



**HAL**  
open science

# Asservissement visuel direct utilisant les décompositions en shearlets et en ondelettes de l'image

Lesley-Ann Duflot

► **To cite this version:**

Lesley-Ann Duflot. Asservissement visuel direct utilisant les décompositions en shearlets et en ondelettes de l'image. Traitement du signal et de l'image [eess.SP]. Université de Rennes, 2018. Français. NNT : 2018REN1S018 . tel-01861464v2

**HAL Id: tel-01861464**

**<https://theses.hal.science/tel-01861464v2>**

Submitted on 30 Aug 2018

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

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



**THÈSE / UNIVERSITÉ DE RENNES 1**  
*sous le sceau de l'Université Bretagne Loire*

pour le grade de  
**DOCTEUR DE L'UNIVERSITÉ DE RENNES 1**

*Mention : Signal, Image, Vision*

**École Doctorale MathSTIC**

présentée par

**Lesley-Ann DUFLLOT**

préparée à l'unité de recherche Inria Rennes - Bretagne Atlantique  
et à l'institut FEMTO-ST / département AS2M à Besançon

---

**Asservissement  
visuel direct  
utilisant les  
décompositions  
en shearlets et  
en ondelettes  
de l'image**

**Thèse soutenue à Rennes  
le vendredi 13 Juillet 2018**

devant le jury composé de :

**Youcef MEZOUAR**

Professeur de l'Université Clermont Auvergne /  
*Rapporteur*

**Christophe DOIGNON**

Professeur de l'Université de Strasbourg / *Rapporteur*

**Marie CHABERT**

Professeure de l'Université de Toulouse / *Examinatrice*

**Eric MARCHAND**

Professeur de l'Université de Rennes 1 / *Examineur*

**Nicolas ANDREFF**

Professeur de l'Université de Franche-Comté / *Invité*

**Alexandre KRUPA**

Chargé de recherche, HDR, INRIA Rennes /  
*Directeur de thèse*

**Brahim TAMADAZTE**

Chargé de recherche, CNRS, institut FEMTO-ST /  
*Co-encadrant de thèse*



*"The good thing about computers is that they do what you tell them to do.  
The bad news is that they do what you tell them to do." – Ted Nelson*





# Remerciements

Je tiens tout d'abord à remercier les membres de mon jury de thèse. Eric Marchand, qui a accepté de le présider, Youcef Mezouar et Christophe Doignon qui m'ont apporté leur expertise pour rapporter sur mon manuscrit, Marie Chabert qui a tenu le rôle d'examinatrice et Nicolas Andreff, qui m'a accueilli chaleureusement dans son équipe bisontine et a fait le déplacement pour assister à ma soutenance. Je les remercie pour l'attention apportée à ce manuscrit, pour leurs remarques et pour être venus en cette période estivale. Je remercie également mon directeur de thèse Alexandre Krupa ainsi que mon co-encadrant Brahim Tamadazte, qui ont formé une parfaite équipe "good-cop/bad-cop" pour me guider, m'orienter, m'aider et me conseiller tout au long de ce parcours qu'est la thèse. Vous avez su me laisser une certaine autonomie tout en étant toujours disponible et de bon conseil, un mélange pas si courant et qui fonctionne bien. Merci enfin, pour les relectures de ce manuscrit et plus largement, les relectures des articles que nous avons eu le plaisir de publier ensemble.

J'adresse mes sincères remerciements à tous les intervenants, professeurs ou collègues qui par leurs enseignements, leurs paroles, leurs écrits, leurs conseils ou leurs critiques ont guidé mes réflexions et m'ont amené petit à petit vers l'obtention de ce doctorat dont je suis très fière. Je remercie également l'Inria Rennes et le FEMTO-ST pour l'accueil et les conditions de travail privilégié qui m'ont été offertes.

Merci à François Chaumette pour m'avoir accueillie à Rennes dans son équipe LAGADIC, qui m'a beaucoup apportée, professionnellement et personnellement. Merci à notre assistante Hélène Delaruée pour son aide lors de la préparation de missions ou, plus récemment, pour l'organisation de la soutenance de cette thèse. Merci à Fabien Spindler qui est toujours disponible en cas de problèmes techniques ainsi qu'à Pierre Chatelain qui m'a beaucoup aidé au début de ma thèse. Je remercie Jason, Thomas, Bryan et Pierre-Marie pour avoir été des super co-bureaux ainsi que Quentin, Nico, Aurélien, Lucas, Vishnu, Fabrizio, Giovanni, Aly, Hadrien, Pol, Flo, Fabien, Julien, Marc, Firas, Rahaf, Clément et Alex pour leur amitié, leur bonne humeur quotidienne, les pauses cafés, les parties de tarot, les vendredi soir festifs, les billards, les campings pré-séminaire au vert, toutes ces petites choses annexes qui mises bout à bout forment mon quotidien et me donnent envie de venir travailler. Je souhaite aux membres de la nouvelle équipe - RAINBOW - de connaître une ambiance aussi fantastique que celle qui a régné ici jusqu'à maintenant.

Merci à Nicolas Andreff pour m'avoir accueillie quelque mois à Besançon dans son équipe MINAROB, à Christine et Patrick de m'avoir hébergée chez eux durant cette période. Merci à Mouloud pour ses précieux conseils ainsi qu'à Kanty, Guillaume, Bassem, Cédric, Benoit, Wissem, Amélie, Elodie, Nathalie, Hector, Tristan et Vladimir pour m'avoir intégrée dans le département AS2M et pour avoir partagé avec moi de nombreux cafés, repas et sorties.

Je tiens également à remercier tous les membres de ma famille, qui forment des soutiens inébranlables. Merci à mes parents, Virginie et Martial, qui m'ont toujours permis d'aller plus loin dans mes études et qui ont cru en moi dès le départ. Merci pour votre éducation, vos valeurs, votre soutien, vos conseils avisés. Merci à Maman d'avoir largement contribué à la relecture de ce manuscrit. Merci à Lætitia dont l'aide a été précieuse lors de certains moments plus compliqués de cette thèse. Merci à mes grands-parents Joëlle, Philippe et Simone qui sont très fiers de moi et à Maurice qui l'aurait été j'en suis sûre. Merci à mon frère Quentin, avec qui je partage beaucoup de choses malgré nos 8 ans de différence (et encore désolée de t'avoir dégoutté des maths) et merci à mes beaux-parents Isabelle et Eric d'être venue à ma soutenance de thèse et d'avoir participé à son organisation.

Enfin, merci à Nicolas. Je n'ai pas trouvé les mots pour te remercier à la fin de ma soutenance. Je m'en excuse mais sache que c'est parce qu'il y avait trop à dire. J'aurai déjà pu dire que tu as toujours été là pour me motiver mais aussi pour apporter ton regard si objectif et rigoureux sur mon travail. J'aurai pu te dire merci de m'avoir supportée dans les moments où l'on est obnubilé par la thèse, ceux où l'on est de mauvaise foi ou encore ceux où l'on veut tout arrêter. Enfin, j'aurai aussi pu dire que cette thèse c'est un peu la tienne car tu partages ma vie et ton soutien m'a permis de faire le maximum et d'avancer dans les bons et les mauvais moments. Ta fierté est la plus belle des récompenses.



# Table des matières

<b>Introduction</b>	<b>3</b>
<b>1 Asservissement visuel</b>	<b>11</b>
1.1 Éléments de vision par ordinateur . . . . .	13
1.1.1 Formation d'une image . . . . .	13
1.1.2 Projection perspective et coordonnées pixelliques . . . . .	15
1.1.3 Coordonnées homogènes et changement de repère . . . . .	16
1.2 Commande . . . . .	17
1.2.1 Primitives visuelles . . . . .	17
1.2.2 Matrice d'interaction . . . . .	20
1.2.3 Analyse de la stabilité . . . . .	22
1.3 Exemples d'asservissements visuels 2D . . . . .	22
1.3.1 Approches géométriques . . . . .	23
1.3.2 Approches directes . . . . .	26
1.4 Conclusion . . . . .	38
<b>2 Espace fréquentiel et représentations multi-échelles</b>	<b>39</b>
2.1 Analyse de Fourier . . . . .	40
2.1.1 Transformée de Fourier continue . . . . .	41
2.1.2 Transformée de Fourier discrète . . . . .	43
2.1.3 Applications et limites de la transformée de Fourier . . . . .	43
2.2 Représentations multi-échelles . . . . .	45
2.2.1 Analyse multi-résolution . . . . .	46
2.2.2 Transformée en ondelettes . . . . .	47
2.2.3 Applications de la transformée en ondelettes . . . . .	54
2.2.4 Transformée en shearlets . . . . .	58
2.2.5 Différentes implémentations des shearlets . . . . .	60
2.2.6 Applications de la transformée en shearlets . . . . .	64
2.3 Conclusion . . . . .	67
<b>3 Asservissement visuel utilisant les shearlets appliqué à l'imagerie échographique</b>	<b>69</b>
3.1 Imagerie médicale . . . . .	70
3.1.1 Imagerie échographique . . . . .	70

3.1.2	Tomographie par cohérence optique . . . . .	80
3.2	Matrice d'interaction numérique . . . . .	85
3.2.1	Concept et implémentation . . . . .	85
3.2.2	Résultats de simulation . . . . .	87
3.2.3	Résultats expérimentaux . . . . .	90
3.3	Matrice d'interaction analytique . . . . .	95
3.3.1	Concept et implémentation . . . . .	95
3.3.2	Résultats de simulation . . . . .	97
3.3.3	Résultats expérimentaux . . . . .	104
3.4	Conclusion . . . . .	106
<b>4</b>	<b>Généralisation à l'asservissement visuel fondé sur les transformées multi-échelles</b>	<b>107</b>
4.1	Principe de l'asservissement visuel fondé sur les transformées multi-échelles	108
4.1.1	Approche utilisant la transformée en ondelettes . . . . .	108
4.1.2	Approche utilisant la transformée en shearlets . . . . .	114
4.2	Simulation et validation expérimentale sur caméra conventionnelle . . . . .	115
4.2.1	Détails d'implémentation . . . . .	116
4.2.2	Domaine de convergence . . . . .	119
4.2.3	Simulation . . . . .	120
4.2.4	Validation expérimentale . . . . .	126
4.3	Validation expérimentale en utilisant l'OCT . . . . .	138
4.3.1	Description du dispositif robotique . . . . .	138
4.3.2	Détails d'implémentation . . . . .	139
4.3.3	Validation expérimentale . . . . .	141
4.4	Conclusion . . . . .	147
<b>5</b>	<b>Shearlets en acquisition comprimée : application à l'OCT</b>	<b>149</b>
5.1	Bases de l'acquisition comprimée . . . . .	150
5.1.1	Problème d'optimisation . . . . .	150
5.1.2	Reconstruction d'un signal . . . . .	153
5.1.3	Acquisition comprimée en imagerie . . . . .	156
5.2	Application à l'imagerie OCT . . . . .	157
5.2.1	Nouvelle formulation de l'algorithme FISTA . . . . .	158
5.2.2	Validation numérique . . . . .	159
5.3	Conclusion . . . . .	166
	<b>Conclusion</b>	<b>167</b>
	<b>Bibliographie</b>	<b>189</b>

# Introduction

## Motivation

L’asservissement visuel est une technique qui utilise des informations visuelles extraites des images, obtenues à l’aide d’un ou plusieurs capteurs, pour contrôler le déplacement d’un robot par une commande en boucle fermée (Hutchinson et al., 1996; Chaumette et Hutchinson, 2006, 2007). Les trente dernières années ont vu apparaître d’importants progrès dans ce domaine avec des applications dans le contexte de la manipulation d’objets industriels (Lippiello et al., 2007), de la robotique médicale (Abolmaesumi et al., 2002; Krupa et al., 2003) ou, plus récemment pour la navigation de drones (Barajas et al., 2013; Máthé et al., 2016). Généralement, l’objectif de l’asservissement visuel est d’effectuer une tâche de positionnement ou de suivi en minimisant la différence entre un vecteur  $\mathbf{s}(t)$  de primitives courantes et un vecteur  $\mathbf{s}^*$  de primitives désirées. Les premières primitives considérées en asservissement visuel étaient de type géométrique tels que les points (Wilson et al., 1996), les lignes (Andreff et al., 2002) ou les moments géométriques (Tahri et al., 2015). Ces approches géométriques dépendent fortement de la capacité de la méthode à détecter, extraire et suivre les primitives visuelles durant la procédure de contrôle du robot.

Récemment, de nouvelles méthodes d’asservissement visuel ont émergé permettant d’éviter la contrainte de détecter et suivre tout au long de la tâche des primitives visuelles géométriques. Ces méthodes, dites directes, utilisent l’information globale contenue dans les pixels de l’image. La plus simple des primitives directes est la photométrie, c’est-à-dire l’ensemble des valeurs des intensités des pixels de l’image (Collewet et Marchand, 2011) mais d’autres, plus complexes, existent comme l’information mutuelle (Dame et Marchand, 2011), la somme des variances conditionnelles de l’image (Richa et al., 2011), la transformée de Fourier (Kallem et al., 2007; Marturi et al., 2016), les mélanges de Gaussiennes (Crombez et al., 2015) ou encore les moments photométriques (Bakthavatchalam et al., 2013). Ces techniques d’asservissement visuel direct sont souvent considérées comme étant plus précises que les méthodes utilisant des primitives visuelles géométriques, mais elles ont cependant un domaine de convergence plus réduit.

Le travail décrit dans cette thèse traite du développement d’une approche d’asservissement visuel direct utilisant des représentations multi-échelles caractérisant les informations spatiales mais également fréquentielles de l’image. Les représentations multi-

échelles d'une image sont obtenues en appliquant successivement des filtres passe-hauts et passe-bas à une image, suivis d'un sous-échantillonnage. Chaque niveau de filtrage contient alors des détails de plus en plus fins au fil des décompositions. En complément de l'obtention de plusieurs niveaux de détails, ces méthodes permettent d'avoir également une représentation parcimonieuse de l'image, c'est-à-dire une représentation dans laquelle la plupart des coefficients ont une valeur proche de zéro. Les coefficients non nuls reflètent alors l'information essentielle de l'image. La première représentation multi-échelle fut la représentation pyramidale laplacienne d'une image, inventée en 1983 par Burt et Adelson (Burt et Adelson, 1983), qui est caractérisée par l'application successive de noyaux Gaussiens sur une image. Ce concept fut ensuite étendu par Simoncelli (Simoncelli et Freeman, 1995) pour améliorer les représentations des courbes et des points d'intérêts. L'analyse harmonique a contribué à l'essor des représentations multi-échelles. Cette discipline découle directement de l'analyse de Fourier et s'intéresse aux représentations des signaux sous forme d'une succession d'ondes. Pour élaborer une représentation multi-échelle intéressante, un système d'ondes optimal est construit pour décrire la structure d'une image. L'élaboration de tels systèmes est un axe de recherche majeur des mathématiques appliquées modernes.

De toute évidence, le concept le plus célèbre en analyse harmonique se trouve dans la théorie des ondelettes. Contrairement à la transformée de Fourier, qui ne contient que des informations fréquentielles sur un signal et *a fortiori* une image, la transformée en ondelettes permet une localisation à la fois en espace et en fréquence de ce signal. Une première définition d'une ondelette fut donnée en 1910 par Haar (Haar, 1910). Cependant, ce sont les travaux de Meyer (Meyer, 1986), de Daubechies (Daubechies, 1988) et de Mallat (Mallat, 1989), dans la fin des années 1980, qui rendirent la transformée en ondelettes discrète applicable dans de nombreux domaines. La transformée en ondelettes d'une image est depuis un outil largement utilisé en traitement d'images pour des applications très variées comme par exemple : la compression (Christopoulos et al., 2000), le filtrage, la détection ou le suivi d'objet (Argoul et al., 1990). Bien que la transformée en ondelettes soit bien adaptée pour décrire les singularités d'un signal à une dimension, elle n'est pas forcément optimale pour décrire des images (Donoho, 2001). Ce constat a motivé de nombreuses recherches sur l'élaboration de transformées multi-échelles sensibles aux hautes fréquences plus localisées dans l'espace 2D que les ondelettes. Les curvelets (Candes et Donoho, 2000), les contourlets (Do et Vetterli, 2003) ou plus récemment, les shearlets (Labate et al., 2005; Kutyniok et Labate, 2012b) sont typiquement des transformées adaptées aux signaux 2D. Ces outils trouvent de nombreuses applications en traitement d'images comme le débruitage (Easley et al., 2009), la restauration d'image (King et al., 2013) ou la détection de contours (Yi et al., 2008; Kutyniok et Petersen, 2015; Reisenhofer et al., 2016). Malgré leur rôle important en traitement d'images, les représentations multi-échelles n'ont été considérées que très tardivement dans une boucle de commande par asservissement visuel. Le premier travail dans cette voie a été présenté en 2016 dans (Ourak et al., 2016c) où un asservissement visuel à 6 degrés de liberté (ddl) utilisant les coefficients obtenus grâce aux filtres passe-bas d'une transformée en ondelettes a été développé.

## Contexte et objectifs

Cette thèse a été réalisée sous la co-supervision de l'équipe Rainbow (ex : Lagadic) de l'Inria Rennes Bretagne Atlantique et de l'équipe MiNaRoB du département AS2M du laboratoire FEMTO-ST de Besançon grâce au financement de la région Bretagne et du projet ANR NEMRO (<http://www.agence-nationale-recherche.fr/Projet-ANR-14-CE17-0013>).

**Projet NEMRO :** Le projet NEMRO traite du diagnostic précoce des maladies neurodégénératives comme Alzheimer ou Parkinson. En effet, différentes études cliniques ont montré une corrélation entre la perte d'odorat et l'apparition de ce type de maladies. L'enjeu est alors de déterminer quel est le lien de causalité, c'est-à-dire si la perte d'odorat est un signe de dégénérescence neuronale ou si l'altération du fonctionnement des neurones entraîne la non-stimulation des cellules olfactives.

Ce projet vise donc au développement d'un endoscope nasal flexible de diamètre millimétrique capable de naviguer dans les fosses nasales jusqu'à atteindre les muqueuses olfactives. Ces muqueuses contiennent des cellules olfactives de quelques dizaines de millimètres directement liées au cerveau. L'endoscope envisagé est un robot flexible avec sur sa partie distale une sonde OCT (Tomographie par cohérence optique) fibrée, permettant la visualisation 3D de chaque muqueuse olfactive afin de pouvoir l'étudier. Ce système sera un moyen de caractérisation dynamique in-vivo et non-invasif permettant d'effectuer des images 3D à haute résolution (biopsies optiques 3D). Il permettra le suivi de l'évolution dans le temps de la perte d'odorat et sa répercussion sur la dégénérescence neuronale grâce à un système fiable à coût réduit.

Des techniques de commande par retour visuel permettront la navigation intranasale, notamment pour la réalisation des biopsies optiques 3D reproductibles, obtenues par le capteur OCT. La reproductibilité permettra le repositionnement sur des zones d'intérêt au cours de l'évolution de la pathologie. Dans le contexte de ce projet, mes travaux portent plus particulièrement sur l'élaboration d'une nouvelle approche d'asservissement visuel utilisant les représentations multi-échelles des images.

L'asservissement visuel sur appareil OCT est un domaine très novateur (Ourak et al., 2016a,b). Cette modalité présente également de nombreuses similitudes avec l'imagerie échographique pour laquelle davantage de travaux ont été réalisés en asservissement visuel (Abolmaesumi et al., 2001; Lee et al., 2007; Krupa et al., 2016) afin de procurer des systèmes d'aide à la visualisation échographique ou au guidage d'outils pour les chirurgiens.

Les résultats présentés dans ce manuscrit sont le fruit de deux ans et demi de recherches, dont deux ans réalisés sur le site de Rennes, plus spécialisé en imagerie échographique, et six mois sur le site de Besançon, plus spécialisé en imagerie par OCT. L'objectif de la thèse était de répondre à la question suivante :

*Comment utiliser les transformées multi-échelles pour améliorer la robustesse aux conditions défavorables des asservissements visuels utilisant des modalités d'images médicales très bruitées comme l'imagerie échographique ou la tomographie par cohérence optique ?*

## Contributions

Les contributions majeures de cette thèse sont les suivantes :

- Une méthode d'asservissement visuel robuste aux bruits en imagerie échographique reposant sur l'utilisation de la transformée en shearlets.
- Une méthode générique d'asservissement visuel reposant sur les transformées multi-échelles.
- Une validation expérimentale de l'asservissement visuel et une comparaison des performances obtenues pour différents types de transformées multi-échelles dans le cas de la commande d'une caméra CCD embarquée sur un robot à 6 ddl.
- Une validation de la méthode d'asservissement visuel utilisant la transformée en ondelettes sous-échantillonnée de l'image fournie par un capteur OCT pour la commande d'un robot à 3 ddl.
- Une comparaison entre différentes transformées multi-échelles pour une tâche d'acquisition comprimée par un capteur OCT.

