stagiaires que j'ai encadrés aussi. Ils m'ont tous permis d'apporter une pierre à l'édifice.

Je tiens à exprimer aussi mes remerciements les plus sincères à Seth Hutchinson, professeur à l'université de l'Illinois à Urbana-Champaign, à El Mustapha Mouaddib, professeur à l'université de Picardie Jules Verne ainsi qu'à Patrick Rives, directeur de recherche au centre INRIA Sophia Antipolis - Méditerranée d'avoir accepté de rapporter ce travail. Je remercie également Patrick Bouthemy, François Chaumette, directeurs de recherche à l'INRIA Rennes - Bretagne atlantique, Alain Trémeau, professeur à l'université Jean Monnet de Saint-Étienne qui m'ont fait le plaisir de participer à ce jury d'habilitation et Michel Dhome, directeur de recherche au CNRS qui m'a fait l'honneur de le présider.

La rédaction de ce document coïncide également avec la fin de mon détachement dans l'équipe projet Lagadic. Je profite de cette occasion pour en remercier tous ses membres et anciens membres que j'ai pu apprécier durant ces trois années et en particulier Éric pour ses conseils avisés et les échanges enrichissants que nous avons eus. Je les remercie tous, pour toutes les discussions passionnantes que nous avons eues, pour leur bonne humeur mais également pour la bonne ambiance qu'ils contribuent, je ne sais trop par quelle alchimie d'ailleurs, à faire régner au sein de ce projet.

Enfin, je remercie Chloé, Maël et Laurick, mes enfants, et Guylaine, mon épouse, d'avoir accepté mon manque de disponibilité lors de certainement trop nombreuses soirées ou week-ends. C'est à eux que je dédie ce document.

---

## Avant-propos

---

Ce document décrit les travaux que j'ai effectués durant les périodes 1999-2005 au Cemagref et 2005-2007 à l'INRIA Rennes - Bretagne Atlantique et plus précisément au sein de l'équipe projet Lagadic.

Le Cemagref, bien qu'étant un ESPT<sup>1</sup>, se singularise fortement de ses homologues comme le CNRS ou l'INRIA par exemple. Il se dit « finalisé » dans le sens où il se restreint à des domaines d'applications bien précis. Dans ces domaines, il a un rôle de prospective et se veut à l'écoute des différents acteurs concernés (pouvoirs publics, centres techniques, industriels, équipementiers, etc.) pour définir des axes de recherche prioritaires. Cette façon de fonctionner n'est malheureusement pas sans conséquence quant à la possibilité de mener une recherche de haut niveau. Le principal obstacle concerne la pérennité des travaux, beaucoup trop courte à mon sens (de l'ordre de trois à cinq ans).

C'est dans ce contexte que ce document tente de synthétiser les recherches que j'ai menées, tout en s'efforçant de dégager un fil conducteur. C'est ainsi que certains travaux n'y sont donc volontairement pas présents, non pas qu'ils soient inintéressants mais ils nuiraient malheureusement à la clarté du document. Ils concernent en particulier le traitement d'images, et plus précisément la sélection de points d'intérêt, dès l'image initiale, dans une optique de suivi d'une telle primitive [*Kermad et Collewet, 2001; Kermad et al., 2002*], la segmentation d'images couleur basée sur l'étude des phénomènes de la transparence [*Gouiffès et al., 2004a; Gouiffès et al., 2004b*], l'extraction et le suivi de contours complexes par contour actif [*Mebarki et al., 2008; Dune et al., 2008*] ainsi que le décodage robuste de motifs de lumière structurée et codée [*Pagès et al., 2005b*].

D'un point de vue pratique, précisons que les références à nos travaux sont signalées en italique.

---

<sup>1</sup>Établissement Public à caractère Scientifique et Technique.





---

## Notations utilisées

---

### Notations générales

- $a, \lambda$  : scalaire noté en minuscule
- $\mathbf{v}$  : vecteur noté en minuscule gras,  $v_i$  désigne la  $i$ -ème composante du vecteur  $\mathbf{v}$
- $\mathbf{M}$  : matrice notée en majuscule gras,  $M_{ij}$  désigne l'élément situé à l'intersection de la  $i$ -ème ligne et de la  $j$ -ème colonne de la matrice  $\mathbf{M}$
- $[\mathbf{v}]_{\times}$  matrice de préproduit vectoriel associée au vecteur  $\mathbf{v}$
- $\mathbf{M}^+$  pseudo-inverse de la matrice  $\mathbf{M}$
- $P$  désigne un point de la scène, il est décrit par le vecteur  $\mathbf{P} = (X_P, Y_P, Z_P)$ . Il se projette dans le capteur en  $p$ , décrit en coordonnées projectives normalisées par le vecteur  $\mathbf{x}_p = (x_p, y_p, 1)$ .
- $M$  est un autre point de la scène, on introduit de façon similaire le vecteur  $\mathbf{X} = (X, Y, Z)$ . Il se projette dans le capteur en  $m$ , décrit en coordonnées projectives normalisées par le vecteur  $\mathbf{x} = (x, y, 1)$ .
- L'estimation d'une grandeur (scalaire ou matricielle) est indiquée grâce au symbole  $\hat{\cdot}$ , par exemple  $\hat{a}$  indique une estimation de la grandeur  $a$ .

### Changements de repère – transformations

- ${}^a\mathbf{M}_b$  : matrice homogène définie par

$${}^a\mathbf{M}_b = \begin{pmatrix} {}^a\mathbf{R}_b & {}^a\mathbf{t}_b \\ \mathbf{0}_{1 \times 3} & 1 \end{pmatrix}$$

où  ${}^a\mathbf{R}_b$  et  ${}^a\mathbf{t}_b$  sont respectivement la matrice de rotation et le vecteur de translation exprimant la transformation entre les repères  $\mathcal{F}_a$  et  $\mathcal{F}_b$ .  ${}^a\mathbf{M}_b$  décrit la localisation du repère  $\mathcal{F}_b$  exprimé dans le repère  $\mathcal{F}_a$ .

- ${}^c\mathbf{M}_o$  : Cas particulier de la matrice de pose :  ${}^c\mathbf{M}_o$  décrit la localisation du repère de la scène dans le repère de la caméra.

### Asservissement visuel

- $\mathbf{v}$  : vitesse de la caméra exprimée dans son repère,  $\mathbf{v} = (\mathbf{v}, \boldsymbol{\omega})$  où :
  - $\mathbf{v} = (v_x, v_y, v_z)$  : vitesse de translation
  - $\boldsymbol{\omega} = (\omega_x, \omega_y, \omega_z)$  : vitesse de rotation
- $\mathbf{s}$  : informations impliquées dans la loi de commande
- $\mathbf{L}_s$  : matrice d'interaction relative à  $\mathbf{s}$
- Les grandeurs calculées en position désirée sont notées à l'aide du symbole \*, par exemple :  $\mathbf{s}^*$  ou  $\mathbf{L}_s^*$ .

# CHAPITRE 1

---

## Introduction

---

Nos travaux concernent le domaine de *l'asservissement visuel*. D'une façon générale, cette technique permet de contrôler le déplacement d'un système dynamique à l'aide d'une ou plusieurs caméras, embarquées ou non [Chaumette et Hutchinson, 2006; Chaumette et Hutchinson, 2007], le système considéré, tout comme la scène observée, pouvant être réel ou virtuel [Marchand, 2004]. Le plus souvent, on s'emploiera en fait à commander un *robot*. Cette approche relève donc naturellement du domaine de la *robotique*. Les non-spécialistes du domaine n'y verront à notre avis pas beaucoup plus clair sans une définition plus précise de ce que l'on entend, en robotique, par le mot « robot » ; le terme robotique étant à accepter ici au sens de discipline scientifique.

En effet, pour le plus grand nombre, il s'agira d'une machine à l'aspect humain destinée au travail. C'est en fait le sens primaire tel qu'il est apparu pour la première fois en 1921 dans la pièce *RUR* de son auteur tchèque Karel Čapek. Ce sigle désignait l'entreprise Rosum's Universal Robots, qui fabriquait à la chaîne des humains organiques artificiels destinés à faire le travail à la place des hommes. Le mot « robot » est donc un néologisme. Il est toutefois issu du tchèque *robota* (corvée) et de même racine que du russe *rabota* (travail) ou *robotnik* (travailleur)<sup>1</sup>. Le terme « robotique », quant à lui, a été introduit dans la littérature par Isaac Asimov en 1942 dans le livre *Runaround*. Il y énonce en particulier les trois célèbres lois de la robotique. Les progrès technologiques aidant, peu à peu la lit-

---

<sup>1</sup>Certains y voient aussi un lien avec les racines « rab », « rob » ou « rabu » de diverses langues slaves signifiant serf ou esclave. Le lien étymologique entre « travail » et « esclave » se retrouve également dans d'autres langues. En grec moderne, par exemple : *douleia*, pour le travail, et *doulos*, pour l'esclave. Plus révélateur encore, *douleu* signifiait « être esclave » en grec ancien et signifie « travailler » aujourd'hui. Concernant le français, « travail » est issu de l'ancien français « tourment », « souffrance » et du latin *tripalium*, ce fameux instrument de torture à trois pieux... (Propos tenus par les romanciers Rémi Bertrand et Alain Van Crugten recueillis sur le site internet de Vincent Engel (<http://www.vincent-engel.com/>), lui-même romancier).

térature et le cinéma de science-fiction conduiront à une autre définition du mot « robot », sans toutefois exclure sa première acception, un robot pouvant être aussi un « humain » entièrement mécanique, c'est-à-dire doté des mêmes facultés de mouvement, de perception et de prise de décisions. Il va sans dire que cette définition ne reflète en rien la réalité et n'est donc aucunement satisfaisante.

Dans l'industrie, le sens le plus conventionnel se rapportera plutôt à celui d'automate programmable, c'est-à-dire à un système mécanique où les tâches à effectuer sont apprises puis répétées quel que soit son environnement. On parle le plus souvent même de « robot industriel ». On le rencontre dans l'industrie manufacturière ; l'exemple le plus connu étant celui de l'industrie automobile où les tâches d'assemblage, de soudage et de peinture sont effectuées exclusivement maintenant par des robots. On retrouve en somme la définition primaire du mot « robot », toute la nuance étant à présent dans la nature même des tâches à effectuer. On y perçoit de ce fait émerger une notion d'autonomie pour réaliser les dites tâches.

La robotique, en tant que discipline scientifique, vise finalement à doter un système mécanique du degré d'autonomie nécessaire à la réalisation d'une tâche donnée. Ce système mécanique se verra donc équipé d'un certain nombre de capteurs, couplés à une certaine puissance de calcul pour exploiter, en temps réel, les mesures fournies par ses capteurs. Ces capteurs sont principalement de deux types :

- *proprioceptifs*, ils doivent permettre en premier lieu de remonter à l'état complet du système à commander. Néanmoins, ils peuvent s'avérer suffisants dans certains cas pour réaliser des tâches très simples comme le suivi d'une trajectoire apprise qu'il s'agira de répéter par exemple.
- *extéroceptifs*, ils doivent permettre d'appréhender l'environnement extérieur au système dans lequel il est amené à évoluer et lui assurer ainsi le degré d'autonomie recherché.

Fortes des informations issues des différents capteurs, la *loi de commande* se chargera, quant à elle, de modifier l'état du système pour l'amener d'un état initial à un état désiré, tout en assurant à la fois son intégrité et celle de son environnement. Environnement à prendre au sens le plus large, il peut s'agir en effet d'autres robots, d'obstacles, éventuellement mobiles, voire d'humains, s'il est amené à interagir ou à coopérer avec eux comme ce peut être le cas en robotique de service.

Les travaux décrits dans ce manuscrit portent d'une façon générale sur les asservissements visuels, c'est-à-dire sur la synthèse de schémas de commande d'un « robot » par bouclage d'informations issues d'un capteur de type caméra vidéo<sup>2</sup>. Plus précisément, ce type d'approche repose sur l'élaboration d'une commande permettant d'amener le robot de sa situation initiale à sa situation désirée (éventuellement fonction du temps pour une

<sup>2</sup>Signalons que ces techniques d'asservissement visuel peuvent également appréhender d'autres capteurs plus complexes comme les capteurs échographique [Krupa et Chaumette, 2006] ou omnidirectionnel [Barreto *et al.*, 2002] par exemple. En outre, elles peuvent très bien se transposer à d'autres capteurs, comme à des capteurs proximétriques pour la commande de robots mobiles [Samson *et al.*, 1991] ou à des capteurs d'altitude utilisés en robotique sous-marine [Santos *et al.*, 1995], on parle alors de commande référencée capteur. Historiquement, c'est même la commande référencée capteur qui est à l'origine de la commande référencée « vision » [Chaumette, 1990].

trajectoire) grâce à des informations  $s$  et  $s^*$  fournies respectivement par une (ou plusieurs) caméra(s) embarquée(s) ou déportée(s) en situations courante et désirée.

Si le capteur « caméra » se particularise par rapport à d'autres capteurs utilisés en robotique (comme par exemple des capteurs proximétriques de type télémètre) de par sa grande richesse d'informations qu'il fournit, et cela à des cadences qui peuvent être maintenant très élevées [Nakabo *et al.*, 2000; Ginhoux *et al.*, 2004], son utilisation n'est toutefois pas sans poser de problèmes théoriques inhérents à sa façon même de percevoir son environnement :

- En premier lieu, ce capteur ne délivre qu'une information bidimensionnelle d'un monde tridimensionnel. En effet, l'image de la scène est obtenue par le biais d'une projection de l'espace 3D sur le plan 2D du capteur. De ce fait, on a longtemps pensé que la reconstruction 3D de la scène observée était un préalable à tout mouvement d'un robot<sup>3</sup>. En fait, comme on le verra dans les prochains chapitres, il est tout à fait possible de s'affranchir de cette phase de reconstruction 3D en utilisant exclusivement des informations visuelles  $s$  issues directement du capteur, donc de type 2D (comme les coordonnées de points par exemple). Il s'agira alors de garantir que la trajectoire du robot générée à partir de ces informations visuelles est satisfaisante.
- En outre, le champ de vision d'un tel capteur est limité. Il faudra donc s'assurer, en premier lieu, que la scène est visible depuis la position initiale du robot. Quel que soit le type d'informations  $s$  choisi (2D, 3D, 2D1/2), le schéma de commande doit ensuite garantir que les mouvements générés par la loi de commande conduiront à conserver dans le champ de vision de la caméra les informations visuelles permettant le calcul de  $s$ , sous peine de mise en échec de la tâche à réaliser.
- La connaissance des informations visuelles désirées  $s^*$  peut s'avérer être également problématique, en particulier dans le cas d'informations de type 2D. Elles sont généralement obtenues par le biais d'un apprentissage, quand celui-ci est possible à réaliser. Dans ce cas, le robot est amené à sa position désirée et la caméra acquiert simplement l'image de la scène. Sinon, elles sont estimées grâce à une projection d'un modèle de l'objet à la pose désirée. Cette approche repose, bien entendu, sur une modélisation précise de l'objet et du dispositif de projection ainsi que sur son étalonnage parfait.
- L'information fournie par un tel capteur est une information de luminance alors que ce qui intéresse le concepteur de la loi de commande est le plus souvent une information de type géométrique. Il s'en suit nécessairement d'importants travaux en *traitement d'images* pour extraire les informations  $s$  à partir de l'image acquise en position initiale, pour les mettre en correspondance avec les informations désirées  $s^*$  mais aussi, et surtout, pour maintenir cette mise en correspondance tout au long du mouvement du robot. Ces différentes tâches ne sont pas simples, en particulier dans le cas où des variations de luminance doivent être considérées, soit parce que la scène n'est pas Lambertienne et / ou parce qu'elle subit des variations de l'éclairage ambiant, voire parce qu'elle se déplace par rapport à la source d'éclairage.

Par ailleurs, mais comme c'est le cas dans tout problème d'asservissement, il faudra également se poser le problème de la stabilité du système bouclé par les informations  $s$ ,

<sup>3</sup>Cette façon de raisonner a conduit à la fin des années 70 aux travaux organisés autour de la structure bien connue « Look and Move » [Rosen, 1976; Tani *et al.*, 1977].

stabilité globale ou pour le moins locale, et cela en présence d'erreurs de mesure ou d'éventuelles imprécisions des modèles utilisés (modèle de la caméra, modèle du robot, modélisation imprécise des informations ou encore estimation erronée de certains paramètres, comme la profondeur de l'objet par exemple). Le concepteur de la loi de commande devra aussi garantir l'absence de singularité dans l'espace de travail tout comme l'absence de minima locaux. Il devra, enfin, assurer un bon découplage des informations  $s$  avec les variables de contrôle garantissant une trajectoire du robot satisfaisante.

Finalement, le but de tout asservissement visuel consiste à sélectionner des mesures décrites par  $s$  et à élaborer une loi de commande permettant le contrôle des degrés de liberté souhaités tout en assurant les différentes contraintes énoncées ci-dessus... Le problème n'est donc assurément pas simple.

**Contributions.** Nos contributions au domaine de l'asservissement visuel ont débuté par le biais d'un projet inscrit au contrat de plan État-Région Bretagne 1994-1998 lors d'une collaboration entre l'INRIA (plus précisément le projet TEMIS) et le Cemagref de Rennes<sup>4</sup>. C'est dans ce cadre que j'ai effectué mes travaux de thèse [Collewet, 1999]. À cette époque, les recherches dans ce domaine portaient, de façon très générale, sur la compréhension et l'amélioration du comportement des différentes techniques d'asservissement visuel et, de ce fait, des scènes très simples étaient utilisées pour s'affranchir des problèmes de traitement d'images que nous avons évoqués plus haut (voir cependant les travaux rapportés dans [Crétual, 1998], également réalisés dans le cadre du projet VIDAC). C'est ainsi que le plus souvent des objets de forme plane et dotés de marqueurs étaient appréhendés ou, le cas échéant, des objets parfaitement connus par le biais d'un modèle CAO par exemple [Marchand *et al.*, 2001; Drummond et Cipolla, 2002]. Le champ applicatif considéré était alors, par voie de conséquence, restreint à celui de l'industrie manufacturière.

De par les domaines d'applications traités au Cemagref, nous avons été confrontés à cette époque à d'autres types de scènes, en particulier issues des secteurs agricole et agroalimentaire. Ce changement du domaine applicatif ne fut pas anodin car se posa très rapidement un certain nombre de problèmes théoriques<sup>5</sup>. Ils peuvent se résumer en se posant la question suivante : *comment réaliser (par asservissement visuel) une tâche de positionnement précis vis-à-vis d'un objet mal connu ?*

C'est en effet le cas de figure que nous avons rencontré dans bon nombre d'applications. Citons par exemple la robotisation de la traite des vaches laitières, où il s'agit de poser quatre gobelets trayeurs sur chacun des trayons de l'animal ; les applications de traçabilité dans les abattoirs où il faut rechercher un tatouage sur une pièce de viande puis, une fois localisé, réaliser une tâche de focalisation pour assurer sa lecture par reconnaissance optique. La même problématique fut également présente pour la réalisation de tâches de caractérisation de produits de type agroalimentaire ; une caméra devait se déplacer par rapport à des objets circulant en vrac sur des convoyeurs pour effectuer certaines mesures (géométriques ou colorimétriques par exemple, en rapport avec la qualité du produit) et aboutir de cette façon à un tri de ces produits. Dans ces différents cas de figure, il s'agissait

<sup>4</sup>Projet VIDAC : VIsion Dynamique Active et Communication : applications à la télésurveillance et à l'industrie agroalimentaire.

<sup>5</sup>Plus généralement, c'est également le cas en robotique non manufacturière [Dauchez, 2000].

en fait de déplacer par asservissement visuel un bras porteur de la caméra et de réaliser ainsi une tâche de positionnement par rapport à un objet mal connu. Ce contexte applicatif explique par ailleurs le fait que nos travaux se restreignent au cas d'une caméra embarquée et jamais déportée. Ils expliquent également pourquoi nous n'avons pas abordé le domaine de la robotique mobile.

S'efforcer de répondre à la question que nous nous sommes posée à la page précédente, à savoir : « *comment réaliser (par asservissement visuel) une tâche de positionnement précis vis-à-vis d'un objet mal connu ?* » fut en fait le fil conducteur des différents travaux rapportés dans ce document. Diverses approches ont été proposées, elles reposent cependant toutes sur la même idée générale : s'affranchir le plus possible de la scène observée ou, en d'autres termes encore, assurer une indépendance la plus grande possible vis-à-vis de cette scène<sup>6</sup>. Dans ces conditions, nos contributions ont porté sur les thèmes suivants :

- En premier lieu la texture des objets à traiter peut être inconnue, ou même trop complexe à exploiter, de telle sorte qu'il faille considérer le contenu de l'image comme inconnu. Du point de vue de l'asservissement visuel, il n'est alors plus permis de faire l'hypothèse de la présence de tel ou tel indice visuel dans l'image. Une manière d'aborder ce problème consiste à proposer certaines informations visuelles non géométriques. En effet, cette façon de procéder conduit à une approche beaucoup moins dépendante de l'apparence des objets observés. De plus, elle évite les différentes phases de mise en correspondance, toujours délicates à réaliser. Deux types d'informations visuelles ont été investigués. Dans le cadre de la thèse d'Ali Alhaj [Alhaj, 2004], nous avons proposé l'utilisation d'informations visuelles basées sur les paramètres d'un modèle de mouvement affine. Lors de ma deuxième année de détachement au sein du projet Lagadic de l'INRIA, nous avons utilisé directement l'information de luminance en supposant la luminance constante pendant toute la durée de la tâche à effectuer.

De même, une autre approche permettant de s'affranchir de la texture de l'objet concerne le couplage de l'asservissement visuel et de la lumière structurée. Dans ce cas, les informations visuelles sont apportées directement par le motif projeté. Nous nous sommes focalisés sur de telles approches dans le cadre de la thèse de Jordi Pagès [Pagès, 2005] en considérant tout aussi bien les cas où l'éclairage était déporté qu'embarqué. Dans le cas déporté, une matrice codée constituée de points lumineux a permis de s'affranchir, de surcroît, de tout algorithme de mise en correspondance. Quant au cas embarqué, un capteur basé sur quatre faisceaux laser a rendu les phases d'extraction et de suivi des indices visuels triviales. Nous avons ensuite conçu la géométrie du capteur de manière à assurer un bon comportement du système bouclé, tant du point de vue de sa stabilité que des trajectoires générées.

- L'objet étant inconnu, il est légitime de considérer le cas où l'image désirée est inconnue et, *a fortiori*, n'importe quelle primitive géométrique 2D en position désirée. En

<sup>6</sup>Certains pourront y voir un conflit avec le concept prôné en vision active [Aloimonos *et al.*, 1987; Bajcsy, 1988; Ballard, 1991] ou en vision intentionnelle [Aloimonos, 1990] selon lequel une solution moins générique, mais plus efficace, doit être recherchée. Néanmoins, ce concept exige, par essence, un but bien précis ainsi qu'un cadre bien déterminé [Aloimonos, 1990]. Les applications que nous venons de citer ne vérifient malheureusement pas ce prérequis.

effet, comme nous l'avons déjà évoqué, ces primitives sont généralement obtenues par apprentissage ou grâce à un modèle 3D de l'objet. Dans le cas des applications qui nous ont intéressées au Cemagref, la variabilité biologique des produits d'une même classe impliquait un positionnement du capteur adapté à chaque produit, un apprentissage n'était donc plus possible. Il s'agit là d'une distinction notable avec les objets rencontrés dans l'industrie manufacturière où, au contraire, ils se doivent de respecter des contraintes strictes de fabrication, en particulier de type géométrique, autorisant ainsi un unique apprentissage valide pour tous les objets d'une même classe. Nous nous sommes donc intéressés au cours de la thèse d'Ali Alhaj à l'utilisation d'une technique de reconstruction 3D locale par *vision dynamique* permettant de s'affranchir de la connaissance de cette image. Son couplage avec la loi de commande a permis de s'assurer, en pratique, que la zone d'intérêt reste visible tout au long de la réalisation de la tâche.

- La géométrie de l'objet peut également être inconnue. Lors de ma première année de détachement au sein du projet Lagadic, nous avons proposé une approche unifiée ne nécessitant pas cette connaissance. Elle est unifiée dans le sens où elle a conduit à la réalisation de tâches de positionnement par rapport à un objet de forme quelconque, aussi bien plan que non, sans avoir, bien évidemment, à passer d'un modèle de l'objet à un autre comme cela a pu être proposé dans la littérature (voir par exemple [Irani et Anandan, 1998; Schindler et Suter, 2006]). Différentes solutions ont été proposées pour garantir la robustesse de notre approche à la forme de l'objet observé.
- Outre les propriétés physiques de l'objet lui-même qui peuvent être inconnues, nous avons considéré que la géométrie de la scène pouvait l'être également, voire qu'elle pouvait être dynamique. Précisons que l'on entend ici par géométrie de la scène, les positions relatives du triplet observateur / objet / éclairage. Considérer une scène dynamique est donc tout à fait légitime dans un contexte d'asservissement visuel. C'est ainsi que nous nous sommes focalisés dans le cadre de la thèse de Michèle Gouiffès [Gouiffès, 2005] sur la modélisation des phénomènes de réflexion. Ce travail a conduit à l'élaboration d'un algorithme de suivi de motifs texturés à partir d'informations de luminance, mais aussi de chrominance, robuste à toute modification de cette géométrie, permettant ainsi une indépendance vis-à-vis de variations de luminance, qu'elles soient dues à l'apparition de spéularités ou même, à un changement d'irradiance de la scène. En outre, à l'occasion de mon détachement au sein du projet Lagadic, nous avons mis à profit cette modélisation pour introduire directement la luminance dans une loi de commande, tout en nous permettant de prendre en compte le caractère dynamique de la scène.

Comme le lecteur peut le constater, les travaux présentés ci-dessus conduisent à une certaine forme d'indépendance. Elle s'exerce néanmoins à divers degrés. Nous avons donc organisé nos contributions suivant ce degré d'indépendance. Ce document se structure ainsi en trois chapitres :

- Le chapitre 2 rapporte tout d'abord les différentes techniques se réclamant de l'asservissement visuel ainsi que les fondements théoriques sur lesquels elles reposent.
- Le chapitre 3 présente nos travaux visant à s'affranchir de l'objet d'intérêt lui-même, c'est-à-dire du contenu des images issues de sa projection sur le capteur de la caméra



mais également de sa géométrie.

- De façon plus générale, le chapitre 4 regroupe nos contributions nous permettant de nous affranchir de la scène en elle-même ; c'est-à-dire de l'objet mais surtout de l'environnement dans lequel il est amené à évoluer et ainsi de pouvoir prendre en compte explicitement le caractère dynamique de cette scène.

Ce document s'achève par un chapitre de conclusion, nous y évoquons également quelques perspectives de recherche.



## CHAPITRE 2

---

### L'asservissement visuel

---

Comme évoqué dans le chapitre introductif à ce manuscrit, le couplage entre une caméra et un robot permet de le déplacer d'une position courante à une position désirée vis-à-vis d'une scène d'intérêt. Les techniques utilisées relèvent de l'asservissement visuel, c'est-à-dire de la commande d'un système en boucle fermée, ici le robot, grâce à des indices visuels extraits d'images acquises par la caméra.

Bien sûr, des approches en boucle ouverte existent [Rosen, 1976; Tani *et al.*, 1977], mais elles sont bien évidemment très peu robustes car beaucoup trop sensibles aux erreurs de calibration de l'ensemble robot / caméra. La première utilisation de la vision en boucle fermée est due à Shirai et Inoue [Shirai et Inoue, 1973] qui décrivent comment un capteur de vision pouvait limiter l'influence de ce type d'erreur et ainsi augmenter la précision du positionnement. On parlait alors de retour par vision mais, c'est à Agin [Agin, 1979]<sup>1</sup> que l'on doit l'apparition du terme asservissement visuel. Depuis, cette approche a reçu ses lettres de noblesse et joue maintenant un rôle important dans la communauté robotique. En effet, les asservissements visuels ont montré qu'ils étaient à même d'appréhender des applications ou des environnements très complexes comme la conduite automobile autonome, la commande de robots sous-marins, de robots à pattes (bipèdes voire plus) ou humanoïdes, de robots pour le handicap, de robots volants comme des dirigeables ou des drones (de type mini avion ou hélicoptère) et de robots médicaux, en plein essor actuellement. Le lecteur trouvera de plus amples descriptions en consultant les actes des différentes Journées Nationales de la Recherche en Robotique (JNRR) par exemple.

Nous décrivons maintenant les fondements sur lesquels s'appuient les techniques d'asservissement visuel<sup>2</sup> et renvoyons le lecteur intéressé par de plus amples détails aux tuto-

---

<sup>1</sup> Cet article a longtemps été attribué par erreur à Hill et Park (voir [Agin, 1985]).

<sup>2</sup> Toutefois, comme évoqué dans le chapitre d'introduction, nous restreignons notre description au cas d'une

riels rapportés dans [Hutchinson *et al.*, 1996; Chaumette et Hutchinson, 2006; Chaumette et Hutchinson, 2007].

Dès à présent, nous faisons la distinction entre contrôler les 6 degrés de liberté du porteur, ce qui permet de réaliser une liaison rigide entre la caméra et la scène, et ne contrôler que  $m$  degrés de liberté ( $m < 6$ ) et ainsi en laisser libres  $6 - m$  pour éventuellement réaliser une autre tâche selon le formalisme de redondance [Espiau *et al.*, 1992] comme nous le verrons ci-après.

## 2.1 Contrôler les 6 degrés de liberté d'un porteur

Le plus souvent, le but de tout asservissement visuel consiste à contrôler les 6 degrés de liberté du porteur en minimisant le signal d'erreur :

$$\mathbf{e} = \mathbf{s}(\mathbf{r}(t), t) - \mathbf{s}^* \quad (2.1)$$

où  $\mathbf{s}$  est un vecteur de dimension  $k$  ( $k \geq 6$ ) décrivant  $k$  informations obtenues pour la situation courante  $\mathbf{r}(t)$  du capteur vis-à-vis de la scène ( $\mathbf{r}$  est donc un élément de  $SE(3)$ ),  $\mathbf{s}^*$  décrivant ces mêmes informations mais obtenues en situation désirée  $\mathbf{r}^*$  ( $\mathbf{s}^* = \mathbf{s}(\mathbf{r}^*)$ ). Comme déjà mentionné dans le chapitre introductif, rappelons que l'obtention de  $\mathbf{s}^*$  n'est pas forcément triviale comme il peut y paraître.

À ce stade, on comprend intuitivement que l'élaboration d'une loi de commande en boucle fermée sur  $\mathbf{s}$  nécessite de connaître la relation qui lie sa variation à celle des variables de commande. Une phase de modélisation de l'interaction du capteur avec l'environnement qu'il visualise s'avère donc nécessaire.

### 2.1.1 Modélisation

Pour ce faire, écrivons la dérivée temporelle de  $\mathbf{s}$  :

$$\dot{\mathbf{s}} = \frac{\partial \mathbf{s}}{\partial \mathbf{r}} \dot{\mathbf{r}} + \frac{\partial \mathbf{s}}{\partial t} = \mathbf{L}_s \mathbf{v} + \frac{\partial \mathbf{s}}{\partial t} \quad (2.2)$$

où  $\mathbf{L}_s$  est la matrice d'interaction bien connue [Espiau *et al.*, 1992], elle est de dimension  $k \times 6$ ;  $\mathbf{v}$  est la vitesse de la caméra exprimée dans son repère, elle peut être décomposée en une composante de translation  $\mathbf{v} = (v_x, v_y, v_z)$  et de rotation  $\boldsymbol{\omega} = (\omega_x, \omega_y, \omega_z)$  telles que  $\mathbf{v} = (\mathbf{v}, \boldsymbol{\omega})$ . Le vecteur  $\partial \mathbf{s} / \partial t$  décrit, quant à lui, un éventuel mouvement de la scène observée.