## Structure du document

L'organisation de ce document s'articule autour de cinq chapitres :

- Dans le premier chapitre, nous rappellerons les éléments de vision par ordinateur nécessaires à la compréhension de l'asservissement visuel. Ensuite, nous présenterons les principes de base de la commande d'un robot par asservissement visuel. Les notions de primitives visuelles, de matrice d'interaction, d'erreur visuelle, ainsi que différents types d'asservissement visuel (2D, 3D,  $2D\frac{1}{2}$ ) y seront décrits. Enfin, la dernière partie de ce chapitre sera consacrée à la présentation de différentes primitives visuelles 2D, en distinguant les approches géométriques et les approches directes. L'asservissement visuel utilisant les points comme information visuelle ainsi que celui utilisant l'information photométrique de l'image sont les deux approches les plus importantes pour la compréhension de nos contributions, même si

d'autres (moments géométriques, information mutuelle, mélanges de gaussiennes) y seront également présentées.

- Dans le second chapitre, nous introduirons des éléments importants de l'analyse de Fourier et de l'analyse harmonique. Une première section sera consacrée à la transformée de Fourier et aux représentations fréquentielles des images, en distinguant les transformées continues et les transformées discrètes. Une deuxième section présentera l'analyse multi-résolution et les transformées multi-échelles avec, d'une part, les transformées en ondelettes et, d'autre part, en shearlets. Ces transformées seront utilisées en tant que primitives visuelles dans les approches de commande par asservissement visuel proposées dans les chapitres 3 et 4.
- Le troisième chapitre s'articulera autour de l'imagerie échographique. Une première section présentera les modalités d'imagerie médicale utilisées dans les travaux de thèse (imagerie échographique et OCT). Une deuxième section exposera notre première contribution à l'asservissement visuel utilisant l'imagerie échographique : l'utilisation des coefficients de la transformée en shearlets d'une image en tant que primitives visuelles avec une estimation numérique de la matrice d'interaction pour réaliser une commande à 6 ddl. La dernière section présentera une formulation analytique de cette matrice d'interaction pour le cas d'une commande à 3 ddl. Les résultats de ces deux contributions ont été validés par simulation et également avec un robot anthropomorphique sur lequel une sonde échographique a été montée.
- Le quatrième chapitre exposera une généralisation des méthodes présentées dans le chapitre 3 afin de pouvoir utiliser diverses transformées multi-échelles en tant que primitives dans une commande d'asservissement visuel. La première section présentera le modèle fondé sur les ondelettes et celui fondé sur les shearlets. La deuxième section introduira les résultats de simulation et expérimentaux obtenus pour une tâche de positionnement d'une caméra CCD embarquée sur un robot à 6 ddl. Ces tests ont permis de comparer les transformées en ondelettes et en shearlets issues de deux variantes d'implémentation : sous-échantillonnées et non sous-échantillonnées. Enfin, la dernière section exposera les résultats obtenus par un asservissement visuel utilisant la transformée en ondelettes sous-échantillonnées de l'image fournie par un dispositif OCT.
- Dans le dernier chapitre, nous présenterons le concept d'acquisition comprimée. Nous étudierons comment accélérer l'acquisition des données d'un appareil OCT grâce à cette technique en comparant différents masques d'acquisition et différentes transformées multi-échelles servant à la reconstruction de l'image.

## Listes des publications

### Article publié dans une revue internationale

L. A. Dufflot, R. Reichenhofer, B. Tamadazte, N. Andreff and A. Krupa, "Wavelet and shearlet-based image representation for visual servoing," *The International Journal of Robotic Research* 2018.

### Articles publiés dans une conférence internationale

L. A. Dufflot, A. Krupa, B. Tamadazte and N. Andreff, "Shearlet-based vs. photometric-based visual servoing for robot-assisted medical applications," *2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Daejeon, 2016, pp. 4099-4104.

L. A. Dufflot, A. Krupa, B. Tamadazte and N. Andreff, "Shearlet transform : A good candidate for compressed sensing in optical coherence tomography," *2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, Orlando, FL, 2016, pp. 435-438.

L. A. Dufflot, A. Krupa, B. Tamadazte and N. Andreff, "Towards Ultrasound-based visual servoing using shearlet coefficients," *2016 IEEE International Conference on Robotics and Automation (ICRA)*, Stockholm, 2016, pp. 3420-3425.

L. A. Dufflot, B. Tamadazte, N. Andreff and A. Krupa, "Wavelet-based visual servoing using OCT images," *2018 IEEE International Conference on Biomedical Robotics and Biomechatronics (BioROB)*.

## Notations

Dans ce document, nous utiliserons des lettres en gras pour représenter les vecteurs (exemple :  $\mathbf{s}$ ) et des lettres en gras et en majuscule pour représenter les matrices (exemple :  $\mathbf{A}$ ). Les valeurs scalaires et les fonctions continues seront représentées par des lettres en italique (par exemple  $x$ , resp.  $f$ ). Les coefficients d'un vecteur ou d'une matrice seront représentés grâce à un indice (par exemple :  $h_n$ , resp.  $A_{n,m}$ ).

Pour une meilleure compréhension du document, le lecteur est invité à se référer au tableau suivant :

Notations	Description
	Ensembles
$\mathbb{R}^N$	Ensemble des nombres réels de dimension $N$
$\mathbb{N}$ et $\mathbb{N}_0$	Ensemble des entiers naturels, respectivement sans et avec 0
$\mathbb{Z}$	Ensemble des entiers relatifs
$L^2(\mathbb{R}^N)$	Espace des fonctions réelles de carré intégrable de dimension $N$
$\ell^2(\mathbb{Z})$	Espace des suites de carrés sommables
	Fonctions et matrices usuelles
$ \cdot $	Valeur absolue
$\lceil \cdot \rceil$	Fonction de seuillage
$\langle \cdot, \cdot \rangle$	Produit scalaire de $L^2$
$\bar{x}$	Complexe conjugué de $x$
$\cdot^\top$	Transposée
$\det \mathbf{A}$	Déterminant de la matrice $\mathbf{A}$
$\mathbf{I}_N$	Matrice identité de $\mathbb{R}^N$
$\mathbf{A}^+$	Pseudo-inverse de Moore-Penrose de $\mathbf{A}$
$\ker$	Noyau
$\Delta \mathbf{I}$	Gradient de l'image $\mathbf{I}$
$\hat{\cdot}$	Transformée de Fourier
$i$	Unité imaginaire
$*$	Convolution
$\binom{n}{k}$	Coefficient binomial ( $k$ parmi $n$ )
	Géométrie projective
$\mathbf{X}_o = (X_o, Y_o, Z_o)$	Coordonnées d'un point dans le repère objet
$\mathcal{F}_o$	Repère objet
$\mathbf{X}_c = (X_c, Y_c, Z_c)$	Coordonnées d'un point dans le repère caméra
$\mathcal{F}_c$	Repère caméra
$\mathcal{P}$	Plan image
$(x, y)$	Coordonnées métriques d'un point
$f$	Distance focale
$(u, v)$	Coordonnées en pixels d'un point
$\boldsymbol{\kappa}$	Matrice des paramètres intrinsèques de la caméra
$\bar{\mathbf{X}}$	Coordonnées homogènes de $\mathbf{X}$
${}^c\mathbf{M}_o$	Transformation homogène de $\mathcal{F}_o$ vers $\mathcal{F}_c$
	Notations usuelles en asservissement visuel
$\mathbf{r}$ et $\mathbf{r}^*$	Poses courantes et désirées, respectivement
$\mathbf{s}$ et $\mathbf{s}^*$	Primitives visuelles courantes et désirées, respectivement
$\mathbf{I}$ et $\mathbf{I}^*$	Images courantes et désirées, respectivement

$\mathbf{v}$	Vecteur de vitesse de la caméra
$v_x, v_y, v_z$	Vitesses linéaires
$\omega_x, \omega_y, \omega_z$	Vitesse angulaires
$\mathbf{e}$	Erreur visuelle
$\mathbf{L}_s$	Matrice d'interaction
$\widehat{\mathbf{L}}_s$	Approximation de la matrice d'interaction
$\mathbf{L}_p$	Matrice d'interaction relative à un point
$\lambda$	Gain de la loi de commande de type Gauss-Newton
$\mathbf{H} = \mathbf{L}_s^\top \mathbf{L}_s$	Matrice Hessienne de $\mathbf{s}$
$\mu$	Paramètre de la loi de commande de type Levenberg-Marquardt
$\mathcal{L}$	Fonction de Lyapunov
$t_x, t_y, t_z$	Translations selon les axes $x, y$ et $z$
$r_x, r_y, r_z$	Rotations selon les axes $x, y$ et $z$
Notations spécifiques à ce manuscrit en asservissement visuel	
$\mathbf{s}_i$ ( $i = [\text{ph}, \text{w}, \text{sh}]$ )	Primitives visuelles photométriques, ondelettes et shearlets
$\mathbf{e}_i$ ( $i = [\text{ph}, \text{w}, \text{sh}]$ )	Vecteur d'erreur visuelle photométrique, ondelettes et shearlets
$\mathbf{L}_{s_{\text{ph}}}$	Matrice d'interaction dans le cas de l'approche photométrique
$\mathbf{L}_{s_{\text{w}}}$ et $\widehat{\mathbf{L}}_{s_{\text{w}}}$	Matrice d'interaction de l'approche ondelette et son approximation
$\mathbf{L}_{s_{\text{sh}}}$ et $\widehat{\mathbf{L}}_{s_{\text{sh}}}$	Matrice d'interaction de l'approche shearlet et son approximation
$\mathbf{L}_p(x, y)$	Matrice d'interaction du point $(x, y)$ dans le cas de l'approche photométrique
$\mathbf{L}_{\text{pw}}(x, y)$	Matrice d'interaction du point $(x, y)$ dans le cas de l'approche ondelette
$\mathbf{L}_{\text{psh}}(x, y)$	Matrice d'interaction du point $(x, y)$ dans le cas de l'approche shearlet
$Z_c$	Profondeur du plan image
$I_t$	Image observée à l'instant $t$ , considérée comme une fonction continue
$I_t(x, y)$	Intensité de l'image au point $(x, y)$ à l'instant $t$
Notations relatives aux représentations multi-échelles	
$D_2$ et $T_{\mathbf{m}}$	Opérateurs de dilatation dyadique et de translation de $L^2(\mathbb{R}^2)$
$\phi_{\mathbf{m}}^{(1)}$	Fonction d'échelle de $L^2(\mathbb{R}^2)$
$\psi_{j, \mathbf{m}}^{(1)}$	Ondelette horizontale de $L^2(\mathbb{R}^2)$ au niveau $j$ tradlatée de $\mathbf{m}$
$\psi_{j, \mathbf{m}}^{(2)}$	Ondelette verticale de $L^2(\mathbb{R}^2)$ au niveau $j$ tradlatée de $\mathbf{m}$
$\psi_{j, \mathbf{m}}^{(3)}$	Ondelette diagonale de $L^2(\mathbb{R}^2)$ au niveau $j$ tradlatée de $\mathbf{m}$
$\phi$	Fonction d'échelle de $L^2(\mathbb{R})$
$\psi$	Fonction génératrice d'ondelettes de $L^2(\mathbb{R})$
$\mathcal{W}_\psi f$	Transformée en ondelettes de la fonction $f \in L^2(\mathbb{R}^2)$
$\mathbf{h}$ et $\mathbf{g}$	Filtres passe-bas et passe-haut, respectivement
$\psi_{j, k, \mathbf{m}}^{(1)}$	Shearlet verticale de $L^2(\mathbb{R}^2)$ au niveau $j$ , avec le découpage $k$ tradlatée de $\mathbf{m}$
$\psi_{j, k, \mathbf{m}}^{(2)}$	Shearlet horizontale de $L^2(\mathbb{R}^2)$ au niveau $j$ , avec le découpage $k$ tradlatée de $\mathbf{m}$
$D_{\mathbf{M}}$	Opérateur de transformation géométrique de $L^2(\mathbb{R}^2)$
$D_{\mathbf{A}}$	Opérateur de dilatation de $L^2(\mathbb{R}^2)$ avec la matrice d'échelle $\mathbf{A}$
$D_{\mathbf{S}_k}$	Opérateur de découpage de $L^2(\mathbb{R}^2)$ avec la matrice de découpage $\mathbf{S}_k$
$\mathcal{S}_\psi f$	Transformée en shearlets de la fonction $f \in L^2(\mathbb{R}^2)$

TABLE 1 – Liste des notations

# Chapitre 1

## Asservissement visuel

L'asservissement visuel est un procédé permettant de contrôler le déplacement d'un robot en utilisant des données visuelles issues d'une ou plusieurs caméras. Une première utilisation d'un capteur de vision en boucle fermée, améliorant la précision de positionnement d'un robot, a été décrite dans les travaux de Shirai et Inoue (Shirai et Inoue, 1973). Cependant, ce n'est que vers la fin des années 1980 que le concept d'asservissement visuel est réellement apparu (Weiss et al., 1987; Feddema et Mitchell, 1989). En règle générale, l'image est fournie par une caméra monoculaire et correspond à une projection en perspective d'un environnement 3D sur un plan 2D (image) qui n'est autre que le capteur électronique de la caméra. L'asservissement visuel peut faire également appel à d'autres types de capteurs comme les caméras stéréoscopiques qui permettent d'avoir deux vues différentes d'une même scène (Alkhalil et Doignon, 2012; Tamadazte et Andreff, 2014), des caméras RGB-D qui donnent, en plus de l'information image 2D, une estimation de la profondeur de chaque point de l'image (Nakhaeinia et al., 2014; Teulière et Marchand, 2014) ou encore, des caméras infra-rouges (Mayya et Doignon, 2011). Le sommaire ci-dessous résume le contenu de ce premier chapitre :

### Sommaire

---

<b>1.1</b>	<b>Éléments de vision par ordinateur</b>	<b>13</b>
1.1.1	Formation d'une image	13
1.1.2	Projection perspective et coordonnées pixelliques	15
1.1.3	Coordonnées homogènes et changement de repère	16
<b>1.2</b>	<b>Commande</b>	<b>17</b>
1.2.1	Primitives visuelles	17
1.2.2	Matrice d'interaction	20
1.2.3	Analyse de la stabilité	22
<b>1.3</b>	<b>Exemples d'asservissements visuels 2D</b>	<b>22</b>
1.3.1	Approches géométriques	23
1.3.2	Approches directes	26
<b>1.4</b>	<b>Conclusion</b>	<b>38</b>

---

L'asservissement visuel est utilisé en robotique pour de nombreuses applications, en particulier la navigation autonome (Usher et al., 2003; Cherubini et Chaumette, 2013), la saisie d'objet (Kragic, 2001; Vahrenkamp et al., 2008) et la robotique médicale (Abolmaesumi et al., 2001; Krupa et al., 2016). La FIGURE 1.1 montre, par exemple, un robot humanoïde s'emparant d'une boîte (Vahrenkamp et al., 2008).

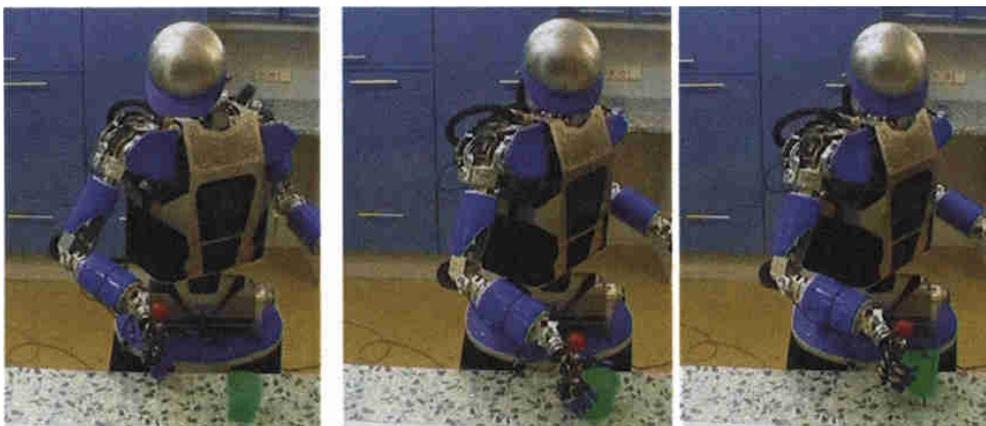


FIGURE 1.1 – Exemple de saisie d'objet par un robot humanoïde (Vahrenkamp et al., 2008).

Pour mettre au point un asservissement visuel, il est nécessaire de choisir des informations visuelles adéquates, appelées primitives visuelles, connaître l'état de celles-ci pendant toute la tâche robotique et avoir un système d'acquisition d'images suffisamment rapide pour agir en temps réel (soit environ 25 images par secondes). La FIGURE 1.2 schématise le principe de fonctionnement d'un asservissement visuel.

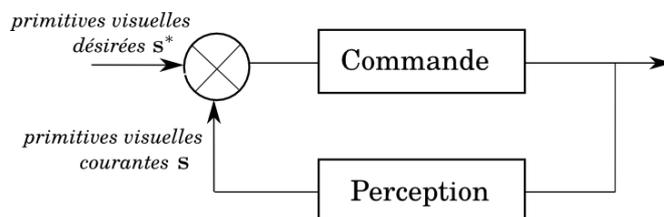


FIGURE 1.2 – Principe de l'asservissement visuel

Dans ce chapitre, nous verrons quelques éléments de base de vision par ordinateur avant de s'intéresser en détails aux deux étapes : perception et commande qui constituent le cœur d'un asservissement visuel.

## 1.1 Éléments de vision par ordinateur

Cette section s'intéresse aux différentes notions mathématiques et géométriques nécessaires pour la compréhension de la suite du manuscrit. Nous nous limiterons dans un premier temps aux images issues de caméras conventionnelles (caméras à lumière blanche). Nous aborderons plus tard, dans un autre chapitre, l'imagerie échographique et celle par OCT qui font parties des objectifs applicatifs des travaux de cette thèse.

### 1.1.1 Formation d'une image

Une image conventionnelle est une représentation en deux dimensions (2D) d'un monde en trois dimensions (3D). La manière la plus simple de comprendre la formation d'une image numérique est de s'intéresser à la chambre noire et à la caméra sténopé. Une chambre noire est un ancien instrument optique qui permet d'obtenir une projection de la lumière sur une surface plane. Elle est composée d'une boîte percée d'un petit trou servant à faire passer la lumière. Une image vient alors se former sur la surface opposée à ce trou. Il s'agit de l'image inversée de la réalité extérieure<sup>1</sup>. Les plus anciennes traces écrites sur la chambre noire dates du IV<sup>ème</sup> siècle. De nombreux grands peintres et artistes de la Renaissance, dont Vermeer<sup>2</sup>, ont également utilisé ce principe pour peindre leurs œuvres (voir FIGURE 1.3). Si du papier photosensible est placé sur la surface plane de la chambre noire, nous pouvons obtenir une photographie après un temps d'impression (exposition) pouvant aller jusqu'à plusieurs heures, c'est le principe de la caméra sténopé. La première photographie connue a été vraisemblablement prise avec un tel système en 1827 par Niépce (voir FIGURE 1.4). Par la suite et progressivement, le trou de la chambre noire a été remplacé par des capteurs dans nos appareils photographiques actuels mais le principe reste globalement le même.

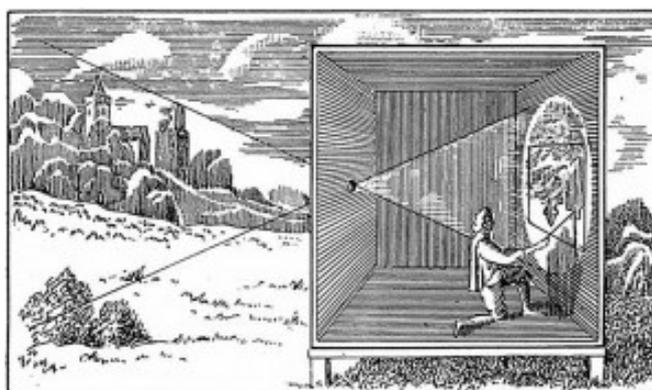


FIGURE 1.3 – Gravure représentant un peintre utilisant une chambre noire.

1. L'œil reçoit également une image inversée du réel et notre cerveau la redresse.

2. voir [https://fr.wikipedia.org/wiki/Chambre\\_noire](https://fr.wikipedia.org/wiki/Chambre_noire)

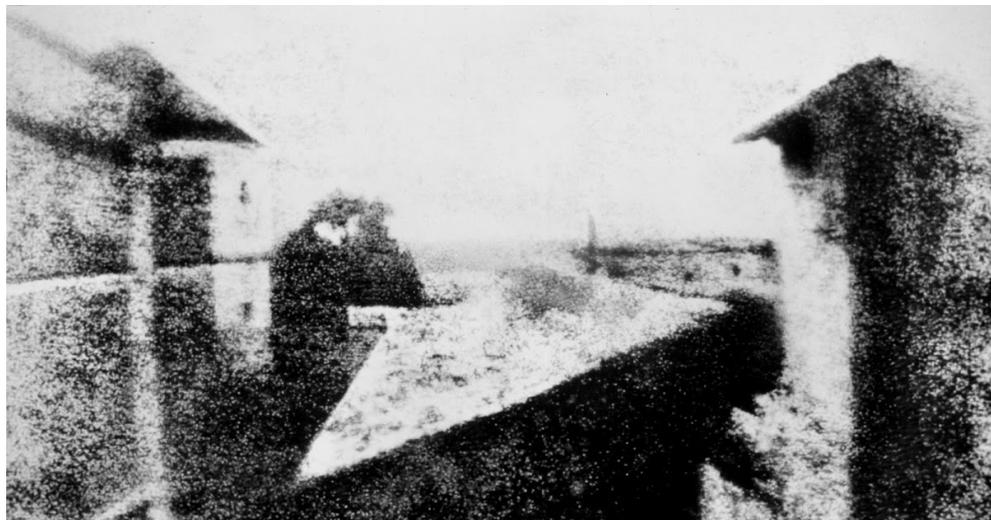
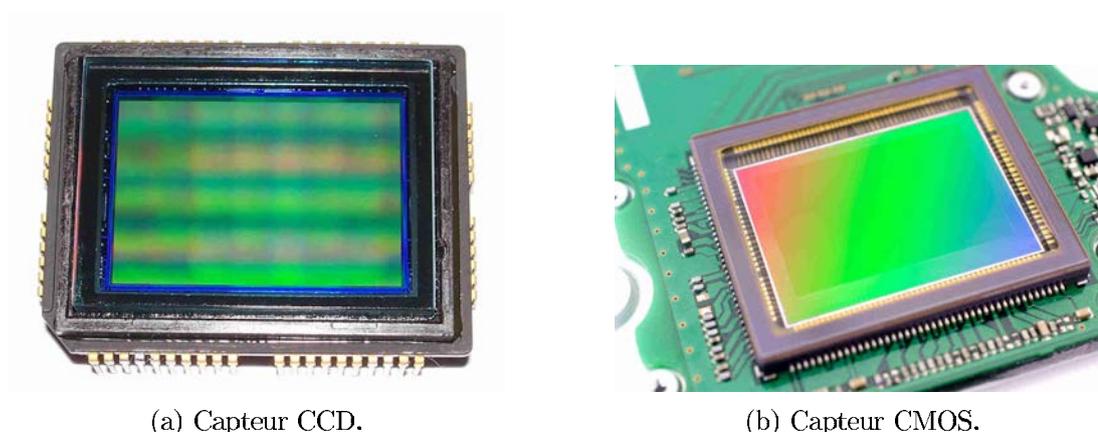


FIGURE 1.4 – *Point de vue du Gras* : la première image permanente réussie et connue de l'histoire de la photographie, prise par l'inventeur français Nicéphore Niépce (1765-1833) en 1826 ou 1827.

De nos jours, les capteurs numériques d'images sont présents dans de nombreux objets de la vie quotidienne : appareil photo, webcam, caméra de surveillance ou encore dans nos téléphones portables. Leur but est de retranscrire, aussi fidèlement que possible, l'image d'un objet éclairé. Si les premières caméras fonctionnaient avec des films photographiques en tant que détecteurs, nous avons vu l'émergence depuis les années 1970 de capteurs optoélectroniques qui sont toujours d'actualité. Le plus ancien est le capteur CCD ("*Charged Coupled Device*"), inventé par Smith et Boyle dans les Laboratoires Bell en 1969 (voir FIGURE 1.5(a)). Ce capteur permet la collecte, le transfert et la conversion de la charge électrique générée par des photons incidents, capturés par les cellules photovoltaïques. Il a valu à ses créateurs le Prix Nobel de physique en 2009. Le deuxième capteur optique majeur est le système CMOS ("*Complimentary Metal Oxyde Semiconductor*") qui est composé de photodiodes, où chaque cellule photovoltaïque possède sa propre collecte et sa propre conversion de la charge générée par les photons incidents (voir FIGURE 1.5(b)). Chacun de ces capteurs possède des avantages et des inconvénients. Le capteur CCD a une meilleure sensibilité à la lumière que le capteur CMOS, cela dit, il consomme plus d'énergie et est plus coûteux à la fabrication. Ce sont ces derniers points qui ont amené les fabricants de smartphones à privilégier l'utilisation du capteur CMOS. Sa structure permet également de réaliser du traitement d'images embarqué, ce qui supprime la différence de qualité d'images avec le capteur CCD.



(a) Capteur CCD.

(b) Capteur CMOS.

FIGURE 1.5 – Exemples de capteurs CCD et CMOS.

### 1.1.2 Projection perspective et coordonnées pixelliques

Soit  $M$ , un point 3D de coordonnées  $\mathbf{X}_o = (X_o, Y_o, Z_o)$  dans le repère objet  $\mathcal{F}_o$  et de coordonnées  $\mathbf{X}_c = (X_c, Y_c, Z_c)$  dans le repère caméra  $\mathcal{F}_c$ . La FIGURE 1.6 illustre la projection perspective, notée  $m$ , de ce point sur un plan image  $\mathcal{P}$ . Cette dernière fournit des coordonnées métriques  $(x, y)$  dans le repère du plan image. L'intersection entre le plan image  $\mathcal{P}$  et l'axe optique correspond au centre optique  $(u_0, v_0)$ . Ce centre optique est situé à une distance  $f$  de la caméra  $C$ , qui s'appelle la *distance focale*. Il est possible alors de calculer les coordonnées du point  $m$  dans le plan image à partir des coordonnées  $(X_c, Y_c, Z_c)$  du point  $M$  dans le repère caméra  $\mathcal{F}_c$ . Ces coordonnées  $(x, y)$  sont obtenues comme suit :

$$\begin{pmatrix} x \\ y \end{pmatrix} = -\frac{f}{Z_c} \begin{pmatrix} X_c \\ Y_c \end{pmatrix} \quad (1.1)$$

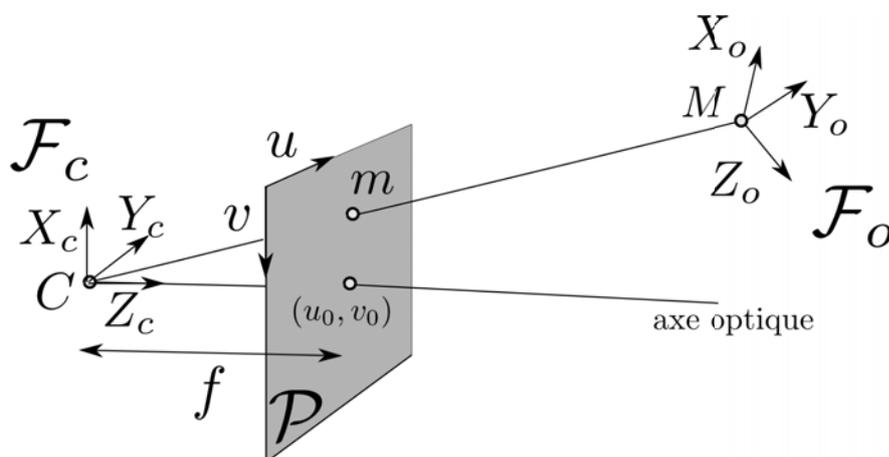


FIGURE 1.6 – Projection perspective d'un point 3D sur le plan image.

Pour obtenir les coordonnées en pixel  $(u, v)$  du point image  $m$ , nous pouvons réécrire la relation (1.1) sous forme linéaire en utilisant les coordonnées homogènes :

$$\begin{pmatrix} u \\ v \\ 1 \end{pmatrix} = \boldsymbol{\kappa} \begin{pmatrix} x \\ y \\ 1 \end{pmatrix} \quad \text{avec} \quad \boldsymbol{\kappa} = \begin{bmatrix} p_x & 0 & u_0 \\ 0 & p_y & v_0 \\ 0 & 0 & 1 \end{bmatrix} \quad (1.2)$$

où  $(u_0, v_0)$  représentent les coordonnées en pixel du centre optique.

Soient  $l_x$  et  $l_y$ , les dimensions d'un pixel en largeur et en hauteur, alors nous pouvons écrire  $p_x = \frac{f}{l_x}$  et  $p_y = \frac{f}{l_y}$ . Nous appelons  $\boldsymbol{\kappa}$ , la matrice des paramètres intrinsèques de la caméra  $(u_0, v_0, p_x, p_y)$ . Cette matrice est calculée hors-ligne à l'aide d'algorithmes d'étalonnage (Heikkila et Silven, 1997; Sturm et Maybank, 1999; Z. Zhang, 2000). Des paramètres de compensation sont parfois considérés afin de corriger les distorsions optiques (tangentiellles, radiales, etc).

### 1.1.3 Coordonnées homogènes et changement de repère

**Définition 1.1. Coordonnées homogènes :** *Un point  $M$  de coordonnées 3D  $\mathbf{X} = (X, Y, Z)$  dans l'espace peut être désigné par ses coordonnées homogènes sous la forme  $\overline{\mathbf{X}} = (X, Y, Z, 1)$ . Les coordonnées homogènes sont pratiques pour appliquer des transformations à un point 3D telles que les changements de repères : les rotations, les translations, les homothéties ou encore les projections. Ces transformations sont alors représentées sous la forme de matrices de taille  $4 \times 4$ .*

Dans l'espace cartésien, un point 3D  $M$  exprimé dans le repère Cartésien  $\mathcal{F}_o$ , par  $\mathbf{X}_o = (X_o, Y_o, Z_o)$ , peut être exprimé dans le repère Cartésien  $\mathcal{F}_c$ , donnant  $\mathbf{X}_c = (X_c, Y_c, Z_c)$  par la relation :

$$\mathbf{X}_c = {}^c\mathbf{R}_o \mathbf{X}_o + {}^c\mathbf{t}_o \quad (1.3)$$

Cependant,  $\mathbf{X}_o \in \mathbb{R}^3$  peut également être décrit avec ses coordonnées homogènes  $\overline{\mathbf{X}}_o = (X_o, Y_o, Z_o, 1)$ . Cette représentation nous permet de réécrire la relation (1.3) sous forme d'une équation linéaire comme suit :

$$\overline{\mathbf{X}}_c = {}^c\mathbf{M}_o \overline{\mathbf{X}}_o \quad \text{avec} \quad {}^c\mathbf{M}_o = \begin{bmatrix} {}^c\mathbf{R}_o & {}^c\mathbf{t}_o \\ 0 & 1 \end{bmatrix} \quad (1.4)$$

La matrice  ${}^c\mathbf{M}_o$  est appelée *transformation homogène* de  $\mathcal{F}_o$  vers  $\mathcal{F}_c$ . Le vecteur  ${}^c\mathbf{t}_o = (t_x, t_y, t_z)^\top$  est la translation nécessaire sur chaque axe pour aller du repère  $\mathcal{F}_c$  au repère  $\mathcal{F}_o$ .  $\mathbf{r} = (r_x, r_y, r_z)^\top$  correspond au vecteur unité de l'axe de rotation et  $\theta \in ]-\pi, \pi[$  à l'angle de rotation entre les deux repères. Nous obtenons alors :

$${}^c\mathbf{R}_o(\mathbf{r}, \theta) = \mathbf{I}_3 + [\mathbf{r}]_\times \sin \theta + [\mathbf{r}]_\times^2 (1 - \cos \theta) \quad (1.5)$$

avec  $\mathbf{I}_3$ , la matrice identité de  $\mathbb{R}^3$  et  $[\mathbf{r}]_{\times}$  la matrice antisymétrique associée au vecteur  $\mathbf{r}$  :

$$[\mathbf{r}]_{\times} = \begin{bmatrix} 0 & -r_z & +r_y \\ +r_z & 0 & -r_x \\ -r_y & +r_x & 0 \end{bmatrix} \quad (1.6)$$

Les notations introduites dans cette section sont primordiales pour comprendre le fonctionnement de la commande en asservissement visuel, ce qui sera le sujet de notre prochaine section.

## 1.2 Commande

Il existe deux configurations possibles pour le couple robot-caméra en asservissement visuel :

**Caméra embarquée ("Eye-in-hand")** : Dans cette configuration, la caméra est montée sur l'effecteur du robot (voir FIGURE 1.7(a)).

**Caméra déportée ("Eye-to-hand")** : Dans cette configuration, la caméra est fixe et observe la scène et le robot (voir FIGURE 1.7(b)).

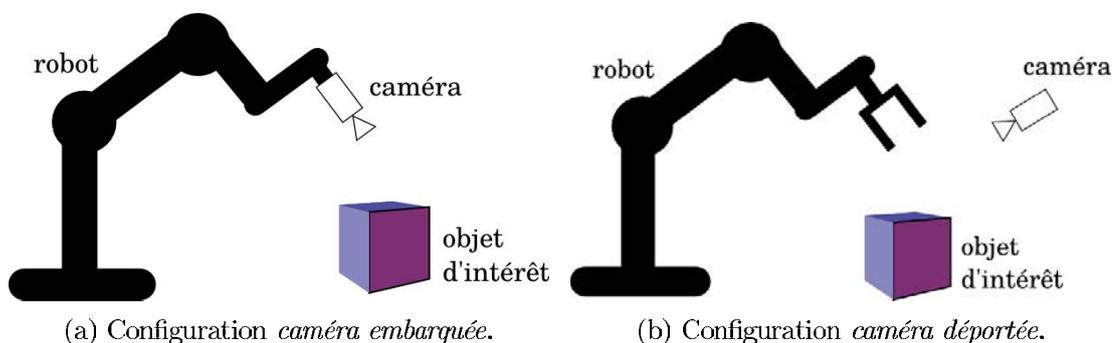


FIGURE 1.7 – Différence entre les configurations *caméra embarquée* et *caméra déportée*. Dans le premier cas le robot fait bouger la caméra qui observe un objet d'intérêt, dans le second cas la caméra est fixe et regarde le robot et l'objet.

En résumé, les données visuelles sont obtenues grâce à une caméra qui est, soit directement montée sur le robot, auquel cas le mouvement du robot implique le déplacement de la caméra, soit fixe observant une scène incluant le robot (en mouvement) et l'objet (en mouvement ou fixe).

### 1.2.1 Primitives visuelles

Pour mettre en œuvre un asservissement visuel, une étape importante consiste à déterminer la relation qui lie la variation temporelle des informations visuelles extraites de l'image aux mouvements du robot ou de la caméra. Le vecteur des primitives visuelles

de la scène à la position désirée est noté  $\mathbf{s}^*$ . Il est extrait sur une image désirée, notée  $\mathbf{I}^*$ . Le vecteur  $\mathbf{s}$  représente celui des primitives visuelles courantes, obtenues à partir de l'image courante  $\mathbf{I}$  de la scène (Chaumette et Hutchinson, 2006). La FIGURE 1.8 montre la structure d'un schéma standard de commande par asservissement visuel. L'objectif est le retour du robot à la position désirée, en utilisant les informations fournies par la caméra utilisée et certaines connaissances sur la scène.

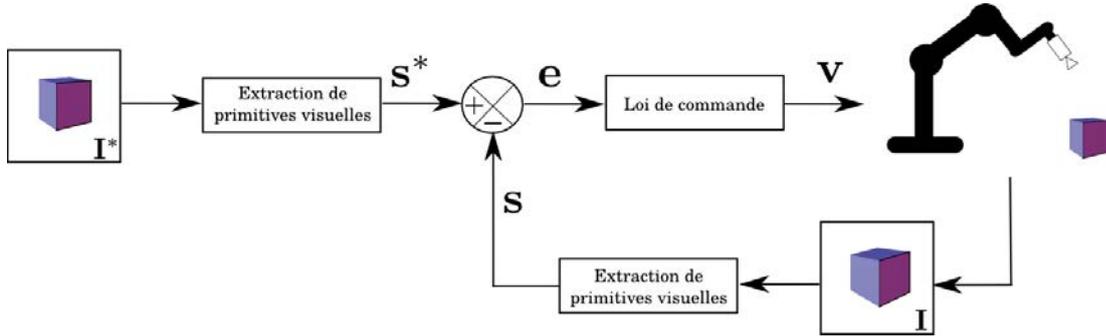


FIGURE 1.8 – Schéma standard de commande d'un asservissement visuel.

Le but de l'asservissement visuel est de minimiser l'erreur  $\mathbf{e}(t)$  entre les primitives visuelles  $\mathbf{s}$  et  $\mathbf{s}^*$ . Cette erreur est définie par :

$$\mathbf{e}(t) = \mathbf{s}(\mathbf{mes}(t), \mathbf{a}) - \mathbf{s}^* \quad (1.7)$$

où le vecteur  $\mathbf{mes}(t)$  représente un ensemble de mesures (souvent des informations géométriques extraites de l'image) et  $\mathbf{a}$  un ensemble de paramètres qui constituent une connaissance supplémentaire sur le système (comme les paramètres intrinsèques de la caméra). La manière dont  $\mathbf{s}$  est définie est liée à un choix de primitives visuelles qui aura une grande influence sur la réussite et les performances de la loi de commande d'asservissement visuel. Il existe trois types d'approches largement étudiées et intimement liées aux choix des primitives visuelles :

**Asservissement visuel 3D** ou "*Position-based visual servoing*" (PBVS) :

Dans cette approche, l'erreur de position de la caméra est calculée dans le repère Cartésien (Allen et al., 1993; Gangloff et al., 1998; Drummond et Cipolla, 1999; Mezouar et al., 2004; Heitzmann et al., 2008). De plus,  $\mathbf{s}$  inclut un ensemble de paramètres 3D, qui sont estimés à partir des données visuelles issues de l'image mais aussi du modèle 3D de l'objet d'intérêt. L'avantage majeur de cette approche est le contrôle de la trajectoire du robot directement dans le repère Cartésien.

Nous notons  $\mathcal{F}_c$  le repère caméra courant,  $\mathcal{F}_{c^*}$  le repère caméra à la position désirée et  $\mathcal{F}_o$  le repère attaché à l'objet d'intérêt. De fait,  $\mathbf{s}$  peut être défini de deux manières. Dans le premier cas,  $\mathbf{s} = ({}^c \mathbf{t}_o, \theta \mathbf{u})$  et  $\mathbf{s}^* = ({}^{c^*} \mathbf{t}_o, 0)$ , la loi de commande offre alors une trajectoire rectiligne dans le repère image. Dans le second cas,

$\mathbf{s} = ({}^c \mathbf{t}_c, \theta \mathbf{u})$  et  $\mathbf{s}^* = \mathbf{0}$ , ici, un découplage entre les mouvements en translation et en rotation est obtenu tandis que la trajectoire de la caméra suit une ligne droite dans le repère monde.

Cependant, l'inconvénient de cette méthode est qu'il n'y a aucun contrôle dans le repère de l'image et les primitives visuelles nécessaires à l'estimation de la pose courante du robot peuvent sortir du champ de vision de la caméra. De plus, cette approche est très sensible aux bruits et aux erreurs d'étalonnage de la caméra et/ou du robot. L'utilisation des asservissements visuels 3D est particulièrement intéressante dans le cas où le capteur visuel utilisé permet d'obtenir des informations d'ordre 3D sur la scène, c'est le cas des caméras stéréoscopiques (Kim et al., 2009; Alkhalil et Doignon, 2012; Doignon, 2015).

### Asservissement visuel 2D ou "*Image-based visual servoing*" (IBVS) :

Dans cette approche,  $\mathbf{s}$  est obtenu directement à partir des données visuelles 2D de l'image (Chaumette et al., 1993; Chesi et al., 2000). Dans ce cas,  $\mathbf{s}$  représente souvent les coordonnées d'un ensemble de points de l'image et  $\mathbf{a}$  les paramètres intrinsèques de la caméra et une estimation de la profondeur  $Z$  des points de l'image. La commande est alors effectuée directement dans l'image et non-plus dans l'espace Cartésien. Les primitives extraites de l'image peuvent être géométriques (points (Wilson et al., 1996), droites (Andreff et al., 2002), moments géométriques d'un contour (Tahri et Chaumette, 2003; Tahri et al., 2015) ou peuvent être globales (analyse en composante principale (Deguchi et Noguchi, 1996), photométrie (Collewet et Marchand, 2011), mélange de Gaussiennes (Crombez et al., 2015), etc). L'avantage de ce type d'asservissement est sa robustesse au bruit et aux erreurs d'étalonnage non seulement de la caméra mais aussi du robot (Espiau, 1994). Une estimation de la profondeur des points de la scène est souvent nécessaire. Néanmoins, seule une connaissance approximative du modèle 3D de l'objet d'intérêt est requise (les travaux de (Malis et al., 2010) montrent que ce n'est pas toujours le cas, en particulier avec les caméras catadioptriques). En revanche, cette approche implique des trajectoires du robot dans l'espace 3D parfois complexes ainsi qu'un risque accru de tomber dans un minimum local (la convergence est assurée d'un point de vue théorique uniquement dans un voisinage de la position désirée). Des exemples de primitives visuelles étudiées pour cette approche seront détaillés dans la section 1.3.