Comme on le verra ci-après, la matrice d'interaction joue un rôle fondamental dans la synthèse de la loi de commande. Aussi a-t-elle été calculée sous forme analytique pour un grand nombre de primitives, aussi bien de type 2D que 3D.

caméra embarquée, la transcription au cas d'une caméra déportée se déduisant généralement facilement du cas précédent. Néanmoins, nous invitons le lecteur intéressé par ce type de configuration à se référer à [Hager *et al.*, 1995; Dornaika, 1995; Horaud *et al.*, 1998; Ruf et Horaud, 1999], voire à [Flandin, 2001] pour la coopération entre une caméra déportée et une caméra embarquée.

- Pour des informations de type 2D : elle a bien sûr été calculée tout d'abord pour la primitive point [Weiss *et al.*, 1987; Espiau *et al.*, 1992] mais elle a pu être également calculée pour toute primitive géométrique paramétrable [Espiau *et al.*, 1992], comme par exemple la droite ou l'ellipse ou pour des primitives plus complexes comme la primitive contour [Colombo et Allotta, 1999; Drummond et Cipolla, 1999], [Collewet et Chaumette, 2000] ou les moments [Chaumette, 2004; Tahri et Chaumette, 2005][Mebarki *et al.*, 2008].
- Pour des informations de type 3D : elle a été obtenue de façon triviale dans le cas d'un point 3D. Elle a été également calculée pour la primitive plan [Motyl, 1992], [Pagès *et al.*, 2006c], pour des primitives associées à une rotation [Malis *et al.*, 1999] ou pour d'autres informations plus complexes localement isomorphes à  $SE(3)$  [Benhimane et Malis, 2007].

Dans le cas d'informations très complexes, et c'est à nos yeux le seul cas où de telles approches se justifient, cette matrice peut-être obtenue par apprentissage, donc sous une forme numérique cette fois. Effectivement, cette façon de faire interdit toute analyse de la stabilité du système bouclé. Par exemple dans [Nayar *et al.*, 1996; Deguchi, 2000], les auteurs ont utilisé pour informations visuelles les composantes de l'image projetée sur une base d'images obtenues par analyse en composante principale de l'image désirée. Elle peut être également estimée en ligne comme dans [Hosoda et Asada, 1994; Jägersand *et al.*, 1997; Piepmeyer *et al.*, 2004] ou directement son inverse [Suh et Kim, 1993] ou sa pseudo-inverse [Lapresté *et al.*, 2004] qui donnent de meilleurs résultats en pratique.

*Remarque* : Il est également possible de considérer des informations visuelles dynamiques, comme les paramètres d'un modèle de mouvement 2D [Odobez et Bouthemy, 1995]. Le principe de la commande consiste alors à contrôler le mouvement de la caméra de telle sorte que le champ de vitesse 2D mesuré atteigne un champ de vitesse désiré [Santos-Victor et Sandini, 1997; Crétual et Chaumette, 2001].

### 2.1.2 Extraction, mise en correspondance et suivi d'informations visuelles

Ces différentes phases sont en fait des points clés sur lesquels repose tout asservissement visuel, que les informations visuelles soient de type 2D ou 3D. En effet, elles sont issues de mesures  $\mathbf{x}(\mathbf{r})$  dans l'image (combinées dans le cas 3D avec des connaissances supplémentaires (paramètres intrinsèques de la caméra, modèle 3D de l'objet par exemple)). Dans tous les cas, une extraction robuste, une mise en correspondance (entre  $\mathbf{x}(\mathbf{r}_t)$  et  $\mathbf{x}^* = \mathbf{x}(\mathbf{r}^*)$  où  $t$  est l'instant discret courant) et un suivi spatio-temporel (entre  $\mathbf{x}(\mathbf{r}_{t-1})$  et  $\mathbf{x}(\mathbf{r}_t)$ ) doivent être réalisés (voir par exemple [Marchand et Chaumette, 2005] pour un état de l'art récent de ce domaine).

Pour des informations visuelles de type 3D, une phase supplémentaire s'avère nécessaire pour aboutir à leur estimation. Dans le cas le plus général une phase de reconstruction 3D est alors utilisée, nous en reparlerons de façon détaillée au chapitre 3. En revanche, lorsqu'un modèle de l'objet et les paramètres de calibration de la caméra sont disponibles, cette estimation peut être obtenue via un calcul de pose, c'est-à-dire une estimation de  $\hat{\mathbf{r}}_t$  soit une

estimation de  ${}^c\mathbf{R}_o$  et de  ${}^c\mathbf{t}_o$ . C'est en fait le cas le plus fréquent rencontré en asservissement visuel 3D. Notons que, très généralement, ces techniques sont de type numérique (voir par exemple [Lowe, 1987; Haralick *et al.*, 1989; Dementhon et Davis, 1995; Marchand et Chaumette, 2002]).

### 2.1.3 Lois de commande

Une fois les informations  $\mathbf{s}$  choisies, le vecteur de consigne  $\mathbf{s}^*$  connu et la matrice d'interaction obtenue, il est possible d'élaborer une loi de commande. Depuis l'article fondateur [Espiau *et al.*, 1992], l'approche la plus souvent retenue s'appuie sur la spécification d'un comportement désiré pour  $\mathbf{e}$  :

$$\dot{\mathbf{e}} = -\lambda\mathbf{e}, \quad (2.3)$$

c'est-à-dire sur une décroissance exponentielle découplée du signal d'erreur  $\mathbf{e}$  (le scalaire positif  $\lambda$  permettant de régler la vitesse de décroissance). En exprimant la dérivée temporelle de  $\mathbf{e}$ , le comportement souhaité (2.3) conduit à la relation suivante :

$$\dot{\mathbf{e}} = \frac{\partial \mathbf{e}}{\partial \mathbf{r}} \dot{\mathbf{r}} = \mathbf{L}_e \mathbf{v} = -\lambda\mathbf{e}, \quad (2.4)$$

dans laquelle le terme de déplacement libre  $\partial \mathbf{e} / \partial t$  a été omis pour alléger les expressions obtenues. Il vient alors la loi de commande :

$$\mathbf{v} = -\lambda \mathbf{L}_e^+ \mathbf{e}. \quad (2.5)$$

Toutefois, comme en pratique, seules des estimations de la matrice d'interaction sont disponibles, la loi de commande s'écrit :

$$\mathbf{v} = -\lambda \widehat{\mathbf{L}_e^+} \mathbf{e}. \quad (2.6)$$

Bien d'autres formulations existent cependant. Certains auteurs abordent le problème de l'élaboration de la loi de commande sous le regard de l'automatique traditionnelle. Par exemple, Weiss a développé une commande anticipative avec adaptation paramétrique du modèle du système à commander [Weiss *et al.*, 1987], Papanikolopoulos une commande de type PI, ou par placement de pôles [Papanikolopoulos *et al.*, 1991]. Plus généralement, des commandes optimales basées sur la minimisation d'un critère temps-énergie ont été mises en œuvre [Hashimoto *et al.*, 1996; Smith *et al.*, 1997] ou encore diverses commandes prédictives [Gangloff *et al.*, 1998; Allibert *et al.*, 2007].

Il est également possible de formuler le problème d'obtention d'une loi de commande en terme d'un problème d'optimisation en cherchant à minimiser, en fonction de  $\mathbf{r}$ , le critère

suivant<sup>3,4</sup> :

$$\mathcal{C}(\mathbf{r}) = \frac{1}{2} \|\mathbf{e}(\mathbf{r})\|^2. \quad (2.7)$$

Cette façon de faire a été retenue par exemple dans [Miura *et al.*, 2002] où la pose  $\mathbf{r}$  à atteindre est calculée à chaque itération  $t$  grâce à une méthode de simplex, tout en prenant en compte diverses contraintes assurant, par exemple, que l'objet d'intérêt ne sorte pas de l'image. Dans [Malis, 2004], des schémas de convergence différentiels ont été proposés, plus à même de se transcrire en loi de commande cinématique. Ils sont du type :

$$\mathbf{r} = \mathbf{r}_t + t_t \mathbf{d}(\mathbf{r}_t) \quad (2.8)$$

où « + » désigne ici l'opérateur consistant à combiner deux changements de repère ;  $\mathbf{r}_t$  exprime la pose courante ;  $t_t$  un scalaire positif (le pas de descente) et  $\mathbf{d}(\mathbf{r}_t)$  une direction de descente assurant la décroissance de (2.7). Dans ce cas, si  $t_t$  est suffisamment faible, de (2.8) on en déduit la loi de commande suivante :

$$\mathbf{v} = \lambda \mathbf{d}(\mathbf{r}) \quad (2.9)$$

où les indices  $t$  ont été omis.

Suivant la direction de descente utilisée, (2.6) se généralise ainsi à :

$$\mathbf{v} = -\lambda \hat{\mathbf{N}}_e \mathbf{e} \quad (2.10)$$

avec

- $\mathbf{N}_e = \mathbf{L}_e^\top$  pour une méthode de gradient. Cette méthode a été appliquée par exemple dans [Hashimoto et Kimura, 1993].

- $\mathbf{N}_e = \mathbf{L}_e^+$  pour une méthode de Gauss-Newton, signalons que l'on reconnaît là, exactement l'approche classique donnée par (2.6). Il est également possible de considérer le cas  $\mathbf{N}_e = \mathbf{L}_{e^*}^+$  [Espiau *et al.*, 1992], très utilisé en pratique.

- $\mathbf{N}_e = (\mathbf{H} + \mu \text{diag}(\mathbf{H}))^{-1} \mathbf{L}_e^\top$  pour une méthode de Levenberg-Marquardt (où  $\mathbf{H} = \mathbf{L}_e^\top \mathbf{L}_e$  est une approximation de la matrice Hessienne de la fonction de coût et  $\mu$  un scalaire positif) [Collewet *et al.*, 2008a; Collewet *et al.*, 2008b].

- $\mathbf{N}_e = (\mathbf{L}_e + \mathbf{L}_{e^*})^+$  pour la méthode ESM proposée dans [Malis, 2004]. Notons que par rapport aux algorithmes classiques de minimisation, cette approche tire profit du fait que l'allure de la fonction de coût est connue au minimum, ce qui n'est que rarement le cas dans d'autres problèmes classiques de minimisation. En pratique, elle s'avère donc beaucoup moins sensible à la présence de minima locaux que les autres approches mentionnées ci-dessus, son domaine de convergence est également plus large.

<sup>3</sup>Le lien entre minimisation d'une fonction de coût et le cas spécifique de la régulation d'une fonction de tâche a été établi dans [Samson *et al.*, 1991] en considérant comme fonction de tâche le gradient de la fonction de coût.

<sup>4</sup>Outre le fait que la forme du critère à minimiser puisse être complexe, le problème est rendu plus compliqué encore par le fait que l'argument permettant la minimisation du critère est en pratique la vitesse de la caméra (dans le cas d'une commande cinématique). De ce fait, l'algorithme employé doit non seulement conduire à la minimisation de (2.7) mais également assurer des trajectoires satisfaisantes de la caméra. Cela réduit malheureusement en pratique l'éventail d'approches potentielles, dont en particulier les techniques de recherche linéaire, de régions de confiance (voir néanmoins [Jägersand, 1996]) ou encore basées sur des méthodes de séparation et d'évaluation (branch and bound) qui nous permettraient d'atteindre un domaine de convergence plus large.

Par conséquent, écrire la loi de commande sous la forme générale (2.10) nous permettra dans la suite de ce document d'exhiber des résultats pour tout un panel de lois de commande.

### 2.1.4 Stabilité

L'étude de la stabilité du système bouclé nécessite d'exprimer la variation temporelle du signal d'erreur donné en (2.1). En le dérivant par rapport au temps, puis en utilisant (2.10), il vient :

$$\dot{\mathbf{e}} = -\lambda \mathbf{L}_e \widehat{\mathbf{N}}_e \mathbf{e} = -\lambda \mathbf{M}_e \mathbf{e} \quad (2.11)$$

avec  $\mathbf{M}_e = \mathbf{L}_e \widehat{\mathbf{N}}_e$  de dimension  $k \times k$ .

Tout d'abord, il faut assurer que le système bouclé ne présente qu'un unique point d'équilibre  $\mathbf{e}^* = \mathbf{0}$ . D'après (2.11), cette condition sera vérifiée si

$$\text{Ker } \mathbf{M}_e = \emptyset, \quad (2.12)$$

c'est-à-dire si les matrices  $\mathbf{L}_e$  et  $\widehat{\mathbf{N}}_e$  sont de rang plein dans tout l'espace de travail.

L'analyse de la convergence de  $\mathbf{e}$  vers  $\mathbf{e}^*$  peut ensuite être réalisée grâce à la méthode de Lyapunov en choisissant comme fonction candidate [Malis et Chaumette, 2002] :

$$\mathcal{L} = \frac{1}{2} \|\mathbf{e}\|^2 \quad (2.13)$$

dont on cherche à exhiber les conditions assurant sa décroissance en fonction du temps. Il vient alors grâce à (2.11) :

$$\dot{\mathcal{L}} = \mathbf{e}^\top \dot{\mathbf{e}} = -\lambda \mathbf{e}^\top \mathbf{M}_e \mathbf{e} \quad (2.14)$$

qui conduit à la condition suffisante de stabilité globale suivante :

$$\mathbf{M}_e > \mathbf{0}. \quad (2.15)$$

Dans ce cas, la décroissance de  $\|\mathbf{e}\|$  vers zéro est donc assurée puisque  $\mathcal{L}$  est décroissante.

Remarquons que l'on peut retrouver de cette façon la contrainte bien connue en optimisation portant sur la direction de descente. En effet, (2.13) n'est autre que la fonction de coût (2.7) qui, en utilisant (2.9), conduit à :

$$\dot{\mathcal{L}} = \lambda \nabla \mathcal{C}(\mathbf{r})^\top \mathbf{d}(\mathbf{r}) \quad (2.16)$$

et qui fournit finalement la contrainte suivante :

$$\nabla \mathcal{C}(\mathbf{r})^\top \mathbf{d}(\mathbf{r}) < 0. \quad (2.17)$$



### 2.1.5 Choix des informations visuelles

Ce choix est prépondérant puisqu'il fixe la forme de la matrice d'interaction et de la matrice  $\widehat{\mathbf{N}}_e$  et donc du comportement du système bouclé<sup>5</sup>. Intéressons-nous tout d'abord au choix de la dimension  $k$  du vecteur  $\mathbf{s}$  comme proposé dans [Chaumette, 2002].

- *Cas où  $\mathbf{s}$  est de dimension 6 ( $k = 6$ ).* En considérant que  $\mathbf{L}_e$  est de rang plein 6, le concepteur de la loi de commande doit choisir également une matrice  $\widehat{\mathbf{N}}_e$  de rang plein 6 (voir les conditions de stabilité (2.12) et (2.15)). En outre, il lui faut assurer que la trajectoire des informations visuelles est satisfaisante dans l'image, c'est-à-dire que la matrice  $\mathbf{M}$  est proche de la matrice identité  $\mathbb{I}_6$  conduisant par voie de conséquence à une même vitesse de décroissance (découplée) pour toute information visuelle  $\dot{\mathbf{s}} \approx -\lambda(\mathbf{s} - \mathbf{s}^*)$  [Chaumette, 2002]. Ce cas peut-être observé par exemple si  $\widehat{\mathbf{N}}_e = \mathbf{L}_e^{-1}$  est choisi.

- *Cas d'informations de type 2D.* En fait, trouver six informations visuelles conduisant à un bon comportement du système bouclé est un des principaux problèmes de l'asservissement visuel, sinon le principal problème. Il est par exemple bien connu que construire  $\mathbf{s}$  à partir des coordonnées cartésiennes de trois points dans l'image fournit une matrice d'interaction présentant des singularités pour certaines configurations bien précises [Michel et Rives, 1993]. De plus, il existe également quatre situations de la caméra par rapport à l'objet conduisant à la même image [Dhome *et al.*, 1989]. La fonction de coût (2.7) présente donc quatre minima globaux. Il est donc important de choisir des informations visuelles assurant un isomorphisme avec l'espace des configurations  $SE(3)$ . En fait, le Graal consisterait à trouver des informations visuelles conduisant à une matrice d'interaction diagonale et, par suite, à la commande d'un système linéaire que l'on sait traiter. La stabilité asymptotique globale serait alors obtenue, tout comme l'isomorphisme. En outre, la trajectoire dans l'espace 3D serait correcte, la matrice d'interaction ne présentant plus alors de non linéarité. C'est précisément le but des travaux rapportés dans [Corke et Hutchinson, 2001; Mahony *et al.*, 2002; Tahri et Chaumette, 2005; Kalleem *et al.*, 2007], [Pagès *et al.*, 2006c].
- *Cas d'informations de type 3D.* Le problème majeur n'est plus ici de choisir six informations visuelles, quoique le problème de visibilité de l'objet et donc de l'obtention des mesures  $\mathbf{x}(\mathbf{r})$  reste ouvert (voir cependant [Morel *et al.*, 2000] où cette contrainte est explicitement prise en compte autant que faire se peut). En effet, on choisit le plus souvent les coordonnées cartésiennes d'un point 3D qui correspondent typiquement à l'origine d'un repère pour contrôler les translations [Wilson *et al.*, 1996; Martinet *et al.*, 1997] et le vecteur  $\mathbf{u}\theta$  caractérisant la rotation à réaliser pour atteindre l'orientation désirée [Malis *et al.*, 1999]. On obtient ainsi une matrice d'interaction très sympathique puisque sa forme est bloc triangulaire et qu'elle ne comporte qu'une singularité bien identifiée (en  $\theta = 2\pi$ ). Par contre, l'analyse de la stabilité pose problème [Chaumette, 1998]. Pour s'en persuader exprimons la matrice d'interaction associée à la pose estimée :

$$\mathbf{L}_{\hat{\mathbf{r}}} = \frac{\partial \hat{\mathbf{r}}}{\partial \mathbf{x}} \mathbf{L}_{\mathbf{x}} \quad (2.18)$$

<sup>5</sup>D'où l'intérêt de la connaissance sous forme analytique de la matrice d'interaction.

où  $\mathbf{L}_x$  est la matrice d'interaction 2D associée à  $x$ . En effet, comme nous l'avons vu à la section 2.1.2, le premier terme n'est pas obtenu sous forme analytique mais par calcul de pose et ne permet donc pas d'exhiber des conditions particulières de convergence. Par ailleurs, cette matrice peut dans certains cas être très mal conditionnée. Citons l'exemple désormais classique où  $s$  est constitué de quatre points coplanaires, bien que, théoriquement, il n'existe qu'une seule valeur de pose conduisant aux valeurs courantes  $s$  [Heraud *et al.*, 1989], différentes valeurs de pose permettent d'obtenir des valeurs quasiment identiques pour  $s$  [Chaumette, 1998]. De faibles différences sur les valeurs de  $s$  conduisent dans ce cas à de grandes variations de  $\hat{r}$ . On suppose donc que  $\hat{r} = r$ , c'est-à-dire que la caméra est parfaitement calibrée, que le modèle de l'objet est également parfaitement connu mais aussi que les mesures  $x$  ne sont entachées d'aucune erreur et que la méthode de calcul de pose est également parfaite. Est-il besoin de préciser que cela n'est jamais le cas en pratique ?

- *Cas d'informations de type 2D1/2.* Nous avons vu dans la section décrivant les asservissements visuels 3D que les matrices d'interaction obtenues conduisaient à de bonnes propriétés de découplage des translations et des rotations mais qu'en revanche aucun contrôle dans l'image ne permettait de garantir que l'objet observé reste dans le champ de vision de la caméra. A l'inverse, les asservissements visuels 2D permettent un contrôle dans l'image mais dans ce cas les translations et les rotations sont en général fortement couplées. Dans [Malis *et al.*, 1999] un découplage partiel est obtenu en combinant des informations visuelles 2D et 3D. En outre, le calcul de pose n'est plus nécessaire [Malis *et al.*, 2000] et comme les matrices d'interaction sont triangulaires par bloc, il a pu être exhibé des conditions analytiques de stabilité asymptotique locale et globale [Malis et Chaumette, 2002].

Signalons une approche similaire dans [Morel *et al.*, 2000] où l'ensemble des points de l'objet observé est contraint à rester dans le champ de vision de la caméra. Cependant, dans ces conditions, la matrice d'interaction n'est plus triangulaire et les résultats analytiques de stabilité sont beaucoup plus difficiles à obtenir. Un schéma de type 2D1/2 est également décrit dans [Andreff *et al.*, 2002] où les coordonnées Pluckeriennes de droites sont utilisées.

- *Cas où  $s$  est de dimension supérieure à 6 ( $k > 6$ ).* Comme on l'a vu dans les paragraphes précédents, ce cas est typique des informations visuelles 2D puisque dans les autres cas nous avons vu qu'il était possible de trouver 6 informations visuelles conduisant à un comportement satisfaisant du système bouclé <sup>6</sup>.

Dans ce cas de figure, la matrice  $\mathbf{M}_e$  de dimension  $k \times k$  impliquée dans l'analyse de la stabilité est au plus de rang 6. Par conséquent, la stabilité globale n'est pas assurée (voir les conditions (2.12) et (2.15)). En effet, toute configuration telle que  $e \in \text{Ker } \mathbf{M}_e$  peut conduire à la présence d'un minimum local. Toutefois, pour que ce minimum local existe vraiment, il faut que cette configuration conduise à une pose de la caméra par rapport à l'objet réalisable. En fait, la présence potentielle de minima locaux est liée à l'existence de mouvements  $e^\perp$  irréalisables dans l'image [Chaumette, 1998]. Ces mouvements sont tels que  $e^\perp \in \text{Ker } \mathbf{L}_e^\top$ , par conséquent des minima locaux peuvent apparaître si  $\text{Ker } \mathbf{L}_e^\top \cap$

<sup>6</sup>Voir néanmoins [Cervera *et al.*, 2003; Schramm *et al.*, 2004] où des approches de type 2D1/2 sont proposées.

$\text{Ker } \mathbf{M}_e \neq \emptyset$  ou si  $\text{Ker } \mathbf{L}_e^\top \cap \text{Ker } \widehat{\mathbf{N}}_e \neq \emptyset$  si  $\mathbf{L}_e$  est de rang plein. C'est par exemple le cas quand  $\widehat{\mathbf{N}}_e = \mathbf{L}_e^\top$  ou quand  $\widehat{\mathbf{N}}_e = \mathbf{L}_e^+$  (puisque  $\text{Ker } \mathbf{L}_e^+ = \text{Ker } \mathbf{L}_e^\top$ ). L'existence de minima locaux a été prouvée dans ce dernier cas alors que dans les mêmes conditions ils sont évités avec  $\widehat{\mathbf{N}}_e = \mathbf{L}_e^{*+}$  [Chaumette, 1998].

Par contre, en introduisant l'erreur  $e' = \widehat{\mathbf{N}}_e e$ , la stabilité locale peut être garantie si les conditions  $\text{Ker } \mathbf{M}_{e'} = \emptyset$  et  $\mathbf{M}_{e'} > \mathbf{0}$ , où  $\mathbf{M}_{e'} = \widehat{\mathbf{N}}_e \mathbf{L}_e$ , sont respectées. Une façon de faire repose alors sur la planification de la trajectoire des informations visuelles désirées contraignant de ce fait s à rester proche de  $s^*$  [Mezouar *et al.*, 2002]. De plus, le domaine de convergence devient alors très étendu en pratique.

Si en théorie le choix  $k > 6$  semble problématique, il est toutefois très utilisé en pratique dans la mesure où, rappelons-le, trouver 6 informations visuelles conduisant à un bon comportement du système bouclé reste un problème ouvert. En outre, utiliser des informations visuelles redondantes permet en général une bonne robustesse vis-à-vis d'éventuelles occultations et réduit la possibilité d'existence de plusieurs minima globaux de la fonction de coût.

## 2.2 Contrôler $m$ degré(s) de liberté d'un porteur ( $m < 6$ )

Nous avons vu dans la section précédente comment il était possible de réaliser une tâche de positionnement par asservissement visuel. De cette façon, une liaison rigide virtuelle entre la scène et le capteur est créée. Cependant, une telle liaison n'est pas toujours souhaitée, on peut vouloir utiliser les degrés de liberté non contraints par la tâche à réaliser pour satisfaire simultanément un autre objectif [Coste-Manière *et al.*, 1995; Santos *et al.*, 1995; Marchand et Hager, 1998; Chaumette et Marchand, 2001; Cadenat *et al.*, 2001], [Collewet et Chaumette, 2002] ou une autre liaison, comme une tâche de positionnement plan à plan [Pagès *et al.*, 2006c] par exemple<sup>7</sup>, exploitant ainsi au mieux les degrés de liberté restant.

Une façon élégante de poser ce problème est de l'inscrire dans le formalisme de redondance de fonctions de tâche [Samson *et al.*, 1991]. On définit une fonction de tâche  $e_1$  prioritaire, dite tâche principale, et de dimension  $m$  que l'on fait cohabiter avec une autre tâche  $e_2$ , dite tâche secondaire. Une fonction de tâche  $e$  permettant de réaliser  $e_2$  sous la contrainte  $e_1 = \mathbf{0}$  s'écrit [Samson *et al.*, 1991] :

$$e = \mathbf{W}^+ e_1 + (\mathbb{I}_6 - \mathbf{W}^+ \mathbf{W}) e_2 \quad (2.19)$$

où  $\mathbf{W}$  est une matrice de dimension  $m \times 6$  de rang plein  $m$  et telle que  $\text{Ker } \mathbf{W} = \text{Ker } \mathbf{J}_1$  où  $\mathbf{J}_1$  est la matrice Jacobienne de  $e_1$  par rapport à  $\mathbf{r}$ . La matrice  $\mathbb{I}_6 - \mathbf{W}^+ \mathbf{W}$  est un opérateur de projection dans le noyau de la tâche  $e_1$ , de ce fait la réalisation de  $e_2$  ne pourra en aucun cas perturber celle de la tâche principale. En pratique,  $\mathbf{J}_1$  n'étant connue que par la biais d'une estimation, le noyau de  $\mathbf{W}$  sera donc forcément différent de celui de  $\mathbf{J}_1$  et, par conséquent,  $e_1$  devra être amené à compenser ces perturbations en générant une commande (voir un exemple dans [Collewet et Chaumette, 2002]).

<sup>7</sup>D'autres exemples de liaisons virtuelles sont données dans [Chaumette *et al.*, 1993].

Comme dans le cas d'une liaison rigide, la régulation de  $e$  à  $0$  est assurée grâce à la loi de commande donnée en (2.10). De plus, la condition de stabilité (2.15) reste valable. Cependant, le calcul de  $L_e$  est maintenant plus complexe, néanmoins il a pu être montré dans [Samson *et al.*, 1991] que si la condition suivante était respectée

$$\mathbf{J}_1 \mathbf{W}^+ > 0 \quad (2.20)$$

alors, dans le cas d'une commande de type Gauss-Newton, (2.15) était le plus souvent satisfaite. On utilise donc (2.20) en pratique pour étudier la stabilité.

### 2.3 Conclusion

Ce chapitre nous a permis de présenter les fondements théoriques sur lesquels reposent les asservissements visuels. Il nous a permis aussi de rappeler les principaux problèmes inhérents à telle ou telle approche. Ces problèmes sont maintenant clairement identifiés et des solutions sont régulièrement proposées en vue de les résoudre. Les trajectoires, aussi bien 2D que 3D, sont de plus en plus satisfaisantes et la convergence des lois de commande obtenue pour des positions toujours plus éloignées de la position désirée. Néanmoins, comme nous le verrons dans le chapitre suivant, d'autres problèmes subsistent. Ils sont dus en grande partie aux scènes que nous avons décidé d'aborder, scènes constituées d'objets hétérogènes à l'intérieur d'une même classe les caractérisant mais, le cas échéant aussi, d'objets de forme inconnue. De plus, contrairement à la grande majorité des travaux en asservissement visuel, nous ne supposons pas la scène Lambertienne, ni même statique.

## CHAPITRE 3

---

### Vers une indépendance vis-à-vis de l'objet observé

---

Le chapitre précédent nous a montré qu'un large panel d'informations visuelles pouvait être utilisé lors de la synthèse d'une loi de commande par asservissement visuel, pourvu que la matrice d'interaction associée à ces informations visuelles était connue. Dans le cas où une connaissance *a priori* de l'objet d'intérêt est disponible, comme c'est le cas des objets manufacturiers, le choix des informations visuelles impliquées dans la loi de commande peut se faire en toute connaissance de cause. De plus, lorsqu'un modèle 3D de l'objet est disponible, la phase de suivi des primitives nécessaires au calcul de la loi de commande peut se révéler être plus robuste à réaliser. Nous pensons notamment aux techniques de suivi 3D basées modèle (voir par exemple [Drummond et Cipolla, 2002; Comport *et al.*, 2006b]). Le cas échéant, la pose même de marqueurs visuels *ad hoc*, peut être envisagée. Dans ces conditions, l'algorithme obtenu (la loi de commande et le traitement d'images associé) peut alors être très efficace car dédié à la scène.

En revanche, dans d'autres secteurs applicatifs, l'objet d'intérêt peut être mal connu. Nous avons déjà cité les domaines de l'agriculture et de l'agro-alimentaire, nous pouvons également citer les domaines de la robotique horticole, médicale, mobile (terrestre en environnement extérieur), sous-marine, aérienne, voire les domaines de la robotique de service ou de la robotique pour le handicap où une interaction avec l'humain est nécessaire. Dans ces conditions, la texture de l'objet d'intérêt, ou son aspect visuel, n'est pas précisément connu, rendant en particulier le choix des informations visuelles problématique. Par ailleurs, l'objet d'intérêt étant inconnu, sa forme l'est également. Dans ce contexte, l'hypothèse très répandue dans la littérature d'une géométrie particulière ne peut plus être formulée. En outre, signalons que nous faisons la distinction ici entre objet complexe et objet inconnu, en effet, l'objet inconnu dont il est question peut très bien s'avérer être un objet de forme simple, comme un plan par exemple. Illustrons ce propos à l'aide de l'application

particulière consistant à caractériser des défauts d'aspect sur une pièce de viande <sup>1</sup>, suivant la localisation du défaut, la surface de l'objet peut tout aussi bien se révéler être plane que non plane.

À ce stade, précisons également que nous n'avons jamais voulu nous placer dans le cadre des techniques basées sur de l'apprentissage hors ligne, cadre que nous avons jugé inapproprié au vu des applications visées (voir le chapitre introductif à ce manuscrit). Néanmoins, le lecteur intéressé est invité à se référer à la section 2.1.1 où de telles approches sont décrites, voir également [Wells *et al.*, 1996] où la réalisation de tâches de positionnement est effectuée à l'aide d'un réseau de neurones.

Notre souci fut de nous affranchir, tout au long des travaux rapportés dans ce chapitre, des deux problèmes évoqués ci-dessus, à savoir l'indépendance vis-à-vis de la texture de l'objet et l'indépendance vis-à-vis de sa géométrie. Plusieurs solutions ont été investiguées, traitant plus précisément de l'un ou de l'autre de ces deux problèmes, voire des deux problèmes simultanément. Ces différents travaux sont décrits à présent.