### Asservissement visuel 2D $\frac{1}{2}$ ou "*Hybrid visual servoing*" :

Ici, des informations à la fois de natures 2D et 3D sont utilisées (Malis et Chaumette, 2000; Abdelkader et al., 2005; Hadj-Abdelkader et al., 2006a). Un schéma de contrôle ne nécessitant pas de modèle 3D de l'objet mais assurant une zone de convergence plus importante que celle d'une approche IBVS, est proposé. L'asservissement visuel 2D $\frac{1}{2}$  s'appuie sur l'estimation du déplacement de la caméra à chaque itération afin d'offrir une loi de commande découplée contrôlant les 6

ddl.  $\mathbf{s}$  est alors défini de la manière suivante :  $\mathbf{s} = [\mathbf{s}_t, \theta \mathbf{u}]$  où  $\mathbf{s}_t$  est un ensemble de primitives 2D permettant le contrôle des mouvements de translations et  $\theta \mathbf{u}$  est la rotation entre les poses courante et désirée, typiquement utilisée dans les approches PBVS. Le vecteur des informations visuelles  $\mathbf{s}$  est donc défini sur la base de données 2D et 3D. La loi de commande ainsi synthétisée permet de séparer la rotation de la translation de la caméra, et d'obtenir un fort découplage de la commande. L'avantage majeur de cette approche se trouve dans le fait que la connaissance du modèle géométrique 3D de l'objet n'est plus nécessaire contrairement à la méthode purement 3D.

Une fois le choix de  $\mathbf{s}$  effectué, il est possible de mettre en place une commande en vitesse pour contrôler les déplacements de la caméra ou de l'effecteur du robot. Le torseur de vitesse, noté  $\mathbf{v}$  est constitué de six composantes, trois linéaires  $v_x, v_y, v_z$  et trois angulaires  $\omega_x, \omega_y, \omega_z$ . Le lien entre  $\mathbf{v}$  et  $\mathbf{s}$  est alors obtenu grâce à la *matrice d'interaction*, qui sera abordée ci-dessous.

### 1.2.2 Matrice d'interaction

**Définition 1.2. Matrice d'interaction :** La relation entre la variation temporelle des primitives visuelles  $\dot{\mathbf{s}} \in \mathbb{R}^k$  et la vitesse de la caméra  $\mathbf{v} \in \mathbb{R}^6$  est exprimée par une matrice  $\mathbf{L}_s \in \mathbb{R}^{k \times 6}$ , appelée matrice d'interaction relative à  $\mathbf{s}$ .

$$\dot{\mathbf{s}} = \mathbf{L}_s \mathbf{v} \quad (1.8)$$

La matrice d'interaction nous permet également de relier la vitesse de la caméra à la variation de l'erreur  $\mathbf{e}$ . En effet, en associant les relations (1.7) et (1.8) et en considérant une configuration *caméra embarquée*, l'expression suivante peut être obtenue :

$$\dot{\mathbf{e}} = \mathbf{L}_e \mathbf{v} \quad (1.9)$$

Par ailleurs, dans la configuration *caméra déportée*, la relation (1.9) devient :

$$\dot{\mathbf{e}} = \mathbf{L}_e {}^c \mathbf{V}_r \mathbf{v} \quad (1.10)$$

avec  ${}^c \mathbf{V}_r$  la matrice qui permet d'exprimer le torseur vitesse  $\mathbf{v}$  de l'effecteur (initialement exprimé dans le repère de base) dans le repère caméra :

$${}^c \mathbf{V}_r = \begin{bmatrix} {}^c \mathbf{R}_r & [{}^c \mathbf{t}_r]_{\times} {}^c \mathbf{R}_r \\ \mathbf{0}_3 & {}^c \mathbf{R}_r \end{bmatrix} \quad (1.11)$$

De plus, dans le cas usuel où  $\dot{\mathbf{s}}^* = 0$ , nous pouvons considérer que  $\mathbf{L}_e = \mathbf{L}_s$ .

Afin d'obtenir une décroissance exponentielle de l'erreur visuelle  $\dot{\mathbf{e}} = -\lambda \mathbf{e}$ , le torseur de vitesses est obtenu (pour la configuration *caméra embarquée*) de la manière suivante :

$$\mathbf{v} = -\lambda \mathbf{L}_s^+ \mathbf{e} \quad (1.12)$$

où  $\mathbf{L}_s^+$  est la pseudo-inverse de Moore-Pensore de  $\mathbf{L}_s$  (voir définition ci-après).

**Définition 1.3. Pseudo-inverse de Moore-Penrose :** La pseudo inverse  $\mathbf{A}^+ \in \mathbb{R}^{m \times k}$  d'une matrice  $\mathbf{A} \in \mathbb{R}^{k \times m}$  de rang  $r$  est la seule matrice vérifiant les conditions suivantes :

- $\mathbf{A}\mathbf{A}^+\mathbf{A} = \mathbf{A}$
- $\mathbf{A}^+\mathbf{A}\mathbf{A}^+ = \mathbf{A}^+$  ( $\mathbf{A}^+$  est un inverse pour le semi-groupe multiplicatif);
- $(\mathbf{A}\mathbf{A}^+)^\top = \mathbf{A}\mathbf{A}^+$  ( $\mathbf{A}\mathbf{A}^+$  est une matrice hermitienne);
- $(\mathbf{A}^+\mathbf{A})^\top = \mathbf{A}^+\mathbf{A}$  ( $\mathbf{A}^+\mathbf{A}$  est également hermitienne).

En pratique,

si  $k = m = r$ , alors  $\mathbf{A}^+ = \mathbf{A}^{-1}$ ,

si  $k > m$  et  $r = m$  alors  $\mathbf{A}^+ = (\mathbf{A}^\top \mathbf{A})^{-1} \mathbf{A}^\top$  et  $\mathbf{A}^+ \mathbf{A} = \mathbf{I}_m$ ,

si  $m > k$  et  $r = k$  alors  $\mathbf{A}^+ = \mathbf{A}^\top (\mathbf{A}^\top \mathbf{A})^{-1}$  et  $\mathbf{A}\mathbf{A}^+ = \mathbf{I}_k$ ,

sinon,  $\mathbf{A}^+$  est obtenue grâce à une décomposition en valeurs singulières (SVD).

Lorsque la matrice  $\mathbf{L}_s$  est carrée et de rang plein, la loi de commande devient :

$$\mathbf{v} = -\lambda \mathbf{L}_s^{-1} \mathbf{e} \quad (1.13)$$

où  $\mathbf{L}_s^{-1}$  est la matrice inverse de  $\mathbf{L}_s$ .

Néanmoins, dans des systèmes réels, il est difficile de connaître parfaitement la matrice  $\mathbf{L}_s$  alors, souvent, nous nous contentons d'utiliser une approximation de  $\mathbf{L}_s$ , notée  $\widehat{\mathbf{L}}_s$  (Hutchinson et al., 1996). La plupart des lois de commande par asservissement visuel sont construites sur ce modèle. Une seconde approche, consiste à utiliser une matrice d'interaction constante calculée à la position désirée  $\mathbf{L}_s^*$  (Espiau et al., 1992). Enfin, les deux approches peuvent être combinées en utilisant  $\widehat{\mathbf{L}}_s^+ = \frac{1}{2}(\widehat{\mathbf{L}}_s + \mathbf{L}_s^*)^+$ .

Il est également possible de modifier la relation (1.12) en utilisant une optimisation de type *Levenberg-Marquardt* (Marquardt, 1963). La loi de commande devient alors :

$$\mathbf{v} = -\lambda (\mathbf{H} + \mu \text{diag}(\mathbf{H}))^{-1} \mathbf{L}_s^\top (\mathbf{s}(t) - \mathbf{s}^*) \quad (1.14)$$

avec  $\mathbf{H} = \mathbf{L}_s^\top \mathbf{L}_s$ , la matrice Hessienne de  $\mathbf{s}$ . Lorsqu'une valeur élevée pour  $\mu$  ( $\mu = 1$ ) est utilisée, l'optimisation correspond à un algorithme de descente de gradient, entraînant des déplacements plus importants lorsque le gradient est faible tandis qu'une valeur faible de  $\mu$  ( $\mu = 10^{-3}$ ) revient à utiliser la technique de type *Gauss-Newton* présentée en (1.12).

### 1.2.3 Analyse de la stabilité

En théorie du contrôle, la notion de stabilité de Lyapunov fait référence pour l'étude du comportement des systèmes dynamiques. L'idée de Lyapunov consiste à dire que si tous les points d'un système démarrent autour d'un point  $P$  et que tous ces points restent autour de  $P$ , alors ce dernier est stable au sens de Lyapunov (Lyapunov, 1992). La méthode de Lyapunov peut être utilisée pour étudier la stabilité d'un système en boucle fermée (Chaumette et Hutchinson, 2006). Pour cela, la fonction de Lyapunov  $\mathcal{L}$  peut être définie comme suit :

$$\mathcal{L} = \frac{1}{2} \|\mathbf{e}(t)\|^2 \quad (1.15)$$

Les conditions permettant à  $\mathcal{L}$  de décroître en fonction du temps sont alors recherchées en calculant la dérivée de  $\mathcal{L}$  :

$$\dot{\mathcal{L}} = \mathbf{e}^\top \dot{\mathbf{e}} \quad (1.16)$$

$$= -\lambda \mathbf{e}^\top \mathbf{L}_s \widehat{\mathbf{L}}_s^+ \mathbf{e} \quad (1.17)$$

Ainsi, pour obtenir la stabilité globale du système, il faut vérifier la condition suivante :

$$\mathbf{L}_s \widehat{\mathbf{L}}_s^+ > 0. \quad (1.18)$$

Dans le cas IBVS, la dimension  $k$  de  $\mathbf{s}$  est souvent supérieure à 6, c'est-à-dire qu'au moins trois primitives visuelles 2D sont considérées,  $\mathbf{L}_s \widehat{\mathbf{L}}_s^+ \in \mathbb{R}^{k \times k}$  peut avoir un rang supérieur à 6 et donc des noyaux non nuls. Les configurations dans lesquelles  $\mathbf{e} \in \ker \mathbf{L}_s^+$  correspondent alors à des minima locaux.

**Définition 1.4. Noyau d'une application linéaire :** Si  $f$  est une application linéaire d'un espace vectoriel  $V$  dans un espace vectoriel  $W$ , alors le noyau de  $f$  est défini par

$$\ker(f) = \{x \in V \mid f(x) = 0_W\} \quad (1.19)$$

où  $0_W$  est l'élément neutre de l'espace vectoriel  $W$ .

La condition (1.18) permet alors seulement de garantir la stabilité asymptotique locale au voisinage de la configuration désirée. En pratique, ce voisinage est souvent assez large. Bien que cette relation paraisse assez simple, son étude reste encore un domaine d'investigation de la communauté de l'asservissement visuel.

Dans le cas PBVS, la stabilité asymptotique globale est obtenue dès lors que  $\mathbf{L}_s \widehat{\mathbf{L}}_s^{-1} = \mathbf{I}_6$  sous la contrainte forte que les coordonnées des différents repères (robot, caméra, objet) soient parfaitement corrects.

## 1.3 Exemples d'asservissements visuels 2D

La notion de primitives visuelles a été brièvement présentée dans la section précédente et le lecteur a déjà pu constater l'importance du choix des primitives en asservissement visuel. Cette section vise à présenter différentes primitives possibles. Dans

le cadre de cette thèse, nous présenterons des primitives visuelles de type 2D. Nous pouvons les classer en deux grandes catégories :

- Les primitives géométriques qui sont extraites de l'images par des procédés de traitement d'image complexes (détection, suivi et mise en correspondance, etc.)
- Les primitives directes qui peuvent être obtenues avec des algorithmes de traitement d'images globaux mais sans étape de suivi visuel.

### 1.3.1 Approches géométriques

Les approches géométriques sont les premières ayant été développées en asservissement visuel. Elles consistent à segmenter l'image pour y détecter des formes géométriques connues (points, segments, ellipses, etc) ou même des contours de forme complexe pour construire la matrice d'interaction associée. Cette section présentera l'élaboration de cette dernière à partir de différentes primitives visuelles.

**Points :** La primitive visuelle la plus simple à considérer est sans doute le point (Chaumette et Hutchinson, 2006). En se mettant dans le cadre d'un asservissement visuel en configuration *caméra embarquée*, un point 3D  $\mathbf{X}$  de coordonnées  $(X_c, Y_c, Z_c)$  exprimé dans le repère caméra  $\mathcal{F}_c$  est projeté dans le plan image  $\mathcal{P}$  en un point 2D, défini de la manière suivante :

$$x = X_c/Z_c = (u - u_0)/p_x \quad (1.20)$$

$$y = Y_c/Z_c = (v - v_0)/p_y \quad (1.21)$$

En considérant  $\mathbf{X}$  comme une primitive visuelle, l'expression suivante peut être obtenue :

$$\dot{\mathbf{X}} = \mathbf{L}_p \mathbf{v} \quad (1.22)$$

et la matrice d'interaction  $\mathbf{L}_p$  relative au point  $\mathbf{X}$  est donnée par :

$$\mathbf{L}_p = \begin{bmatrix} -1/Z_c & 0 & x/Z_c & xy & -(1+x^2) & y \\ 0 & -1/Z_c & y/Z_c & 1+y^2 & -xy & -x \end{bmatrix} \quad (1.23)$$

La valeur de  $Z_c$  correspond à la profondeur du point par rapport au repère caméra. Une approximation de sa valeur est souvent utilisée. De plus, il faut au moins trois points pour définir une commande à 6 ddl. Pour cela, une concaténation de trois matrices  $\mathbf{L}_{p1}$ ,  $\mathbf{L}_{p2}$  et  $\mathbf{L}_{p3}$ , correspondant à trois points  $\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3$  est réalisée pour construire la matrice d'interaction associée (Michel et Rives, 1993). Cependant, plus de trois points sont en général utilisés pour obtenir de la redondance d'informations visuelles dans la commande, comme le montre la FIGURE 1.9, issue d'un exemple de la librairie ViSP (Visual Servoing Platform) (Marchand et al., 2005).

D'autres approches géométriques utilisant des primitives visuelles de formes simples ont été développées en considérant des segments, des droites, des sphères, des cercles



(a) Début du positionnement    (b) Positionnement en cours    (c) Fin du positionnement

FIGURE 1.9 – Exemple d'une tâche d'asservissement visuel avec 4 points.

ou encore des cylindres. Ces dernières sont présentées dans (Chaumette et al., 1993). Même si généralement les coordonnées cartésiennes sont utilisées dans la modélisation de la matrice d'interaction, des coordonnées cylindriques peuvent être également utilisées (Iwatsuki et Okiyama, 2005) ou encore polaires comme dans (Hadj-Abdelkader et al., 2006b). Enfin, certaines approches ont utilisé des moments géométriques en tant qu'informations visuelles (Chaumette, 2004; Mebarki et al., 2010; Tahri et al., 2015). Cette dernière approche fera l'objet du prochain paragraphe.

**Moments géométriques :** Développée pour la première fois en 2003 par Tahri (Tahri et Chaumette, 2003), l'approche utilisant les moments géométriques a été améliorée par la suite dans (Tamtsia et al., 2013) et (Tahri et al., 2015).

**Définition 1.5. Moments 2D d'ordre  $i + j$  :** Soit  $S$  la surface couverte par un objet sur une image, alors les moments d'ordre  $i + j$  de  $S$  ( $i, j \in \mathbb{N}$ ) sont donnés par :

$$m_{i,j} = \int \int_S x^i y^j dx dy \quad (1.24)$$

où  $(x, y)$  sont les coordonnées des points de  $S$ .

Le moment d'ordre 0 ( $m_{0,0}$ ) de  $S$  correspond à son aire et ceux d'ordre 1 ( $m_{0,1}, m_{1,0}$ ) sont reliés aux coordonnées du barycentre  $(x_g, y_g)$  de  $S$  par la relation :

$$x_g = \frac{m_{1,0}}{m_{0,0}} \quad \text{et} \quad y_g = \frac{m_{0,1}}{m_{0,0}} \quad (1.25)$$

L'idée est alors d'utiliser des primitives visuelles construites à partir de moments de différents ordres et de formuler la matrice d'interaction associée afin de réaliser la tâche d'asservissement visuel souhaitée (voir FIGURE 1.10). Des combinaisons de certains moments, pouvant être centrés, sont définis afin d'être invariants à certaines transformations (translation, rotation, changement d'échelle). Dans Tahri et Chaumette (2003), le vecteur  $\mathbf{s}$  est choisi de la manière suivante :

$$\mathbf{s} = [x_n, y_n, a_n, s_4, s_5, \theta] \quad (1.26)$$

où  $a_n = Z^* \frac{m_{0,0}^*}{m_{0,0}}$ ,  $x_n = a_n x_g$ ,  $y_n = a_n y_g$ ,  $\theta$  est l'orientation principale de l'objet et  $(s_4, s_5)$  sont deux invariants aux translations, aux rotations et aux changements d'échelle servant

à contrôler les rotations selon les axes  $x$  et  $y$ . Le choix des primitives  $(s_4, s_5)$ , calculées avec des moments d'ordre 3 à 5, dépendent de la forme de l'objet d'intérêt. Ce point constitue son principal inconvénient.

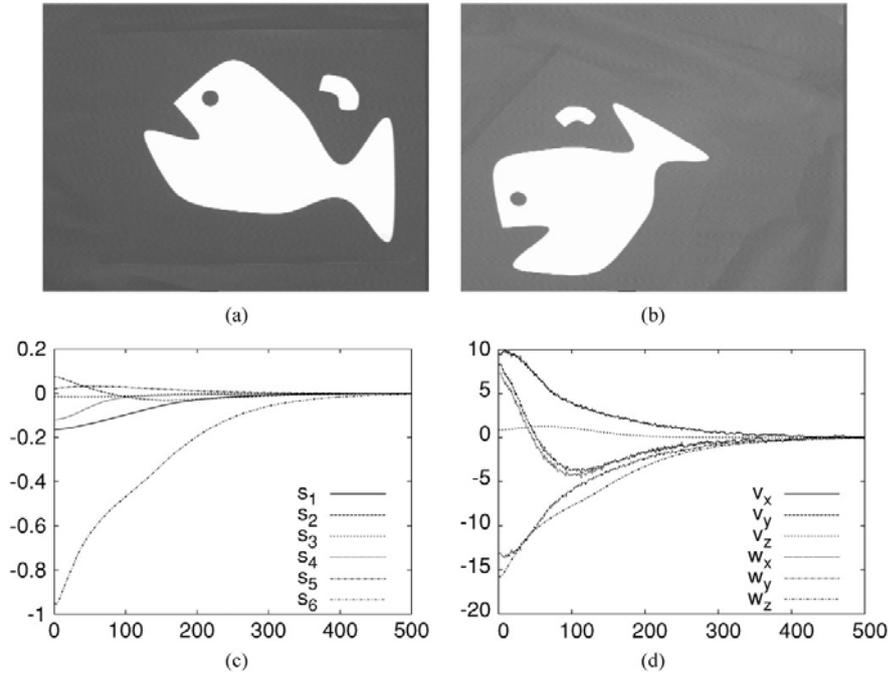


FIGURE 1.10 – Asservissement visuel utilisant les moments géométriques : (a) image désirée (b) image initiale (c) erreur sur chaque composante de  $s$  (d) vitesse appliquée à la caméra (cm/s et °/s) (Chaumette, 2004).

Une nouvelle sorte de moments, appelés "moments décalés" (*shifted moments*) est utilisée dans (Tahri et al., 2015) pour définir  $(s_4, s_5)$ .

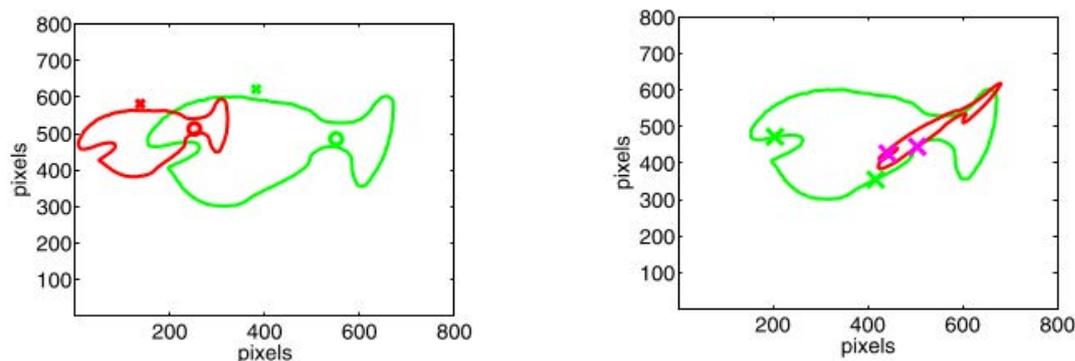
**Définition 1.6. Moments décalés 2D d'ordre  $i + j$  :** Soit  $S$  la surface couverte par un objet sur une image, alors les moments décalés d'un vecteur  $(x_g - x_{sh}, y_g - y_{sh})$  d'ordre  $i + j$  de  $S$  ( $i, j \in \mathbb{N}$ ) sont donnés par :

$$\mu_{i,j} = \int \int_S (x - x_g + s_{sh})^i (y - y_g + y_{sh})^j dx dy \quad (1.27)$$

où  $(x, y)$  sont les coordonnées des points de  $S$  et  $(x_g, y_g)$  est son centre de gravité.

L'utilisation des moments décalés permet d'éviter l'obtention de moments nuls dus aux éventuelles symétries de l'objet. Néanmoins, pour préserver les invariances aux transformations (translations, rotations et changements d'échelle) les points de décalage  $(x_{sh}, y_{sh})$  doivent être choisis avec précaution. Deux choix sont considérés, le premier fondé sur les moments de l'objet, le deuxième fondé sur des points du contour de l'objet (voir FIGURE 1.11). Ce deuxième choix impose la détection et le suivi de deux points

mais permet des performances exceptionnelles pour une approche IBVS. De plus, dans les deux cas, l'utilisation de ces moments décalés permet d'obtenir une définition pour  $(s_4, s_5)$  quelque soit la forme de l'objet d'intérêt.



(a) Vecteur de décalage fondé sur les moments.

(b) Vecteur de décalage fondé sur des points du contour.

FIGURE 1.11 – Illustration de la méthode des moments décalés avec les deux manières de définir le vecteur de décalage. La position initiale est représentée en rouge, la position désirée en vert. La loi de commande a permis la convergence de manière très précise dans les deux cas (Tahri et al., 2015).

Ces techniques s'appuyant sur des primitives géométriques ont l'avantage de posséder un domaine de convergence assez large. Cependant, elles reposent sur des algorithmes de détection, de sélection, de suivi et de mise en correspondance de ces informations visuelles. Cela impose de disposer d'une image qui possède des entités géométriques bien définies. De plus, l'efficacité de ces méthodes est directement dépendante de la qualité des algorithmes de suivi visuel utilisés. Afin de s'affranchir des algorithmes de suivi visuel, d'autres formes de primitives ont été considérées pour introduire la notion d'asservissement visuel.

### 1.3.2 Approches directes

Les approches d'asservissement visuel direct utilisent, non pas des formes ou des points d'intérêt pour réaliser une tâche de positionnement, mais plutôt toute l'information contenue dans l'image, on parle alors d'informations globales de l'image. Les premières approches directes sont apparues en 1996 avec des méthodes reposant sur l'analyse en composantes principales (Nayar et al., 1996; Deguchi et Noguchi, 1996) permettant de contrôler 2 ddl en réduisant le dimensionnement des données consistant l'information visuelle. En 2007, d'autres méthodes sont apparues avec d'une part une méthode utilisant uniquement les homographies (Benhimane et Malis, 2007) et d'autre part une méthode fondée sur les noyaux de convolution (Kallem et al., 2007). En 2008, l'asservissement visuel fondé sur l'intensité des pixels de toute l'image, associée à la

contrainte du flux optique, a été mise en œuvre (Collewet et al., 2008; Collewet et Marchand, 2011). Il faut noter que cette contrainte du flux optique stipule que l'intensité d'un point image est indépendante du temps, ce qui n'est pas toujours le cas lors du changement des conditions d'illumination dans la scène par exemple. D'autres approches plus robustes à ce type de conditions défavorables sont apparues utilisant les gradients de l'image (Marchand et Collewet, 2010), l'information mutuelle (Dame et Marchand, 2011), la somme de variances conditionnelles d'une image (Richa et al., 2011; Delabarre et Marchand, 2012), les modèles de mélanges de Gaussiennes (Crombez et al., 2015) ou encore plus récemment les histogrammes (Bateux et Marchand, 2017). Une technique utilisant les moments photométriques a également été mise en œuvre dans (Bakthavatchalam et al., 2013). Enfin, l'utilisation récente de méthodes d'apprentissage de type réseaux de neurones a été étudiée dans (Bateux et al., 2017).

Dans cette section nous détaillerons essentiellement la méthode photométrique et quelques approches dérivées de cette dernière.

**Photométrie :** Dans l'asservissement visuel photométrique, introduit dans (Collewet et Marchand, 2011), la primitive visuelle utilisée est constituée d'un ensemble fini d'intensités lumineuses des  $N$  pixels de l'image  $(x_n, y_n)_{n \leq N} \subset \mathbb{R}^2$  à l'instant  $t$  :

$$\mathbf{s}_{\text{ph}}(t) = \left( I(x_1, y_1, t), I(x_2, y_2, t), \dots, I(x_N, y_N, t) \right)^\top. \quad (1.28)$$

L'image  $\mathbf{I}$  est composée de valeurs d'intensité  $I(\mathbf{x}, t)$ , avec  $\mathbf{x} \in \mathbb{Z}^2$  les coordonnées spatiales de ces valeurs.

Pour atteindre une position désirées  $\mathbf{s}_{\text{ph}}^*$ , le robot doit annuler progressivement l'erreur définie par :

$$\mathbf{e}_{\text{ph}}(t) = \mathbf{s}_{\text{ph}}(t) - \mathbf{s}_{\text{ph}}^* \quad (1.29)$$

Pour ce faire, il est nécessaire de relier la variation temporelle des intensités des pixels de l'image aux mouvements de la caméra. Dans (Collewet et Marchand, 2011), les auteurs établissent cette relation grâce à une matrice d'interaction  $\mathbf{L}_{\text{sph}}$  définie comme :

$$\frac{d\mathbf{s}_{\text{ph}}(t)}{dt} = \mathbf{L}_{\text{sph}}(t)\mathbf{v}(t). \quad (1.30)$$

Sous réserve de respecter la continuité temporelle de la luminance introduite dans (Horn et Schunck, 1981), et en notant  $\nabla I_t(x, y)$  la dérivée spatiale de l'image  $\mathbf{I}$  en chaque point de l'image de coordonnées  $(x, y)$  à l'instant  $t$ , l'équation de contrainte du flux optique peut s'écrire :

$$\frac{\partial I(x, y, t)}{\partial t} = -\left( \nabla I_t(x, y) \right)^\top \left( \frac{dx}{dt}, \frac{dy}{dt} \right)^\top. \quad (1.31)$$

En utilisant le vecteur de vitesse  $\mathbf{v}$  de la caméra, la relation (1.31) peut s'écrire

$$\frac{\partial I(x, y, t)}{\partial t} = -\left( \nabla I_t(x, y) \right)^\top \mathbf{L}_{\text{p}}(x, y)\mathbf{v}(t), \quad (1.32)$$

où  $\mathbf{L}_p(x, y)$  représente la matrice d'interaction d'un point 2D, défini en (1.23).

Les relations (1.30) et (1.32) permettent alors de calculer la matrice d'interaction  $\mathbf{L}_{\text{sph}}$  pour tous points-pixels à l'instant  $t$  :

$$\mathbf{L}_{\text{sph}}(t) = - \begin{bmatrix} \left( \nabla I_t(x_1, y_1) \right)^\top \mathbf{L}_p(x_1, y_1) \\ \vdots \\ \left( \nabla I_t(x_N, y_N) \right)^\top \mathbf{L}_p(x_N, y_N) \end{bmatrix} \in \mathbb{R}^{N \times 6} \quad (1.33)$$

Finalement, le vecteur de vitesses de la caméra  $\mathbf{v}$  est obtenu grâce à une optimisation de type *Levenberg-Marquardt* :

$$\mathbf{v} = -\lambda(\mathbf{H} + \mu \text{diag}(\mathbf{H}))^{-1} \mathbf{L}_{\text{sph}}^\top \mathbf{e}_{\text{ph}} \quad (1.34)$$

où  $\lambda > 0$  est un gain positif,  $\mu$  est un facteur de poids.

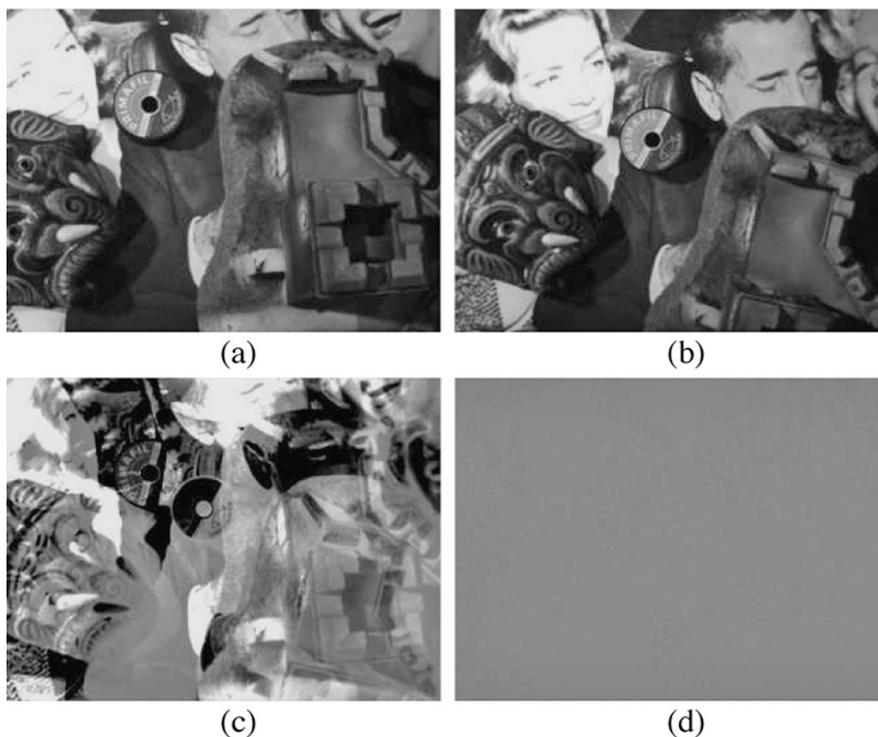


FIGURE 1.12 – Expérience de robustesse à l'approximation de la profondeur  $Z_c$  de l'image ; (a) image initiale (b) image finale (c)  $\mathbf{I} - \mathbf{I}^*$  à la position initiale (d)  $\mathbf{I} - \mathbf{I}^*$  à la position finale (Collewet et Marchand, 2011).

Grâce à l'utilisation d'informations redondantes et d'une absence d'étapes de traitement d'images complexes (détection, segmentation, mise en correspondance, etc), cette méthode s'avère d'une grande précision. Dans (Collewet et Marchand, 2011), elle est

comparée à un asservissement visuel géométrique utilisant une extraction de points d'intérêt par un descripteur de type SURF. Elle a démontré une précision dix fois supérieure à l'asservissement visuel 2D standard, ceci dans des conditions nominales. Cette méthode s'avère également robuste aux occultations et ne nécessite pas d'estimation précise de la profondeur  $Z_c$  lors de la tâche de positionnement (voir FIGURE 1.12). Cependant, la non-linéarité de la relation (1.33) impose l'utilisation d'une méthode d'optimisation non-linéaire (ici la méthode *Levenberg-Marquardt*).

**Information mutuelle :** L'asservissement visuel utilisant l'information mutuelle est une des premières méthodes directes robuste aux occultations et aux changements d'illumination. Pour cela, l'entropie de Shannon est utilisée, issue de la théorie de l'information (Shannon, 1948).

**Définition 1.7. Entropie de Shannon :** L'entropie  $H(\mathbf{I})$  est une mesure de variabilité d'une variable aléatoire, ici une image notée  $\mathbf{I}$ . Si  $i$  est une valeur possible de  $\mathbf{I}(i)$  ( $i \in [0, 255]$ ) (l'intervalle  $[0, 255]$  représentant les niveaux de gris d'un pixel de l'image) et  $p_{\mathbf{I}}(i)$  est la distribution de probabilité de  $i$ , alors l'entropie de Shannon de la variable  $\mathbf{I}$  est donnée par :

$$H(\mathbf{I}) = - \sum_{i=0}^{255} p_{\mathbf{I}}(i) \log(p_{\mathbf{I}}(i)) \quad (1.35)$$

**Définition 1.8. Entropie conjointe :** L'entropie conjointe  $H(\mathbf{I}, \mathbf{I}^*)$  entre deux variables, ici l'image courante  $\mathbf{I}$  et l'image désirée  $\mathbf{I}^*$ , représente la variabilité entre ces deux variables. L'entropie conjointe est alors définie par :

$$H(\mathbf{I}, \mathbf{I}^*) = - \sum_{i=0}^{255} \sum_{j=0}^{255} p_{\mathbf{I}\mathbf{I}^*}(i, j) \log(p_{\mathbf{I}\mathbf{I}^*}(i, j)) \quad (1.36)$$

où  $p_{\mathbf{I}\mathbf{I}^*}(i, j)$  est la distribution de probabilité conjointe de  $i$  et  $j$ .

**Définition 1.9. Information mutuelle :** L'information mutuelle entre les images  $\mathbf{I}$  et  $\mathbf{I}^*$  est alors donnée par :

$$MI(\mathbf{I}, \mathbf{I}^*) = H(\mathbf{I}) + H(\mathbf{I}^*) - H(\mathbf{I}, \mathbf{I}^*) \quad (1.37)$$

Il est alors possible d'obtenir l'information mutuelle entre l'image courante et l'image désirée en utilisant la formule :

$$MI(t) = \sum_{ij} p_{\mathbf{I}\mathbf{I}^*}(i, j) \log \left( \frac{p_{\mathbf{I}\mathbf{I}^*}(i, j)}{p_{\mathbf{I}}(i)p_{\mathbf{I}^*}(j)} \right) \quad (1.38)$$

La FIGURE 1.13 compare les fonctions de coût de la méthode photométrique et de l'information mutuelle. Ces fonctions sont définies par la minimisation de  $\mathcal{C} = \frac{1}{2} \|\mathbf{e}\|^2$  dans le cas de la photométrie et la maximisation de  $MI(\mathbf{I}, \mathbf{I}^*)$  dans le cas de l'information mutuelle.

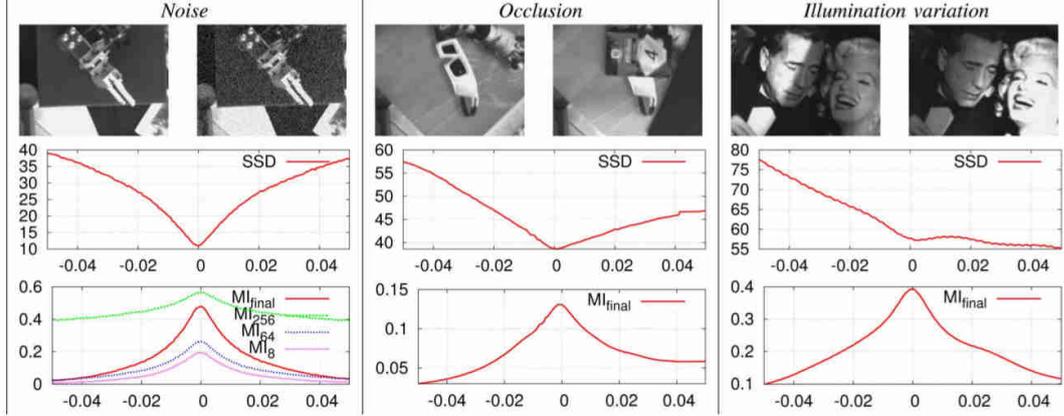


FIGURE 1.13 – Évaluation de l'asservissement visuel utilisant l'information mutuelle  $MI_i$  en présence de bruit, d'occultations et de changements d'illumination, en utilisant  $i$  intervalles de niveaux de gris. La deuxième et la troisième ligne représentent, respectivement, la fonction de coût photométrique et celle de l'information mutuelle (Dame et Marchand, 2011).

Ainsi, l'information mutuelle permet de mesurer la similarité entre deux images et peut être appliquée à des modalités d'imagerie différentes du moment qu'il existe un recouvrement suffisant entre l'image courante et l'image désirée. La FIGURE 1.14, montre un asservissement visuel associé à une planification de trajectoire où l'image courante est obtenue par une caméra monoculaire tandis que l'image désirée provient d'une carte géographique. Néanmoins, un inconvénient de cette méthode est qu'elle nécessite l'utilisation de courbes de régression (B-spline) pour calculer la matrice d'interaction.

**Somme des variances conditionnelles :** La somme des variances conditionnelles d'une image est une autre méthode directe reposant sur la théorie de l'information. Élaborée pour des tâches de suivi par (Richa et al., 2011), elle fut ensuite étendue à des tâches de positionnement dans (Delabarre et Marchand, 2012).

**Définition 1.10. Somme des variances conditionnelles d'une image :** La somme des variances conditionnelles d'une image (SVC) est définie de la manière suivante :

$$SCV = \sum_{\mathbf{x}} \left( \mathbf{I}(\mathbf{x}) - \hat{\mathbf{I}}(\mathbf{x}) \right) \quad (1.39)$$

Le terme  $\hat{\mathbf{I}}$  est obtenu grâce à l'espérance  $\mathcal{E}$  calculée grâce à une distribution de probabilité jointe entre  $\mathbf{I}$  et  $\mathbf{I}^*$ .

$$\hat{\mathbf{I}} = \mathcal{E}(\mathbf{I}(\mathbf{x}) | \mathbf{I}^*(\mathbf{x})) \quad (1.40)$$

Cette méthode consiste à ne plus considérer les pixels de l'image comme primitives visuelles mais les variances conditionnelles pour chaque pixel. La matrice d'interaction

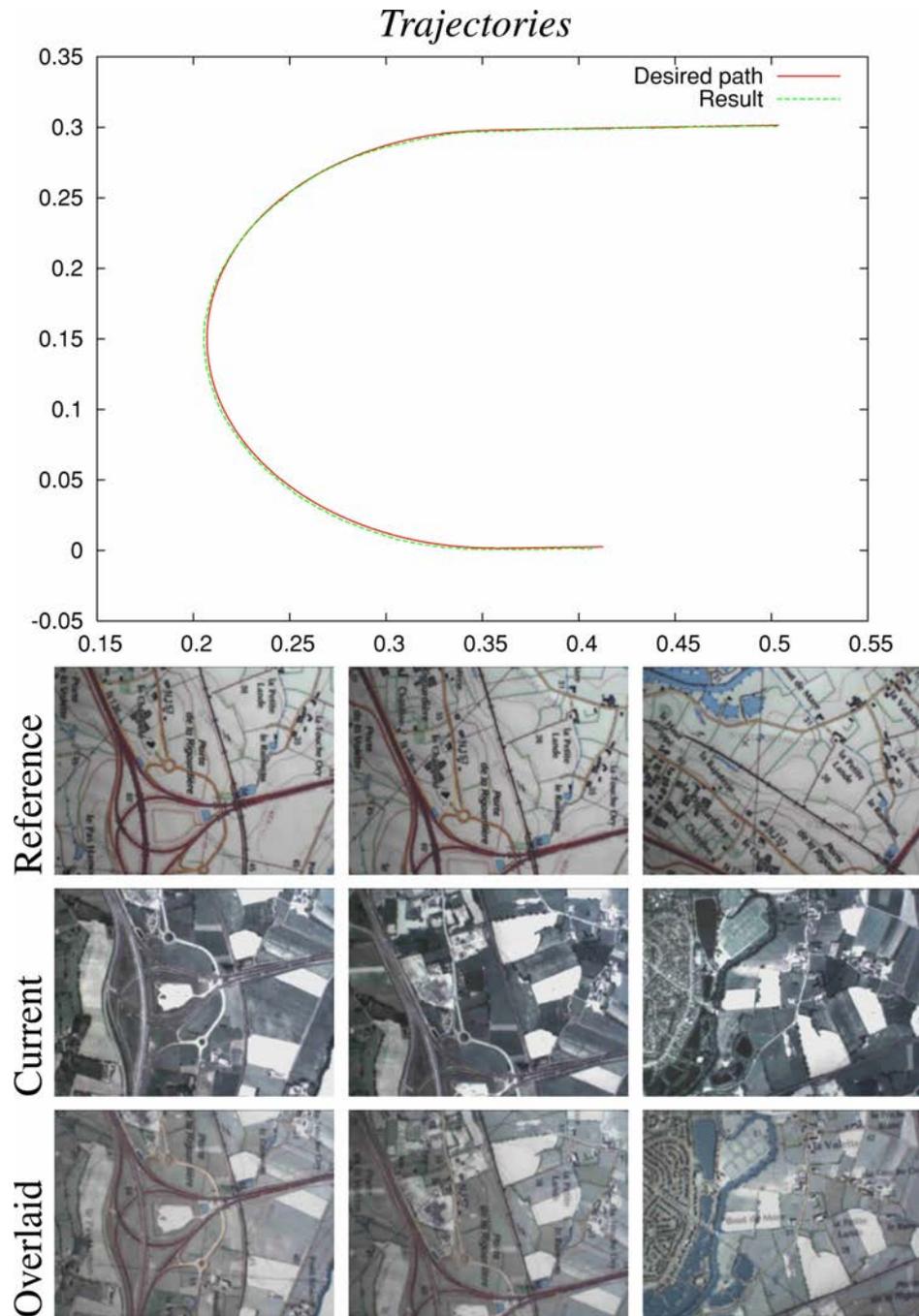


FIGURE 1.14 – Exemple d'asservissement visuel utilisant l'information mutuelle dans une tâche de navigation. Un chemin est appris avec une carte géographique et la tâche d'asservissement est réalisée sur des images de cartographies issues d'images satellites. (Dame et Marchand, 2011).

utilisée par cette méthode ne nécessite que le calcul du gradient de l'image, ce qui la rend plus simple d'utilisation que la méthode reposant sur l'information mutuelle. D'un autre côté, elle s'avère peu robuste aux changements locaux d'illumination.