### 3.1 Indépendance vis-à-vis de la texture de l'objet

L'indépendance vis-à-vis de l'aspect visuel de l'objet se révèle problématique dans le sens où un grand nombre d'indices visuels ne peut plus être considéré permettant, éventuellement, l'élaboration d'informations plus évoluées (voir la section 2.1.1). Il s'agit essentiellement d'indices visuels de type géométrique comme des droites, des segments ou des contours. En revanche, la primitive géométrique « point » peut être considérée comme intéressante, dans le sens où elle sera présente dans un grand nombre de scènes. Notons que cette primitive est également générique au sens de l'asservissement visuel et ce choix n'est donc pas pénalisant en soi. En effet, comme nous l'avons vu au chapitre 2, elle peut tout aussi bien être utilisée pour des approches 2D [Papanikolopoulos, 1995; Kelly *et al.*, 2000; Comport *et al.*, 2006a] que 2D1/2 [Malis *et al.*, 1999] ou 3D [Wilson *et al.*, 1996; Martinet *et al.*, 1997], ces dernières approches étant le plus souvent basées sur un calcul d'homographie ou de pose.

Paradoxalement, l'indépendance recherchée peut être obtenue, en considérant le signal de luminance, c'est-à-dire l'image courante en elle-même, ou un sous-ensemble de cette image, conduisant ainsi à des informations visuelles de nature non géométrique. Certains travaux relatent cette façon de procéder; citons par exemple [Deguchi et Noguchi, 1996; Nayar *et al.*, 1996; Deguchi, 2000; Kallem *et al.*, 2007]. Soulignons que ces approches sont très intéressantes dans la mesure où, de surcroît, elles ne requièrent aucune phase de mise en correspondance (voir également [Abdul Hafez *et al.*, 2008]) ainsi qu'un effort de traitement d'images vraiment minime. Ce signal de luminance peut également conduire à l'obtention d'informations de nature dynamique. Citons par exemple l'utilisation en robotique mobile des paramètres d'un modèle de mouvement 2D, la plupart du temps affine, où différentes tâches d'alignement de l'axe optique de la caméra avec la direction du mouvement [Sundareswaran *et al.*, 1996] ou avec la direction de la normale

<sup>1</sup>Travaux réalisés dans le cadre d'une convention Cemagref / INRIA / OFIVAL (Office national interprofessionnel des viandes de l'élevage et de l'aviculture) / Olympig (industriel du secteur de la viande).

à un objet plan [Questa *et al.*, 1995; Santos-Victor et Sandini, 1997] sont considérées, et ce, sans requérir une quelconque phase de reconstruction 3D. Ces différentes tâches ne contrôlent toutefois qu'un faible nombre de degrés de liberté du système (voir également [Crétual et Chaumette, 1998] où une tâche de suivi de cible est rapportée). En revanche, dans [Benhimane et Malis, 2007], les 6 degrés de liberté sont contrôlés. Pour ce faire, une loi de commande basée sur l'homographie liant l'image courante à l'image désirée est exploitée. De même, dans [Crétual et Chaumette, 2001], un important travail de modélisation a permis d'obtenir la matrice d'interaction sous forme analytique relative aux paramètres d'un modèle de mouvement quadratique et de s'affranchir ainsi de la texture de la scène observée. Signalons, en outre, que cette approche ne nécessite pas la connaissance de l'image désirée, mais uniquement du champ de vitesse désiré.

Remarquons que toutes les approches que nous venons d'évoquer reposent sur l'hypothèse d'un objet texturé. Néanmoins, elles n'en font aucune quant au contenu explicite de l'image courante.

Une autre approche, bien connue en robotique, et permettant de s'affranchir grandement de la texture de l'objet d'intérêt consiste à utiliser un éclairage structuré<sup>2</sup>. De cette façon, les informations visuelles choisies par le concepteur de la loi de commande sont introduites artificiellement sur l'objet contournant ainsi le problème posé. Ici aussi, signalons que le traitement d'images peut être grandement facilité, en particulier lorsque l'objectif de la caméra est équipé d'un filtre accordé à la longueur d'onde de la source lumineuse. Précisons néanmoins que si les phases d'extraction et de suivi sont simplifiées, la phase de mise en correspondance initiale demeure. En outre, il est possible d'estimer la profondeur en tout point éclairé par triangulation pourvu que le couple caméra / source lumineuse soit étalonné. Ce type d'approche a été [Agin, 1985] et reste très largement utilisé en reconstruction 3D [Watanabe *et al.*, 2007], mais beaucoup moins en asservissement visuel (voir cependant [Urban *et al.*, 1994; Khadraoui *et al.*, 1996; Andreff *et al.*, 2002; Krupa *et al.*, 2003]).

Après avoir évoqué ces deux grands types d'approches, nous détaillons à présent nos contributions s'y réclamant.

### 3.1.1 Utilisation d'un éclairage structuré

L'utilisation de dispositifs à base de lumière structurée est en fait très répandue en robotique. Cette popularité est en partie due au fait qu'il est très facile d'obtenir des informations de profondeur. En effet, la profondeur d'un point éclairé de la scène peut être obtenue par triangulation entre le faisceau émis et la droite de visée de la caméra. L'alternative consistant à utiliser deux caméras est généralement plus complexe, la phase d'appariement s'avérant plus délicate à réaliser.

La littérature sur ce sujet rapporte différents types de configuration matériel, de un à plusieurs faisceaux laser, de un à plusieurs plans, mobile(s) le cas échéant, permettant

<sup>2</sup>Notons cependant que ce type d'approche n'est pas envisageable sur des objets purement spéculaires. En effet, théoriquement le rayon réfléchi ne peut être vu que dans une unique direction d'observation. En pratique, l'existence d'un lobe spéculaire permet de l'observer dans une zone de l'espace beaucoup moins restrictive.

ainsi une reconstruction rapide d'un ensemble de points de la scène. Très tôt de tels dispositifs ont donc investi le domaine de la robotique et en particulier celui de la robotique mobile. L'ensemble caméra / dispositif lumineux est alors embarqué sur le robot autorisant la réalisation de tâches de navigation et d'évitement d'obstacles (voir par exemple [Dubrawski et Siemiatkowska, 1998; Haverinen et Roning, 1998; Kim et Cho, 2001; Surmann *et al.*, 2003; Kondo et Tamaki, 2004]). De tels dispositifs sont également embarqués sur des robots manipulateurs pour la réalisation de tâches de positionnement comme dans [Kemotsu et Kanade, 1995; Niel *et al.*, 2004] ou pour des tâches de suivi de profil par exemple [Kim *et al.*, 1999; Amin-Nejad *et al.*, 2003], voire d'inspection et d'usinage [Kwok *et al.*, 1998].

Néanmoins, lorsque des tâches plus complexes sont envisagées, un éclairage total de la scène est requis. C'est ainsi que des premières solutions ont été avancées pour éclairer totalement l'objet grâce à l'utilisation d'une grille lumineuse [Hall *et al.*, 1982]. Cette façon de faire nécessitait toutefois un effort important de traitement d'images afin de retrouver dans l'image les plans responsables du motif observé et de réaliser ensuite la phase de triangulation. Des solutions ont dès lors été proposées pour coder le motif projeté et s'affranchir ainsi le plus possible du traitement d'images [Inokuchi *et al.*, 1984], tout en permettant une mise en correspondance robuste (voir la figure 3.1). Depuis, la littérature rapporte un grand nombre de travaux sur ce sujet, le lecteur trouvera dans [Salvi *et al.*, 2004] un état de l'art récent de ces différentes techniques. C'est ainsi que des tâches, ou des scènes, plus complexes peuvent être appréhendées comme par exemple en robotique de service où une collaboration homme / robot est considérée [Lee *et al.*, 2004], en robotique médicale [Albitar *et al.*, 2007] dans le but d'assister le praticien dans ses actes chirurgicaux, ou encore dans les cas où une reconnaissance d'objet est nécessaire avant toute tâche de préhension [Stahs et Wahl, 1992; Alshawish et Allen, 1995; Hattori et Sato, 1995].

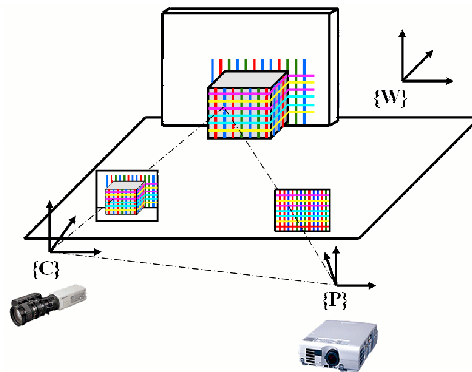


FIG. 3.1: Schéma de principe d'un dispositif utilisant un motif de lumière codée. Un projecteur éclaire la scène selon un motif lumineux bien défini. La profondeur d'un point éclairé est obtenue par triangulation, comme dans le cas classique où deux caméras stéréoscopiques sont utilisées. Toutefois, la mise en correspondance est grandement simplifiée grâce à la connaissance du motif lumineux qui permet de caractériser un point par rapport à son voisinage.



Toutefois, les travaux évoqués ci-dessus relèvent tous de l'asservissement visuel 3D<sup>3</sup> et reposent donc sur un parfait étalonnage de l'ensemble caméra / dispositif de projection tant du point de vue des paramètres intrinsèques qu'extrinsèques. De plus, comme nous l'avons vu dans le chapitre 2, il se pose également le problème de la visibilité des informations visuelles tout au long du mouvement lorsque de telles lois de commande 3D sont envisagées. Nous nous sommes donc focalisés dans le cadre de la thèse de Jordi Pagès sur des schémas de commande de type 2D, évitant ainsi ce type de problèmes. Deux cas de figure ont été traités, le cas où l'éclairage est déporté puis le cas où il est embarqué, la caméra étant portée par le bras manipulateur dans les deux cas.

### 3.1.1.1 Utilisation d'un éclairage déporté et codé

Précisons, tout d'abord, que nous nous sommes restreints dans ce travail au cas d'un objet immobile par rapport au dispositif d'éclairage. Ce cadre nous a donc permis d'employer un dispositif de lumière codée nous permettant de nous affranchir totalement de la texture de l'objet observé et de faciliter grandement le processus de mise en correspondance tout en le robustifiant. En particulier, des scènes très peu texturées ont pu être prises en compte<sup>4</sup>.

Comme nous l'avons déjà évoqué, bon nombre de méthodes de codage ont été proposées dans la littérature depuis les travaux précurseurs décrits dans [Inokuchi *et al.*, 1984]. Signalons seulement qu'elles sont principalement de trois types, soit basées sur un multiplexage temporel du motif lumineux, soit basées sur un codage spatial, ou encore basées sur un codage direct, grâce à la couleur le plus souvent. En pratique, seules les deux dernières approches peuvent être utilisables en robotique puisqu'une unique image est requise pour le décodage du motif. En effet, il est alors possible de considérer des scènes dynamiques, soit parce que la scène l'est, soit parce que le capteur est en mouvement, voire les deux le cas échéant. Toutefois, ces différentes approches sont dédiées au problème de la reconstruction 3D. Très généralement, la problématique consiste à obtenir un motif où le nombre de symboles ou de couleurs utilisées est minimum, tout en cherchant à augmenter la résolution obtenue et la vitesse du décodage (voir la figure 3.2). De ce fait, l'utilisation de bandes lumineuses est à la base des motifs les plus usités [Salvi *et al.*, 2004]. En effet, elles permettent d'allier, tout à la fois, vitesse et robustesse du décodage et reconstruction dense.

**Choix des informations visuelles et de la matrice d'interaction.** Même si cette primitive est rencontrée dans de nombreux motifs, l'utilisation de bandes lumineuses en tant que primitive impliquée dans une loi de commande s'avère beaucoup plus problématique. En effet, il faudrait alors envisager une primitive de type « ligne courbe » (si l'on veut pouvoir appréhender des objets de forme non plane) et donc obtenir la matrice d'interaction associée. De plus, la pertinence d'une telle primitive vis-à-vis du comportement de la loi

<sup>3</sup>Voir cependant [Kahane et Rosenfeld, 2004] où des heuristiques sont utilisées conduisant à une approche pouvant être considérée comme de type 2D.

<sup>4</sup>Le contexte applicatif de ces travaux concernait la réalisation d'une tâche de focalisation par asservissement visuel en vue de la lecture d'un identifiant présent sur une pièce de viande. Ce projet a été réalisé en collaboration avec les sociétés Olympig, Marconi (industriel du marquage) et avec l'aide de l'OFIVAL.

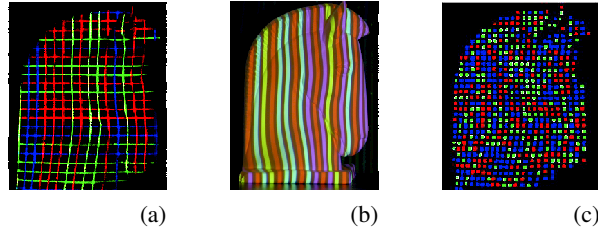


FIG. 3.2: Exemples de codage spatial : (a) Utilisation d'une grille [Salvi *et al.*, 1998]; (b) Utilisation de lignes [Monks *et al.*, 1992]; (c) Utilisation d'une matrice de points [Morano *et al.*, 1998]. Dans les cas (a) et (b), le décodage est obtenu grâce à la donnée de la couleur de la ligne d'intérêt et de celles des lignes adjacentes (le même système de codage est utilisé dans le cas (a) pour les lignes horizontales et verticales). Dans le cas (c), le décodage est obtenu grâce à la propriété d'unicité de chaque sous matrice de taille  $3 \times 3$ .

de commande reste à avérer. En revanche, il est beaucoup plus aisé de s'intéresser à un motif tel que celui conduisant à la figure 3.2c. En effet, il est très simple d'utiliser en asservissement visuel des points, quitte à construire ensuite des informations visuelles ayant un comportement beaucoup plus satisfaisant, comme par exemple les moments associés à ces points [Tahri et Chaumette, 2005]. Néanmoins, nous avons simplement considéré dans ce travail les coordonnées dans l'image de chaque point  $\mathbf{x}_i$  décodé, soit

$$\mathbf{s} = (\mathbf{x}_1, \dots, \mathbf{x}_k). \quad (3.1)$$

Dans ce cas, la matrice d'interaction  $\mathbf{L}_s$  associée à  $\mathbf{s}$  s'obtient de façon triviale à partir de la matrice d'interaction  $\mathbf{L}_{\mathbf{x}_i}$  associée au point  $\mathbf{x}_i$ , ce qui nous a permis d'utiliser directement (2.10). Plus précisément, la méthode de Gauss-Newton ayant été retenue, nous avons ici  $\mathbf{N}_e = \mathbf{L}_e^+ = \mathbf{L}_s^+$ .

**Choix du motif lumineux.** Le but recherché ici est de projeter un motif lumineux conduisant à une mise en correspondance (entre les informations visuelles initiales et désirées mais aussi entre les informations visuelles courantes et désirées) rapide et robuste. La stratégie retenue est basée sur la théorie des matrices parfaites pseudo-aléatoires [MacWilliams et Sloane, 1976]<sup>5</sup>. De telles matrices se caractérisent par le fait que chaque sous-matrice  $n \times n$  les composant est unique pour un alphabet donné (valeur pouvant être prise par chaque élément de la matrice). De plus, à chaque sous-matrice peut être associé un code définissant leur position absolue dans la matrice et permettant ainsi une mise en correspondance rapide. Quant à la robustesse, elle est obtenue de par la conception du motif lui-même [Griffin *et al.*, 1992; Morano *et al.*, 1998; Albitar *et al.*, 2007]. En ligne, elle est obtenue très généralement grâce à l'utilisation d'heuristiques qui exploitent la redondance d'informations (voir par exemple [Pagès *et al.*, 2006a; Pagès *et al.*, 2006b]). La plus courante se base sur le fait que la position de chaque point d'une sous-matrice de taille  $n \times n$  peut être obtenue grâce à  $n^2$  matrices, le décodage s'appuie alors sur des techniques de

<sup>5</sup>C'est en fait une extension au cas 2D des séquences pseudo-aléatoires de type De Bruijn, utilisées pour générer les motifs des figures 3.2a et 3.2b.

vote. Dans le cadre de l'asservissement visuel, nous avons cherché une plus grande robustesse encore qu'en reconstruction 3D en imposant une même réponse pour les  $n^2$  matrices. Nous pensons en effet qu'une mise en correspondance erronée peut avoir des conséquences beaucoup plus dramatiques qu'en reconstruction 3D, comme par exemple une trajectoire fortement discontinue du robot <sup>6</sup>.

**Résultats.** En pratique, le motif lumineux a été généré grâce à l'algorithme proposé dans [Morano *et al.*, 1998] (voir la figure 3.2c) conduisant à une matrice de points de taille  $20 \times 20$  pour un alphabet de dimension 3 et des sous-matrices de taille  $3 \times 3$ . Le temps de décodage des 400 points est de l'ordre de 30 ms (sur un Pentium cadencé à 2 Ghz), soit inférieur à la cadence vidéo. La figure 3.3 rapporte le comportement de la loi de commande lors de la réalisation d'une tâche de positionnement par rapport à un cylindre elliptique <sup>7</sup>. Signalons que seuls 148 points ont satisfait les contraintes de mise en correspondance robustes. Par ailleurs, notons que dans la mesure où des informations visuelles de type points ont été utilisées, la loi de commande a un comportement très classique, comme la présence de couplages par exemple. Néanmoins, la convergence est assurée malgré le fait que la matrice d'interaction utilisée ait été estimée en position désirée et en supposant toutes les profondeurs égales. Nous tenons à rappeler ici qu'un comportement plus satisfaisant encore aurait pu être obtenu en estimant, par exemple, la profondeur en chaque point par triangulation ou encore en choisissant des informations visuelles plus pertinentes, comme nous l'avons déjà mentionné. Il est clair que toutes les potentialités de cette approche n'ont donc pas été exploitées même si, en pratique, la contrainte de rigidité entre le dispositif lumineux et l'objet pour lequel un positionnement est requis rend l'approche utile uniquement pour des robots non répétables.

### 3.1.1.2 Utilisation d'un éclairage embarqué

L'utilisation d'un éclairage structuré, codé et déporté répond parfaitement au problème de l'indépendance recherchée vis-à-vis de la texture de l'objet d'intérêt. De plus, cette solution permet également une mise en correspondance robuste et rapide des informations visuelles. Toutefois, cette approche nécessite un éclairage fixe par rapport à la scène, contrainte qui limite ainsi son champ applicatif. L'alternative consiste bien entendu à embarquer le dispositif d'éclairage. Cependant, nous n'avons pas envisagé une telle solution, que nous avons jugée trop difficile à mettre en œuvre en pratique compte tenu de l'encombrement et du poids actuel des projecteurs. Nous avons recouru à l'utilisation de diodes laser qui permettent tout autant de s'affranchir de la texture de l'objet que de simplifier le traitement d'images.

Comme nous l'avons vu au début de la section 3.1.1, cette façon de faire est très répandue en robotique. Néanmoins, les approches retenues sont généralement de type 3D nécessitant une phase d'étalonnage préalable, indispensable au processus de triangulation.

<sup>6</sup>D'où l'intérêt d'utiliser un asservissement visuel robuste aux erreurs d'extraction et de mise en correspondance des informations visuelles [Comport *et al.*, 2006a] quand une solution basée sur un éclairage codé ne peut être retenue.

<sup>7</sup>Le lecteur pourra trouver d'autres résultats expérimentaux dans [Pagès, 2005; Pagès *et al.*, 2006a; Pagès *et al.*, 2006b].

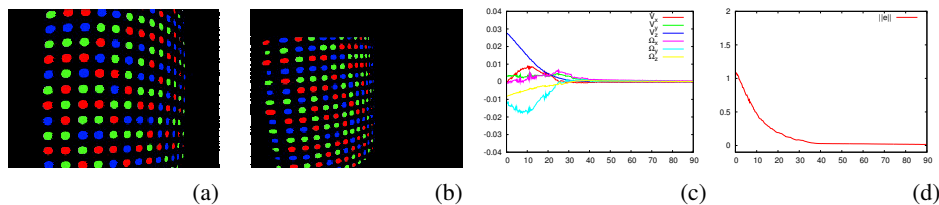


FIG. 3.3: Tâche de positionnement par rapport à un cylindre elliptique : (a) image désirée ; (b) image initiale ; (c) vitesse de la caméra (ms/s ou rad/s) en fonction du temps (en s) ; (d) Norme du signal d'erreur en fonction du temps (en s).

Effectivement, la littérature ne rapporte que très peu de travaux concernant le couplage de la lumière structurée et d'un asservissement visuel 2D. Citons toutefois Andreff *et al.* [Andreff *et al.*, 2002] qui ont introduit dans leur schéma de commande un faisceau laser dans le but de contrôler la distance de la caméra à l'objet. De façon similaire, Krupa *et al.* [Krupa *et al.*, 2003] ont fixé un faisceau laser à un instrument chirurgical de manière à contrôler sa distance par rapport à un organe. Par ailleurs, la plupart des travaux recensés dans la littérature utilisent la lumière structurée pour le contrôle d'un unique axe. Très peu de travaux ont traité le cas de plusieurs axes. La contribution majeure dans ce domaine est due à Motyl *et al.* [Khadraoui *et al.*, 1996]. Les auteurs ont en effet modélisé la variation des informations visuelles, obtenues par projection d'un plan laser sur des objets plans ou sphériques, dans le but de réaliser une tâche de positionnement par asservissement visuel.

Somme toute, à la lecture de ces différents travaux, nous avons été convaincus que toutes les potentialités relatives au couplage de la lumière structurée et de l'asservissement visuel n'avaient pas encore été pleinement exploitées. Plus précisément, l'utilisation d'un tel couplage n'avait pas encore été vu comme un moyen d'aboutir à une loi de commande optimale au sens défini au chapitre 2, c'est-à-dire essentiellement en terme de stabilité, de découplage des informations visuelles mais également en terme de trajectoire satisfaisante de la caméra. En fait, pouvoir choisir la géométrie du dispositif lumineux (en terme de nombre de faisceaux laser mais aussi de position et d'orientation) introduit une certaine liberté que nous avons cherchée à exploiter au mieux.

**Définition de la tâche à réaliser.** Traiter ce problème dans le cas général est, néanmoins, loin d'être trivial. Nous avons donc considéré en premier lieu la réalisation d'une tâche de positionnement plan à plan. Cette tâche consiste à amener le plan du capteur de façon parallèle à celui de l'objet supposé plan, elle constitue donc une liaison virtuelle de classe 3 telle que les mouvements fronto-parallèles ainsi que la rotation autour de l'axe optique laissent les informations visuelles choisies invariantes [Chaumette *et al.*, 1993]. Trois informations visuelles conduisant à une matrice d'interaction de rang plein sont donc à définir.

*Remarque :* D'autres tâches ont été étudiées, des objets non plans ont été également considérés, cependant nous ne présenterons pas ces travaux ici et renvoyons le lecteur intéressé à [Pagès, 2005].

**Choix du dispositif lumineux.** Considérons au préalable la matrice d'interaction  $\mathbf{L}_x$  relative aux coordonnées dans l'image du point  $\mathbf{x} = (x, y)$  décrivant la projection du point  $\mathbf{X}$  de l'espace 3D obtenu par intersection du faisceau laser et du plan de l'objet<sup>8</sup>. Cette matrice étant en fait de rang 1 [Andreff *et al.*, 2002][Pagès *et al.*, 2006c], théoriquement 3 faisceaux laser sont donc suffisants. Néanmoins, une configuration avec 4 faisceaux laser (voir la figure 3.4.a) a conduit à de meilleurs résultats. De façon plus précise, les 4 lasers sont positionnés dans le cas idéal de façon symétrique par rapport au centre optique de la caméra et sont orientés parallèlement à la normale au plan de l'objet en position désirée (voir la figure 3.4.b).

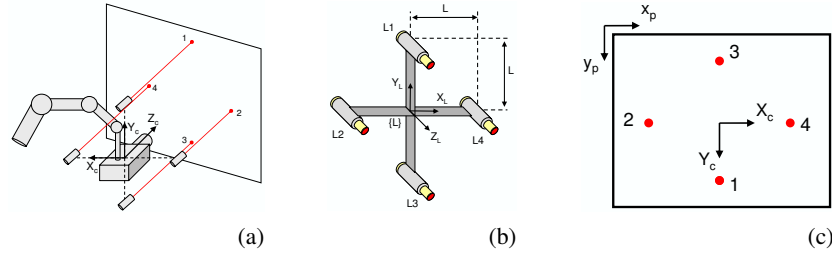


FIG. 3.4: Architecture du capteur caméra / dispositif lumineux en configuration idéale : (a) positionnement du capteur laser par rapport à la caméra ; (b) capteur constitué de 4 faisceaux laser ; (c) image en position désirée.

**Choix des informations visuelles et matrice d'interaction associée.** Les informations visuelles les plus intuitives sont obtenues en considérant les coordonnées dans l'image des 4 points éclairés par les faisceaux laser conduisant à la figure 3.4.c en position désirée. Toutefois, ce choix est loin de conduire à un comportement optimal de la loi de commande [Pagès *et al.*, 2004]. Nous avons considéré en revanche les informations visuelles suivantes<sup>9</sup> :

$$\mathbf{s} = \frac{1}{2} (y_1^{-1} - y_3^{-1}, y_1^{-1} + y_3^{-1}, x_2^{-1} + x_4^{-1}). \quad (3.2)$$

qui conduisent à la matrice d'interaction suivante :

$$\mathbf{L}_s = \begin{bmatrix} \alpha/L & \beta/L & -1/L & -\beta\gamma/L & \alpha\gamma/L & 0 \\ 0 & 0 & 0 & -(1 + \beta^2) & \alpha\beta & -\alpha \\ 0 & 0 & 0 & -\alpha\beta & 1 + \alpha^2 & \beta \end{bmatrix} \quad (3.3)$$

où  $L$  est la distance séparant la position d'un laser avec le centre du dispositif lumineux (voir la figure 3.4.b) et  $\alpha$ ,  $\beta$  et  $\gamma$  les paramètres décrivant l'équation du plan de l'objet

<sup>8</sup>Il est intéressant de noter ici que, contrairement au cas classique où le point  $\mathbf{x}$  décrit simplement la projection perspective d'un point  $\mathbf{X}$  de l'espace, tous les termes de la matrice d'interaction contiennent maintenant des informations 3D comme la position du laser, la distance de ce point au point  $\mathbf{X}$  mais également les paramètres décrivant le plan de l'objet [Pagès *et al.*, 2006c]. Cette matrice est donc beaucoup plus complexe à évaluer.

<sup>9</sup>Notons que les grandeurs  $y_1, y_3, x_2$  ou  $x_4$  ne sont jamais nulles, exceptées dans le cas dégénéré où la caméra est à l'infini.

tel que  $Z = \alpha X + \beta Y + \gamma$ . Comme on peut le constater, cette matrice d'interaction est très intéressante. Elle est toujours de rang plein 3. De plus, elle est fortement découplée. D'autre part, on peut montrer que les informations visuelles choisies peuvent se réécrire sous la forme :

$$\mathbf{s} = (\gamma/L, \beta, \alpha). \quad (3.4)$$

Dans le cas idéal, le choix des informations visuelles (3.2) est donc particulièrement judicieux puisqu'il permet d'estimer la totalité des termes de la matrice d'interaction à partir d'informations 2D uniquement (et de la distance  $L$ ). Ce qui n'est que rarement le cas en asservissement visuel classique puisque des informations de profondeur sont requises, faute de quoi la stabilité du système bouclé peut ne plus être garantie [Malis et Rives, 2003].

**Analyse de la stabilité.** Cette analyse s'appuie sur l'étude de la matrice  $\mathbf{M}_e$  (voir l'équation (2.11)) qui dépend de la loi de commande utilisée. Nous nous sommes restreints ici au cas où la loi de commande est de type Gauss-Newton. Relatons tout d'abord le cas où le dispositif formé des quatre faisceaux laser est positionné exactement au centre optique de la caméra et lorsque la distance  $L$  est parfaitement connue. Deux cas de figure peuvent être envisagés.

- *Utilisation de la matrice d'interaction en position courante.* En se rappelant que (3.2) permet d'obtenir (3.4), nous pouvons considérer que  $\widehat{\mathbf{L}}_s \approx \mathbf{L}_s$ , la matrice  $\mathbf{M}_e$  vaut alors l'identité conduisant à une décroissance exponentielle découplée du signal d'erreur  $\mathbf{e}$  et par suite des informations visuelles et cela dans tout l'espace de travail. Le système est donc globalement asymptotiquement stable.
- *Utilisation de la matrice d'interaction en position désirée.* La matrice  $\mathbf{L}_{s^*}$  s'écrivant très simplement, elle conduit à une matrice  $\mathbf{M}_e$  suffisamment simple pour que le système (2.11) puisse être résolu de manière analytique [Pagès et al., 2005a; Pagès et al., 2006c]. Ces résultats ont tout d'abord montré, qu'ici aussi, le système bouclé était globalement asymptotiquement stable. De plus, il a pu être montré que les deux dernières composantes du vecteur  $\mathbf{e}$  étaient monotones. Par contre, la première peut, dans certains cas, passer par un extremum avant de tendre ensuite vers zéro<sup>10</sup>. Par ailleurs, compte tenu de la simplicité de la matrice  $\mathbf{L}_{s^*}$ , les vitesses de la caméra sont directement proportionnelles aux composantes du vecteur  $\mathbf{e}$  et conduisent donc à des composantes monotones pour les vitesses de rotation  $\omega_x$  et  $\omega_y$ ,  $v_z$  pouvant, de la même façon que pour la première composante de  $\mathbf{e}$ , ne pas être monotone. Les trajectoires observées sont donc extrêmement intéressantes, en particulier beaucoup plus intéressantes que dans le cas où une matrice d'interaction estimée en position courante est utilisée. Ce résultat n'est pas surprenant en soi, il est en effet bien connu que le comportement du système bouclé peut être très différent selon que la matrice d'interaction est calculée en position courante ou désirée [Chaumette, 1998]. Ce comportement est clairement mis en évidence sur la figure 3.5 où les deux lois de commande sont comparées.

<sup>10</sup>Bien que le système bouclé soit globalement asymptotiquement stable (au sens de Lyapunov), cette propriété n'implique pas que  $\|\mathbf{e}(t)\|$  soit décroissant quel que soit  $t$ .

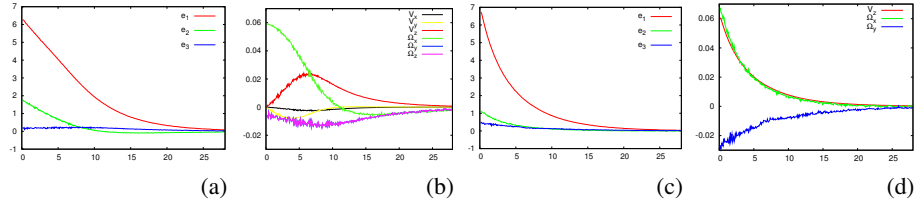


FIG. 3.5: Utilisation de la matrice d'interaction en position courante ou désirée. (a) et (b) utilisation de  $\mathbf{L}_S$ . (a) composantes du signal d'erreur en fonction du temps (en s) ; (b) vitesse de la caméra (ms/s ou rad/s) en fonction du temps (en s) ; (c) et (d) utilisation de  $\mathbf{L}_{S^*}$ . (c) composantes du signal d'erreur en fonction du temps (en s) ; (d) vitesse de la caméra (ms/s ou rad/s) en fonction du temps (en s).

**Analyse de la stabilité en présence d'erreur d'étalonnage.** Plusieurs types d'erreur d'étalonnage peuvent être considérés, le plus souvent on s'intéresse aux erreurs concernant les paramètres intrinsèques de la caméra [Espiau, 1993]. Toutefois, nous avons montré qu'un mauvais positionnement du dispositif lumineux par rapport au cas modélisé pouvait conduire à des erreurs de positionnement plus importantes [Pagès *et al.*, 2006c]. Aussi avons-nous dû introduire les informations visuelles suivantes :

$$\mathbf{s}' = \frac{1}{2} \left( y_1'^{-1} - y_3'^{-1}, y_1'^{-1} + y_3'^{-1}, x_2'^{-1} + x_4'^{-1} \right). \quad (3.5)$$

Sans entrer dans le détail des calculs, elles s'obtiennent grâce à une transformation affine de l'image désirée en une image qui serait obtenue en l'absence d'erreur de positionnement du dispositif lumineux. Signalons que dans le cas où cette erreur est nulle, nous avons bien sûr  $\mathbf{s}' = \mathbf{s}$  de telle sorte que les résultats précédemment établis restent valables.

Ces nouvelles informations visuelles ont conduit à une loi de commande très robuste [Pagès *et al.*, 2006c], mais l'analyse de la stabilité globale s'est révélée être malheureusement trop complexe. Nous présentons donc ici des résultats concernant la stabilité asymptotique locale conduisant à l'étude de la matrice  $\mathbf{M}^* = \mathbf{L}_{S^*} \mathbf{L}_S^+$  où  $\mathbf{L}_{S^*}$  est la matrice d'interaction calculée en prenant en compte le désalignement du dispositif laser. Compte tenu de la complexité des calculs, nous avons dû considérer séparément les mouvements de translation du dispositif laser des mouvements de rotation. Dans ces conditions, la stabilité locale a pu être prouvée quelles que soient les valeurs de ces translations ; concernant les rotations, la stabilité locale a pu être prouvée pour une valeur de rotation autour des axes  $X$ ,  $Y$  ou  $Z$  dans l'intervalle  $[-\pi/2; \pi/2]$ . Le domaine de convergence est donc extrêmement large. Ces résultats très intéressants sont illustrés sur la figure 3.6 où une tâche de positionnement a été réalisée en présence d'un fort désalignement du système mécanique liant les lasers entre eux (voir la figure 3.6a). Même dans ces conditions, un comportement exponentiel découplé des vitesses de la caméra peut être observé, ainsi qu'une trajectoire satisfaisante de la caméra.

De plus, nous avons également prouvé dans [Pagès *et al.*, 2005a] que la stabilité locale étaient également obtenue en présence d'erreur de calibration des paramètres intrinsèques de la caméra.

*Remarque :* Une approche également très robuste utilisant des informations visuelles complètement différentes (basées sur la surface du quadrilatère construit à partir des 4

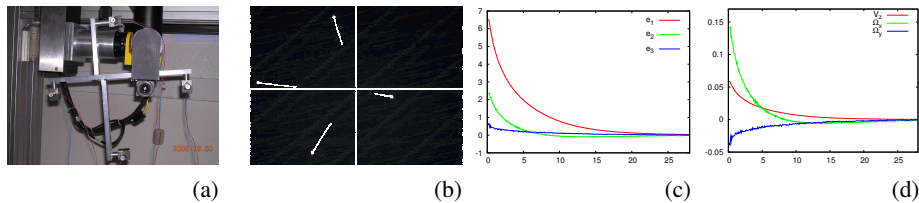


FIG. 3.6: Tâche de positionnement dans le cas d'un désalignement important. (a) vue du dispositif de projection désaligné; (b) image finale (avec la trajectoire des faisceaux laser); (c) composantes du signal d'erreur en fonction du temps (en s); (d) vitesse de la caméra (ms/s ou rad/s) en fonction du temps (en s).

points dans l'image issus des points éclairés par les faisceaux laser et de 2 différences d'angles issus des cotés de ce quadrilatère) a été présentée dans [Pagès et al., 2004; Pagès et al., 2005a], toutefois aucune propriété de convergence globale n'a pu être exhibée lors d'un alignement parfait. En revanche, la stabilité locale a pu être obtenue dans les mêmes conditions que ci-dessus concernant la modélisation du désalignement du dispositif laser.

**Comparaison avec une approche 3D.** Au lieu d'utiliser (3.5) pour estimer les paramètres du plan de l'objet, il est bien évidemment possible de les obtenir par triangulation. Plus précisément, nous avons utilisé une méthode permettant de prendre en compte, dans une certaine mesure, d'éventuelles erreurs d'alignement du dispositif lumineux [Pagès et al., 2006c]. Elle aboutit toutefois à un système d'équations non linéaires. En pratique, cette approche s'est révélée être moins robuste que l'approche 2D. En effet, la résolution du système d'équations peut converger vers un minimum local. Elle est également plus sensible au bruit présent dans l'image. De plus, aucune propriété de convergence ne peut être établie, la méthode étant basée sur une approche numérique de résolution du système d'équations.

### 3.1.2 Utilisation de la luminance

Nous avons recensé au début de ce chapitre plusieurs solutions cherchant à nous affranchir de l'apparence de l'objet observé dont, d'une façon générale, l'utilisation d'informations visuelles non géométriques. Cependant, contrairement à la plupart des travaux mentionnés à la section 3.1, notre approche est radicalement différente. En effet, le but consiste ici à utiliser directement la luminance en tant qu'informations visuelles impliquées dans une loi de commande, et non plus en tant que signal permettant la construction d'une information visuelle particulière, comme une homographie ou les paramètres d'un modèle de mouvement par exemple. Cette façon de procéder garantit, dans une certaine mesure, l'indépendance recherchée ; nous y reviendrons ci-après.

Toutefois, concevoir une loi de commande basée sur de telles informations visuelles est incontestablement une tâche délicate. En effet, la valeur de la luminance mesurée en un point dépend de phénomènes physiques complexes relatifs à l'interaction de la lumière et de la matière. Comme nous le verrons au chapitre 4, un grand nombre de paramètres



influent directement sur la lumière réfléchiée par la surface d'un objet dans une direction donnée. Citons par exemple la géométrie de la scène, le nombre de sources lumineuses, leur intensité, leur température de couleur mais également la nature physique même de l'objet. C'est ainsi que les techniques retenues en asservissement visuel basés sur des informations de luminance s'appuient sur des techniques d'apprentissage permettant d'éviter la connaissance de ces divers paramètres impliqués dans la mesure de la luminance [Deguchi et Noguchi, 1996; Nayar *et al.*, 1996; Deguchi, 2000]. Cette approche est toutefois problématique. Tout d'abord, elle requiert en pratique un grand nombre d'images de la scène pour diverses positions de la caméra. Elle est également problématique dans le sens où, à chaque nouvel objet ou, d'une manière plus générale à chaque modification d'un des éléments constituant le couple lumière / matière, une nouvelle phase d'apprentissage s'avère nécessaire et ce, même si le type d'informations visuelles reste inchangé. En fait, on ne sait pas, indépendamment de l'objet, exprimer la variation de telles informations par rapport au mouvement du porteur ; c'est-à-dire exprimer la matrice d'interaction de manière analytique. Par ailleurs, une approche intéressante, considérant également l'intensité lumineuse, a été proposée dans [Kallem *et al.*, 2007]. Elle est basée sur une méthode de noyaux opérant sur différents signaux permettant d'atteindre un haut degré de découplage de la loi de commande. Toutefois, seules les translations et la rotation suivant l'axe optique ont été considérées alors que notre approche considère les 6 degrés de liberté. Signalons également les travaux rapportés dans [Silveira et Malis, 2007a] où un asservissement visuel basé homographie est présenté. Dans ce travail, les intensités de l'image sont utilisées pour estimer une homographie entre les images courante et désirée par calcul d'une mise en correspondance, homographie qui est ensuite utilisée dans la loi de commande. Bien que, comme dans notre cas, l'intensité de l'image soit à la base de cette approche, un effort important de traitement d'images doit être réalisé pour obtenir cette homographie.

Nous avons voulu explorer, au cours de ma deuxième année de détachement, l'utilisation d'une matrice d'interaction obtenue sous forme analytique. Le but est bien sûr de nous affranchir des contraintes liées aux approches basées apprentissage que nous venons d'évoquer. De plus, nous avons également cherché à éviter l'estimation des nombreux paramètres que nous avons mentionnés. Cela n'est bien sûr possible que sous certaines hypothèses conduisant par voie de conséquence à autant de contraintes à respecter<sup>11</sup>.

*Remarque :* Utiliser la luminance comme informations visuelles dans un schéma de commande est également très intéressant d'un point de vue traitement d'images. En effet, ces traitements se trouvent réduits à leur minimum, les différentes phases de mise en correspondance n'étant plus à réaliser. Nous tenons à insister ici sur le fait que cette approche évite en particulier le problème très délicat de la mise en correspondance entre les informations visuelles initiales et désirées. C'est un problème très complexe et qui est malheureusement un prérequis à tout asservissement visuel.

---

<sup>11</sup>Lever ces hypothèses est néanmoins possible, mais exige une modélisation poussée du signal de luminance ainsi qu'un effort de traitement d'images pour estimer certains paramètres. Dans la mesure où cette modélisation ne dépend pas uniquement de l'objet d'intérêt, mais également de son environnement, ces travaux sont présentés dans le prochain chapitre.

**Hypothèses.** Nous nous plaçons tout d'abord dans le cas où les informations visuelles désirées  $\mathbf{I}^*$  sont connues grâce à un apprentissage. Cette apprentissage de l'image désirée  $\mathbf{I}^*$  pouvant être réalisé une bonne fois pour toutes, il faut pouvoir garantir que cette image restera invariante au cours du temps. En particulier, mais comme c'est déjà le cas en pratique avec la plupart des informations visuelles géométriques, les sources lumineuses, en position et en intensité, devront être inchangées. En outre, dans un souci de simplicité, nous nous plaçons dans le cas d'une scène supposée Lambertienne. En effet, dans ce cas, le résultat obtenu sera indépendant de la position de l'observateur. En revanche, il ne sera valide que pour un nombre réduit de matériaux, en l'occurrence les matériaux diélectriques inhomogènes opaques parfaits (voir le chapitre 4).

**Synthèse de la loi de commande.** Comme nous l'avons vu au chapitre 2, l'obtention de la loi de commande requiert le calcul de la matrice d'interaction. Ce calcul a été réalisé dans [Marchand, 2007]<sup>12</sup> sous l'hypothèse d'un modèle Lambertien et d'un éclairage fixe par rapport à la scène. De ce fait, cette matrice étant connue, une loi de commande est donnée simplement par (2.10). Cependant, nous avons montré dans [Collewet et al., 2008a; Collewet et al., 2008b] que les lois de commande utilisées habituellement pour traiter des informations visuelles de nature géométrique au mieux convergent, mais en générant des mouvements 3D inappropriés de la caméra, ou au pire divergent. Une loi de commande dédiée à la luminance a donc été proposée. Elle s'écrit formellement comme une loi de commande basée sur une méthode de Levenberg-Marquardt (voir la section 2.1.3); toutefois, le paramètre  $\mu$  est réglé de façon complètement différente. Sans entrer dans des considérations trop techniques, cette méthode exploite la forme particulière de la fonction de coût. En outre, des résultats de simulation ont montré que la forme de cette fonction de coût ne dépendait pas trop de l'objet observé tant que l'image ne présentait pas un motif périodique ou une forte variation des composantes du gradient spatial de l'image. Elle présente toujours une vallée très abrupte au milieu d'un plateau en pente douce et de pente non constante (voir la figure 3.7). Cette forme particulière s'explique très bien grâce à une étude locale de la fonction de coût par le biais de l'étude des valeurs propres de la matrice  $\mathbf{H}^*$  impliquée dans la loi de commande. En effet, cette étude permet d'exhiber des directions selon lesquelles le coût décroît soit de façon très faible lorsque la valeur propre associée est très faible (le fond de la vallée), ou inversement de façon importante lorsque la valeur propre est élevée (la direction orthogonale à la vallée), c'est-à-dire dans une direction très voisine de  $\nabla \mathcal{C}(\mathbf{r})$ . En pratique, une forte valeur de  $\mu$  est tout d'abord utilisée pour atteindre aisément le fond de la vallée puisque la loi de commande se comporte comme une approche de type gradient. Une valeur intermédiaire est ensuite utilisée pour atteindre le minimum avec un taux de convergence plus élevé tout en assurant une trajectoire 3D satisfaisante, tout au moins beaucoup plus satisfaisante qu'avec les lois de commande classiquement utilisées. Concernant la stabilité du système, elle ne peut qu'être locale puisque  $k > 6$  (voir la section 2.1.5). En pratique, le domaine de convergence est de l'ordre d'une vingtaine de degrés pour les rotations et d'une à quelques dizaines de centimètres pour les translations.

<sup>12</sup>Dans [Marchand, 2007], cette matrice d'interaction a été utilisée pour contrôler la position de la caméra et d'une source lumineuse afin de réaliser un éclairage optimal de la scène, sans toutefois réaliser une tâche de positionnement rigide basée sur la luminance.

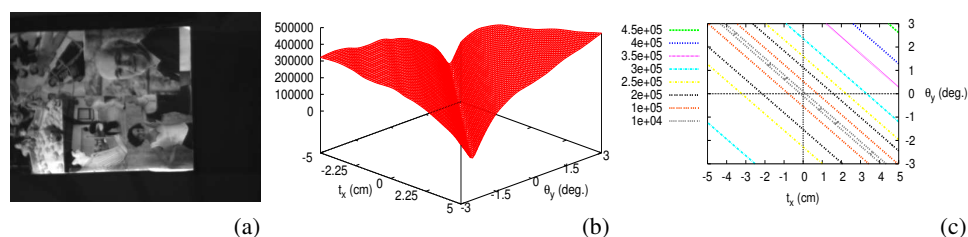


FIG. 3.7: Fonction de coût : (a) Objet observé (plan) ; (b) Allure de la fonction de coût : par rapport à sa position souhaitée, la caméra a subi une translation d'axe  $x$  de valeur  $t_x$  et une rotation d'axe  $y$  de valeur  $\theta_y$ . Cette figure montre clairement la présence d'une vallée étroite ; (c) Iso-contours dans le sous-espace  $(t_x, \theta_y)$  : la fonction de coût décroît rapidement dans une direction orthogonale aux iso-contours, cette direction est voisine de celle du gradient du coût.

**Résultats.** Différentes tâches de positionnement ont été réalisées (voir un exemple sur la figure 3.8) et ce, vis-à-vis de différentes textures validant le fait que l'allure de la fonction de coût ne dépend pas intrinsèquement de l'image. Cette fonction de coût est en revanche très sensible à la pose et de fait, des erreurs de positionnement extrêmement faibles ont toujours été obtenues. De plus, cette approche s'est révélée être très peu sensible aux occultations partielles de la scène (voir la figure 3.9) et à de fortes approximations des profondeurs, nécessaires pour le calcul de la matrice d'interaction. En outre, même si son calcul a été réalisé à partir d'hypothèses Lambertiennes, les expérimentations ont montré que de très faibles erreurs de positionnement pouvaient être obtenues même dans le cas d'une scène non Lambertienne. Rappelons une nouvelle fois, qu'excepté le calcul des gradients spatiaux, aucun autre traitement d'images n'est requis.

## 3.2 Indépendance vis-à-vis de la texture de l'objet et de sa géométrie

Les travaux rapportés dans cette section ont été initiés au Cemagref dans le cadre de la thèse d'Ali Alhaj et poursuivis au sein du projet Lagadic lors de ma première année de détachement. Ils concernent toujours la réalisation de tâches de positionnement vis-à-vis d'objets dont la texture est inconnue, nous avons supposé de plus leur forme inconnue et avons donc pris en compte explicitement ce fait. En effet, l'approche est ici foncièrement différente de celles se reposant sur la robustesse éventuelle de la loi de commande aux imprécisions sur l'estimation des profondeurs comme cela est fait dans la grande majorité des cas. Les objets en question sont en fait des produits carnés et les applications sont celles décrites dans le chapitre introductif à ce manuscrit. Cependant, il pourrait très bien s'agir d'autres applications non manufacturières comme celles mentionnées au début de ce chapitre, comme le positionnement d'un outil par rapport à un organe en robotique médicale par exemple. Dans ces conditions, nous avons supposé également qu'un apprentissage de l'image désirée n'était pas possible. Les travaux rapportés ci-après visent ainsi deux objectifs, s'affranchir de la géométrie de l'objet et de la connaissance de l'image désirée.

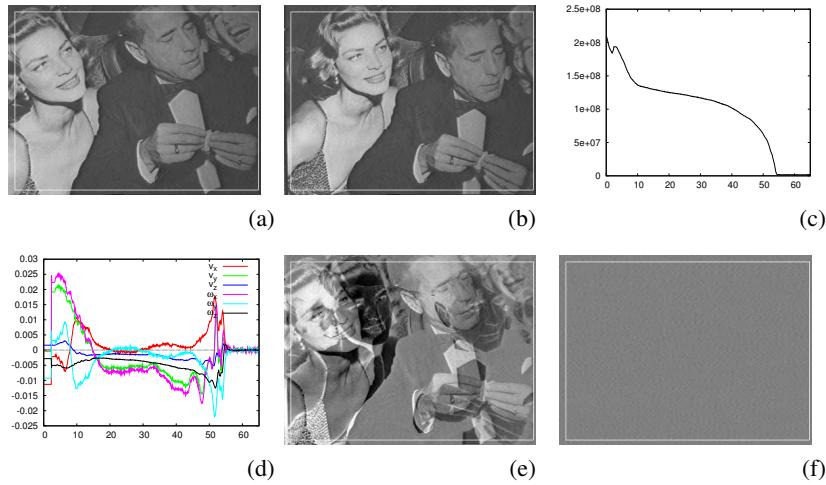


FIG. 3.8: Exemple de tâche de positionnement (axes des  $x$  en seconde). (a) Image initiale, (b) Image désirée, (c) Fonction de coût, (d) Vitesse de la caméra (m/s ou rad/s), (e)  $I - I^*$  pour la position initiale, (f)  $I - I^*$  pour la position finale.

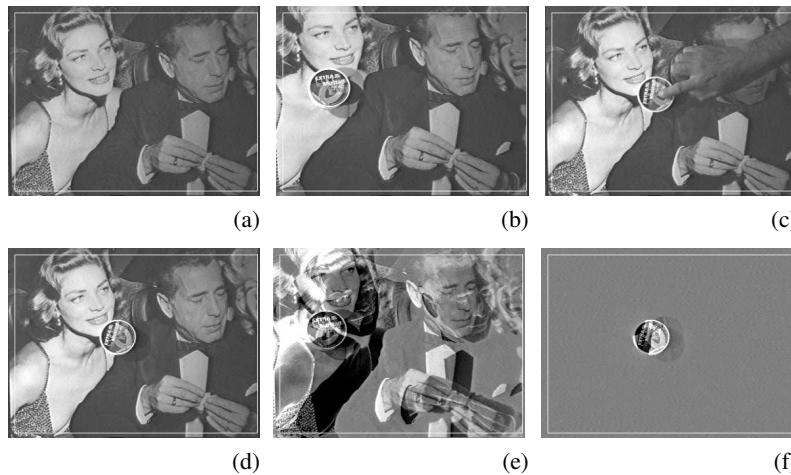


FIG. 3.9: Exemple d'une tâche de positionnement en présence d'occultations. Après avoir appris l'image désirée, un objet a été ajouté à la scène, il a ensuite été déplacé manuellement pendant la réalisation de la tâche. (a) Image apprise, (b) Image initiale, (c) Image intermédiaire, (d) Image finale, (e)  $I - I^*$  pour la position initiale, (f)  $I - I^*$  pour la position finale.

Le fait de ne pas disposer de cette connaissance en asservissement visuel est, on s'en doute, très problématique. En effet, il est dans ce cas impossible d'utiliser une approche classique basée sur un asservissement visuel 2D ou sur un asservissement visuel 3D basé sur un calcul de pose, le modèle de l'objet observé n'étant non plus disponible. De même, les approches dites « sans modèle » (voir par exemple [Malis *et al.*, 2000]) ne sont non plus envisageables, l'image désirée étant nécessaire. De la même façon, les différents travaux basés sur des informations visuelles de nature dynamique [Questa *et al.*, 1995; Sundareswaran *et al.*, 1996; Santos-Victor et Sandini, 1997; Créteil et Chaumette, 2001] supposent, en l'état, que la scène observée soit plane.

Cette problématique, véritable verrou à l'élargissement du champ applicatif de l'asservissement visuel à la robotique non manufacturière, a été initiée lors de mes travaux de thèse [Collewet, 1999]. La maximisation de la surface dans l'image du triangle formé à partir de trois points d'intérêt avait été utilisée pour la réalisation de tâches de positionnement vis-à-vis d'un objet plan [Collewet et Chaumette, 2002]<sup>13</sup>. En effet, comme nous l'avons vu au début de ce chapitre à la section 3.1, la mesure de type « point » constitue l'unique mesure géométrique exploitable dans une image permettant une indépendance par rapport à la texture de l'objet observé. À cet égard, précisons que les approches évoquées à la section 3.1 ne sont pas utilisables, car reposant toutes sur la connaissance de l'image désirée, inconnue ici.

Le travail rapporté maintenant est basé sur une reconstruction locale par vision dynamique parallèlement à la réalisation d'une tâche de positionnement par asservissement visuel 2D1/2. Plus précisément, l'approche a consisté à contrôler l'orientation du capteur par rapport à la normale au plan tangent en un point de l'objet observé. Ce point pouvant correspondre, par exemple, au centre d'une zone d'intérêt dans l'image. En pratique, il s'agissait d'une zone de l'image correspondant à un défaut d'aspect présent sur une pièce de viande<sup>14</sup>. L'approche présentée ici est donc à la fois indépendante par rapport à la texture de l'objet et à sa forme.

Les techniques de reconstruction 3D s'appuient le plus souvent sur l'utilisation de plusieurs caméras [Ayache et Lustman, 1987; Faugeras, 1993] mais ce choix implique l'appariement des données perçues au sein des différentes images acquises, ce que nous voulons éviter. La vision dynamique est une alternative à cette façon de procéder. En effet, plutôt qu'exploiter les images issues de plusieurs caméras, il est possible de n'en déplacer qu'une seule et d'obtenir ainsi une séquence d'images pour laquelle il est beaucoup plus aisé de réaliser la phase de mise en correspondance. Cette façon de faire est en fait très proche des techniques dites d'obtention de la « structure à partir du mouvement » où il s'agit, à partir de la mesure du mouvement 2D, de retrouver des informations 3D de la scène observée ainsi que le déplacement réalisé par le capteur<sup>15</sup>.

L'obtention de la structure à partir du mouvement a été largement traitée dans la littérature faisant émerger deux grandes familles d'approches selon la manière dont est appré-

<sup>13</sup>Le contexte applicatif de cette figure concernait la réalisation d'un démonstrateur en vue du tri à l'arrivée en salaisonnerie de jambons frais.

<sup>14</sup>Ces travaux ont été réalisés dans le cadre d'une convention Cemagref / INRIA / OFIVAL / Olympig en vue de la caractérisation de défauts d'aspect sur des pièces de viande.

<sup>15</sup>En vision dynamique, ce déplacement peut être considéré comme connu, ou tout au moins bien estimé, car issu d'une loi de commande.

hendé le mouvement 2D, de façon continue ou discrète.

- *Les approches continues* s'appuient sur une formulation différentielle du mouvement 2D et nécessitent donc, le plus souvent, la connaissance du flot optique [Prazdny, 1981; Bruss et Horn, 1983; Adiv, 1985; Sundareswaran, 1991; Heeger et Jepson, 1992; Gupta et Kanal, 1995; Irani *et al.*, 1997; Srinivasan, 2000]<sup>16</sup> voire de ces dérivées spatiales jusqu'à l'ordre deux dans le cas où une reconstruction locale est réalisée (on cherche alors l'orientation du plan tangent et les termes de courbure en un point) [Longuet-Higgins et Prazdny, 1980; Waxman et Ullman, 1985; Waxman *et al.*, 1987; Subbarao, 1988; Calway, 2005]. Ces approches aboutissent dans le cas général à la résolution d'équations non linéaires et nécessitent le plus souvent des hypothèses quant à la scène observée, en particulier la présence ou non de surfaces planes [Adiv, 1985; Negahdaripour, 1992; Irani *et al.*, 1997; Calway, 2005]. De plus, l'unicité de la solution n'est pas garantie, que l'objet observé soit plan ou non (voir [Subbarao, 1988] où ces différents cas d'ambiguïté sont recensés).
- *Les approches discrètes* reposent, quant à elles, sur un processus de mise en correspondance de primitives d'une image à l'autre. Ce type d'approche conduit donc nécessairement à une reconstruction éparse (voir néanmoins les techniques d'ajustement de faisceaux où une reconstruction quasi-dense est effectuée [Hartley, 1994; Lhuillier et Quan, 2002]). Dans le cas de la primitive point<sup>17</sup> des formulations linéaires peuvent être obtenues à l'aide de la décomposition de la matrice essentielle [Hartley, 1997] ou fondamentale [Luong et Faugeras, 1996] dès la donnée de 8 points appariés. Cependant, il est à noter la présence de cas dégénérés, en particulier lorsque la scène est plane [Longuet-Higgins, 1984] conduisant à l'utilisation d'approches spécifiques (voir par exemple [Irani et Anandan, 1998; Schindler et Suter, 2006]).

En outre, ce type d'approche devient très instable quand la caméra est soumise à de faibles mouvements entre les deux images considérées ; ce qui est malheureusement le cas dans une optique de commande puisque l'on cherche, au contraire, une fréquence d'acquisition des images la plus élevée possible. Une solution à ce problème est de considérer les deux images en question comme l'image courante et l'image désirée [Malis *et al.*, 2000] ou d'exploiter toute la séquence d'images [Soatto et Perona, 1998].

Contrairement à la plupart des autres approches basées également sur l'utilisation du mouvement 2D [Bouthemy et François, 1993; Questa *et al.*, 1995; Colombo *et al.*, 1995; Colombo et Allotta, 1999; Sundareswaran *et al.*, 1996; Santos-Victor et Sandini, 1997], recourir à la vision dynamique permet de contrôler les 6 degrés de liberté mais également de nous affranchir de la connaissance de l'image désirée et bien sûr, de la structure de la scène, et pour partie également de sa texture. Nous nous sommes focalisés, en effet, sur l'étude

<sup>16</sup>Voir cependant l'utilisation des méthodes directes permettant d'éviter le recours au calcul du flot optique [Negahdaripour et Horn, 1987; Horn et Weldon, 1988; Taalebinezhad, 1992; Stein et Shashua, 2000], ces approches se révèlent toutefois peu précises en pratique. Voir également l'utilisation de méthodes hybrides dans le sens où elles combinent à la fois des techniques continue et discrète [Rives et Espiau, 1987; Espiau et Rives, 1987; Xie et Rives, 1989]. En effet, des primitives particulières sont utilisées, comme dans le cas des approches discrètes pour effectuer la mise en correspondance, évitant ainsi le calcul du flot optique.

<sup>17</sup>Nous n'évoquerons pas les travaux concernant d'autres primitives dans la mesure où, par essence, elles nécessitent des hypothèses quant à l'objet observé que nous ne voulons pas faire.

d'un modèle de mouvement unifié, à même de considérer à la fois des scènes plane ou non plane, sans avoir bien évidemment à faire appel à des mécanismes de sélection qui seraient fonction de la scène observée [Irani et Anandan, 1998; Schindler et Suter, 2006]. Notre approche permet donc d'atteindre une forte indépendance vis-à-vis de l'objet observé.

**Reconstruction de la normale.** Le calcul de la normale unitaire  $\tilde{\mathbf{n}}$  en un point  $P$  ne pose pas de problème en lui-même, il s'obtient à partir des paramètres  $\alpha$  du plan tangent en  $P$  lorsqu'on l'écrit sous la forme  $1/Z = \alpha^\top \mathbf{x}$  où  $\mathbf{x}$  décrit les coordonnées projectives normalisées dans l'image d'un point  $m$  autour de  $p$  (voir les notations introduites à la page 9). En effet,  $\tilde{\mathbf{n}}$  s'écrit simplement dans ce cas  $\tilde{\mathbf{n}} = -\alpha / \|\alpha\|$ . Par contre, obtenir l'équation du plan tangent n'est pas trivial. Ce problème n'est toutefois pas nouveau (voir le bref rappel concernant les approches continues de structure à partir du mouvement). Néanmoins, ces travaux considèrent comme acquis le calcul du mouvement 2D alors qu'en pratique il n'en est rien, en particulier s'il doit être calculé au voisinage d'un point  $P$  de forte courbure. De même, l'utilisation d'une approche discrète est problématique, puisque elle repose sur une mise en correspondance de primitives éparses.

Réécrivons tout d'abord la profondeur en un point  $M$  du plan tangent sous la forme  $1/Z = \beta^\top \mathbf{u}$  où  $\mathbf{u} = (u, v, 1)$  avec  $u = x - x_p$  et  $v = y - y_p$  et où  $\beta$  s'écrit simplement en fonction de  $\mathbf{x}_p$  et de  $\alpha$  (voir également les notations page 9). Nous exprimons alors la profondeur en tout point  $M$  de l'objet d'intérêt sous la forme

$$\frac{1}{Z} = \beta^\top \mathbf{u} + \varphi(\mathbf{u}) \quad (3.6)$$

où la fonction  $\varphi(\mathbf{u})$  décrit uniquement les termes d'ordre supérieur à 1. Elle vérifie de plus  $\varphi(\mathbf{0}) = \mathbf{0}$  et s'annule pour un objet plan.

- *Utilisation d'une approche continue.* En utilisant l'équation du mouvement projeté [Longuet-Higgins et Prazdny, 1980] et en y introduisant (3.6), on obtient une expression générale du mouvement 2D pour tout point de l'objet. Toutefois, dans un premier temps, nous n'avons considéré que les points proches du point  $p$ . Il est alors possible de réécrire le mouvement 2D sous la forme d'un mouvement affine :

$$\dot{\mathbf{x}} = \mathbf{M} \mathbf{u} \quad (3.7)$$

où la matrice  $\mathbf{M}$  dépend uniquement de la projection  $p$  du point  $P$ , de la vitesse  $\mathbf{v}$  de la caméra et de la structure  $\beta$ . Cette structure s'obtient alors facilement par résolution au moindres carrés d'un système linéaire [Collewet et Chaumette, 2008] conduisant ensuite trivialement à la normale unitaire  $\tilde{\mathbf{n}}$ .

*Remarque :* Si dans  $\mathbf{M}$  on fait coïncider  $p$  avec le point principal, on retrouve la formulation habituelle du modèle de mouvement affine [Koenderink et Van Doorn, 1991]. Notre formulation est donc plus générale, mais surtout elle conduit à un modèle de mouvement beaucoup plus précis dès lors que  $p$  est loin du point principal (voir un exemple très parlant dans [Collewet et Chaumette, 2008]).