**Mélanges de Gaussiennes :** Une autre méthode directe se démarque par ses performances, il s'agit de l'approche de mélanges de Gaussiennes (Crombez et al., 2015) dans laquelle chaque pixel  $\mathbf{u} = (u, v)$  est représenté par une distribution Gaussienne, de la forme :

$$g(\mathbf{u}, \mathbf{u}_0, I_u, \lambda_g) = \exp\left(-\frac{(u - u_0)^2 + (v - v_0)^2}{2\lambda_g^2 I_u^2}\right) \quad (1.41)$$

où  $\mathbf{u}_0 = (u_0, v_0)$  représente les coordonnées du centre de chaque Gaussienne calculée dans l'image et  $\lambda_g$  est un paramètre contrôlant l'extension de la Gaussienne. L'effet du paramètre  $\lambda_g$  sur le mélange de Gaussiennes est visible sur le FIGURE 1.15. Le mélange de Gaussiennes en un point  $\mathbf{u}$  est alors donné par :

$$gm(\mathbf{u}, \mathbf{I}) = \sum_x g(\mathbf{u}, \mathbf{u}_0, I_u, \lambda_g) \quad (1.42)$$

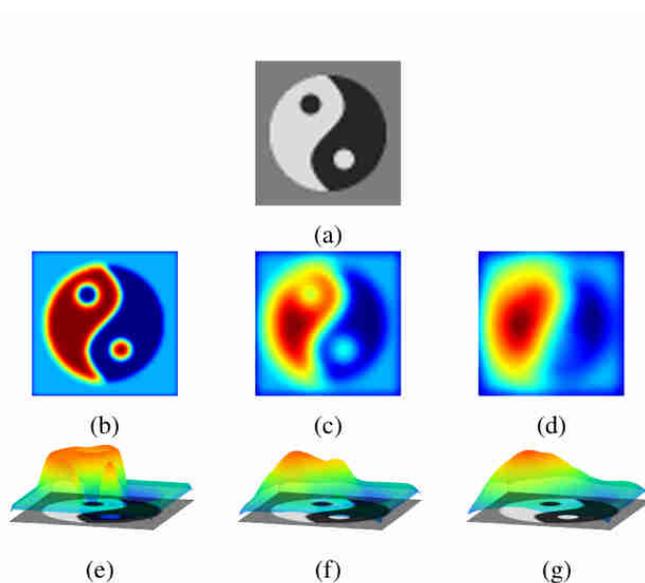


FIGURE 1.15 – Influence de  $\lambda_g$  (a) image en niveaux de gris (b) représentation 2D et (e) 3D du mélange de Gaussiennes pour  $\lambda_g = 0.01$ , (c)-(f) pour  $\lambda_g = 0.03$  et (d)-(g) pour  $\lambda_g = 0.05$  (Crombez et al., 2015).

Cette approche présente un large domaine de convergence et permet de faire un asservissement visuel même lorsqu'il y a un faible recouvrement photométrique entre l'image courante et l'image désirée (voir FIGURE 1.16).

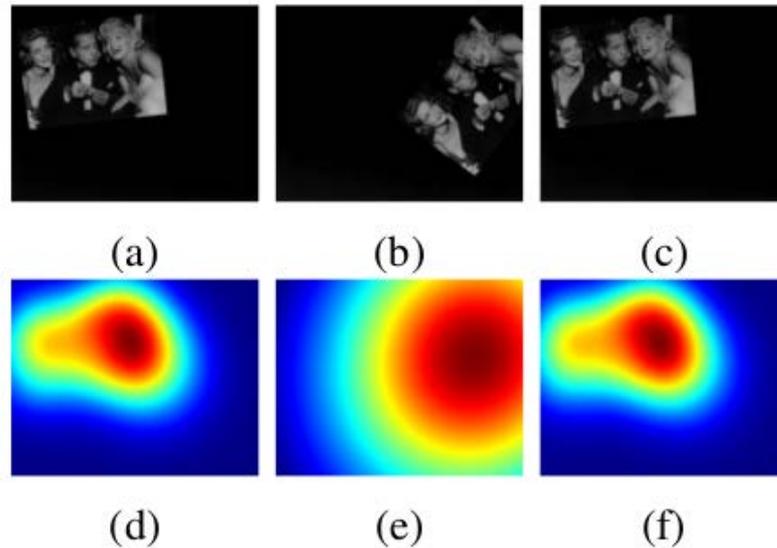


FIGURE 1.16 – Image et mélange de Gaussiennes correspondantes, (a)-(d) à la position désirée, (b)-(e) à la position initiale, (c)-(f) à la position finale. (Crombez et al., 2015).

Certaines approches ont utilisé des informations fréquentielles et non temporelles dans la boucle de commande. L'apport du domaine fréquentiel a été mis à profit en recalage d'image grâce au travail de Reddy et Chatterji (Reddy et Chatterji, 1996) dès le milieu des années 1990. Son utilisation en asservissement visuel a été décrite ultérieurement pour réaliser une tâche de positionnement avec une approche s'appuyant sur les noyaux de convolution et la transformée de Fourier. Ce sera le sujet du prochain paragraphe.

**Méthode à noyaux :** Les méthodes utilisant la transformée de Fourier en asservissement visuel sont de deux types. Les premières utilisent la transformée de Fourier pour paramétrer un contour, et sont dans ce cas des méthodes géométriques (Collewet, 1999; Li et al., 2011). Les secondes utilisent la transformée de Fourier en tant qu'information globale. C'est notamment le cas de la méthode à noyaux de convolution (Kallem et al., 2007) ou *Kernel-based Visual Servoing*. Cette approche utilise la projection de l'image sur un noyau de convolution  $K$  pour déterminer les déplacements à appliquer au robot. Néanmoins, il faut noter que dans ce travail uniquement 4 ddl sont contrôlés (les trois translations  $(t_x, t_y, t_z)$  et la rotation autour de l'axe optique  $(r_z)$ ).

La valeur  $\nu$  de la projection du signal  $S$  sur un noyau  $K$  peut être définie de la manière suivante :

$$\nu(t) = \int_{\mathbb{R}^2} K(\mathbf{x})S(\mathbf{x}, t)dx \quad (1.43)$$

Si nous considérons  $\nu^*$  la valeur de la projection du signal correspondant à la position

désirée alors l'erreur à minimiser pour réaliser la tâche de positionnement est  $\nu - \nu^*$ .

Dans le cas d'une translation selon les axes  $x$  et  $y$ ,  $S$  est simplement l'image  $\mathbf{I}$ . Sous réserve de disposer d'un noyau  $K$  lisse et différentiable ( $\nabla K = \frac{\partial K}{\partial \mathbf{x}}$ ), les vitesses de translations  $\dot{\mathbf{t}}_{\mathbf{x}}$  et  $\dot{\mathbf{t}}_{\mathbf{y}}$  peuvent être obtenues de la manière suivante :

$$(\dot{\mathbf{t}}_{\mathbf{x}}, \dot{\mathbf{t}}_{\mathbf{y}})^{\top} = -(\nu_{t_{xy}} - \nu_{t_{xy}}^*) \int_{\mathbb{R}^2} \nabla K(\mathbf{x}) I(\mathbf{x}, t) d\mathbf{x} \quad (1.44)$$

Pour connaître la vitesse linéaire de la translation  $t_z$  et la vitesse angulaire de la rotation  $r_z$  le long de l'axe optique,  $S$  est pris comme la transformée de Fourier, notée  $F$ , de l'image  $\mathbf{I}$ . En notant  $\xi \in \mathbb{Z}^2$  les index des valeurs de la transformée de Fourier de l'image, les deux relations suivantes sont obtenues :

$$\dot{t}_z = (\nu_{t_z} - \nu_{t_z}^*) \int_{\mathbb{R}^2} K'(\xi) \xi F(\xi, t) d\xi \quad (1.45)$$

$$\dot{r}_z = -(\nu_{r_z} - \nu_{r_z}^*) \int_{\mathbb{R}^2} K'(\xi R_{-\frac{\pi}{2}} \xi) F(\xi, t) d\xi \quad (1.46)$$

où,  $R_{-\frac{\pi}{2}}$  est l'opérateur de translation suivant :

$$R_{\theta} = \begin{pmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{pmatrix} \quad (1.47)$$

Cette approche est très élégante du point de vue théorique. Cependant, cette loi de commande préconise un contrôle au fur et à mesure des différents ddl (en effet, il est conseillé de contrôler d'abord  $t_z$  avant  $t_x$  et  $t_y$ ). De plus, seuls quatre ddl sont commandés car les noyaux utilisés ne permettent pas de définir les rotations  $r_x$  et  $r_y$ . Enfin, le choix du noyau à utiliser peut s'avérer difficile. Un mauvais choix peut entraîner des problèmes d'instabilité et de convergence (Kallem, 2009). Une technique assez similaire a été utilisée dans (Marturi et al., 2016) pour une tâche d'asservissement visuel utilisant un microscope électronique à balayage comme capteur visuel. La méthode propose également une manière de déterminer les deux degrés de liberté manquants à l'aide d'une transformée de Fourier sphérique. La FIGURE 1.17 montre les résultats de cette méthode dans des conditions nominales d'utilisation.

**Moments photométriques :** Cette méthode présentée dans (Bakthavatchalam et al., 2013) réunit l'approche fondée sur les moments géométriques et la méthode photométrique. Elle présente également quelques similarités avec la méthode à noyaux de convolution présentée ci-dessus. En d'autres termes, cette méthode permet d'associer les avantages d'une méthode géométrique, comme le large domaine de convergence, à ceux d'une méthode directe, c'est-à-dire la redondance d'informations visuelles et donc la précision.

**Définition 1.11. Moments photométriques d'ordre  $i + j$  :** Soit  $\mathbf{I}$  une image, alors les moments d'ordre  $i + j$  de  $\mathbf{I}$  ( $i, j \in \mathbb{N}$ ) sont donnés par la relation :

$$m_{ij} = \int \int x^i y^j I(x, y, t) dx dy \quad (1.48)$$

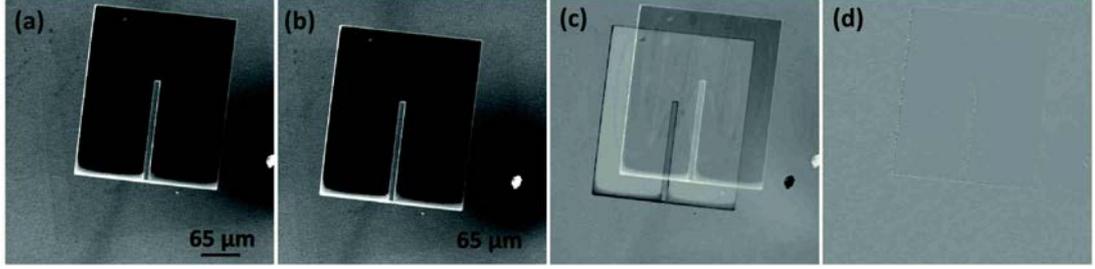


FIGURE 1.17 – Asservissement visuel pour un tâche de nanopositionnement en condition normale (a) image désirée (b) image courante (c) différence initiale entre les deux images (d) différence finale entre les deux images (Marturi et al., 2016).

L'ensemble  $\{x^i, y^j\}$  forme une base polynomiale sur laquelle les moments de l'image sont obtenus par projection. Ces moments ne sont pas orthogonaux et contiennent une redondance de l'information comme dans toutes les approches directes. Si l'approche des moments géométriques reposait sur une étape de segmentation de contour, celle de la méthode des moments photométriques permet de dépasser cette limitation en considérant toute l'information disponible dans l'image. La dérivée des moments photométriques est alors obtenue par la formule :

$$\dot{\mathbf{L}}_{ij} = \int \int x^i y^j \frac{\partial I(x, y, t)}{\partial t} dx dy \quad (1.49)$$

En exprimant la dérivée des moments photométriques en fonction du torseur des vitesses  $\mathbf{v}$  de la caméra, la formule suivante est alors obtenue :

$$\dot{m}_{ij} = \mathbf{L}_{ij} \mathbf{v} \quad (1.50)$$

En reliant la profondeur de la scène aux coordonnées des points du plan image par la formule  $\frac{1}{Z_c} = Ax + By + C$  (Chaumette, 2004), où  $A$ ,  $B$ ,  $C$  sont les paramètres de ce plan, la matrice d'interaction  $\mathbf{L}_{ij}$  peut être uniquement définie en fonction des paramètres  $A$ ,  $B$  et  $C$  et des moments d'ordre  $i + j + 1$ . Les moments photométriques peuvent être également perçus comme des noyaux de type  $x^i y^j$ .

**Transformée en ondelettes :** Si les méthodes se servant de la transformée de Fourier utilisent des primitives visuelles fréquentielles pour réaliser un asservissement visuel, elles n'en restent pas moins limitées puisqu'elles ne prennent pas en compte les informations spatiales contenues dans l'image. C'est pour cette raison qu'une première approche exploitant les ondelettes est apparue en 2016 (Ourak et al., 2016c). La transformée en ondelettes permet en effet une représentation simultanée dans l'espace et en fréquences d'une image.

Une décomposition en ondelettes d'une image permet d'obtenir quatre sous-images (notées  $\Gamma_0$ ,  $\Gamma_H$ ,  $\Gamma_V$  et  $\Gamma_D$ ).  $\Gamma_0$  représente une approximation de l'image,  $\Gamma_H$  les détails horizontaux de l'image,  $\Gamma_V$  les détails verticaux et  $\Gamma_D$  les détails diagonaux comme le montre la FIGURE 1.18.

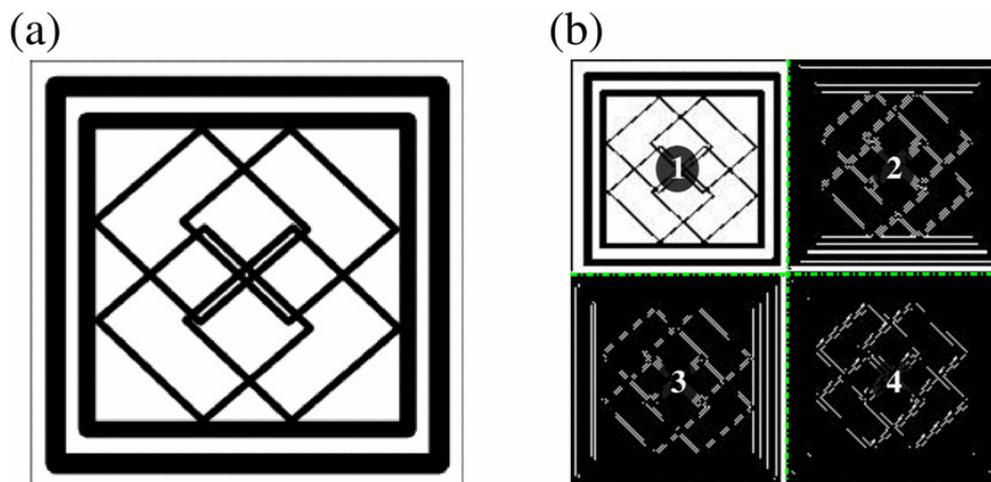


FIGURE 1.18 – (a) Image originale (b) décomposition en ondelettes avec en (1)  $\Gamma_0$ , en (2)  $\Gamma_V$ , en (3)  $\Gamma_H$  et en (4)  $\Gamma_D$  (Ourak et al., 2016c).

Les primitives visuelles utilisées sont les coefficients d'ondelettes de  $\Gamma_0$ , tandis que ceux de  $\Gamma_H$  et  $\Gamma_V$  sont dérivés afin de remplacer les gradients utilisés dans l'équation (1.31) de l'approche photométrique (voir FIGURE 1.19). Cette première formulation des coefficients d'ondelettes dans un asservissement visuel a montré un comportement intéressant de la loi de commande en termes de robustesse aux occultations partielles et aux changements d'illumination. Le vecteur des primitives visuelles ne contient qu'une partie limitée des coefficients de la transformée en ondelettes de l'image mais cette méthode ouvre la voie vers l'utilisation des transformées multi-échelles en asservissement visuel.

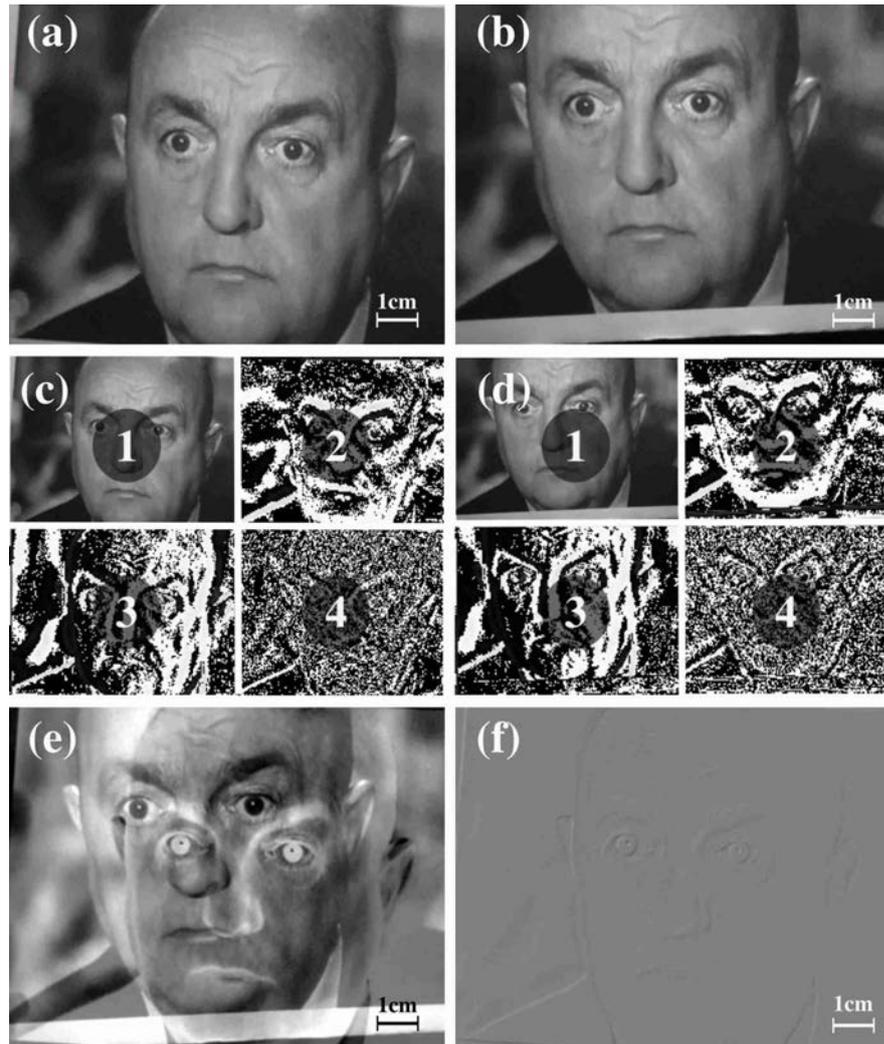


FIGURE 1.19 – Résultats d'une tâche de positionnement à 6 ddl réalisée par un asservissement visuel dont les primitives visuelles sont les coefficients d'ondelettes (Ourak et al., 2016c).

## 1.4 Conclusion

Nous avons pu voir dans ce chapitre les éléments de base de la vision par ordinateur, de la formation d'une image par la projection 2D d'un monde en 3D et de quelques outils géométriques (coordonnées cartésiennes, changement de repère, coordonnées homogènes). Nous nous sommes ensuite intéressés à la partie commande, en distinguant deux configurations possibles en asservissement visuel : la *caméra embarquée* et la *caméra déportée*. Nous avons également abordé le concept de la matrice d'interaction qui permet de relier la variation temporelle des primitives visuelles au torseur cinématique de la caméra. La dernière section de ce chapitre nous a permis de présenter quelques types de primitives visuelles en distinguant les approches géométriques (nécessitant des algorithmes de suivi visuel) et les approches directes (utilisant des primitives visuelles de type globales). La méthode photométrique présente l'avantage d'être très précise grâce à la redondance dans les données utilisées. Cependant, son principal défaut est son manque de robustesse lorsqu'il y a des perturbations lumineuses dans l'image. D'autres méthodes directes corrigent les inconvénients de la méthode photométrique, comme celle utilisant l'information mutuelle. Robuste aux occultations partielles et aux changements d'illumination, cette dernière méthode est également très intéressante lorsque différentes modalités d'images sont considérées dans la même boucle de commande. La méthode de mélange de Gaussiennes est également une alternative intéressante à la méthode photométrique puisqu'elle permet une convergence de la tâche d'asservissement visuel même lorsqu'il y a peu de recouvrement photométrique entre l'image courante et l'image désirée. Nous avons également vu des méthodes à noyaux de convolution, qui présentent une formulation théorique fondée sur la transformée de Fourier. Enfin, une première méthode tirant parti des ondelettes a été présentée, ouvrant la voie vers l'utilisation des représentations multi-échelles en asservissement visuel.