- *Utilisation d'une approche discrète.* En utilisant un raisonnement similaire au cas continu, il est possible d'exhiber de la même façon un modèle local de mouvement

affine

$$\mathbf{x}_{t+1} = \mathbf{M} \mathbf{u}_t \quad (3.8)$$

reliant le mouvement 2D entre deux images acquises aux instants  $t$  et  $t+1$ , la caméra ayant subi une rotation  $\mathbf{R}$  et une translation  $\mathbf{t}$  entre ces deux instants consécutifs à la loi de commande ;  $\mathbf{R}$  et  $\mathbf{t}$  sont donc supposés connus car s'exprimant en fonction de  $\mathbf{v}$ . Comme dans le cas continu, la matrice  $\mathbf{M}$  dépend du point  $p$  et de la structure  $\beta$ , elle dépend également de  $\mathbf{v}$  mais de manière indirecte par le biais de  $\mathbf{R}$  et  $\mathbf{t}$ . En revanche, l'obtention de  $\tilde{\mathbf{n}}$  est cette fois plus complexe, un système d'équation non linéaires étant alors à résoudre pour obtenir  $\beta$  [Collewet et Chaumette, 2008].

La même remarque que dans le cas continu peut être également formulée.

Qu'une approche continue ou discrète soit considérée, un modèle de mouvement 2D unifiée a été obtenu. En revanche, notre approche est locale. Nous avons toutefois montré qu'il était possible d'élargir son domaine de validité.

**Élargissement du domaine de validité du modèle de mouvement unifiée.** L'approche proposée repose sur la modélisation de l'erreur  $\mathbf{E}(\mathbf{u})$  définie comme la différence point à point entre le mouvement 2D exact (obtenu en considérant les profondeurs modélisées par (3.6)) et le mouvement donné par (3.8). Cette erreur s'écrit de façon très générale sous la forme de la somme de deux termes, un premier dû à la contribution des points appartenant au plan tangent, et un second dû aux points n'y appartenant pas. En pratique, le premier terme est très inférieur au second dès que la courbure de l'objet devient importante [Collewet et Chaumette, 2008]. Par ailleurs, il est même souvent très faible [Koenderink et Van Doorn, 1991; Negahdaripour, 1992], il a donc été négligé. Dans ces conditions, l'erreur  $\mathbf{E}(\mathbf{u})$  devient [Collewet et Chaumette, 2008] :

$$\mathbf{E}(\mathbf{u}) = \varphi(\mathbf{u}) \xi(\mathbf{u}, \mathbf{R}, \mathbf{t}, \beta, \varphi(\mathbf{u})). \quad (3.9)$$

Élargir le domaine de validité du modèle de mouvement unifié revient donc à minimiser (3.9) en tout point  $\mathbf{u}$  de la zone d'intérêt<sup>18</sup>. Deux stratégies distinctes ont été étudiées.

- *Utilisation de la vision active.* La première approche a consisté à minimiser  $\xi(\cdot)$ , le second terme impliqué dans  $\mathbf{E}(\mathbf{u})$ . Pour ce faire, nous avons étudié les conditions portant sur  $\mathbf{R}$  et  $\mathbf{t}$  conduisant à une valeur minimum de  $\xi(\cdot)$  et par suite de (3.9). En effet, imposer certaines contraintes sur le mouvement de la caméra est bien connu pour améliorer la précision de la reconstruction [Aloimonos *et al.*, 1987; Bajcsy, 1988; Ballard, 1991; Espiau et Rives, 1987; Xie et Rives, 1989; Chaumette *et al.*, 1996; Marchand et Chaumette, 1996]. En ce qui nous concerne, dans la mesure où  $\mathbf{R}$  reste bornée à une valeur faible entre les instants  $t$  et  $t+1$  (et ce, même si la rotation à réaliser pour atteindre la position désirée est importante), la condition de minimisation de  $\mathbf{E}(\mathbf{u})$  est simplement  $t_z = 0$ .
- *Utilisation d'une sélection de points.* Nous avons cherché un lieu de points  $\mathcal{L}(\mathbf{u})$ , le cas échéant loin du point  $p$ , pour lequel les erreurs  $\mathbf{E}(\mathbf{u})$  sont faibles. Ce lieu a

<sup>18</sup>Cette zone est en pratique la fenêtre dans laquelle le mouvement 2D est calculé.



ensuite été exploité lors du calcul du mouvement 2D. Pour simplifier le problème, nous avons considéré la classe d'objets telle que

$$\varphi(\mathbf{u}) = e_\beta(\mathbf{u}) = \beta_{20}u^2 + \beta_{11}uv + \beta_{02}v^2 \quad (3.10)$$

où les  $\beta_{ij}$  dépendent de la pose de la caméra par rapport à l'objet et de la courbure en  $P$ . Signalons que  $e_\beta(\mathbf{u})$  peut tout aussi bien être vu comme les termes d'ordre deux de la profondeur exacte en  $\mathbf{u}$  (donnée par (3.6)). De ce fait, (3.10) est valide pour tout objet tel que sa surface en  $P$  soit au moins  $\mathcal{C}^2$ . Nous nous sommes donc intéressés aux isocontours  $e_\beta(\mathbf{u}) = c$  où  $c$  est une constante arbitrairement faible. Sans entrer dans les détails, nous avons montré dans [Collewet et Chaumette, 2008] que, suivant les valeurs propres de la matrice associée à la forme quadratique définie par (3.10), un tel lieu  $\mathcal{L}(\mathbf{u})$  pouvait exister. Effectivement, il peut prendre selon les cas la forme d'une droite, voire de deux droites et au pire d'un unique point (le point  $p$ ). En conséquence, bien que le modèle de mouvement unifié ait été établi grâce à une modélisation locale, il a un caractère global pour tout point  $\mathbf{u} \in \mathcal{L}(\mathbf{u})$  quand  $\mathcal{L}(\mathbf{u})$  ne se réduit pas à au point  $p$ .

En pratique, la sélection des points appartenant à  $\mathcal{L}(\mathbf{u})$  est réalisée pendant le calcul du mouvement 2D à l'aide d'une technique statistique d'estimation robuste (les M-estimateurs [Huber, 1981]). On cherche donc la matrice  $\mathbf{M}$  telle que :

$$\widehat{\mathbf{M}} = \arg \min_{\mathbf{M}} \sum_{m_t \in W} \rho \left( \frac{I_t(m_t) - I_{t+1}(\delta(m_t, \mathbf{M}))}{\sigma} \right) \quad (3.11)$$

où la fonction vectorielle  $\delta(\cdot)$  décrit le mouvement 2D entre les images  $t$  et  $t+1$  selon le modèle affine unifié,  $I(\cdot)$  la fonction de luminance,  $\sigma$  la variance des résidus vérifiant le modèle affine et  $\rho(\cdot)$  une fonction robuste. La technique des moindres carrés itératifs pondérés a été utilisée pour résoudre (3.11) [Huber, 1981]. Pour ce faire une fonction d'influence [Hampel *et al.*, 1986] très stricte, dans le sens où elle s'annule si les résidus d'intensité sont trop élevés, a été utilisée. Il s'agit de la fonction de Tuckey bi-pondéré [Huber, 1981].

**Résultats.** Les techniques décrites ci-dessus ont permis la réalisation de tâches de positionnement par rapport à des objets plans ou non. Dans le cas d'objets plans, les deux techniques sont équivalentes. Par contre, concernant les objets non plans, ces techniques n'ont plus un comportement similaire. En effet, tant que les termes  $\beta_{ij}$  impliqués dans (3.10) sont faibles l'utilisation de la vision active fournit de bons résultats, dans le cas contraire de plus fortes erreurs peuvent être rencontrées. A contrario, l'utilisation des M-estimateurs permet une plus grande indépendance vis-à-vis de la valeur de ces termes. La figure 3.10 montre un exemple de tâche de positionnement vis-à-vis d'un objet imprécisément connu et de forte courbure.

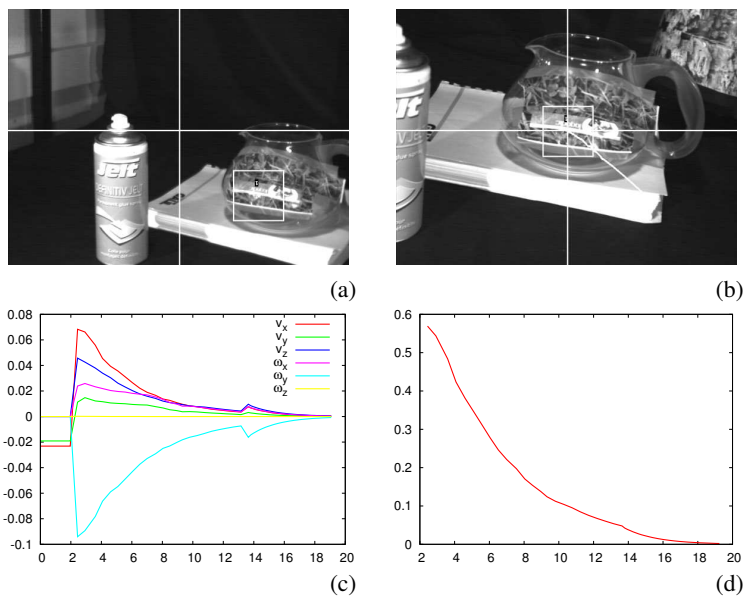


FIG. 3.10: Tâche de positionnement par rapport à un objet de forte courbure. (a) image initiale ; (b) image désirée ; (c) vitesse de la caméra (ms/s ou rad/s) en fonction du temps (en s) ; (d) Norme du signal d'erreur  $e$  en fonction du temps (en s).

## CHAPITRE 4

---

### Vers une indépendance vis-à-vis de la scène observée

---

L'objet du chapitre précédent fut de présenter certaines techniques d'asservissement visuel présentant, à divers degrés, une indépendance vis-à-vis de la texture de l'objet d'intérêt et, le cas échéant, vis-à-vis de sa forme également. En revanche, ces approches reposent sur l'hypothèse implicite d'une scène Lambertienne, ce qui peut être considéré comme une limite importante à ces approches. D'une façon générale, rares sont les travaux en asservissement visuel permettant de prendre en compte des variations brusques d'intensité des sources lumineuses, voire de directions mobiles par rapport à la scène d'intérêt, scène pouvant être constituée, le cas échéant, de matériaux spéculaires comme c'est le cas des produits carnés [Gouiffès *et al.*, 2004c]. Le but des travaux rapportés dans ce chapitre a consisté à s'affranchir de ces différents phénomènes. Comme nous le verrons ci-après, ils dépendent de la nature physique de l'objet, mais en grande partie aussi de la géométrie de la scène par le biais des positions relatives du triplet observateur / objet / éclairage.

Une façon de pallier ce problème pourrait consister à considérer des indices visuels de nature géométrique ; ce type d'indice procurant une relative indépendance à la géométrie de la scène. En effet, dans ces conditions, le problème se pose en terme de traitement d'images qui s'avère, en l'occurrence, robuste aux variations de luminance. Cela s'explique simplement par le fait que ces algorithmes exploitent, pour l'essentiel, les gradients spatiaux du signal de luminance permettant le suivi par exemple de droite ou de segments [Crowley *et al.*, 1992; Gennery, 1992; Boukir *et al.*, 1998], de coins [Sinclair *et al.*, 1993; Hager *et al.*, 1995], d'ellipses [Vincze, 2001] ou de contours complexes [Kass *et al.*, 1987; Terzopoulos et Metaxas, 1991; Bascle *et al.*, 1994; Precioso *et al.*, 2005; Pressigout et Marchand, 2007]/[Mebarki *et al.*, 2008]. Précisons que ces algorithmes deviennent plus robustes encore dès lors que l'on dispose d'un modèle 3D de l'objet [Drummond et Cipolla, 2002; Comport *et al.*, 2006b], autorisant par exemple des occultations partielles.

Cependant, pour préserver l'indépendance vis-à-vis de l'objet d'intérêt, acquise au chapitre précédent, nous n'avons pas considéré de tels indices visuels. Les travaux rapportés dans ce chapitre concernent donc exclusivement le signal de luminance, plus précisément l'image elle-même, ou un sous-ensemble de celle-ci. Dans ces conditions, l'invariance recherchée par rapport au cas général d'une scène dynamique a nécessité l'obtention d'une modélisation poussée de ce signal. Elle a été obtenue en considérant un modèle plus réaliste que celui de Lambert. Forts de cette modélisation, deux champs applicatifs complètement différents ont été investigués :

- Le premier porte sur l'amélioration des algorithmes de mise en correspondance de motifs texturés. En effet, comme déjà mentionné, c'est une primitive de bas niveau dont dépend un grand nombre d'algorithmes de vision par ordinateur et, en conséquence, certaines approches relevant de l'asservissement visuel. De tels motifs peuvent, en effet, conduire à la mise en correspondance de points d'intérêt, de régions ou encore au calcul des paramètres d'un modèle de mouvement paramétrique, informations dont dépend un grand nombre de techniques d'asservissement visuel largement évoquées dans ce document. Citons de façon non exhaustive [Papanikolopoulos, 1995; Comport *et al.*, 2006a; Benhimane et Malis, 2007; Questa *et al.*, 1995; Santos-Victor et Sandini, 1997; Crétual et Chaumette, 2001][Collewet *et Chaumette*, 2002; Collewet *et Chaumette*, 2008]. Notre contribution à ce thème concerne principalement la capacité à pouvoir traiter des images acquises dans des conditions complexes d'illumination (changement d'intensité de l'éclairage, mouvement de l'observateur et / ou de l'éclairage et apparition de spécularités) sans connaissance précise de la scène observée et sans avoir à estimer en ligne un trop grand nombre de paramètres, toujours dans une optique de calculs pouvant s'effectuer proche de la cadence vidéo, le but final étant la réalisation d'une tâche d'asservissement visuel. Des approches basées sur la luminance, sur la couleur ainsi que sur l'utilisation de corrections photométriques de certains invariants couleur ont été étudiées.
- Le deuxième champ applicatif concerne l'extension aux objets non Lambertiens de l'asservissement visuel basé sur la luminance décrit à la section 3.1.2. C'est, en effet, une approche que nous jugeons prometteuse vu les précisions de positionnement obtenues ainsi que la robustesse obtenue dans certaines configurations (erreurs sur les profondeurs, prise en compte des occultations). Notre contribution à ce thème porte principalement sur l'obtention analytique de la matrice d'interaction associée à la luminance et ce, dans le cas le plus général d'un objet non Lambertien. Comme nous le verrons, ces travaux nous ont permis de traiter le cas complexe d'une scène dynamique.

## 4.1 Modélisation de l'interaction lumière / matière

### 4.1.1 Modélisation de la réflexion

Chaque objet d'une scène réfléchit la lumière qu'il reçoit en modifiant sa couleur ou son intensité selon un modèle de réflexion. Plus précisément, un tel modèle permet de dé-

crire quantitativement la lumière réfléchie en fonction de divers paramètres caractérisant la scène, il permet donc d'établir la relation liant la *radiance* émise par un élément de surface en fonction de l'*irradiance* qu'il reçoit. En pratique, les paramètres impliqués dans ce processus sont nombreux, si nombreux même, que plusieurs modèles de réflexion ont été établis. Toutefois, le phénomène de la réflexion pouvant s'expliquer grâce à la théorie des ondes électromagnétiques, la perméabilité magnétique, la permittivité électrique et la conductivité du matériau permettent de décrire les propriétés optiques du matériau considéré. De ce point de vue, on peut considérer quatre grandes classes de matériaux :

- *Les matériaux conducteurs*. Pour ce type de matériaux, l'onde incidente est rapidement atténuée conduisant essentiellement à des phénomènes de surface. De ce cas, la rugosité joue un rôle prépondérant ; une surface très lisse aura un comportement proche de celui d'un miroir et la lumière sera réfléchie dans une seule direction. Le comportement du matériau est alors spéculaire et la première loi de réflexion de Descartes peut être appliquée en pratique. Quant aux matériaux plus rugueux, ils diffuseront la lumière autour de cette direction privilégiée, formant un lobe spéculaire.
- *Les matériaux diélectriques* (comme le verre par exemple). Ces matériaux sont peu conducteurs et laissent pénétrer très profondément l'onde incidente. Dans ces conditions, le modèle de réflexion doit tenir compte de phénomènes physiques dans le matériau lui-même, et dépend donc fortement de ce matériau.
- *Les matériaux homogènes* (comme par exemple le verre ou les métaux). Ces matériaux se caractérisent par un indice de réfraction constant à l'intérieur du matériau même. Les phénomènes de réfraction et de diffusion sont prépondérants.
- *Les matériaux inhomogènes* (comme par exemple le plastique, le papier, certains tissus). Ces matériaux sont composés de particules ayant des indices de réfraction différents. La réflexion diffuse caractérise ce type de matériaux.

Le modèle de Lambert [Lambert, 1760] permet d'expliquer le comportement des diélectriques inhomogènes opaques [Reichmann, 1973]. Il se caractérise par un unique terme de réflexion diffuse et ne prend donc pas en compte la direction de l'observateur. Le modèle de Beckmann-Spizzichino [Beckmann et Spizzichino, 1987] est basé sur les lois de l'électromagnétisme et sur une modélisation des aspérités de surface. Il permet donc de prendre en compte les matériaux conducteurs ou non, qu'ils soient lisses ou non. Le modèle de Torrance-Sparrow [Torrance et Sparrow, 1967], quant à lui, est basé sur les équations de l'optique géométrique, il se restreint par conséquent au cas où la taille des aspérités de surface excède largement la longueur d'onde de la lumière incidente. Ce modèle n'est donc valable que pour les matériaux rugueux (conducteurs ou non). Signalons également le modèle de Nayar [Nayar *et al.*, 1991] qui peut se concevoir comme une synthèse des deux modèles précédents tout comme le modèle de Shafer [Shafer, 1985] revisité ensuite par [Healey, 1989]. Ces modèles sont toutefois quantitatifs, ils expriment la radiançe d'une surface comme la somme d'un terme de réflexion diffuse et d'un terme de réflexion spéculaire. En outre, ces auteurs émettent l'hypothèse selon laquelle chaque terme de réflexion peut s'écrire comme le produit d'un terme ne dépendant que de facteurs géométriques et d'un autre terme fonction de la longueur d'onde. Ce type de modèle est à la base des travaux portant sur la recherche d'invariants couleur. Citons également le modèle de Phong [Phong, 1975] ; contrairement aux modèles rapportés jusqu'à présent, ce modèle ne repose

pas sur les lois de la physique, c'est un modèle issu de la communauté de la synthèse d'images. Il a toutefois le mérite d'être à la fois simple et très réaliste.

### 4.1.2 Modélisation de la luminance

Les différents modèles rappelés dans la section précédente permettent d'exprimer la radiance spectrale  $\mathcal{L}(\lambda)$  émise par une surface. Toutefois, la luminance étant une grandeur photométrique, c'est la sensation perçue par l'observateur, qui reçoit une certaine irradiance, qui nous intéresse en réalité. Dans le cas d'un dispositif optique de type capteur caméra, Horn [Horn, 1986] a montré que l'irradiance  $\mathcal{I}(\lambda)$  s'écrit :

$$\mathcal{I}(\lambda) = K_c \mathcal{L}(\lambda). \quad (4.1)$$

Nous supposons que  $K_c$  est un terme constant, ce qui est une hypothèse généralement admise comme valide (cas d'un objectif constitué de lentilles parfaites ou dotées du traitement de surface *ad hoc*). Dans ces conditions, en notant  $p$  le point dans le capteur issu de la projection du point  $P$  de la scène, la luminance  $I(p)$  s'écrit :

$$I(p) = K_c \int_{\lambda_{min}}^{\lambda_{max}} \mathcal{S}(\lambda) \mathcal{L}(\lambda, P) d\lambda \quad (4.2)$$

où  $\mathcal{L}(\lambda, P)$  est la radiance émise par un élément de surface en  $P$  et où  $\mathcal{S}(\lambda)$  décrit la sensibilité spectrale du capteur dans l'intervalle  $[\lambda_{min}, \lambda_{max}]$  (lorsque l'œil humain est le capteur considéré, la fonction d'efficacité lumineuse spectrale  $\mathcal{V}(\lambda)$  se substitue à  $\mathcal{S}(\lambda)$ ). Par suite, en utilisant explicitement un modèle de réflexion, il est possible de réécrire (4.2) sous la forme [Gouiffès, 2005] :

$$I(p) = k_d(p)a(p) \cos \theta_i(P) + K_s(p)h(P, \theta_r, \rho) \quad (4.3)$$

où  $k_d$  et  $K_s$  sont des gains dépendants du dispositif optique mais aussi de l'irradiance perçue par l'élément de surface en  $P$ ;  $a(p)$  est un terme dépendant du coefficient de réflectance spectrale diffuse en  $P$  (appelé aussi albedo), de l'irradiance spectrale en  $P$  et de la sensibilité spectrale du capteur en  $p$ . Les angles  $\theta_i$ ,  $\theta_r$  et  $\rho$  représentent respectivement les angles que forment avec la normale  $\mathbf{n}$  en  $P$ , la direction  $\mathbf{L}$  de la source lumineuse, la direction  $\mathbf{V}$  de l'observateur et la bissectrice  $\mathbf{B}$  entre les vecteurs  $\mathbf{V}$  et  $\mathbf{L}$  (voir la figure 4.1). La fonction  $h$  représente la contribution des termes spéculaires, elle dépend du modèle de réflexion utilisée.

### 4.1.3 Modélisation de la couleur

On sait depuis les travaux d'Helmholtz (1866) que la perception humaine de la couleur se caractérise par son aspect tridimensionnel. En synthèse d'image additive, les trois couleurs primaires normalisées sont le rouge, le vert et le bleu auxquelles sont respectivement associées les composantes trichromatiques spectrales ou fonctions colorimétriques spectrales  $\bar{r}(\lambda)$ ,  $\bar{v}(\lambda)$  et  $\bar{b}(\lambda)$  qui permettent de déterminer les quantités respectives  $(R, G, B)$  des trois primaires qu'il faut mélanger pour reproduire chaque couleur monochromatique

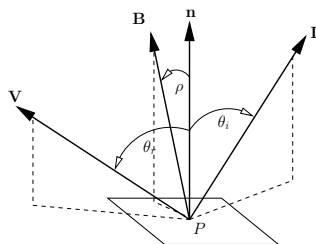


FIG. 4.1: Angles et vecteurs impliqués dans la description du phénomène de la réflexion.

de longueur d'onde  $\lambda$ . L'obtention de la couleur à l'aide d'un capteur tri-CCD est basée sur cette façon de faire en remplaçant les fonctions colorimétriques spectrales par les sensibilités spectrales  $\mathcal{S}_R(\lambda)$ ,  $\mathcal{S}_V(\lambda)$  et  $\mathcal{S}_B(\lambda)$  des trois capteurs. De ce fait, la relation (4.3) se généralise au cas de la couleur en écrivant  $\mathbf{C}(p) = (R, G, B)$  avec  $R$ ,  $G$  et  $B$  obtenus grâce à (4.2) en considérant les sensibilités spectrales propres à chacun des trois capteurs :

$$\mathbf{C}(p) = \mathbf{k}_d(p) \cdot \mathbf{a}(p) \cos \theta_i(P) + \mathbf{K}_s(p) h_f(P, \theta_r, \rho) \quad (4.4)$$

où «  $\cdot$  » exprime ici le produit d'Hadamard.

## 4.2 Application à la mise en correspondance de motifs texturés

Les sections précédentes nous ont permis de modéliser les signaux de luminance et de couleur, nous détaillons maintenant nos travaux concernant l'application de ces différentes modélisations au problème particulier de la mise en correspondance de primitives basées texture. Le problème se pose ici en terme d'estimation d'une transformation 2D permettant de recaler au mieux un motif de référence par rapport à un motif courant. Comme nous l'avons déjà mentionné, notre contribution ne porte pas sur l'algorithme de recalage en lui-même, mais sur une modélisation des variations d'illumination entre les deux motifs. Notre apport concerne donc tout aussi bien les algorithmes de mise en correspondance de primitives, où le recalage doit être effectué à partir d'images issues d'une même caméra, que les algorithmes d'appariement où des caméras différentes sont utilisées. Néanmoins, l'application présentée ici, et qui nous a permis de valider nos développements théoriques, concerne plus précisément le suivi de points d'intérêt ; toutefois, des motifs plus larges ont également été considérés.

Vu l'importance d'une mise en correspondance précise de deux primitives dans un algorithme de haut niveau, la littérature est, on s'en doute, très abondante sur le sujet et nous ne tenterons pas de dresser un état de l'art sur ce sujet, ce dernier n'étant, de surcroît, pas l'objet de notre contribution. Précisons néanmoins que parmi les deux grandes familles d'approche, minimisation d'une mesure de dissimilarité (SAD, SSD, ZSAD, ZSSD, ...) ou maximisation d'une mesure de similarité (CC, NCC, ZNCC, ...) (voir une étude de ces

différentes approches dans [Giachetti, 2000]), les approches de type SSD (pour Sum-of-Squared Differences), ou leur version robuste, sont les plus utilisées ; elles permettent en effet l'utilisation de méthodes de minimisation efficace (voir [Nickels et Hutchinson, 2002] pour une étude détaillée de ce type d'approche). En conséquence, nous avons validé nos résultats en les comparant avec des travaux relevant de cette technique.

Deux approches ont été investiguées, une première considère le cas de motifs décrits exclusivement par des informations de luminance alors que la deuxième considère aussi des informations de chrominance.

### 4.2.1 Utilisation de la luminance

Une façon efficace de résoudre le problème de mise en correspondance est de le traduire en terme d'une fonction de coût  $\mathcal{C}(\boldsymbol{\mu})$  à minimiser où  $\boldsymbol{\mu}$  est un vecteur paramétrisant le modèle de mouvement donné par la fonction  $\delta$ , i.e.  $m' = \delta(m, \boldsymbol{\mu})$  où  $m$  et  $m'$  représentent respectivement la projection d'un même point  $M$  de la scène dans les images  $I$  et  $I'$ . Nous notons  $\mathcal{W}$  le domaine de  $I$  sur lequel opère  $\delta$ . Dans ces conditions  $\mathcal{C}(\boldsymbol{\mu})$  s'écrit [Hager et Belhumeur, 1998] :

$$\mathcal{C}(\boldsymbol{\mu}) = \frac{1}{2} \|\mathbf{e}(\boldsymbol{\mu})\|^2 \quad (4.5)$$

où  $\mathbf{e}(\boldsymbol{\mu})$  est un vecteur colonne dont chaque composante  $e_m(\boldsymbol{\mu})$  s'écrit, pour tout  $m \in \mathcal{W}$ ,  $e_m(\boldsymbol{\mu}) = I'(\delta(m, \boldsymbol{\mu})) - f(I(m))$  ;  $f$  étant une fonction scalaire dépendant du modèle d'illumination et tel que  $f(I(m)) = I'(m')$ . La minimisation de (4.5) repose le plus souvent sur une méthode de Gauss-Newton pour obtenir  $\boldsymbol{\mu}^*$ , le minimiseur de (4.5). Pour ce faire, (4.5) est réécrit en remplaçant  $\mathbf{e}(\boldsymbol{\mu})$  par son développement en série de Taylor tronqué au 1<sup>er</sup> ordre autour d'une approximation  $\hat{\boldsymbol{\mu}}$  de  $\boldsymbol{\mu}^*$  tel que  $\boldsymbol{\mu}^* = \hat{\boldsymbol{\mu}} + \delta\boldsymbol{\mu}$ , conduisant à la solution recherchée :

$$\delta\boldsymbol{\mu} = -\lambda \hat{\mathbf{J}}^+ \mathbf{e}(\hat{\boldsymbol{\mu}}) \quad (4.6)$$

où  $\lambda$  est un scalaire positif et  $\hat{\mathbf{J}}$  une approximation de  $\mathbf{J}$ , la matrice Jacobienne de  $\mathbf{e}$  par rapport à  $\boldsymbol{\mu}$ . Notons à ce propos l'approche proposée dans [Jurie et Dhome, 2002] où un apprentissage de  $\mathbf{J}^+$  est réalisé, aboutissant à de bien meilleurs résultats que l'approche originale [Hager et Belhumeur, 1998].

Par ailleurs, le lecteur pourra constater une certaine analogie entre les équations (4.5) et (4.6) avec le problème d'obtention d'une loi de commande en asservissement visuel (voir les équations (2.7) et (2.10), page 23). À ce titre, il est possible d'établir ici aussi une matrice  $\mathbf{M}(\mathbf{e})$  décrivant les conditions de convergence de  $\mathbf{e}(\boldsymbol{\mu})$  vers  $\mathbf{0}$  au même titre que les conditions données en (2.12) et (2.15). Il en résulterait des contraintes fonction notamment des gradients spatiaux de l'image  $I'$ . L'analogie s'arrête toutefois ici. En effet, en asservissement visuel, des déplacements importants peuvent être réalisés pour atteindre la pose désirée alors qu'en suivi de primitives les déplacements sont plus faibles. De plus, et surtout, la trajectoire dans l'espace des paramètres  $\boldsymbol{\mu}$  nous important beaucoup moins que celle de la caméra lors d'une tâche d'asservissement visuel, la gamme d'approches permettant la minimisation de (4.5) est plus importante ici. Il serait tout à fait possible, et même souhaitable, d'utiliser des méthodes d'optimisation beaucoup plus efficaces que la méthode de Gauss-Newton, comme celles déjà signalées à la section 2.1.3 (les méthodes de régions de



confiance ou basées sur des méthodes de séparation et d'évaluation par exemple) ; voir dans cet esprit [Baker et Matthews, 2004] où d'autres méthodes de minimisation de (4.5) sont utilisées, elles restent néanmoins très classiques ; voir également [Silveira et Malis, 2007b] où une méthode du second ordre ayant un taux de convergence plus élevé que la méthode de Gauss-Newton est utilisée (l'algorithme ESM [Malis, 2004]).

Historiquement, les premiers travaux dans ce domaine sont dus à Kanade, Lucas et Tomasi [Lucas et Kanade, 1981; Tomasi et Kanade, 1991]. Leur approche repose sur la conservation de la luminance en tout point  $m$  de  $\mathcal{W}$  [Horn et Schunck, 1981] si bien que la fonction  $f$  impliquée dans  $e$  est simplement la fonction identité  $I(m) = I'(m')$ . Par ailleurs, le modèle de déplacement retenu dans leur cas se réduit à une translation 2D. Dans le cas général, ce modèle n'est donc pas valide, il peut néanmoins être considéré comme viable dans le cas d'images monoculaires acquises à fréquence élevée. Les travaux qui ont suivi ont donc naturellement porté sur la prise en compte de modèles de déplacement plus complexes, affine [Shi et Tomasi, 1994], quadratique [Odobez et Bouthemy, 1995], voire homographique [Buenaposada et Baumela, 2002; Jurie et Dhome, 2002; Benhimane et Malis, 2004][Collewet et al., 2004]. Néanmoins, toutes ces approches reposent sur l'hypothèse de conservation de la luminance qui, nous l'avons vu, n'est valide que pour des matériaux diélectriques inhomogènes opaques<sup>1</sup>. En outre, cette hypothèse est violée dès lors que l'objet n'est plus fixe par rapport à la direction de la source lumineuse. Bien évidemment, elle est encore violée dans le cas d'une variation d'intensité de l'éclairage ambiant.

Une façon d'aborder ce problème consiste à le placer dans le cadre plus large de l'estimation robuste. En effet, la minimisation de (4.5) ne conduit à une estimation  $\mu^*$  non biaisée que sous l'hypothèse de résidus  $e_m(\mu)$  indépendants et normalement distribués [Hampel et al., 1986]. L'utilisation d'une fonction robuste [Huber, 1981] permet de prendre en compte l'occurrence éventuelle de données aberrantes [Odobez et Bouthemy, 1995; Hager et Belhumeur, 1998; Black et Jepson, 1998; Black et al., 2000; Tarel et al., 2002; Vacchetti et al., 2004] comme par exemple la présence d'occultations partielles ou de variations inexplicables du signal de luminance. Il est toutefois plus rigoureux de considérer un modèle d'illumination plus valide que l'identité pour la fonction  $f$ . C'est ainsi, qu'en pratique, un modèle de variation affine est abondamment utilisé dans la littérature [Negahdaripour, 1998; Tommasini et al., 1999; Jin et al., 2001; Kim et al., 2005] conduisant à :

$$f(I(m)) = \lambda I(m) + \eta. \quad (4.7)$$

Précisons que dans la grande majorité des cas, le calcul des paramètres  $\lambda$  et  $\eta$  est issu d'une simple normalisation photométrique des images  $I$  et  $I'$ , alors que dans [Negahdaripour, 1998; Jin et al., 2001] par exemple, ils sont calculés pour minimiser (4.5) au même titre que  $\mu$ .

Le modèle (4.7) est toutefois encore très simple. Néanmoins, il peut être considéré comme localement valide [Gouiffès et al., 2006a], d'où son utilisation très fréquente dans le contexte de l'appariement ou du suivi de points d'intérêt. Dans le cas de motifs plus larges, des approches plus complexes sont utilisées. Elles sont quasiment toutes basées sur une phase d'apprentissage hors ligne. Plus précisément, elles reposent sur le fait que

<sup>1</sup>De plus, la pratique montre que l'utilisation exclusive de la composante Lambertienne impliquée dans la radiance en un point n'est plus valide pour une source ayant un angle d'incidence rasant ( $\theta_i \approx \pi/2$ ) en ce point.

l'image d'une scène Lambertienne, et en l'absence d'ombres, s'exprime de manière linéaire grâce à la connaissance de trois images acquises pour le même point de vue, mais dans des conditions variables d'illumination [Woodham, 1980; Ikeuchi, 1981]. Cette propriété est mise à profit par exemple dans [Hallinan, 1994; Hager et Belhumeur, 1998; La Cascia *et al.*, 2000]. Ce type d'approche est généralisé dans [Black *et al.*, 2000] au cas de scènes non Lambertiennes, sans toutefois reposer sur une explication physique du phénomène de la réflexion. Précisons que cette dernière approche nécessite également une phase d'apprentissage hors ligne. À l'inverse, dans [Silveira et Malis, 2007b], aucune phase d'apprentissage n'est nécessaire. La fonction  $f$  est décrite sous la forme d'une fonction paramétrique de vecteur  $\gamma$  entraînant une réécriture de (4.5) en fonction de  $\gamma$  également. L'algorithme ESM [Malis, 2004] permet d'obtenir à la fois  $\mu$  et  $\gamma$  par minimisation du nouveau critère. Cependant, vu la grande dimension de l'espace des paramètres, cette méthode n'est exploitable que sur des fenêtres  $\mathcal{W}$  de taille importante. De plus, la fonction paramétrique opérant dans ce cas sur une grande partie de l'image, l'hypothèse implicite d'un seul matériau présent dans la scène doit être formulée sous peine d'augmenter encore la dimension du vecteur  $\gamma$ .

**Compensation photométrique.** Dans le cas de fenêtres de taille plus réduite, comme c'est le cas en suivi de points d'intérêt ou de motif, le modèle (4.7) demeure extrêmement simple et n'est en particulier pas valide en cas d'occurrences de spécularités comme nous le verrons ci-après. Par contre, puisque nous disposons d'une modélisation de la luminance par le biais de l'équation (4.3), il est possible d'établir la relation exacte liant la variation de la luminance en un point entre deux instants  $t$  et  $t + 1$ . Deux formulations sont possibles [Gouiffès *et al.*, 2006a]. La première s'écrit :

$$f(I(m)) = I(m) + \psi(m). \quad (4.8)$$

Cette formulation est toutefois plus adaptée au cas où l'irradiance en  $P$  est constante et lorsque la scène est immobile par rapport à l'éclairage. En effet, dans ce cas, la fonction  $\psi(m)$  est relativement simple, elle ne dépend plus que de la fonction de spécularité  $h$  aux instants  $t$  et  $t + 1$ . Dans le cas général, elle dépend malheureusement du coefficient de réflexion diffuse  $a(m)$  qui peut varier sur  $\mathcal{W}$ , on se restreindra donc dans ce cas à une faible taille de fenêtre. La seconde formulation, en revanche, n'en dépend plus et s'écrit :

$$f(I(m)) = \lambda(m)I(m) + \eta(m) \quad (4.9)$$

où  $\lambda(m)$  et  $\eta(m)$  ne dépendent plus que de la géométrie de la scène (y compris la forme de la surface observée), de la fonction de spécularité  $h$  et d'une éventuelle variation d'irradiance en  $P$ . Cette seconde formulation généralise la modélisation utilisée habituellement (relation (4.7)) et montre qu'elle n'est valide que très localement, en théorie uniquement en  $p$ . Toutefois, son domaine de validité s'étend dans le cas d'une scène Lambertienne et si la surface est localement plane (voir [Gouiffès *et al.*, 2006a]).

Dans le cas de fenêtres de taille importante, il n'est pas possible d'estimer en pratique les paramètres  $\lambda(m)$  et  $\eta(m)$ , ils dépendent en effet d'un trop grand nombre de paramètres, de surcroît, de manière non linéaire. En revanche, si  $\mathcal{W}$  est suffisamment petite, il est possible d'estimer ces fonctions par leur développement en série de Taylor en  $p$  conduisant à un modèle local de variation de la luminance et ne comportant qu'un très petit nombre

de paramètres à identifier<sup>2</sup>. Une démarche similaire peut également être adoptée pour le modèle donné en (4.8). Dans ces conditions, la première formulation devient

$$f(I(m)) = I(m) + \boldsymbol{\alpha}^\top \mathbf{u} \quad (4.10)$$

où  $\mathbf{u} = (x - x_p, y - y_p, 1)$  alors que la seconde devient :

$$f(I(m)) = \boldsymbol{\lambda}^\top \mathbf{u} I(m) + \boldsymbol{\eta}^\top \mathbf{u}. \quad (4.11)$$

Les paramètres  $\boldsymbol{\alpha}$  ou  $\boldsymbol{\lambda}$  et  $\boldsymbol{\eta}$  sont ensuite estimés en ligne simultanément à  $\boldsymbol{\mu}$  dans le cadre de la minimisation de (4.5).

La validité de ce dernier modèle (4.11) a été discutée en détails dans [Gouiffès *et al.*, 2006b] rapportant certains cas de figure plus intéressants que d'autres (par exemple dans le cas où la direction d'observation ou la direction d'éclairage coïncide avec la normale en  $P$ ).

**Résultats.** L'apport des modèles (4.10) et (4.11) a été validé dans le cadre du suivi de points d'intérêt, et cela pour des tailles de fenêtres variables. Cinq méthodes ont été comparées : [Shi et Tomasi, 1994] (méthode « C »), [Tommasini *et al.*, 1999] (méthode « N »), [Jin *et al.*, 2001] (méthode « J »), l'utilisation de (4.10) [Gouiffès *et al.*, 2006a] (méthode « P<sub>3</sub> ») et l'utilisation de (4.11) [Gouiffès *et al.*, 2006b] (méthode « P<sub>6</sub> »). Plusieurs critères de comparaison ont été utilisés. Le premier concerne le nombre de points perdus au cours du mouvement. Un point est dit perdu si, une fois la minimisation de (4.5) réalisée, la valeur  $\mathcal{C}(\hat{\boldsymbol{\mu}}^*)$  reste supérieure à un seuil donné. C'est un critère très souvent utilisé, on peut montrer en effet que lorsque le seuil choisi est faible,  $\mathcal{C}(\hat{\boldsymbol{\mu}}^*)$  est directement lié à l'erreur de localisation  $\|\mathbf{m}' - \boldsymbol{\delta}(\mathbf{m}, \hat{\boldsymbol{\mu}}^*)\|$ . Un autre critère à retenir consiste à comparer, pour les différentes méthodes, l'évolution au cours du temps des résidus  $\mathcal{C}(\hat{\boldsymbol{\mu}}^*)$  calculés au même point  $p$  en fonction de  $t$ . En outre, nous nous sommes intéressés au conditionnement des différentes matrices à inverser impliquées dans (4.6). Une approche utilisant les compléments de Schur a permis de comparer ces matrices. Sans surprise, les méthodes C, N et P<sub>3</sub> ont un bon conditionnement, par contre les méthodes J et P<sub>6</sub> ont un moins bon conditionnement, celui de J étant toutefois le plus mauvais. Les matrices impliquées dans les méthodes P<sub>6</sub> et J ont donc été reconditionnées. Dans ces conditions, la méthode P<sub>6</sub> s'est révélée être la méthode la plus performante des quatre autres méthodes et ce, quelles que soient les causes de variation de l'illumination (variation d'irradiance ou apparition de spécularités), pourvu que la taille de la fenêtre  $\mathcal{W}$  soit suffisamment importante. En revanche, pour des tailles de fenêtre réduites, c'est la méthode P<sub>3</sub> qui est la plus valide. Ces résultats ont été validés pour différentes formes de surface ainsi que pour différents types de matériaux (des exemples de scènes sont rapportés sur les figures 4.2.a, 4.2.b, 4.3.a et 4.3.b). Par ailleurs, concernant les temps de calcul, celui de la méthode P<sub>3</sub> est identique à ceux des méthodes N ou J, celui de P<sub>6</sub> étant très légèrement supérieur à ceux des autres méthodes. Le détail de ces résultats ainsi que d'autres résultats peuvent être obtenus dans [Gouiffès, 2005; Gouiffès *et al.*, 2006a; Gouiffès *et al.*, 2006b].

<sup>2</sup>Cette façon de faire permet également de conserver le caractère local de la luminance (voir l'équation (4.3)) et de considérer ainsi le cas de plusieurs matériaux présents dans la scène.

### 4.2.2 Utilisation de la couleur

L'intérêt d'utiliser la couleur en traitement d'images n'est plus à prouver. La couleur a en particulier un pouvoir plus discriminant que l'utilisation de la luminance par le simple fait que deux points de même luminance peuvent très bien se différencier grâce à leurs attributs de chrominance. Elle permet également de simplifier fortement le traitement d'images, et en particulier la phase de segmentation d'un objet d'intérêt, quand la couleur est une caractéristique connue de l'objet ou du fond. Citons par exemple, la segmentation de joueurs de football [Vandenbroucke *et al.*, 2003], de mauvaises herbes en agriculture [Manh *et al.*, 2001], d'outils en robotique médicale [Wei *et al.*, 1997] ou d'identifiants grâce à la couleur de l'encre utilisée [Gouiffès *et al.*, 2004a; Gouiffès *et al.*, 2004b]. La couleur est également très employée pour la réalisation de tâches de suivi de régions ou d'objets dans des séquences d'images, comme par exemple le suivi de visage [Pérez *et al.*, 2002; Stern et Efros, 2005; Pizarro *et al.*, 2008] ou des mouvements humains [Wu et Huang, 2000].

En revanche, la couleur est très peu utilisée dans le contexte du suivi de points d'intérêt, voir néanmoins [Heigl *et al.*, 1999] où les auteurs ont étendu l'algorithme KLT à ce cas. Les résultats sont bien sûr à son avantage. Ils montrent, en outre, qu'une taille de fenêtre  $\mathcal{W}$  plus faible peut-être utilisée. À notre sens, la réticence à utiliser la couleur dans ce cas de figure est due à l'*a priori* concernant une durée de calcul qui serait prohibitive. Compte tenu de la puissance de calcul des machines actuelles, nous pensons que cet argument n'est plus fondé.

**Utilisation d'invariants couleur.** Nous avons vu dans la section 4.2.1 qu'il était possible de compenser les variations d'illumination dans une image de luminance grâce à la modélisation des phénomènes de réflexion. Toutefois, l'approche obtenue est locale. En revanche, les invariants couleur ont dans certaines conditions bien connues un caractère global. Les propriétés souhaitées concernent l'invariance par rapport à la géométrie de la scène, c'est-à-dire vis-à-vis d'un déplacement de l'objet, de la caméra ou de l'éclairage, mais également vis-à-vis d'une variation d'irradiance de la surface observée. C'est ainsi qu'une abondante littérature porte sur l'utilisation d'invariants couleur pour la recherche de motifs acquis préalablement dans des conditions différentes d'acquisition [Ohba *et al.*, 2000; Drew *et al.*, 1998; Alferez et Wang, 1999; Kim *et al.*, 2001; Gevers, 2001; Levine et Bhattacharyya, 2005] que le motif observé; voir également [Mileva *et al.*, 2007] où une récente contribution des invariants couleur au calcul du flot optique est rapportée. Néanmoins, chaque invariant a son propre domaine de validité, dépendant à la fois du type de matériaux de la surface observée et de la couleur de l'illuminant (blanc ou coloré). Les invariants les plus usités reposent sur le modèle dichromatique de Shafer évoqué à la section 4.1.1.

Dans le cas d'un objet parfaitement Lambertien et d'un illuminant parfaitement blanc (l'irradiance est alors indépendante de la longueur d'onde), il est aisé de montrer à l'aide du modèle de Shafer que tout rapport de composantes  $RGB$  est invariant vis-à-vis des modifications d'irradiance et de la géométrie, conduisant aux invariants normalisés par les normes  $L_1$  ou  $L_2$ . Ces invariants sont cependant sensibles au bruit dans le cas de faibles valeurs de luminance; ce que corrigent en partie les attributs  $c_1 c_2 c_3$  [Gevers et Smeul-

ders, 1997]. La littérature rapporte bien d'autres invariants encore mais leur utilisation est basée sur des hypothèses d'utilisation plus restrictives. Dans le cas où l'illuminant est coloré, un rapport de composantes  $RGB$  restera invariant à la géométrie de la scène, mais plus aux variations d'irradiance, de fortes hypothèses sont alors nécessaires pour assurer une invariance par rapport aux variations d'irradiance (capteur à bande étroite, illuminant Planckian<sup>3</sup>) [Gevers et Smeulders, 1997; Finlayson *et al.*, 2000].

Il existe également des invariants dédiés aux matériaux conducteurs, toutefois, à notre connaissance, ils se restreignent tous au cas d'un illuminant blanc. Dans ces conditions, tout rapport de différences de coordonnées couleur est invariant à la fois aux variations d'irradiance éventuelles et à la géométrie. Ainsi la teinte  $H$  est un invariant bien connu. De même, les attributs  $l_1 l_2 l_3$  conduisent à un espace de couleur invariant [Gevers et Smeulders, 1997; Gevers *et al.*, 1998]. Comme dans le cas des matériaux Lambertiens, ils s'avèrent bruités dans le cas d'une faible luminance mais également dans le cas de couleurs peu saturées ( $R \approx G \approx B$ ). D'autres attributs ont été également proposés, comme les invariants différentiels par exemple [Geusebroek *et al.*, 2001], cependant une connaissance *a priori* de la scène observée est requise (type de matériaux, surface plane ou non, surface terne ou non, etc.).

#### **Compensation photométrique des invariants dédiés aux surfaces Lambertiennes.**

En pratique, les invariants dédiés aux matériaux conducteurs, susceptibles donc de modéliser les phénomènes spéculaires, se sont révélés présenter une forte perte de leur pouvoir discriminant en présence de deux couleurs « proches » [Gouiffès, 2005]. Nous avons donc proposé d'utiliser la méthode  $P_3$  (voir la section 4.2.1) sur chaque composante d'un invariant dédié aux objets Lambertiens, conduisant ainsi à une correction des effets introduits par l'apparition de spéularités.

**Résultats.** Deux cas de figure ont été principalement étudiés, le cas de scènes comportant des objets Lambertiens ou non, soumis à un illuminant blanc d'intensité lumineuse variable (voir un exemple de scènes sur les figures 4.2.a et 4.2.b) et le cas de scènes dédiées plus particulièrement aux objets spéculaires, l'illuminant étant toujours blanc mais d'intensité constante (voir les figures 4.3.a et 4.3.b). Un exemple de résultat est donné dans les tableaux rapportés dans les figures 4.2.c et 4.3.c où les scènes respectivement décrites dans les figures 4.2.a et 4.3.a ont été impliquées. Dans les deux cas, le seuil de réjection des points a été fixé à une valeur conduisant un comportement très sévère. Dans tous les cas de figure, l'apport de la compensation photométrique par la méthode  $P_3$  a amélioré de façon très significative le comportement de l'algorithme de suivi. Dans le premier cas, les attributs  $c_1 c_2 c_3$  ainsi que ceux basés sur la norme  $L_2$  se sont avérés avoir un comportement très satisfaisant, surclassant l'utilisation de la luminance ou des seules composantes  $RGB$ . Dans le second cas, l'apport de la compensation photométrique s'est avéré être indispensable. En outre, dans tous les cas, l'utilisation de la luminance a toujours conduit à de moins bons résultats que l'utilisation d'attributs couleur. Par ailleurs, notons que l'utilisation d'un

<sup>3</sup>Un illuminant est dit Planckian (ou Planckien) si son spectre d'émission correspond à celui d'un corps noir à température de couleur donnée.

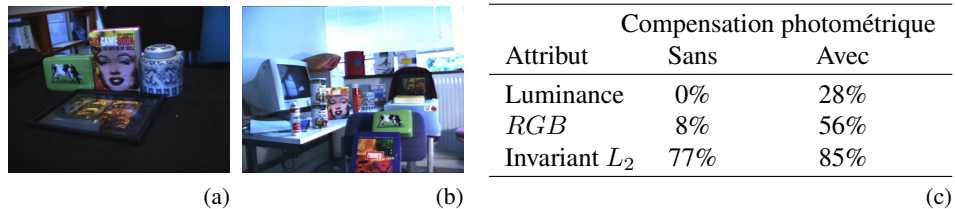


FIG. 4.2: Exemple de scènes comportant des matériaux de différents types soumis à une variation de l'intensité lumineuse : (a) et (b). Pourcentage de points suivis pour différents attributs dans le cas où l'attribut considéré est compensé ou non par la méthode P3 : (c). Les objets étant spéculaires, la correction photométrique améliore fortement le pourcentage de points suivis, et cela quel que soit l'attribut considéré. De plus, l'intensité lumineuse étant variable, l'utilisation d'un invariant couleur est pleinement justifiée.

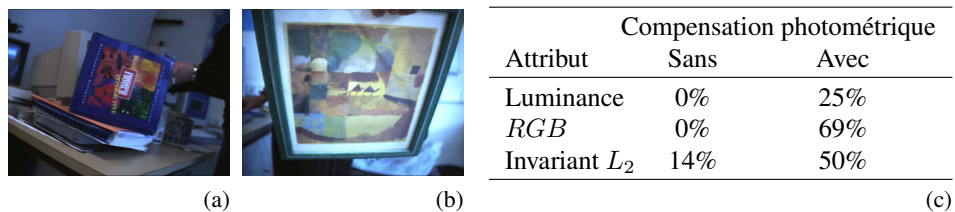


FIG. 4.3: Exemple de scènes comportant des objets spéculaires, l'intensité lumineuse étant maintenue constante : (a) et (b). Pourcentage de points suivis pour différents attributs dans le cas où l'attribut considéré est compensé ou non par la méthode P3 : (c). Les objets étant spéculaires, la correction photométrique améliore ici aussi fortement le pourcentage de points suivis, et cela toujours quel que soit l'attribut considéré. En revanche, l'intensité lumineuse étant constante, l'utilisation d'un invariant couleur n'est maintenant plus justifiée.

invariant couleur par rapport aux composantes *RGB* n'est justifié qu'en cas d'une variation notable de l'illumination.

### 4.3 Application à l'asservissement visuel basé sur la luminance

Nous avons déjà abordé l'étude de l'asservissement visuel basé sur la luminance au chapitre précédent (voir la section 3.1.2) et les différents travaux s'y rapportant. Le but recherché ici est identique, il s'agit d'utiliser la luminance en tout point de l'image comme informations visuelles d'un schéma d'asservissement visuel. L'originalité de notre approche par rapport aux travaux existants repose sur le calcul analytique de la matrice d'interaction. Rappelons qu'elle était apprise jusqu'à présent, avec les limitations que nous avons déjà soulignées. Pour ce faire, nous avons considéré à la section 3.1.2 le modèle de réflexion le plus simple, le modèle de Lambert. Nous avons également supposé un objet fixe par rapport à la source lumineuse. Ces deux hypothèses nous ont conduits à une formulation

extrêmement simple de la matrice d'interaction, ne dépendant pas, par exemple, de la position de cette source lumineuse. En revanche, l'utilisation de ce modèle est soumise à des hypothèses très restrictives concernant le type de matériau constituant les objets présents dans la scène. Ils se doivent être des diélectriques inhomogènes opaques parfaits.

Nous avons donc considéré un modèle de réflexion plus réaliste, il a conduit à une expression générale de la luminance donnée en (4.3). Toutefois, pour être réellement utilisable, il faudrait être en mesure d'estimer certains paramètres de cette expression, en particulier  $k_d(p)a(p)$  (dénommé ci-après  $K_d(p)$ ) mais aussi d'autres paramètres impliquées dans la fonction  $h$  [Tominaga, 1996; Tan et Ikeuchi, 2005]. Dans ce premier travail, nous les avons estimés grâce à la connaissance *a priori* que nous avons de la scène. Nous sommes toutefois conscients qu'en pratique une phase d'estimation de ces paramètres sera nécessaire dans le futur. Par ailleurs, nous avons utilisé un modèle d'illumination relativement simple, il s'agit du modèle de Phong, déjà évoqué à la section 4.1.1. Toutefois, l'utilisation d'un modèle basé sur la physique de la réflexion, comme celui de Torrance-Sparrow, semble possible. Dans ces conditions, le seul obstacle à l'utilisation de la luminance en asservissement visuel reste le calcul analytique de la matrice d'interaction pour une géométrie quelconque.

**Calcul de la matrice d'interaction.** À partir du modèle de Phong, il est possible d'obtenir la matrice d'interaction en tout point  $m$  de l'image, elle s'écrit dans le cas général sous la forme suivante [Collewet et Marchand, 2008b] :

$$\mathbf{L}_I = -\nabla I^\top \mathbf{L}_x + nK_s (\mathbf{R}^\top \mathbf{V})^{n-1} \mathbf{L}_1^\top + K_d \mathbf{L}_2^\top \quad (4.12)$$

où le vecteur  $\mathbf{R}$  décrit la direction idéale de réflexion spéculaire ;  $n$  est un paramètre dépendant de la rugosité du matériaux utilisé et permet de modéliser la largeur du lobe spéculaire autour de  $\mathbf{R}$ . Les matrices  $\mathbf{L}_1$  et  $\mathbf{L}_2$  ont été calculées dans le cas général, elles dépendent de la géométrie de la scène ainsi que de forme de la surface en  $M$ . Nous avons donc considéré deux cas particuliers. Le cas où la source lumineuse est localisée à l'infini et le cas où elle est embarquée sur la caméra. De plus, nous avons supposé une scène plane et avons calculé la matrice d'interaction en position désirée, soit dans la position où les plans de la scène et du capteur sont parallèles et situés à une distance  $Z^*$ . Dans ce cas la matrice d'interaction s'écrit simplement. Nous donnons à titre d'exemple le résultat obtenu dans le cas où la source lumineuse est embarquée sur la caméra (voir également [Collewet et Marchand, 2008a] où le modèle de Blinn-Phong a été utilisé) :

$$\mathbf{L}_1^\top = \frac{1}{\|\mathbf{x}\|} \begin{pmatrix} \frac{1}{\|\mathbf{x}\|^2} \frac{x}{Z^*} & \frac{1}{\|\mathbf{x}\|^2} \frac{y}{Z^*} & \frac{-1}{\|\mathbf{x}\|^2} \frac{x^2 + y^2}{Z^*} & y & -x & 0 \end{pmatrix} \quad (4.13)$$

et  $\mathbf{L}_2^\top = \mathbf{0}$ .

**Résultats.** La connaissance analytique de la matrice d'interaction nous a permis de réaliser diverses tâches de positionnement tout en prenant en compte l'occurrence éventuelle de spéularités dans les images. La figure 4.4 illustre ce cas de figure, l'éclairage

étant solidaire de la caméra. Comme dans le cas de scènes Lambertiennes (voir la section 3.1.2), la réalisation de tâches de positionnement a toujours conduit à des erreurs très faibles. La matrice d'interaction donnée en (4.12) a permis également de réaliser des tâches de suivi de cible. La figure 4.5 rapporte un exemple de tâche où l'objet était solidaire d'un rail motorisé, la caméra étant également embarquée.

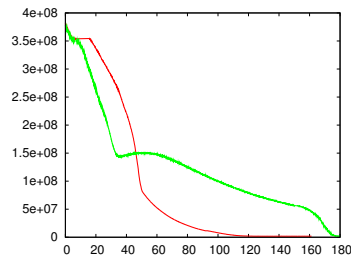
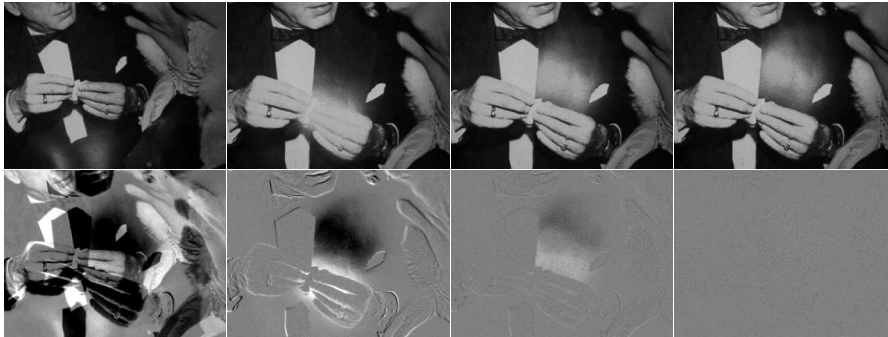


FIG. 4.4: Exemple de tâche de positionnement dans le cas où la source lumineuse est montée sur la caméra. Prise en compte des spécularités. Première ligne : image initiale, deux images intermédiaires, Image finale ; Deuxième ligne : erreur correspondante  $\mathbf{I} - \mathbf{I}^*$  ; Troisième ligne : comparaison des fonctions de coût dans le cas de l'utilisation d'une matrice d'interaction calculée sous des hypothèses Lambertiennes (en vert) ou à l'aide du modèle de Phong (en rouge).



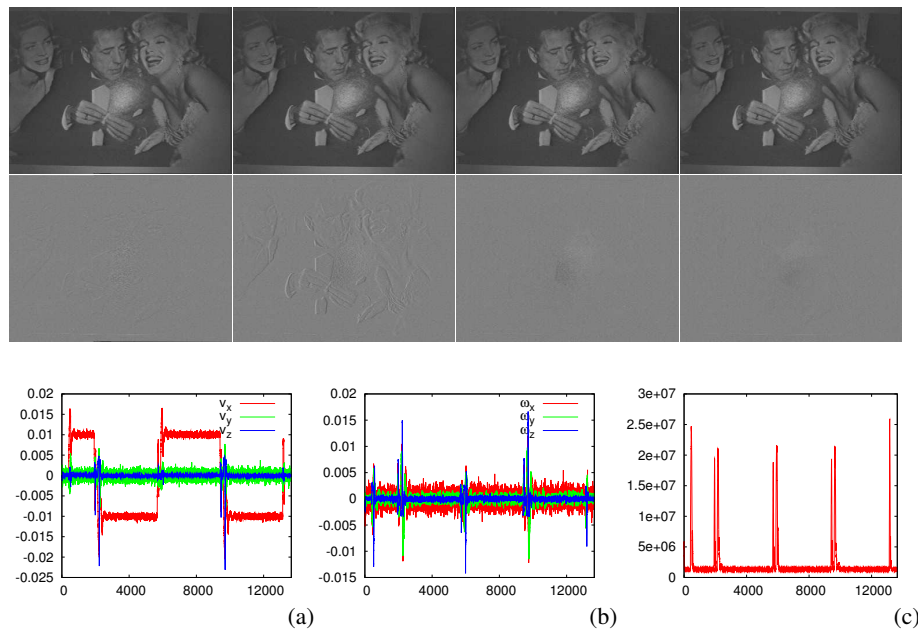


FIG. 4.5: Tâche de suivi d'une cible. Première ligne : images acquises par la caméra ; Deuxième ligne : erreur correspondante  $\mathbf{I}-\mathbf{I}^*$ . Troisième ligne (axe des  $x$  en nombre d'images) : (a) vitesse de translation de la caméra (ms/s) ; (b) vitesse de rotation de la caméra (rad/s) ; (c) fonction de coût.



## CHAPITRE 5

---

### Conclusion et perspectives

---

Les travaux rapportés dans ce manuscrit concernent, d'une manière générale, le domaine de l'asservissement visuel. Comme nous l'avons vu tout au long de ce document, cette technique permet le contrôle d'un système dynamique grâce à des lois de commande en boucle fermée sur des informations visuelles. Même si plusieurs problèmes d'ordre théorique ont été évoqués au chapitre qui lui a été réservée, cette approche se comporte de façon extrêmement satisfaisante en pratique. Elle est, par exemple, robuste dans une certaine mesure à l'imprécision de divers modèles, modèle de l'objet, du capteur mais également du robot. En revanche, nous avons vu que la réalisation d'une tâche robotique vis-à-vis d'un objet d'intérêt non précisément connu, de par sa texture et / ou de sa forme, voire de par la matière même le constituant, s'avérait problématique. De même, nous avons vu que la méconnaissance de l'environnement extérieur dans lequel cet objet était plongé, était également problématique.

Les travaux présentés dans ce document ont permis d'apporter des solutions communes à ces deux cas de figure, conduisant à une certaine indépendance par rapport à la scène observée ; y compris dans le cas d'une scène dynamique, incluant donc toute variation d'illumination au niveau du capteur. Dans ces conditions, nous pensons avoir apporté une contribution importante au domaine de l'asservissement visuel en élargissement considérablement son champ applicatif, des scènes mal connues peuvent maintenant être considérées. Plusieurs approches ont été proposées dans ce but :

- Nous avons tout d'abord investigué l'apport de l'éclairage structuré à l'asservissement visuel. Ce type d'approche a permis de nous affranchir grandement de la texture de l'objet observé, tout en rendant la phase d'extraction des indices visuels très simple. De plus, dans le cas d'un éclairage codé, toutes les phases de mise en corres-

pondance sont extrêmement robustes de par l'approche même utilisée. Par ailleurs, outre l'indépendance obtenue vis-à-vis de la texture de l'objet, une loi de commande particulièrement optimisée a pu être élaborée dans le cas où l'éclairage était embarqué.