## Chapitre 2

# Espace fréquentiel et représentations multi-échelles

La décomposition d'images selon d'autres bases que la base spatiale, permet de représenter l'information visuelle d'une manière différente rendant possible l'accès à certains détails. La fin du chapitre 1 a présenté quelques travaux utilisant ces représentations dans la boucle de commande d'un asservissement visuel. Le chapitre 2 est consacré à l'introduction des représentations fréquentielles (transformée de Fourier) ainsi que celles permettant une expression simultanée des informations visuelles dans l'espace temps-fréquence comme les transformées en ondelettes. Ces éléments fondamentaux vont nous permettre, par la suite, de bien situer nos contributions scientifiques dans l'utilisation des représentations multi-échelles en asservissement visuel. Une première section de ce chapitre sera consacrée à l'analyse de Fourier. Cette théorie est le concept précurseur des représentations multi-échelles qui feront l'objet d'une seconde section. Nous y détaillerons la théorie des ondelettes 1D et 2D ainsi qu'une transformée 2D découlant des ondelettes : les shearlets. Le sommaire ci-dessous présente la structure de ce chapitre :

### Sommaire

---

<b>2.1</b>	<b>Analyse de Fourier</b> . . . . .	<b>40</b>
2.1.1	Transformée de Fourier continue . . . . .	41
2.1.2	Transformée de Fourier discrète . . . . .	43
2.1.3	Applications et limites de la transformée de Fourier . . . . .	43
<b>2.2</b>	<b>Représentations multi-échelles</b> . . . . .	<b>45</b>
2.2.1	Analyse multi-résolution . . . . .	46
2.2.2	Transformée en ondelettes . . . . .	47
2.2.3	Applications de la transformée en ondelettes . . . . .	54
2.2.4	Transformée en shearlets . . . . .	58
2.2.5	Différentes implémentations des shearlets . . . . .	60
2.2.6	Applications de la transformée en shearlets . . . . .	64
<b>2.3</b>	<b>Conclusion</b> . . . . .	<b>67</b>

---

## 2.1 Analyse de Fourier

Né en 1768 et rapidement orphelin, Joseph Fourier (voir FIGURE 2.1) se démarque par son intelligence, en particulier en mathématiques et en sciences physiques. Malgré une enfance très modeste, il bénéficiera des changements amenés par la révolution française et des transformations de la société pour produire un travail scientifique avant-gardiste. En effet, à cette époque où les domaines de recherche étaient très cloisonnés, son approche à l'intersection des sciences de la physique et des mathématiques, rendaient les idées de Fourier difficilement acceptables par la communauté scientifique. Sa "*théorie analytique de la chaleur*" est considérée comme son œuvre majeure, même si elle traite des sciences physiques, les apports mathématiques de ce travail sont considérables (Fourier, 1822). Néanmoins, les séries et la transformée de Fourier furent très largement critiquées à l'époque par Laplace, Lagrange et Poisson. Ce n'est que depuis la fin du XX<sup>ème</sup> siècle que l'analyse de Fourier a été revue à sa juste valeur. La porosité croissante entre les différents domaines scientifiques et l'émergence du numérique en sont les raisons principales. En effet, si l'on retrouve l'analyse de Fourier en théorie des nombres, en acoustique et même en cristallographie, c'est en traitement du signal qu'elle est omniprésente. Plus qu'une théorie, l'analyse de Fourier constitue un véritable outil pour les ingénieurs et les scientifiques d'aujourd'hui, et c'est ce qui rend son travail particulièrement moderne.<sup>1</sup>



FIGURE 2.1 – Portrait de Joseph Fourier (1768-1830).

---

1. Fourier est également l'un des premiers à avoir proposé une théorie de l'effet de serre, base de la climatologie.

### 2.1.1 Transformée de Fourier continue

L'analyse de Fourier repose sur le constat que toute fonction périodique peut se décomposer sous forme de séries trigonométriques (*séries de Fourier*) exprimées sous la forme suivante :

$$f(x) = \sum a_n \cos nx + b_n \sin nx \quad (2.1)$$

$a_0$  constitue alors la composante continue,  $a_1$  et  $b_1$  les composantes fondamentales et  $a_n$  et  $b_n$  ( $n > 1$ ) les harmoniques.

Par ailleurs, si ce genre de décompositions est courante à l'époque de Fourier, puisque déjà largement utilisé par Euler par exemple, la véritable contribution de Fourier a été d'étendre cette représentation à toutes les fonctions intégrables sur  $\mathbb{R}$ , donnant lieu à la *transformée de Fourier*.

**Définition 2.1. Transformée de Fourier :** Si  $f$  est une fonction intégrable sur  $\mathbb{R}$  et  $i$  est l'unité imaginaire, sa transformée de Fourier est la fonction  $\mathcal{F}(f) = \hat{f}$  donnée par la formule :

$$\mathcal{F}(f) : \xi \mapsto \hat{f}(\xi) = \int_{-\infty}^{+\infty} f(t) e^{-i\xi t} dt \quad (2.2)$$

où  $t \in \mathbb{R}$  représente le temps et  $\xi$  représente la fréquence.

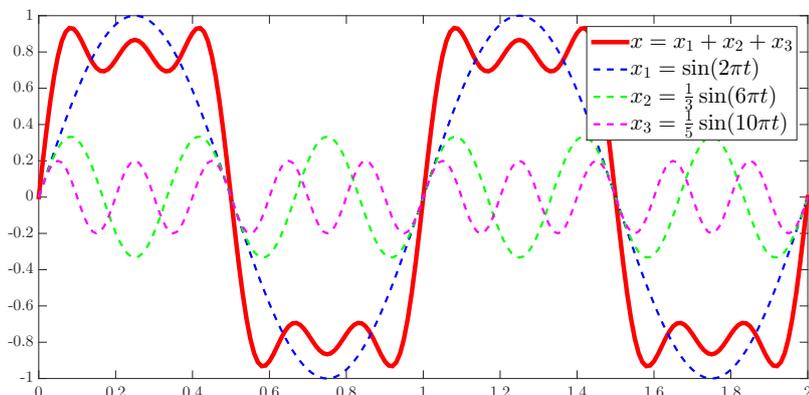
**Définition 2.2. Transformée de Fourier inverse :** Si la transformée de Fourier de  $f$ , notée  $\hat{f}$ , est elle-même une fonction intégrable, la formule dite de transformée de Fourier inverse, notée  $\mathcal{F}^{-1}$ , est donnée par la formule :

$$f(t) = \mathcal{F}^{-1}(\hat{f})(t) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \hat{f}(\xi) e^{+i\xi t} d\xi \quad (2.3)$$

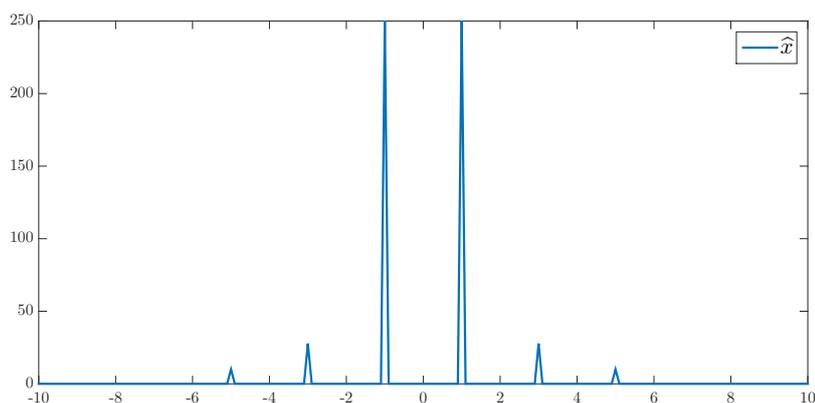
Nous dirons alors que  $f$  est définie dans le domaine temporel et que  $\hat{f}$  est définie dans celui des fréquences. Par exemple, la fonction échelon peut être décomposée comme une somme infinie de fonctions sinus. La FIGURE 2.2 (a) montre comment la somme de trois fonctions sinusoïdales approxime la fonction échelon. La FIGURE 2.2 (b) représente la transformée de Fourier de notre somme, nous observons des pics dont l'abscisse correspond à la fréquence de chacun des trois sinus utilisés pour approximer la fonction échelon. Augmenter le nombre de fonctions sinus dans la somme permettrait une approximation de plus en plus fine de la fonction échelon et ferait apparaître de plus en plus de pics dans sa transformée de Fourier.

**Théorème 2.1. Égalité de Parseval :** La transformée de Fourier préserve la norme.

$$\|f\|_2 = \|\hat{f}\|_2, \quad i.e. \quad \int_{-\infty}^{+\infty} |f(t)|^2 dt = \int_{-\infty}^{+\infty} |\hat{f}(\xi)|^2 d\xi \quad (2.4)$$



(a) Approximation d'une fonction échelon.



(b) Transformée de Fourier correspondante.

FIGURE 2.2 – Décomposition d'une fonction échelon avec 3 sinusoides et transformée de Fourier de cette même fonction.

Le théorème de Parseval signifie que l'énergie totale d'un signal ne change pas selon la représentation choisie : fréquentielle ou temporelle.

**Définition 2.3. Produit de convolution :** *Le produit de convolution est un opérateur bilinéaire et un produit commutatif, noté  $*$ , qui à deux fonctions  $f$  et  $g$  fait correspondre une autre fonction  $f * g$  définie par :*

$$(f * g)(t') = \int_{-\infty}^{+\infty} f(t' - t)g(t) dt = \int_{-\infty}^{+\infty} f(t)g(t' - t) dt \quad (2.5)$$

Il est à noter également qu'un certain nombre de transformations dans le domaine temporel peuvent être facilement retranscrites dans le domaine fréquentiel. Le TABLEAU 2.1 illustre ces équivalences. Par exemple, la contraction du signal dans le domaine temporel implique une dilatation du signal dans le domaine fréquentiel.

TABLE 2.1 – Équivalences entre domaine temporel et fréquentiel grâce à la transformée de Fourier.

Transformation	Domaine temporel	Domaine spectral
Convolution	$(f * g)(t)$	$\hat{f}(\xi) \cdot \hat{g}(\xi)$
Multiplication	$(f \cdot g)(t)$	$\frac{1}{2\pi} (\hat{f} * \hat{g})(\xi)$
Translation temporelle	$f(t - t_0)$	$\hat{f}(\xi) \cdot e^{-i\xi t_0}$
Modulation temporelle	$f(t) \cdot e^{it\xi_0}$	$\hat{f}(\xi - \xi_0)$
Dilatation	$f\left(\frac{t}{a}\right)$	$ a  \hat{f}(a\xi)$
Dérivation temporelle	$f'(t)$	$i\xi \hat{f}(\xi)$
Dérivation fréquentielle	$t \cdot f(t)$	$\frac{i}{2\pi} \hat{f}'(\xi)$

### 2.1.2 Transformée de Fourier discrète

La transformée de Fourier discrète permet d'évaluer une représentation spectrale d'un signal échantillonné sur une fenêtre de temps finie. En considérant un signal comme un échantillon, et donc comme une suite  $\zeta$  de  $N$  termes  $\zeta(0) \dots \zeta(N-1)$ , sa transformée de Fourier discrète est donnée par la formule suivante :

$$\hat{\zeta}(k) = \sum_{n=0}^{N-1} \zeta(n) e^{-2i\pi k \frac{n}{N}} \quad \text{pour} \quad 0 \leq k < N \quad (2.6)$$

Le résultat est une suite  $\hat{\zeta}$  de  $N$  termes correspondants à une approximation de la transformée de Fourier continue de ce signal aux points de fréquence  $f_k = \frac{kf_e}{N}$   $k \in [0, N-1]$ , avec  $f_e$  la fréquence d'échantillonnage du signal. La transformée inverse est alors donnée par la formule ci-dessous :

$$\zeta(n) = \frac{1}{N} \sum_{k=0}^{N-1} \hat{\zeta}(k) e^{2i\pi n \frac{k}{N}} \quad (2.7)$$

Pour un signal échantillonné à la fréquence  $f_e$ , la transformée de Fourier discrète renseigne sur toutes les fréquences situées entre  $-\frac{f_e}{2}$  et  $\frac{f_e}{2}$ .

En traitement du signal, l'algorithme de transformée de Fourier rapide (FFT) est largement employé depuis les années 1960 (Cooley et Tukey, 1965). Cet algorithme utilise la récursivité pour obtenir la représentation fréquentielle d'un signal de taille  $N$  avec une complexité de  $N \times \log N$  contre  $N^2$  itérations en utilisant une méthode de transformée non optimisée.

### 2.1.3 Applications et limites de la transformée de Fourier

En traitement d'images, des transformées de Fourier 2D sont utilisées. L'information temporelle devient ainsi spatiale. La FIGURE 2.3 montre la transformée de Fourier d'une image 2D. L'énergie est très concentrée sur le centre de l'image, c'est-à-dire sur les basses

fréquences. Pour cette raison, il est intéressant de représenter la transformée de Fourier de l'image en échelle logarithmique (voir FIGURE 2.3 (c)) afin de mieux distinguer les différentes fréquences de l'image.

**Définition 2.4. Transformée de Fourier 2D :** si  $f$  est une fonction de  $\mathbb{R}^2$  dans  $\mathbb{R}$  de variable  $(x, y)$ , alors sa transformée de Fourier au point  $(\xi_1, \xi_2)$  dans l'espace fréquentiel est définie par :

$$\hat{f}(\xi_1, \xi_2) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) \cdot e^{-i(\xi_1 x + \xi_2 y)} dx dy. \quad (2.8)$$

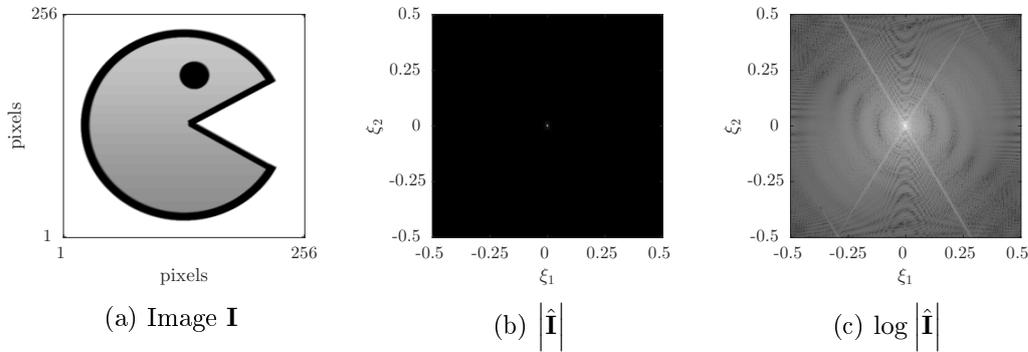


FIGURE 2.3 – Illustration de la transformée de Fourier sur une image 2D. Les basses fréquences de l'image **I** se situent proche du centre de  $|\hat{\mathbf{I}}|$ .

La transformée de Fourier est largement utilisée dans différents domaines de l'imagerie. En effet, la plupart des dégradations (fenêtrage, échantillonnage, bruit) subies par une image lors de son acquisition s'expriment de manière simple dans l'espace de Fourier. Par exemple, une image floue peut être considérée comme la convolution d'une image nette avec un filtre Gaussien. Or, la convolution dans le domaine spatial correspond à une simple multiplication dans le domaine fréquentiel. Étant donné que la complexité d'une convolution est de  $N^2P$  avec  $N$  le nombre de pixels de l'image et  $P$  la taille du support du noyau de la convolution, il est alors plus intéressant de faire cette convolution dans le domaine de Fourier dès lors que  $P > \log(N)$ . Certaines convolutions dans l'espace de Fourier permettent également d'éviter des phénomènes de repliement de spectres ou de réaliser un filtrage passe-haut ou passe-bas sur l'image.

La FIGURE 2.4 montre que la transformée de Fourier ne prend pas en compte la localisation spatiale des informations de l'image. La transformée de l'image représentée sur la FIGURE 2.4 (a) est la même que celle montrée en FIGURE 2.4 (c) alors que le carré représenté sur ces deux images a été translaté sur les axes  $x$  et  $y$ . En 1946, Gabor a proposé une approche découpant le signal en plusieurs parties. Cette méthode est appelée transformée de Fourier à court terme ou encore, transformée de Fourier à fenêtre glissante (Gabor, 1946). L'application de la transformée de Fourier sur une section locale

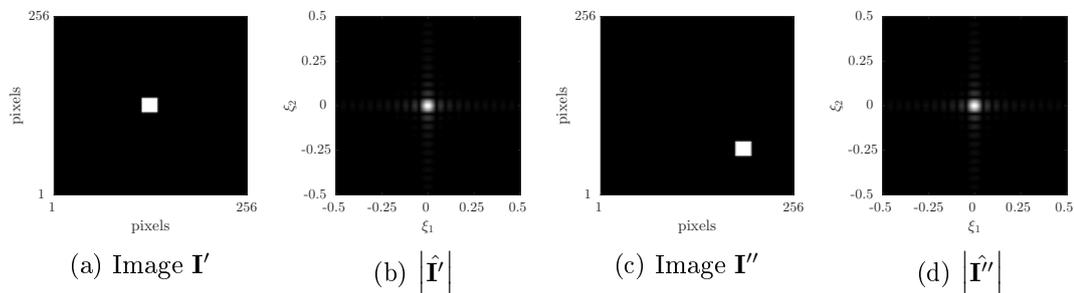


FIGURE 2.4 – Résultat de la transformée de Fourier d'un motif en carré placé à différentes positions dans l'image. La translation du carré blanc dans l'image n'a aucun impact sur sa transformée de Fourier.

d'un signal, appelée fenêtre, permet alors de réunir des informations fréquentielles sur différentes portions du signal. Cependant, la taille fixe de la fenêtre empêche une bonne visualisation des basses-fréquences (respectivement des hautes-fréquences) dans le cas d'une fenêtre plutôt étroite (respectivement large) et induit souvent des problèmes sur les bords de l'image. La base de cosinus locaux s'appuie sur la même idée, elle est notamment utilisée pour la compression JPEG, permet d'éviter les effets de bords et découpe l'image en sous-images afin de donner des informations de localisation en espace et non plus seulement en fréquence. Certains signaux correspondent à une succession de phases stationnaires où la fréquence est stable sur une longue durée, et de phases transitoires, où il y a de fortes variations de fréquences sur une courte durée. C'est pour cette raison qu'est apparue l'idée d'adapter la dimension des fenêtres temps/fréquence.

## 2.2 Représentations multi-échelles

Pendant longtemps l'analyse de Fourier a été la seule technique permettant la décomposition d'un signal discret en fréquences et sa reconstruction sans pertes. Si Gabor a apporté une modification intéressante à la transformée de Fourier en utilisant une fenêtre glissante, ce n'est qu'en 1984 que Morlet et Grossmann introduisent la transformée en ondelettes, terme français qui deviendra "*wavelets*" en anglais. Entre 1984 et 1988 se développèrent ainsi les bases de la théorie des ondelettes. Meyer rassembla en 1986 toutes les découvertes précédentes sur le sujet (Meyer, 1986) et définit les ondelettes orthogonales, tandis que Mallat fit le lien entre les ondelettes et l'analyse multi-résolution (Mallat, 1989). Enfin, Daubechies mis au point une famille d'ondelettes orthogonales qui reste une des plus utilisées aujourd'hui (Daubechies, 1988). C'est notamment ces ondelettes, dites de Daubechies, qui sont utilisées dans le standard JPEG 2000 (Christopoulos et al., 2000). La théorie des ondelettes a eu des répercussions majeures dans de nombreux domaines allant du codage numérique à la détection d'ondes gravitationnelles, ce qui a valu à Meyer le prix Abel en 2017. Tandis que la transformée en ondelettes a été intensément étudiée et mise en application depuis les années 1980, d'autres transformées similaires ont par la suite été développées comme les *curvelets*

(Candes et Donoho, 2000) ou les *contourlets* (Do et Vetterli, 2003). Les *shearlets*, qui sont plus récentes, ont été introduites pour la première fois en 2005 (Labate et al., 2005).

Dans cette thèse, deux représentations multi-échelles seront utilisées dans le cadre de la mise en œuvre de lois de commande par asservissement visuel. Nous nous contenterons donc de présenter ces deux transformées, à savoir les *ondelettes* et les *shearlets*.

### 2.2.1 Analyse multi-résolution

L'analyse multi-résolution (AMR) est une méthode utilisée dans la plupart des transformées en ondelettes discrètes (Digital Wavelet Transform - DWT). Elle fut introduite en 1988 par Mallat et Meyer (Mallat, 1989) mais découle directement des méthodes pyramidales inventées par Burt et Adelson (Burt et Adelson, 1983).

**Définition 2.5.** *Représentation pyramidale d'une image* : En traitement d'images, la représentation pyramidale d'une image est une décomposition multi-échelle dans laquelle l'image est successivement sous-échantillonnée et lissée. On obtient alors plusieurs niveaux de décomposition (voir FIGURE 2.5).