- L'utilisation d'une reconstruction 3D locale par vision dynamique, autour d'une zone d'intérêt, nous a permis de nous affranchir également de la texture de l'objet observé, en exploitant les paramètres d'un modèle de mouvement unifié. Ce modèle est dit unifié dans la mesure où des scènes aussi bien planes que non planes peuvent être appréhendées, sans aucun mécanisme de sélection de modèles. En outre, cette approche a permis, comme nous l'avons vu, de nous affranchir de l'image désirée qui s'avère être problématique à obtenir quand l'objet n'est pas précisément connu et lorsqu'un apprentissage n'est pas possible.
- Pour terminer la synthèse de nos contributions, l'indépendance par rapport à la scène observée (modification de sa géométrie et prise en compte de variations éventuelles d'illumination) a pu être obtenue grâce à l'utilisation d'une modélisation exploitable du phénomène de la réflexion. Cette modélisation a conduit à l'élaboration d'un algorithme de suivi de motifs texturés, robuste à toute variation d'irradiance au niveau du capteur, c'est-à-dire à la dynamique de la scène. Dans ce contexte, nous avons pu observer l'apport indéniable de certains invariants couleur.

En outre, la prise en compte du phénomène de la réflexion, a donné lieu à une approche réellement novatrice<sup>1</sup> en asservissement visuel, il s'agit de l'introduction directement dans un schéma de commande 2D du signal de luminance et ce, pour une géométrie de la scène quelconque (au sens des positions relatives du triplet observateur / objet / éclairage). En effet, la matrice d'interaction associée à la luminance a pu être obtenue sous une forme analytique dans ce cas. Ce type d'approche permet, ici aussi, d'assurer une certaine indépendance vis-à-vis de la texture de l'objet. De plus, cette approche a permis d'éviter toutes les phases de mise en correspondance, y compris la phase très délicate de mise en correspondance entre les informations visuelles initiales et désirées.

Les perspectives à ces différents travaux restent toutefois nombreuses, nous citons ici celles qui nous paraissent les plus pertinentes.

Nous avons présenté dans ce manuscrit une première approche de couplage de l'asservissement visuel avec un dispositif de lumière codée. Néanmoins, les motifs présents dans la littérature, dont celui que nous avons utilisé, ont été conçus dans un but de reconstruction 3D et aucunement en vue d'une utilisation en asservissement visuel.

- Une première perspective à ce travail concerne l'optimisation de la conception de tels motifs, optimisation en terme de robustesse vis-à-vis de rotations importantes de la caméra par rapport à l'objet éclairé, en particulier suivant l'axe optique, et plus généralement vis-à-vis de toute déformation du motif dans l'image. En effet, les motifs actuels, et en particulier ceux basés sur une matrice parfaite, sont sensibles à

<sup>1</sup>En exceptant les approches basées sur des techniques d'apprentissage [Deguchi et Noguchi, 1996; Nayar *et al.*, 1996; Deguchi, 2000], difficilement exploitables en pratique.

de telles déformations dans le sens où le décodage peut être mis en échec, rendant toute phase d'appariement impossible.

- Dans le même ordre d'idée, bien que nous n'ayons jamais voulu formuler une telle hypothèse jusqu'à présent, un modèle de l'objet éclairé offrirait à n'en pas douter des perspectives de recherche particulièrement intéressantes. Dans ces conditions, comme nous l'avons fait pour le dispositif d'éclairage constitué de quatre faisceaux laser et pour un objet plan, il serait possible de concevoir un motif optimal dédié à certaines formes d'objets simples, tels des cylindres, des sphères, voire des formes quadratiques par exemple. Le but recherché est d'aboutir, comme dans le cas des objets plans, à une loi de commande « optimale » garantissant tout aussi bien la convergence vers la position désirée qu'une trajectoire satisfaisante de la caméra. De tels motifs pourraient être définis à partir de critères géométriques dans l'image, comme le parallélisme entre deux droites, la concentricité de deux cercles, voire des contraintes sur des moments par exemple.
- L'hypothèse de la connaissance d'un modèle de la scène observée est également intéressante dans le cas d'un éclairage déporté et codé. En effet, cette hypothèse est en fait tout à fait réaliste dans certains domaines d'application, comme en robotique chirurgicale mini-invasive. Le modèle 3D de l'organe d'intérêt est, en effet, obtenu lors d'examens pré-opératoires à l'aide d'autres modalités d'imagerie. De plus, la géométrie de la scène étant entièrement maîtrisée, il serait possible de concevoir un motif à projeter dédié à la forme de l'organe à traiter. Ce motif serait optimal tout à la fois en terme de propriétés désirées pour la loi de commande, mais aussi en terme de déformations minimum du motif observé dans l'image. Il est en effet possible, connaissant la scène sur laquelle le motif doit être projeté, de projeter un motif tel qu'il apparaîtra non déformé dans l'image (voir par exemple [Harville *et al.*, 2006]). Le problème dans sa globalité est certes complexe, mais sa résolution conduirait à un comportement que l'on pourrait qualifier d'optimal en terme de robustesse de la loi de commande et de la trajectoire de la caméra, tout en assurant une mise en correspondance rapide et robuste des informations visuelles.

D'un tout autre point de vue, nous avons vu que l'utilisation du modèle de réflexion de Phong, plus réaliste que le modèle de Lambert, a conduit à la conception d'un asservissement visuel basé sur la luminance. L'exploitation de ce signal s'est révélée très prometteuse. En effet, ce type d'informations visuelles s'est avéré être très peu sensible aux occultations partielles de la scène ainsi qu'à de fortes approximations des profondeurs nécessaires pour le calcul de la matrice d'interaction. Rappelons également que cette approche évite toute phase de mise en correspondance. Cependant, elle conduit à un domaine de convergence de la loi de commande beaucoup plus étroit qu'avec des informations visuelles géométriques. Un travail est donc nécessaire à ce niveau pour que cette approche soit pleinement exploitable. Nous pensons notamment à l'utilisation des techniques de régularisation qui pourraient contraindre le système à vérifier certaines contraintes, portant en particulier sur la trajectoire de la caméra. Une autre façon de procéder pourrait consister, bien sûr, à ne plus utiliser directement le signal de luminance mais une transformation de celui-ci. Nous pensons là à des techniques basées noyaux [Kalle *et al.*, 2007] ou à l'utilisation des moments n'opérant plus sur une image binaire (comme cela est le cas par exemple dans [Tahri

et Chaumette, 2005]) mais sur l'image de luminance directement.

Nous envisageons également d'étudier l'apport de notre approche au calcul du flot optique lorsque les scènes observées sont soumises à des variations complexes d'illumination. En effet, grâce au modèle de Phong, il est également possible de réécrire l'équation de contrainte du flot optique dans le cas d'une scène non Lambertienne. À notre connaissance, la littérature sur ce sujet ne rapporte que le cas d'utilisation du modèle de réflexion de Lambert. Une première formulation a été présentée en ce sens dans [Collewet et Marchand, 2008a] mais n'a pu encore être validée dans le cadre spécifique du calcul du flot optique.

Toutefois, qu'il s'agisse de l'asservissement visuel ou du calcul du flot optique, l'utilisation d'un modèle de réflexion, à même de modéliser l'occurrence de spécularités, requiert la connaissance d'un certain nombre de paramètres que nous avons ajustés manuellement pour l'instant. Ces paramètres étant dépendants de la scène observée, une estimation en ligne de ces paramètres serait préférable. Certains auteurs relatent de tels travaux (voir par exemple [Tominaga, 1996; Tan et Ikeuchi, 2005]). Néanmoins, les approches proposées n'exploitent qu'un faible nombre d'images. Dans le contexte de l'asservissement visuel, il serait tout à fait envisageable de généraliser ces travaux à un plus grand nombre d'images, quitte à fournir une première estimation de ces paramètres qui seraient raffinés ensuite au cours du mouvement.

Cependant, une alternative à ces travaux existe. Elle consiste à utiliser non plus l'information de luminance, mais des invariants couleur. En effet, certains invariants relatés dans ce document se sont avérés performants tout en s'écrivant de façon très simple. Le calcul de la matrice d'interaction de telles informations visuelles ne devrait donc pas poser de problèmes particuliers. Il s'en suivrait alors une invariance vis-à-vis d'une scène dynamique (variations d'irradiance de la scène et occurrence de spécularités). Cette approche permettrait en particulier de traiter simplement le cas de figure où les informations visuelles courantes sont obtenues dans d'autres conditions d'irradiance que les informations visuelles désirées. Quant à l'application des invariants couleur au calcul du flot optique, elle a été rapportée très récemment dans [Mileva *et al.*, 2007]. Les résultats présentés montrent clairement l'intérêt d'utiliser les invariants couleur et nous confirment dans l'idée que leur utilisation serait également judicieuse dans le cadre de l'asservissement visuel.

Pour terminer ces perspectives, nous nous proposons d'investiguer une toute nouvelle application de l'asservissement visuel. Il s'agit du contrôle des écoulements des fluides. C'est un domaine de recherche récent et très actif dans lequel nous souhaitons nous impliquer par le biais d'une collaboration Cemagref / CNRS / INRIA / Université de Rennes 1. Dans la grande majorité des cas, il s'agit de contrôler la transition laminaire / turbulente de la couche limite dont on souhaite avancer ou reculer le seuil. Les retombées potentielles sont nombreuses. Par exemple, supprimer ou retarder le phénomène de décollement de la couche limite autour d'un profil permet d'améliorer la portance, de réduire le bruit et la traînée des engins aéronautiques conduisant ainsi à une réduction de la consommation de carburant. En revanche, dans d'autres domaines d'application, comme en génie chimique par exemple, il peut être avantageux de favoriser les phénomènes de turbulence afin d'augmenter les échanges de chaleur ou de différents produits chimiques et ainsi augmenter la

vitesse de la réaction. Le même but est également recherché dans le domaine militaire pour augmenter la furtivité. Cependant, il se pose actuellement le problème de l'introduction de capteurs dans l'écoulement permettant de synthétiser une loi de commande sur des informations en boucle fermée. En effet, ils se doivent d'être non intrusifs. Malheureusement, de tels capteurs ne sont actuellement pas disponibles. Seuls des capteurs disposés le long de la (ou des) paroi(s) où s'écoule le fluide le sont [Breuer, 2002; Collis *et al.*, 2004]. L'idée est alors de remonter à l'état complet du système grâce à une modélisation utilisant les seules mesures disponibles sur la (ou les) paroi(s) [Höpfner *et al.*, 2005]. Pour pallier ce problème, nous comptons utiliser un dispositif de visualisation<sup>2</sup> de l'écoulement et ainsi remonter au champ de vitesse des particules [Corpetti *et al.*, 2002; Corpetti *et al.*, 2006]. De cette façon, il sera possible de proposer une loi de commande basée directement sur des informations visuelles issues du dispositif de visualisation et pouvant caractériser globalement l'écoulement, c'est-à-dire de réaliser un asservissement visuel. La tâche sera, à n'en pas douter ardue, les informations visuelles à considérer ici étant de nature dynamique mais ces travaux nous permettrait d'aborder des champs applicatifs complètement vierges.

---

<sup>2</sup>Il peut s'agir d'un système similaire à celui intervenant dans la mesure de vélocimétrie laser par images de particules (PIV). C'est-à-dire, une caméra, un ensemenement de fines particules et un dispositif d'illumination par laser, ou alors des techniques de type Schlieren ou infra-rouge selon l'écoulement étudié.





---

## Bibliographie

---

- Abdul Hafez, A.H., S. Achar et C.V. Jawahar (2008). Visual servoing based on gaussian mixture models. Dans : *IEEE Int. Conf. on Robotics and Automation, ICRA'08*. Pasadena, California. pp. 3225–3230.
- Adiv, G. (1985). Determining three-dimensional motion and structure from optical flow generated by several moving objects. *IEEE Trans. on Pattern Analysis and Machine Intelligence* **7**(4), 384–401.
- Agin, G.J. (1979). Real time control of a robot with a mobile camera. Dans : *9th Int. Symp. on Industrial Robots*. Washington, USA. pp. 233–246.
- Agin, G.J. (1985). Calibration and use of a light stripe range sensor mounted on the hand of a robot. Dans : *IEEE Int. Conf. on Robotics and Automation, ICRA'85*. St Louis, Missouri, USA. pp. 680–685.
- Albitar, C., P. Graebbling et C. Doignon (2007). Robust structured light coding for 3d reconstruction. Dans : *IEEE Int. Conf. on Computer Vision*. Rio de Janeiro, Brésil.
- Alferez, R. et Y.F. Wang (1999). Geometric and illumination invariants for object recognition. *IEEE Trans. on Pattern Analysis and Machine Intelligence* **21**(6), 505–536.
- Alhaj, A (2004). Apports de la vision dynamique en asservissement visuel. Thèse de doctorat. Université de Rennes I. Rennes.
- Allibert, G., E. Courtial et Y. Touré (2007). A flat model predictive controller for trajectory tracking in image based visual servoing. Dans : *7th IFAC Symposium on Nonlinear Control Systems, NOLCOS'07*. Pretoria, South Africa.
- Aloimonos, Y. (1990). Purposive and qualitative active vision. Dans : *IAPR Int. Conf. on Pattern Recognition, ICPR'90*. Vol. 1. Atlantic City, New Jersey. pp. 346–360.
- Aloimonos, Y., I. Weiss et A. Bandopadhyay (1987). Active vision. *Int. Journal of Computer Vision* **1**(4), 333–356.

- Alshawish, H. M. et C. R. Allen (1995). 3d object recognition using coded light projection for robot assembly applications. Dans : *IEEE Int. Conf. on Industrial Electronics*. Vol. 2. pp. 1240–1247.
- Amin-Nejad, S., J. S. Smith et J. Lucas (2003). A visual servoing system for edge trimming of fabric embroideries by laser. *Mechatronics* **13**(6), 533–551.
- Andreff, N., B. Espiau et R. Horaud (2002). Visual servoing from lines. *Int. Journal of Robotics Research* **21**(8), 669–699.
- Ayache, N. et F. Lustman (1987). Fast and reliable passive stereovision using three cameras. Dans : *IEEE Int. Workshop on Industrial Applications of Machine Vision and Machine Intelligence*. Tokyo, Japon.
- Bajcsy, R. (1988). Active perception. *Proceedings of the IEEE* **76**(8), 996–1005.
- Baker, S. et I. Matthews (2004). Lucas-kanade 20 years on : a unifying framework. *Int. Journal of Computer Vision* **56**(3), 221–255.
- Ballard, D.H. (1991). Animate vision. *Artificial Intelligence* **48**(1), 57–86.
- Barreto, J.P., F. Martin et R. Horaud (2002). Visual servoing/tracking using central catadioptric images. Dans : *Int. Symp. on Experimental Robotics, ISER'02*. Bombay, India. pp. 863–869.
- Basclé, B., P. Bouthemy, N. Deriche et F. Meyer (1994). Tracking complex primitives in an image sequence. Dans : *IAPR Int. Conf. on Pattern Recognition*. Jerusalem. pp. 426–431.
- Beckmann, P. et A. Spizzichino (1987). *The scattering of electromagnetic waves from rough surfaces*. 2 ed.. Artech House Inc.
- Benhimane, S. et E. Malis (2004). Real-time image-based tracking of planes using efficient second-order minimization. Dans : *IEEE/RSJ Int. Conf. on Intelligent Robots Systems*. Vol. 1. Sendai, Japon. pp. 943–948.
- Benhimane, S. et E. Malis (2007). Homography-based 2d visual tracking and servoing. *Int. Journal of Computer Vision* **26**(7), 661–676. Numéro spécial IJCV/IJRR sur la vision pour la robotique.
- Black, M.J., D.J. Fleet et Y. Yacoob (2000). Robustly estimating changes in image appearance. *Computer Vision and Image Understanding* **78**, 8–31.
- Black, M.J. et A.D. Jepson (1998). Eigentracking : Robust matching and tracking of articulated objects using a view-based representation. *Int. Journal of Computer Vision* **26**(1), 63–84.
- Boukir, S., P. Bouthemy, F. Chaumette et D. Juvin (1998). A local method for contour matching and its parallel implementation. *Machine Vision and Application* **10**(5/6), 321–330.
- Bouthemy, P. et E. François (1993). Motion segmentation and qualitative dynamic scene analysis from an image sequence. *Int. Journal of Computer Vision* **10**(2), 157–182.
- Breuer, K.S. (2002). Sensors, actuators and algorithms for practical implementations of turbulence boundary layer control. Dans : *3rd Japan Symposium on Turbulence and Flow Control*. Tokyo, Japon.

- Bruss, A.R. et B.K.P. Horn (1983). Passive navigation. *Computer Vision, Graphics, and Image processing* **21**(1), 3–20.
- Buenaposada, J. et L. Baumela (2002). Real-time tracking and estimation of plane pose. Dans : *IAP Int. Conf. on Pattern Recognition, ICPR'02*. Vol. 2. Québec, Canada. pp. 697–700.
- Cadenat, V., Soueres et M. Courdesses (2001). Using system redundancy to perform a sensor-based navigation task amidst obstacles. *Int. Journal of Robotics and Automation* **16**(2), 61–73.
- Calway, A. (2005). Recursive estimation of 3d motion and surface structure from local affine flow parameters. *IEEE Trans. on Pattern Analysis and Machine Intelligence* **27**(4), 562–574.
- Cervera, E., A.P. del Pobil, F. Berry et P. Martinet (2003). Improving image-based visual servoing with three-dimensional features. *Int. Journal of Robotics Research* **22**(10–11), 821–840.
- Chaumette, F. (1990). La relation vision-commande : théorie et application à des tâches robotiques. PhD thesis. Université de Rennes I, IRISA.
- Chaumette, F. (1998). Potential problems of stability and convergence in image-based and position-based visual servoing. Dans : *The Confluence of Vision and Control* (D.J. Kriegman, G. Hager and A.S. Morse, (Eds.)). pp. 66–78. LNCIS Series, No 237, Springer-Verlag.
- Chaumette, F. (2002). Asservissement visuel. Dans : *La commande des robots manipulateurs* (W. Khalil, (Ed.)). Chap. 3, pp. 105–150. Traité IC2. Hermès.
- Chaumette, F. (2004). Image moments : a general and useful set of features for visual servoing. *IEEE Trans. on Robotics* **20**(4), 713–723.
- Chaumette, F. et E. Marchand (2001). A redundancy-based iterative scheme for avoiding joint limits : Application to visual servoing. *IEEE Trans. on Robotics and Automation* **17**(5), 719–730.
- Chaumette, F. et S. Hutchinson (2006). Visual servo control, Part I : Basic approaches. *IEEE Robotics and Automation Magazine* **13**(4), 82–90.
- Chaumette, F. et S. Hutchinson (2007). Visual servo control, Part II : Advanced approaches. *IEEE Robotics and Automation Magazine* **14**(1), 109–118.
- Chaumette, F., P. Rives et B. Espiau (1993). Classification and realization of the different vision-based tasks. Dans : *Visual Servoing* (K. Hashimoto, (Ed.)). Vol. 7. pp. 199–228. World Scientific Series in Robotics and Automated Systems. Singapour.
- Chaumette, F., S. Boukir, P. Bouthemy et D. Juvin (1996). Structure from controlled motion. *IEEE Trans. on Pattern Analysis and Machine Intelligence* **18**(5), 492–504.
- Collewet, C. (1999). Contributions à l'élargissement du champ applicatif des asservissements visuels 2D. Thèse de doctorat. Université de Rennes I. Rennes.
- Collewet, C., A. Alhaj et F. Chaumette (2004). Model-free visual servoing on complex images based on 3d reconstruction. Dans : *IEEE Int. Conf. on Robotics and Automation, ICRA'04*. Vol. 1. Nouvelle Orléans, LA. pp. 751–756.

- Collewet, C., E. Marchand et F. Chaumette (2008a). Asservissement visuel basé sur des informations photométriques. Dans : *16e congrès francophone AFRIF-AFIA Reconnaissance des Formes et Intelligence Artificielle, RFIA'08*. Amiens, France.
- Collewet, C., E. Marchand et F. Chaumette (2008b). Visual servoing set free from image processing. Dans : *IEEE Int. Conf. on Robotics and Automation, ICRA'08*. Pasadena, Californie.
- Collewet, C. et E. Marchand (2008a). Modeling complex luminance variations for target tracking. Dans : *IEEE Int. Conf. on Computer Vision and Pattern Recognition, CV-PR'08*. Anchorage, Alaska.
- Collewet, C. et E. Marchand (2008b). Photometric visual servoing. Rapport de recherche n° 6631. INRIA.
- Collewet, C. et F. Chaumette (2000). A contour approach for image-based control of objects with complex shape. Dans : *IEEE/RSJ Int. Conf. on Intelligent Robots and Systems*. Takamatsu, Japon. pp. 751–756.
- Collewet, C. et F. Chaumette (2002). Positioning a camera with respect to planar objects of unknown shape by coupling 2d visual servoing and 3d estimations. *IEEE Trans. on Robotics and Automation* **18**(3), 322–333.
- Collewet, C. et F. Chaumette (2008). Visual servoing based on structure from controlled motion or on robust statistics. *IEEE Trans. on Robotics* **24**(2), 318–330.
- Collis, S. S., R. D. Joslin, A. Seifert et V. Theofilis (2004). Flow control : new challenges for a new renaissance. *Progress in Aerospace Sciences* **4**, 237–289.
- Colombo, C., B. Allotta et P. Dario (1995). Affine visual servoing : a framework for relative positioning with a robot. Dans : *IEEE Int. Conf. on Robotics and Automation*. Nagoya, Japon. pp. 464–471.
- Colombo, C. et B. Allotta (1999). Image-based robot task planning and control using a compact visual representation. *IEEE Trans. on Systems, Man, and Cybernetics, Part A : Systems and Humans* **29**(1), 92–100.
- Comport, A.I., E. Marchand et F. Chaumette (2006a). Statistically robust 2d visual servoing. *IEEE Trans. on Robotics* **22**(2), 415–421.
- Comport, A.I., E. Marchand, M. Pressigout et F. Chaumette (2006b). Real-time markerless tracking for augmented reality : the virtual visual servoing framework. *IEEE Trans. on Visualization and Computer Graphics* **12**(4), 615–628.
- Corke, P.I. et S. Hutchinson (2001). A new partitioned approach to image-based visual servo control. *IEEE Trans on Robotics and Automation* **17**(4), 507–515.
- Corpetti, T., D. Heitz, G. Arroyo, E. Mémin et A. Santa-Cruz (2006). Fluid experimental flow estimation based on an optical-flow scheme. *Experiments in fluids* **40**(1), 80–97.
- Corpetti, T., E. Mémin et P. Pérez (2002). Dense estimation of fluid flows. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **24**(3), 365–380.
- Coste-Manière, E., P. Couvignou et P. Kohsla (1995). Visual servoing in the task-function framework : a contour following task. *Journal of Intelligent and Robotics Systems* **12**, 1–21.

- Crowley, J.L, P. Stelmazyk et P. Puget (1992). Measurement and integration of 3D structures by tracking edge lines. *Int. Journal of Computer Vision* **8**(1), 29–52.
- Crétual, A. (1998). Asservissement visuel à partir d'informations de mouvement dans l'image. Thèse de doctorat. Université de Rennes I. Rennes.
- Crétual, A. et F. Chaumette (1998). Image-based visual servoing by integration of dynamic measurements. Dans : *IEEE Int. Conf. on Robotics and Automation*. Vol. 3. Louvain, Belgique. pp. 1994–2001.
- Crétual, A. et F. Chaumette (2001). Visual servoing based on image motion. *Int. Journal of Robotics Research* **20**(11), 857–877.
- Dauchez, P., (Ed.) (2000). *Applications non manufacturières de la robotique*. Hermès. Paris. Traité IC2.
- Deguchi, K. (2000). A direct interpretation of dynamic images with camera and object motions for vision guided robot control. *Int. Journal of Computer Vision* **37**(1), 7–20.
- Deguchi, K. et T. Noguchi (1996). Visual servoing using eigenspace method and dynamic calculation of interaction matrices. Dans : *IEEE Int Conf on Pattern Recognition*. pp. 302–306.
- Dementhon, D. et L. Davis (1995). Model-based object pose in 25 lines of codes. *Int. J. of Computer Vision* **15**, 123–141.
- Dhome, M., M. Richetin, J.-T. Lapresté et G. Rives (1989). Determination of the attitude of 3D objects from a single perspective view. *IEEE Trans. on Pattern Analysis and Machine Intelligence* **11**(12), 1265–1278.
- Dornaika, F. (1995). Contributions à l'intégration vision/robotique : calibrage, localisation et asservissement. Thèse de doctorat. Institut National Polytechnique de Grenoble. Grenoble.
- Drew, M.S., J. Wei et Z-N. Li (1998). On illumination invariance in color object recognition. *Pattern Recognition* **31**(8), 1077–1087.
- Drummond, T. et R. Cipolla (1999). Visual tracking and control using lie algebras. Dans : *IEEE Int. Conf. on Computer Vision and Pattern Recognition*. Vol. II. Fort Collins, Colorado, USA. pp. 652–657.
- Drummond, T. et R. Cipolla (2002). Real-time visual tracking of complex structures. *IEEE Trans. on Pattern Analysis and Machine Intelligence* **24**(7), 932–946.
- Dubrawski, A. et I. Siemiatkowska (1998). A method for tracking the pose of a mobile robot equipped with a scanning laser range finder. Dans : *IEEE Int. Conf. on Robotics and Automation*. Vol. 3. Louvain, Belgique. pp. 2518–2523.
- Dune, C., E. Marchand, C. Collewet et C. Leroux (2008). Active rough shape estimation of unknown objects. Dans : *IEEE/RSJ Int. Conf. on Intelligent Robots and Systems, IROS'08*. Nice, France. pp. 3622–3627.
- Espiau, B. (1993). Effect of camera calibration errors on visual servoing in robotics. Dans : *Int. Symposium on experimental Robotics, ISER'93*. Kyoto.
- Espiau, B. et P. Rives (1987). Closed-loop recursive estimation of 3d features for a mobile vision system. Dans : *IEEE Int. Conf. on Robotics and Automation*. Vol. 4. Raleigh, NC. pp. 1436–1443.

- Espiau, B., F. Chaumette et P. Rives (1992). A new approach to visual servoing in robotics. *IEEE Trans. on Robotics and Automation* **8**(3), 313–326.
- Faugeras, O. (1993). *Three-dimensional computer vision : a geometric viewpoint*. MIT Press. Cambridge, Massachusetts.
- Finlayson, G.D., S. D. Hordley, J. Marchant et C. Onyango (2000). Color invariance at a pixel. Dans : *British Machine Vision Conference*. Bristol, UK.
- Flandin, G. (2001). Modélisation probabiliste et exploration visuelle autonome pour la reconstruction de scènes inconnues. Thèse de doctorat. Université de Rennes 1. Rennes.
- Gangloff, J., M. de Mathelin et G. Abba (1998). 6 dof high speed dynamic visual servoing using gpc controllers. Dans : *IEEE Int. Conf. on Robotics and Automation, ICRA'98*. pp. 2008–2013.
- Gennery, D.B. (1992). Visual tracking of known three-dimensional objects. *Int. Journal of Computer Vision* **7**(3), 243–270.
- Geusebroek, J.-M., R. Van den Boomgaard, A.W.M. Smeulders et A.W.M. Geerts (2001). Color invariance. *IEEE Trans. on Pattern Analysis and Machine Intelligence* **23**(12), 1338–1350.
- Gevers, T. (2001). Robust histogram construction from color invariants. Dans : *IEEE Int. Conf. on Computer Vision*. Vol. 1. Vancouver. pp. 615–620.
- Gevers, T., A. W. M. Smeulders et H. M. G. Stokman (1998). Photometric invariant region detection. Dans : *British Machine Vision Conference*. Southampton UK.
- Gevers, T. et A.W.M. Smeulders (1997). Object recognition based on photometric color invariants. Dans : *Scandinavian Conference on Image Analysis*. Lappeenranta, Finlande.
- Giachetti, A. (2000). Matching techniques to compute image motion. *Image and Vision Computing* **18**, 247–260.
- Ginhoux, R., J.A. Gangloff, M.F. De Mathelin, L. Soler, M.M.A. Sanchez et J. Marescaux (2004). Beating heart tracking in robotic surgery using 500 hz visual servoing, model predictive control and an adaptive observer. Dans : *IEEE Int. Conf. on Robotics and Automation, ICRA'04*. Vol. 1. Nouvelle Orléans, USA. pp. 274–279.
- Gouiffès, M. (2005). Apports de la couleur et des modèles de réflexion pour l'extraction et le suivi de primitives. Thèse de doctorat. Université de Poitiers. Poitiers.
- Gouiffès, M., C. Collewet, C. Fernandez-Maloigne, A. Trémeau et P. Loisel (2004a). Invariants couleur dédiés à la transparence de l'encre. Application à la segmentation de marquages. *Traitement du signal* **21**(5), 439–452. Numéro spécial sur l'image numérique couleur.
- Gouiffès, M., C. Collewet, C. Fernandez-Maloigne et A. Trémeau (2004b). Color segmentation of inked characters : application to meat traceability control. Dans : *11th IEEE Int. Conf. on Image Processing, ICIP'2004*. Singapore.
- Gouiffès, M., C. Collewet, C. Fernandez-Maloigne et A. Trémeau (2006a). Feature point tracking : robustness to specular highlights and lighting variations. Dans : *European Conf. on Computer Vision, ECCV'2006*. Vol. 3954 of LNCS. Graz, Austria. pp. 92–93.

- Gouiffès, M., C. Collewet, C. Fernandez-Maloigne et A. Trémeau (2006b). A photometric model for specular highlights and lighting changes. Application to feature points tracking. Dans : *13th IEEE Int. Conf. on Image Processing, ICIP'2006*. Atlanta, USA.
- Gouiffès, M., P. Marty-Mahé, C. Fernandez-Maloigne et A. Trémeau (2004c). Reflectance measuring of an anisotropic surface : the ham. Dans : *2nd European Conf. on Colour in Graphics, Imaging, and Vision, CGIV'2004*. Aix-la-Chapelle, Allemagne.
- Griffin, P.M., L.S. Narasimhan et S.R. Yee (1992). Generation of uniquely encoded light patterns for range data acquisition. *Pattern Recognition* **25**(6), 609–616.
- Gupta, N.C. et L.N. Kanal (1995). 3-d motion estimation from motion field. *Artificial Intelligence* **78**(1-2), 45–86.
- Hager, G. et P. Belhumeur (1998). Efficient region tracking with parametric models of geometry and illumination. *IEEE Trans. on Pattern Analysis and Machine Intelligence* **20**(10), 1025–1039.
- Hager, G., G. Grunwald et K. Toyama (1995). Feature-based visual servoing and its application to telerobotics. In V. Graefe, editor, *Intelligent Robotic Systems*, Elsevier, Amsterdam.
- Hall, E.L., J.B. Tio, C.A. Mac Pherson et F.A. Sadjadi (1982). Measuring curved surfaces for robot vision. *Computer* **15**, 42–54.
- Hallinan, P.W. (1994). A low-dimensional representation of human faces for arbitrary lighting conditions. Dans : *IEEE Int. Conf. on Computer Vision and Pattern Recognition*. Seattle, Washington. pp. 995–999.
- Hampel, F., E.M. Ronchetti, P.J. Rousseeuw et W.A. Stahel (1986). *Robust Statistics : The Approach Based on Influence Functions*. John Wiley and Sons. New York.
- Haralick, R., H. Joo, C. Lee, X. Zhuang, V. Vaidya et M. Kim (1989). Pose estimation from corresponding point data. *IEEE Trans on Systems, Man and Cybernetics* **19**(6), 1426–1445.
- Hartley, R. (1997). In defense of the eight-points algorithm.. Dans : *IEEE Trans. on Pattern Analysis and Machine Intelligence*. Vol. 19. pp. 580–593.
- Hartley, R.I. (1994). Self-calibration from multiple views with a rotating camera. Dans : *European Conf. on Computer Vision*. pp. 471–478.
- Harville, M., B. Culbertson, I. Sobel, D. Gelb, A. Fitzhugh et D. Tanguay (2006). Practical methods for geometric and photometric correction of tiled projector displays on curved surfaces. Dans : *IEEE International Workshop on Projector-Camera Systems, PROCAM'2006*. New York, USA.
- Hashimoto, K. et H. Kimura (1993). Lq optimal and non-linear approaches to visual servoing. Dans : *Visual Servoing* (K. Hashimoto, (Ed.)). Vol. 7. pp. 165–198. World Scientific Series in Robotics and Automated Systems. Singapour.
- Hashimoto, K., T. Ebine et H. Kimura (1996). Visual servoing with hand-eye manipulator — optimal control approach. *IEEE Trans. on Robotics and Automation* **12**(5), 766–774.
- Hattori, K. et Y. Sato (1995). Handy rangefinder for active robot vision. Dans : *IEEE Int. Conf. on Robotics and Automation*. Vol. 2. Nagoya, Japon. pp. 1423–1428.

- Haverinen, J. et J. ; Roning (1998). An obstacle detection system using a light stripe identification based method. Dans : *IEEE Int. Joint Symp. on Intelligence and Systems*. Vol. 1. Rockville, MD, USA. pp. 232–236.
- Healey, G.E. (1989). Using color for geometry insensitive segmentation. *Journal of the Optical Society of America* **6**, 920–937.
- Heeger, D.J. et A.D. Jepson (1992). Subspace methods for recovering rigid motion I : Algorithms and implementation. *Int. Journal of Computer Vision* **7**(2), 95–117.
- Heigl, B., D. Paulus et H. Niemann (1999). Tracking points in sequences of color images. Dans : *German-Russian workshop Pattern Recognition and Image Understanding*. Herrsching, Germany. pp. 70–77.
- Høepffner, J., M Chevalier, T. Bewley et D. S. Henningson (2005). State estimation in wall-bounded flow systems. part 1. perturbed laminar flows. *Journal of Fluid Mechanics* **534**, 263–294.
- Horaud, R., B. Conio, O. Leboulleux et B. Lacolle (1989). An analytic solution for the perspective 4-points problem. *Computer Vision, Graphics and Image Processing* **47**(1), 33–44.
- Horaud, R., F. Dornaika et B. Espiau (1998). Visually guided object grasping. *IEEE Trans. on Robotics and Automation* **14**(4), 525–532.
- Horn, B.K.P. (1986). *Robot vision*. The MIT Press. Cambridge.
- Horn, B.K.P. et B.G. Schunck (1981). Determining optical flow. *Artificial Intelligence* **17**(1-3), 185–203.
- Horn, B.K.P. et E.J. Weldon (1988). Direct methods for recovering motion. *Int. Journal of Computer Vision* **2**(1), 51–76.
- Hosoda, K. et M. Asada (1994). Versatile visual servoing without knowledge of true jacobian. Dans : *IEEE/RSJ Int. Conf on Intelligent Robots and Systems, IROS'94*. Munich, Germany. pp. 186–193.
- Huber, P.-J. (1981). *Robust Statistics*. Wiler, New York.
- Hutchinson, S., G. Hager et P. Corke (1996). A tutorial on visual servo control. *IEEE Trans. on Robotics and Automation* **12**(5), 651–670.
- Ikeuchi, K. (1981). Determining surface orientations of specular surfaces by using the photometric stereo method. *IEEE Trans. on Pattern Analysis and Machine Intelligence* **13**(6), 661–669.
- Inokuchi, S., K. Sato et F. Matsuda (1984). Range imaging system for 3-D object recognition. Dans : *IAPR Int. Conf. on Pattern Recognition*. pp. 806–808.
- Irani, M, b. Rousso et S. Peleg (1997). Recovery of ego-motion using region alignment. *IEEE Trans. on Pattern Analysis and Machine Intelligence* **19**(3), 268–272.
- Irani, M. et P. Anandan (1998). A unified approach to moving object detection in 2d and 3d scenes. *IEEE Trans. on Pattern Analysis and Machine Intelligence* **20**(6), 577–589.
- Jägersand, M. (1996). Visual servoing using trust region methods and estimation of the full coupled visual-motor jacobian. Dans : *IASTED Applications of Robotics and Control*. Orlando, USA. pp. 105–108.



- Jägersand, M., O. Fuentes et R. Nelson (1997). Experimental evaluation of uncalibrated visual servoing. Dans : *IEEE Int. Conf. on Robotics and Automation, ICRA'97*. Vol. 3. Albuquerque, NM. pp. 2874–2880.
- Jin, H., P. Favaro et S. Soatto (2001). Real-Time feature tracking and outlier rejection with changes in illumination. Dans : *IEEE Int. Conf. on Computer Vision*. pp. 684–689.
- Jurie, F. et M. Dhome (2002). Hyperplane approximation for template matching. *IEEE Trans. on Pattern Analysis and Machine Intelligence* **24**(7), 996–1000.
- Kahane, B. et Y. Rosenfeld (2004). Real-time 'sense-and-act' operation for construction robots. *Automation in Construction* **13**(6), 751–764.
- Kallem, V., M. Dewan, J.P. Swensen, G.D. Hager et N.J. Cowan (2007). Kernel-based visual servoing. Dans : *IEEE/RSJ Int. Conf. on Intelligent Robots and System, IROS'07*. San Diego, USA.
- Kass, M., A. Witkin et D. Terzopolous (1987). Snakes : Active contour models. Dans : *IEEE Int. Conf. on Computer Vision*. London, UK. pp. 259–268.
- Kelly, R., R. Carelli, O. Nasisi, B. Kuchen et F. Reyes (2000). Stable visual servoing of camera-in-hand robotic systems. *IEEE/ASME Transactions on Mechatronics* **5**(1), 39–48.
- Kemmotsu, K. et T. Kanade (1995). Uncertainty in object pose determination with three light-stripe range measurements. *IEEE Trans. on Robotics and Automation* **11**(5), 741–747.
- Kermad, C., C. Collewet et K. Chehdi (2002). Apport de la sélection de primitives point à un suivi robuste. Dans : *13<sup>ème</sup> Congrès Francophone AFRIF-AFIA de Reconnaissance des Formes et Intelligence Artificielle, RFIA'2002*. Vol. 2. Angers, France. pp. 539–548.
- Kermad, C. et C. Collewet (2001). Improving feature tracking by robust points of interest selection. Dans : *6th Int. Fall Workshop on Vision, Modeling, and Visualization, VMV'2001*. Stuttgart, Germany.
- Khadraoui, D., G. Motyl, P. Martinet, J. Gallice et F. Chaumette (1996). Visual servoing in robotics scheme using a camera/laser-stripe sensor. *IEEE Trans. on Robotics and Automation* **12**(5), 743–750.
- Kim, J.Y., C.Y. Kim, K.S. Seo et I.S. Kweon (2001). Color indexing using chromatic invariant. *Pattern Recognition* **34**(6), 1189–1197.
- Kim, K. H. et H. S. Cho (2001). Range and contour fused environment recognition for mobile robot. Dans : *IEEE Int. Conf. on Multisensor Fusion and Integration for Intelligent Systems*. Vol. 1. Baden-Baden, Allemagne. pp. 183–188.
- Kim, P., S. Rhee et C. H. Lee (1999). Automatic teaching of welding robot for free-formed seam using laser vision sensor. *Optics and Lasers in Engineering* **31**(3), 173–182.
- Kim, Y-H., A.M. Martinez et A.C. Kak (2005). Robust motion estimation under varying illumination. *Image and Vision Computing* **23**(4), 365–375.
- Koenderink, J.J. et A.J. Van Doorn (1991). Affine structure from motion. *Journal of the Optical Society of America* **8**(2), 377–385.

- Kondo, H. et U. Tamaki (2004). Navigation of an auv for investigation of underwater structures. *Control engineering practice* **12**(12), 1551–1559.
- Krupa, A. et F. Chaumette (2006). Guidance of an ultrasound probe by visual servoing. *Advanced Robotics, Special issue on « Selected papers from IROS'05 »* **20**(11), 1203–1218.
- Krupa, A., J. Gangloff, C. Doignon, M. de Mathelin, G. Morel, J. Leroy, L. Soler et J. Marescaux (2003). Autonomous 3D positioning of surgical instruments in robotized laparoscopic surgery using visual servoing. *IEEE Trans. on Robotics and Automation* **19**(5), 842–853.
- Kwok, K. S., C. S. Loucks et B.J. Driessen (1998). Rapid 3-d digitizing and tool path generation for complex shapes. Dans : *IEEE Int. Conf. on Robotics and Automation*. Vol. 4. Louvain, Belgique. pp. 2789–2794.
- La Cascia, M., S. Sclaroff et V. Athitsos (2000). Fast, reliable head tracking under varying illumination : An approach based on registration of texture-mapped 3D models. *IEEE Trans. on Pattern Analysis and Machine Intelligence* **22**(4), 322–336.
- Lambert, J.H. (1760). *Photometria sive de mensura de gratibus luminis, colorum et ombrae*. Eberhard Klett. Augsburg.
- Lapresté, J.-T., F. Jurie, M. Dhome et F. Chaumette (2004). An efficient method to compute the inverse jacobian matrix in visual servoing. Dans : *IEEE Int. Conf. on Robotics and Automation, ICRA'04*. La Nouvelle-Orléans, Louisiane, USA.
- Lee, S., J. Choi, D. Kim, b. Jung, J. Na et H. Kim (2004). An active 3d robot camera for home environment. *Proc. of the IEEE* **1**, 477–480.
- Levine, M.D. et J. Bhattacharyya (2005). Detecting and removing specularities in facial images. *Computer Vision and Image Understanding* **100**(3), 330–356.
- Lhuillier, M. et L. Quan (2002). Quasi-dense reconstruction from image sequence. Dans : *European Conf. on Computer Vision*. pp. 125–139.
- Longuet-Higgins, H.C (1984). The visual ambiguity of a moving plane. Dans : *IEEE Proc. of the Royal Society of London*. Vol. B223. pp. 165–175.
- Longuet-Higgins, H. C. et K. Prazdny (1980). The interpretation of a moving retinal image. Dans : *Proceedings of the Royal Society of London*. Vol. B208. pp. 385–397.
- Lowe, D.G. (1987). Three-dimensional object recognition from single two-dimensional images. *Artificial Intelligence* **31**(3), 355–394.
- Lucas, B.D. et T. Kanade (1981). An iterative image registration technique with an application to stereo vision. Dans : *Int. Joint Conf. on Artificial Intelligence, IJCAI'81*. pp. 674–679.
- Luong, Q-T. et O. Faugeras (1996). The fundamental matrix : Theory, algorithms, and stability analysis. *Int. Journal of Computer Vision* **17**(1), 43–75.
- MacWilliams, F. J. et N. J. A. Sloane (1976). Pseudorandom sequences and arrays. *Proc. of the IEEE* **64**(12), 1715–1729.
- Mahony, R., P. Corke et F. Chaumette (2002). Choice of image features for depth-axis control in image-based visual servo control. Dans : *IEEE/RSJ Int. Conf. on Intelligent Robots and Systems*. Lausanne, Suisse.

- Malis, E. (2004). Improving vision-based control using efficient second-order minimization techniques. Dans : *IEEE Int. Conf. on Robotics and Automation, ICRA'04*. Vol. 2. pp. 1843–1848.
- Malis, E. et F. Chaumette (2002). Theoretical improvements in the stability analysis of a new class of model-free visual servoing methods. *IEEE Trans. on Robotics and Automation* **18**(2), 176–186.
- Malis, E. et P. Rives (2003). Robustness of image-based visual servoing with respect to depth distribution errors. Dans : *IEEE Int. Conf. on Robotics and Automation*. Vol. 2. Taipei, Taiwan. pp. 1056–1061.
- Malis, E., F. Chaumette et S. Boudet (1999). 2 1/2 D visual servoing. *IEEE Trans. on Robotics and Automation* **15**(2), 238–250.
- Malis, E., F. Chaumette et S. Boudet (2000). 2 1/2 D visual servoing with respect to unknown objects through a new estimation scheme of camera displacement. *Int. Journal of Computer Vision* **37**(1), 79–97.
- Manh, A.-G., G. Rabatel, L. Assemat et M.-J. Aldon (2001). Automation and emerging technologies weed leaf image segmentation by deformable templates. *Journal of Agricultural Engineering Research* **80**(2), 139–146.
- Marchand, E. (2004). Contributions à la commande d'une caméra réelle ou virtuelle dans des mondes réels ou virtuels. Habilitation à diriger les recherches. Université de Rennes. Rennes.
- Marchand, E. (2007). Control camera and light source positions using image gradient information. Dans : *IEEE Int. Conf. on Robotics and Automation*. Rome, Italie.
- Marchand, E. et F. Chaumette (1996). Controlled camera motions for scene reconstruction and exploration. Dans : *IEEE Int. Conf. on Computer Vision and Pattern Recognition, CVPR'96*. San Francisco, Californie. pp. 169–176.
- Marchand, E. et F. Chaumette (2002). Virtual visual servoing : a framework for real-time augmented reality. Dans : *EUROGRAPHICS'02 Conf. Proceeding* (G. Drettakis and H.-P. Seidel, (Eds.)). Vol. 21(3) of *Computer Graphics Forum*. Saarebrücken, Allemagne. pp. 289–298.
- Marchand, E. et F. Chaumette (2005). Feature tracking for visual servoing purposes. *Robotics and Autonomous Systems* **52**(1), 53–70. special issue on “Advances in Robot Vision”, D. Kragic, H. Christensen (Eds.).
- Marchand, E. et G.-D. Hager (1998). Dynamic sensor planning in visual servoing. Dans : *IEEE Int. Conf. on Robotics and Automation*. Vol. 3. Leuven, Belgium. pp. 1988–1993.
- Marchand, E., P. Bouthemy et F. Chaumette (2001). A 2D-3D model-based approach to real-time visual tracking. *Image and Vision Computing* **19**(13), 941–955.
- Martinet, P., N. Daucher, J. Gallice et M. Dhome (1997). Robot control using monocular pose estimation. Dans : *Workshop on new trends in image-based robot servoing, IROS'97*. Grenoble. pp. 1–12.
- Mebarki, R., A. Krupa et C. Collewet (2008). Automatic guidance of an ultrasound probe by visual servoing based on b-mode image moments. Dans : *Int. Conf. on Medical Image Computing and Computer Assisted Intervention (MICCAI)*. New York, USA.

- Mezouar, Y., A. Remazeilles, P. Gros et F. Chaumette (2002). Image interpolation for image-based control under large displacement. Dans : *IEEE Int. Conf. on Robotics and Automation, ICRA'02*. Vol. 3. Washington DC, USA. pp. 3787–3794.
- Michel, H. et P. Rives (1993). Singularities in the determination of the situation of a robot effector from the perspective view of 3 points. Rapport de recherche N° 1850. INRIA.
- Mileva, Y., A. Bruhn et J. Weickert (2007). Illumination-robust variational optical flow with photometric invariants. *Pattern Recognition* **4713**, 152–162.
- Miura, K., J. Gangloff et M. De Mathelin (2002). Robust and uncalibrated visual servoing without Jacobian using a simplex method. Dans : *IEEE/RSJ Int. Conf. on Intelligent Robots and System, IROS'02*. Lausanne, Suisse.
- Monks, T. P., J. N. Carter et C. H. Shadle (1992). Colour-encoded structured light for digitisation of real-time 3D data. Dans : *IEEE Int. Conf. on Image Processing*. pp. 327–30.
- Morano, R. A., C. Ozturk, R. Conn, S. Dubin, S. Zietz et J. Nissanov (1998). Structured light using pseudorandom codes. *IEEE Trans. on Pattern Analysis and Machine Intelligence* **20**(3), 322–327.
- Morel, G., T. Leibezeit, J. Szewczyk, S. Boudet et J. Pot (2000). Explicit incorporation of 2d constraints in vision-based control of robot manipulators. Dans : *Proc. Int. Symp. on Experimental Robotics*. Vol. 250 of *LNCIS*. Sydney. pp. 99–108.
- Motyl, G. (1992). Couplage d'une caméra et d'un faisceau laser en commande référencée vision. PhD thesis. Université Blaise Pascal. Clermont-Ferrand. LASMEA.
- Nakabo, Y., M. Ishikawa, H. Toyoda et S. Mizuno (2000). 1ms column parallel vision system and its application of high speed target tracking. Dans : *IEEE Int. Conf. on Robotics and Automation, ICRA'00*. San Francisco, CA. pp. 650–655.
- Nayar, S.K., K. Ikeuchi et T. Kanade (1991). Surface reflection : Physical and geometrical perspectives. *IEEE Trans. on Pattern Analysis and Machine Intelligence* **13**(7), 611–634.
- Nayar, S.K., S.A. Nene et H. Murase (1996). Subspace methods for robot vision. *IEEE Trans. on Robotics and Automation* **12**(5), 750–758.
- Negahdaripour, S. (1992). Motion recovery from image sequence using only first order optical flow information. *Int. Journal of Computer Vision* **9**(3), 163–184.
- Negahdaripour, S. (1998). Revised definition of optical flow : Integration of radiometric and geometric cues for dynamic scene analysis. *IEEE Trans. on Pattern Analysis and Machine Intelligence* **20**(9), 961–979.
- Negahdaripour, S. et B.K.P. Horn (1987). Direct passive navigation. *IEEE Trans. on Pattern Analysis and Machine Intelligence* **9**(1), 168–176.
- Nickels, K. et S. Hutchinson (2002). Estimating uncertainty in ssd-based feature tracking. *Image and Vision Computing* **20**(1), 47–58.
- Niel, A., P. Sommer, S. H. Kolpl et Y. Lypetsky (2004). High precision measurement of free formed parts using industrial robots. Dans : *IEEE Int. Workshop on Robot Sensing*. Graz, Autriche. pp. 79–84.

- Odobez, J.-M. et P. Bouthemy (1995). Robust multiresolution estimation of parametric motion models. *Journal of Visual Communication and Image Representation* **6**(4), 348–365.
- Ohba, K., Y. Sato et K. Ikeuchi (2000). Appearance-based visual learning and object recognition with illumination invariance. *Machine Vision and Applications* **12**, 189–196.
- Pagès, J. (2005). Assisted visual servoing by means of structured light. Thèse de doctorat. Universités de Rennes I et de Girona. Rennes.
- Pagès, J., C. Collewet, F. Chaumette et J. Salvi (2004). Plane-to-plane positioning from image-based visual servoing and structured light. Dans : *IEEE/RSJ Int. Conf. on Intelligent Robots and Systems, IROS'2004*. Sendai, Japon.
- Pagès, J., C. Collewet, F. Chaumette et J. Salvi (2005a). Visual servoing by means of structured light for plane-to-plane positioning. Rapport de Recherche 5579. INRIA.
- Pagès, J., C. Collewet, F. Chaumette et J. Salvi (2006a). An approach to visual servoing based on coded light. Dans : *IEEE Int. Conf. on Robotics and Automation, ICRA'2006*. Orlando, USA.
- Pagès, J., C. Collewet, F. Chaumette et J. Salvi (2006b). A camera-projector system for robot positioning by visual servoing. Dans : *IEEE International Workshop on Projector-Camera Systems, PROCAM'2006*. New York, USA.
- Pagès, J., C. Collewet, F. Chaumette et J. Salvi (2006c). Optimizing plane-to-plane positioning tasks by image-based visual servoing and structured light. *IEEE Trans. on Robotics* **22**(5), 1000–1010.
- Pagès, J., J. Salvi, C. Collewet et J. Forest (2005b). Optimised De Bruijn patterns for one-shot shape acquisition. *Image and Vision Computing* **23**(8), 707–720.
- Papanikolopoulos, N., P.K. Khosla et T. Kanade (1991). Vision and control techniques for robotic visual tracking. Dans : *IEEE Int. Conf. on Robotics and Automation, ICRA'91*. Vol. 1. Sacramento, Californie. pp. 857–864.
- Papanikolopoulos, N.P. (1995). Selection of features and evaluation of visual measurements during robotic visual servoing tasks. *Journal of Intelligent and Robotic Systems* **13**, 279–304.
- Phong, B.T. (1975). Illumination for computer generated pictures. *Communication of the ACM* **18**(6), 311–317.
- Piepmeyer, J.A., G. V. McMurray et Lipkin H. (2004). Uncalibrated dynamic visual servoing. *IEEE Trans. on Robotics and Automation* **20**(1), 143–147.
- Pizarro, D., J. Peyras et A. Bartoli (2008). Light-invariant fitting of active appearance models. Dans : *IEEE Int. Conf. on Computer Vision and Pattern Recognition, CVPR'08*. Anchorage, Alaska. pp. 1–6.
- Prazdny, K. (1981). Determining the instantaneous direction of motion from optical flow generated by a curvilinear moving observer. *Computer Graphics and Image processing* **17**, 238–248.
- Precioso, F., M. Barlaud, T. Blu et M. Unser (2005). Robust real-time segmentation of images and videos using a smoothing-spline snake-based algorithm. *IEEE Trans. on Image Processing*.

- Pressigout, M. et E. Marchand (2007). Real-time hybrid tracking using edge and texture information. *Int. Journal of Robotics Research* **26**(7), 689–713.
- Pérez, P., C. Hue, J. Vermaak et M. Gangnet (2002). Color-based probabilistic tracking. Dans : *European Conf. on Computer Vision*. Copenhague, Danemark. pp. 661–675.
- Questa, P., E. Grossmann et G. Sandini (1995). Camera self orientation and docking maneuver using normal flow. Dans : *SPIE AeroSense'95*. Orlando, Floride, USA.
- Reichmann, J. (1973). Determination of absorption and scattering coefficients for non homogeneous media. *Applied Optics* **12**, 1811–1815.
- Rives, P. et B. Espiau (1987). Estimation récursive de primitives 3D au moyen d'une caméra mobile. Rapport de Recherche 652. INRIA-IRISA.
- Rosen, C. (1976). Machine intelligence research applied to industrial automation. Rapport de recherche. SRI International. Sixth report.
- Ruf, A. et R. Horaud (1999). Rigid and articulated motion seen with an uncalibrated stereo rig. Dans : *IEEE Int. Conf. on Computer Vision*. Corfu, Greece. pp. 789–796.
- Salvi, J., J. Batlle et E. Mouaddib (1998). A robust-coded pattern projection for dynamic 3d scene measurement. *Pattern Recognition Letters* **19**(11), 1055–1065.
- Salvi, J., J. Pagès et J. Batlle (2004). Pattern codification strategies in structured light systems. *Pattern Recognition* **37**(4), 827–849.
- Samson, C., M. Le Borgne et B. Espiau (1991). *Robot Control : the Task Function Approach*. Clarendon Press, Oxford, Royaume Uni.
- Santos, A., B. Espiau, P. Rives, D. Simon et V. Rigaud (1995). Sensor-based control of holonomic autonomous underwater vehicles. Rapport de Recherche 2609. INRIA.
- Santos-Victor, J. et G. Sandini (1997). Visual behaviors for docking. *Computer Vision and Image Understanding* **67**(3), 223–238.
- Schindler, K. et D. Suter (2006). Two-view multibody structure-and-motion with outliers through model selection. *IEEE Trans. on Pattern Analysis and Machine Intelligence* **28**(6), 983–995.
- Schramm, F., G. Morel, A. Micaelli et A. Lottin (2004). Extended 2d visual servoing. Dans : *IEEE Int. Conf. on Robotics and Automation, ICRA'04*. Vol. 1. Nouvelle Orléans, USA. pp. 267–273.
- Shafer, S.A. (1985). Using color to separate reflection components. *Color research and application* **10**(4), 210–218.
- Shi, J. et C. Tomasi (1994). Good features to track. Dans : *IEEE Int. Conf. on Computer Vision and Pattern Recognition, CVPR'94*. Seattle, Washington. pp. 593–600.
- Shirai, Y. et H. Inoue (1973). Guiding a robot by visual feedback in assembling tasks. *Pattern Recognition* **5**, 99–108.
- Silveira, G. et E. Malis (2007a). Direct visual servoing with respect to rigid objects. Dans : *IEEE/RSJ Int. Conf. on Intelligent Robots and System, IROS'07*. San Diego, USA.
- Silveira, G. et E. Malis (2007b). Real-time visual tracking under arbitrary illumination changes. Dans : *IEEE Int. Conf. on Computer Vision and Pattern Recognition*. Minneapolis, USA.

- Sinclair, D., A. Blake, S. Smith et C. Rothwell (1993). Planar region detection and motion recovery. *Image and Vision Computing* **11**, 229–234.
- Smith, C.E., S.A. Brandt et N.P. Papanikolopoulos (1997). Eye-in-hand robotic tasks in uncalibrated environments. *IEEE Trans. on Robotics and Automation* **13**(6), 903–914.
- Soatto, S. et P. Perona (1998). Reducing "structure from motion" : A general framework for dynamic vision part 1 : Modeling. *IEEE Trans. on Pattern Analysis and Machine Intelligence* **20**(9), 933–942.
- Srinivasan, S. (2000). Extracting structure from optical flow using the fast error search technique. *Int. Journal of Computer Vision* **37**(3), 203–230.
- Stahs, T. et F. Wahl (1992). Fast and versatile range data acquisition in a robot work cell. Dans : *IEEE/RSJ Int. Conf. on Intelligent Robots and Systems*. Raleigh, NC. pp. 1169–1174.
- Stein, G.P. et A. Shashua (2000). Model-based brightness constraints : On direct estimation of structure and motion. *IEEE Trans. on Pattern Analysis and Machine Intelligence* **22**(9), 992–1015.
- Stern, H. et B. Efron (2005). Adaptive color space switching for tracking under varying illumination. *Image and Vision Computing* **23**(3), 353–364.
- Subbarao, M. (1988). Interpretation of image flow : Rigid curved surfaces in motion. *Int. Journal of Computer Vision* **2**, 77–96.
- Suh, I.H. et T.W. Kim (1993). Visual servoing of robot manipulators by fuzzy membership function based neural networks. Dans : *Visual Servoing* (K. Hashimoto, (Ed.). Vol. 7. pp. 285–315. World Scientific Series in Robotics and Automated Systems. Singapour.
- Sundareswaran, V. (1991). Egomotion from global flow field data. Dans : *IEEE Workshop on Visual Motion*. pp. 140–145.
- Sundareswaran, V., P. Bouthemy et F. Chaumette (1996). Exploiting image motion for active vision in a visual servoing framework. *Int. Journal of Robotics Research* **15**(6), 629–645.
- Surmann, H., A. Nüchter et J. Hertzberg (2003). An autonomous mobile robot with a 3d laser range finder for 3d exploration and digitalization of indoor environments. *Robotics and Autonomous Systems* **45**(3-4), 181–198.
- Taalabinezhad, M.A. (1992). Direct recovery of motion and shape in the general case by fixation. *IEEE Trans. on Pattern Analysis and Machine Intelligence* **14**(8), 847–853.
- Tahri, O. et F. Chaumette (2005). Point-based and region-based image moments for visual servoing of planar objects. *IEEE Trans. on Robotics* **21**(6), 1116–1127.
- Tan, R.T. et K. Ikeuchi (2005). Separating reflection components of textured surfaces using a single image. *IEEE Trans. on Pattern Analysis and Machine Intelligence* **27**(2), 178–193.
- Tani, K., M. Abe, K. Tanie et T. Ohno (1977). High precision manipulators with visual sense. Dans : *7th Int. Symp. on industrial robotics*. Tokyo, Japon. pp. 561–568.
- Tarel, J.-P., S.-S. Ieng et P. Charbonnier (2002). Using robust estimation algorithms for tracking explicit curves. Dans : *European Conf. on Computer Vision*. pp. 492–507.

- Terzopoulos, D. et D. Metaxas (1991). Dynamic 3d models with local and global deformations : deformable superquadrics. *IEEE Trans. on Pattern Analysis and Machine Intelligence* **13**, 703–713.
- Tomasi, C. et T. Kanade (1991). Detection and tracking of point features. Rapport de Recherche CMU-CS-91-132. Carnegie Mellon University.
- Tominaga, S. (1996). Surface reflectance estimation by the dichromatic model. *Color Research and Application* **21**(2), 104–114.
- Tommasini, T., A. Fusiello, E. Trucco et V. Roberto (1999). Improving feature tracking with robust statistics. *Pattern Analysis & Applications* **2**, 312–320.
- Torrance, K.E. et E.M. Sparrow (1967). Theory for off-specular reflection from roughened surfaces. *Journal of the Optical Society of America* **57**, 1105–1114.
- Urban, J. P., G. Motyl et J. Gallice (1994). Realtime visual servoing using controlled illumination. *Int. Journal of Robotics Research* **13**(1), 93–10.
- Vacchetti, L., V. Lepetit et P. Fua (2004). Stable real-time 3d tracking using online and offline information. *IEEE Trans. on Pattern Analysis and Machine Intelligence* **26**(10), 1385–1391.
- Vandenbroucke, N., L. Macaire et J.-G Postaire (2003). Color image segmentation by pixel classification in an adapted hybrid color space : application to soccer image analysis. *Computer Vision and Image Understanding* **90**(2), 190–216.
- Vincze, M. (2001). Robust tracking of ellipses at frame rate. *Pattern Recognition* **34**(2), 487–498.
- Watanabe, Y., T. Komuro et M. Ishikawa (2007). 955-fps real-time shape measurement of a moving/deforming object using high-speed vision for numerous-point analysis. *IEEE Int. Conf. on Robotics and Automation, ICRA'07* pp. 3192–3197.
- Waxman, A.M., B.K. Parsi et M. Subbarao (1987). Closed-form solutions to image flow equations for 3D structure and motion. *Int. Journal of Computer Vision* **1**(3), 239–258.
- Waxman, A.M. et S. Ullman (1985). Surface structure and three-dimensional motion from image flow kinematics. *Int. Journal of Robotics Research* **4**(3), 72–94.
- Wei, G-Q., K. Arbter et G. Hirzinger (1997). Real-time visual servoing for laparoscopic surgery. controlling robot motion with color image segmentation. *IEEE Engineering in Medicine and Biology Magazine* **16**(1), 40–45.
- Weiss, L.E., A.C. Sanderson et C.P. Neuman (1987). Dynamic sensor-based control of robots with visual feedback. *IEEE Journal of Robotics and Automation* **3**(5), 404–417.
- Wells, G., C. Venaille et C. Torras (1996). Promising research vision-based robot positioning using neural networks. *Image and Vision Computing* **14**(10), 715–732.
- Wilson, W., C. Hulls et G. Bell (1996). Relative end-effector control using cartesian position-based visual servoing. *IEEE Trans. on Robotics and Automation* **12**(5), 684–696.
- Woodham, R.J. (1980). Photometric method for determining surface orientation from multiple images. *Optical Engineering* **19**(1), 139–144.



- Wu, Y. et T.S. Huang (2000). View-independant recognition of hand postures. *IEEE Int. Conf. on Computer Vision and Pattern Recognition* **2**, 88–94.
- Xie, M. et P. Rives (1989). Toward dynamic vision. Dans : *IEEE Workshop on Interpretation of 3D Scenes*. Austin, Texas. pp. 91–99.