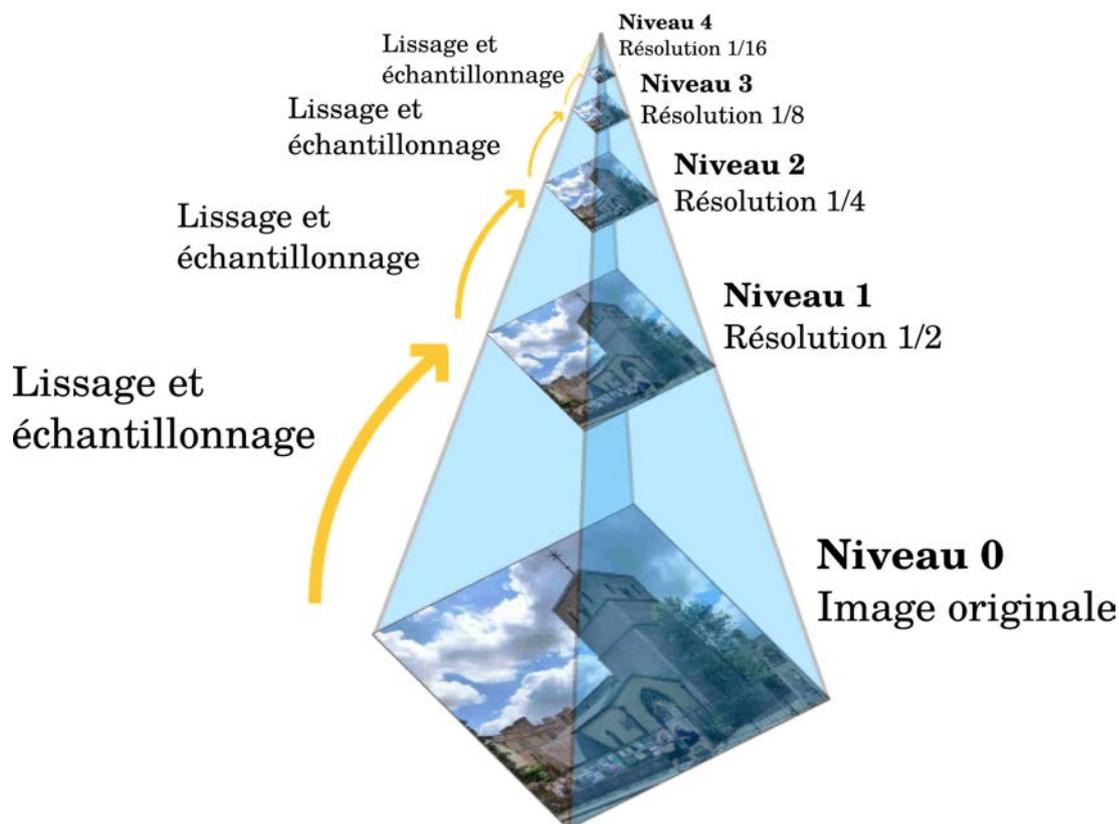


FIGURE 2.5 – Représentation pyramidale d'une image avec 5 niveaux de décomposition.

Soit  $\mathcal{E} = \mathcal{L}^2(\mathbb{R})$  l'espace vectoriel des fonctions réelles à carré intégrable, la structure AMR consiste à projeter une fonction  $f$  sur des sous-espaces vectoriels de  $\mathcal{E}$  définis d'une manière bien précise, décrite ci-dessous.

**Définition 2.6. Analyse multi-résolution :** Une analyse multi-résolution de l'espace vectoriel  $\mathcal{E}$  est définie comme la suite de sous-espaces vectoriels  $V^j$  vérifiant les conditions suivantes :

- $\forall j \in \mathbb{Z}, V^j \subset V^{j+1}$  : les sous-espaces sont emboîtés,
- $\bigcup_{j \in \mathbb{Z}} V_j = 0$  : les sous-espaces ne se recoupent pas,
- $\bigcap_{j \in \mathbb{Z}} V_j = \mathcal{E}$  : l'union des sous-espaces forme  $\mathcal{E}$ ,
- $\forall j, \exists (\phi_m^j)_{k \in \mathbb{Z}}$  formant une base orthonormée de  $V^j$ ,
- $\forall j \in \mathbb{Z}, f(\cdot) \in V^j \iff f(2 \times \cdot) \in V^{j+1}$  :  $V^{j+1}$  est l'image de  $V^j$  par une dilatation de facteur 2,
- $\forall (j, m) \in \mathbb{Z}^2, f(\cdot) \in V^j \iff f(\cdot - 2^j m) \in V^j$  :  $V^j$  est invariant par translation de  $2^j$ .

Puisque  $V^j \subset V^{j+1}$ ,  $\exists W^j$  tel que  $V^{j+1} = V^j \oplus W^j$ . Avec  $\oplus$  l'opérateur de supplémentaire orthogonal. Cela revient à dire que  $W^j$  est le supplémentaire orthogonal de  $V^j$  dans  $V^{j+1}$ . Pour former une AMR, il faut alors trouver une famille de  $m$  fonctions  $(\psi_m^j)$  formant une base de  $W^j$ , sachant que la famille de  $m$  fonctions  $(\phi_m^j)$  forme une base de  $V^j$ . Pour cela, on se sert de la relation suivante :

$$V^{j+1} = V^1 \oplus W^1 \oplus W^2 \oplus \dots \oplus W^j \quad (2.9)$$

Chaque fonction  $f \in V^{j+1}$  possède une unique décomposition de la forme :

$$f(t) = \alpha_1 \phi_1^1 + \sum_{i=1}^j \sum_{m \in \mathbb{Z}} \beta_m^i \psi_m^i(t) \quad (2.10)$$

La fonction  $\phi_1^1$ , appelée fonction d'échelle dans la théorie des ondelettes, est suffisante pour déterminer l'emboîtement des espaces vectoriels  $V^1 \subset V^2 \subset \dots \subset V^{j+1}$ . Former une base d'ondelettes de niveau  $j$  consistera alors à définir un ensemble de fonctions  $(\psi_m^j)$  formant une base orthonormale de  $W^j$  en respectant la condition  $V^{j+1} = V^j \oplus W^j$ .

## 2.2.2 Transformée en ondelettes

La transformée en ondelettes permet une localisation dans l'espace et en fréquence mais plus encore car elle comporte des notions de multi-résolution permettant d'affiner la localisation spatiale de la fonction.

**Base d'ondelettes 1D :** Une base d'ondelettes dans l'espace des fonctions réelles de carré intégrable  $\mathcal{L}^2(\mathbb{R})$  est réalisée en translatant et en dilatant une fonction particulière appelée "ondelette mère"  $\psi$ . Le but est d'obtenir des fonctions situées sur des supports temporels et fréquentiels différents. Pour cela, des opérateurs de dilatation et de translation, dit *dyadiques*, sont utilisés :

$$\psi_m^j(t) = \sqrt{2^j} \psi(2^j t - m) \quad (2.11)$$

où  $j$  correspond à la dilatation de l'ondelette et  $m$  à sa translation temporelle. La FIGURE 2.6 montre la façon dont les différentes bases de représentation du signal découpent le plan temps/fréquence. La base des ondelettes permet un découpage dont les résolutions changent à la fois dans la dimension temporelle mais aussi dans les fréquences. La FIGURE 2.7 nous montre comment la dilatation  $j$  et la translation  $m$  des ondelettes permettent à cette ondelette de s'inscrire dans une portion du plan temps/fréquence.

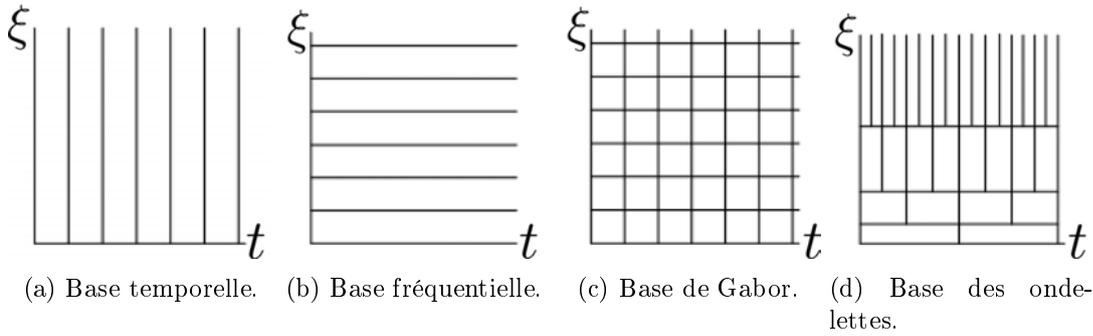


FIGURE 2.6 – Découpage du plan temps/fréquence en fonction de différentes bases de représentations.

Deux exemples de fonctions ondelettes mères sont les *ondelettes de Haar* (Haar, 1910) et les *ondelettes de Daubechies* (Daubechies, 1988), illustrées sur la FIGURE 2.8 avec leurs fonctions d'échelles correspondantes.

**Définition 2.7. Transformée en ondelettes 1D :** La transformée en ondelettes  $\mathcal{W}_\psi f$  d'une fonction  $f \in L^2(\mathbb{R})$ , selon une fonction mère  $\psi \in L^2(\mathbb{R})$  est donnée par le produit scalaire suivant :

$$(\mathcal{W}_\psi f)(j, m) = \langle f, \psi_m^j \rangle = \int_{\mathbb{R}} f(t) \overline{\psi_m^j(t)} dt, \quad (2.12)$$

où  $j \in \mathbb{Z}$  définit l'échelle,  $m \in \mathbb{Z}$  la localisation de l'ondelette  $\psi_m^j$  et  $\overline{\psi_m^j}$  représente son complexe conjugué.

Il est à remarquer que dilater et translater la fonction génératrice  $\psi$  changent la fréquence et le degré de la localisation. Les plus hautes fréquences correspondent à de

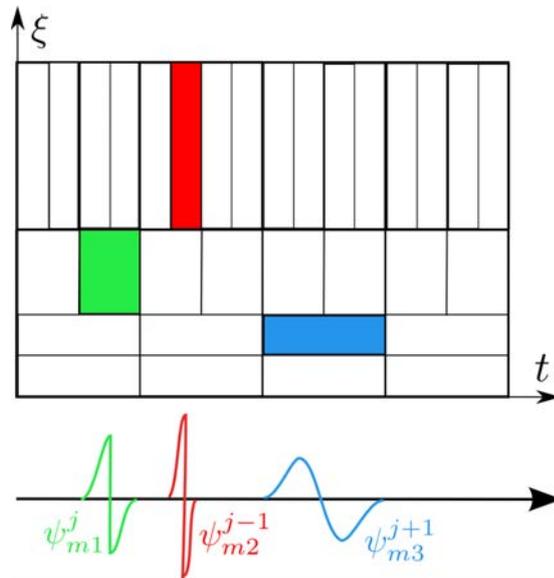
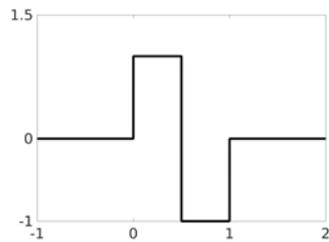
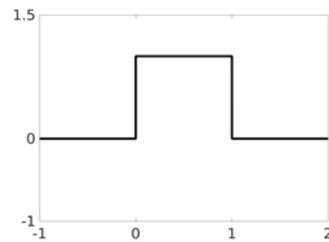


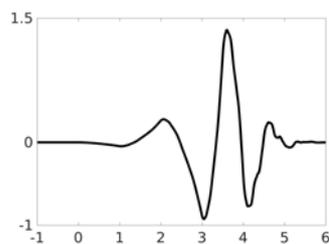
FIGURE 2.7 – Exemple d'ondelettes traduites et dilatées.



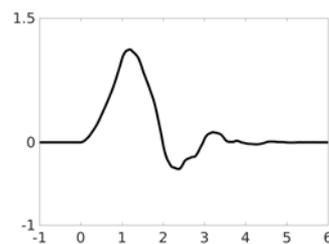
(a) Ondelette de Haar.



(b) Fonction d'échelle de Haar.



(c) Ondelette de Daubechies d'ordre 4 (db4).



(d) Fonction d'échelle de Daubechies d'ordre 4 (db4).

FIGURE 2.8 – Fonctions ondelettes mères de Haar et de Daubechies et leurs fonctions d'échelles correspondantes.

plus hauts degrés de localisation et les plus basses fréquences à des fonctions moins localisées spatialement. Ce comportement caractérise la différence fondamentale entre les ondelettes et la transformée de Fourier à fenêtre glissante (Allen, 1977).

Dans le domaine discret, nous utilisons des filtres pour définir la fonction d'échelle  $\phi \in L^2(\mathbb{R})$  et la fonction ondelette  $\psi \in L^2(\mathbb{R})$ . En effet, en respectant la structure AMR, une ondelette de niveau  $j$  est définie par rapport à celle de niveau  $j - 1$  grâce aux filtres  $\mathbf{h}$  et  $\mathbf{g}$ , en respectant les relations suivantes :

$$\phi^{j-1}(t) = \sum_m h_m \phi_m^j(t) = \sum_m h_m \sqrt{2^j} \phi(2^j t - m) \quad (2.13)$$

$$\psi^{j-1}(t) = \sum_m g_m \psi_m^j(t) = \sum_m g_m \sqrt{2^j} \psi(2^j t - m) \quad (2.14)$$

Le filtre défini par la transformée de Fourier de  $\mathbf{h}$  est un filtre passe-bas encodant les approximations du signal (basses fréquences). Par ailleurs, celui défini par la transformée de Fourier de  $\mathbf{g}$  est un filtre passe-haut décrivant les détails du signal (hautes fréquences). La transformée en ondelettes discrète peut alors être exprimée comme étant l'application successive de filtres passe-haut et passe-bas suivie d'une opération d'échantillonnage (voir FIGURE 2.9). Par exemple, pour les ondelettes de Haar,  $\mathbf{h} = (\frac{1}{\sqrt{2}}; \frac{1}{\sqrt{2}})$  et  $\mathbf{g} = (\frac{1}{\sqrt{2}}; -\frac{1}{\sqrt{2}})$ . Le filtre passe-bas permet de capturer les basses fréquences mais possède une faible résolution temporelle tandis que le filtre passe-haut permet de capturer les hautes fréquences avec une bonne résolution temporelle.

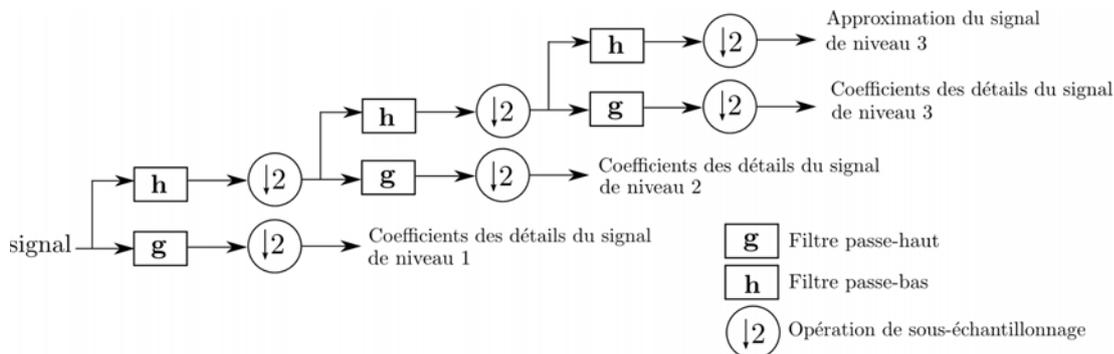


FIGURE 2.9 – Schéma de fonctionnement des filtres passe-haut et passe-bas pour la transformée en ondelettes.

**Base d'ondelettes 2D :** La transformée en ondelettes, dans l'espace des fonctions réelles à deux dimensions de carré intégrable  $L^2(\mathbb{R}^2)$ , est réalisée en translatant et en dilatant une famille de fonctions, appelées fonctions génératrices. Ces dernières sont définies comme le produit entre une fonction d'échelle  $\phi \in L^2(\mathbb{R})$  et une ondelette mère  $\psi \in L^2(\mathbb{R})$ .

**Définition 2.8. Fonctions génératrices d'ondelettes 2D :** A partir de la fonction d'échelle  $\phi \in L^2(\mathbb{R})$  et de l'ondelette mère  $\psi \in L^2(\mathbb{R})$ , nous pouvons définir un système

de fonctions génératrices d'ondelettes sur  $L^2(\mathbb{R}^2)$  :

$$\begin{aligned}\phi^{(1)}(u, v) &= \phi(u)\phi(v), \\ \psi^{(1)}(u, v) &= \phi(u)\psi(v), \\ \psi^{(2)}(u, v) &= \psi(u)\phi(v), \\ \psi^{(3)}(u, v) &= \psi(u)\psi(v),\end{aligned}\tag{2.15}$$

Cela signifie que la transformée en ondelettes 2D consiste en l'application successive de deux fonctions 1D; la première sur les lignes et la deuxième sur les colonnes de l'image<sup>1</sup>. En 2D, les opérateurs de dilatation et de translation sont définis de la manière suivante :

**Définition 2.9. Opérateur dyadique de dilatation :** Soit une fonction  $f$  de  $L^2(\mathbb{R}^2)$  et  $(u, v)$  les coordonnées pixeliques d'un point de l'image, l'opérateur de dilatation dyadique  $D_2$  de  $f$  est défini par :

$$D_2 f(u, v) = 2f(2u, 2v)\tag{2.16}$$

**Définition 2.10. Opérateur de translation :** Soit une fonction  $f$  de  $L^2(\mathbb{R}^2)$ ,  $(u, v)$  les coordonnées pixeliques d'un point de l'image et  $\mathbf{m} = (m_1, m_2)$  un vecteur de translation, l'opérateur de translation  $T_{\mathbf{m}}$  de  $f$  est défini par :

$$T_{\mathbf{m}} f(u, v) = f(u - m_1, v - m_2)\tag{2.17}$$

Il suffit alors d'opérer les transformations présentées en (2.16) et (2.17) sur les fonctions du système (2.15) pour obtenir une base de transformée en ondelettes 2D.

**Définition 2.11. Base de transformée en ondelettes 2D :** Soit  $\phi^{(1)}$  une fonction d'échelle 2D et  $\psi^{(1)}, \psi^{(2)}, \psi^{(3)}$  les fonctions ondelettes générant les détails verticaux, horizontaux et diagonaux de l'image,  $j$  représente le niveau de décomposition (c'est-à-dire le nombre de fois où l'opérateur  $D_2$  est appliqué) et  $\mathbf{m} \in \mathbb{Z}^2$  le vecteur de translation, alors l'union suivante nous donne un système de transformée en ondelettes 2D :

$$\begin{aligned}\left\{ \phi^{(1)} = T_{\mathbf{m}} \phi^{(1)} : \mathbf{m} \in \mathbb{Z}^2 \right\} \cup \\ \left\{ \psi_j^{(l)} = D_2^j T_{\mathbf{m}} \psi^{(l)} : j \in \mathbb{N}_0, \mathbf{m} \in \mathbb{Z}^2, l \in \{1, 2, 3\} \right\},\end{aligned}\tag{2.18}$$

Comme l'illustre la FIGURE 2.10, ces transformations permettent de localiser les fonctions génératrices dans des portions du domaine fréquentiel de l'image ouvrant la possibilité à la formulation d'une analyse multi-résolution.

---

1. Dans le chapitre 4, la fonction  $\phi^{(1)}$  sera notée  $\psi^{(0)}$  pour davantage de généralité.

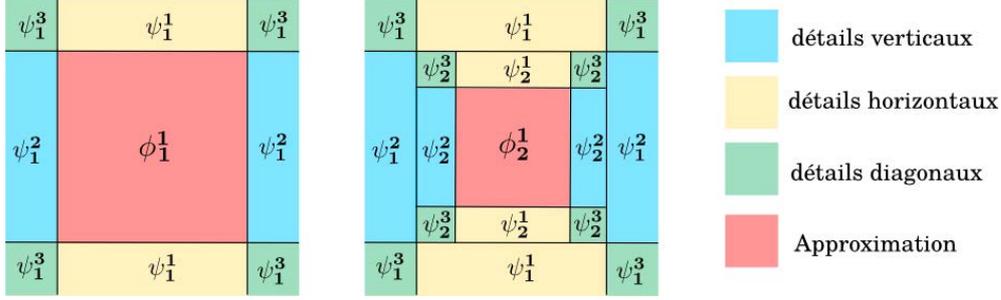


FIGURE 2.10 – Découpage du domaine fréquentiel en ondelettes 2D.

**Définition 2.12. Transformée en ondelettes 2D :** La transformée en ondelettes  $\mathcal{W}_\psi f$  d'une fonction de carré intégrable  $f \in L^2(\mathbb{R}^2)$ , selon une fonction génératrice  $\psi \in L^2(\mathbb{R}^2)$ , est donnée par le produit scalaire suivant :

$$(\mathcal{W}_\psi f)(j, \mathbf{m}) = \langle f, \psi_{j, \mathbf{m}} \rangle = \iint_{\mathbb{R}^2} f(u, v) \overline{D_2^j T_{\mathbf{m}} \psi(u, v)} du dv, \quad (2.19)$$

où  $j \in \mathbb{Z}$  définit l'échelle et  $\mathbf{m} \in \mathbb{Z}^2$  la localisation de l'ondelette  $\psi_{j, \mathbf{m}}$ .

**Transformée en ondelettes d'une image :** La FIGURE 2.11 montre la décomposition en ondelettes d'une image. La transformée en ondelettes de l'image avec un niveau de décomposition  $j = 1$  est visible sur la FIGURE 2.11(a) donnant quatre sous-images :  $\langle I, \phi_1^1 \rangle$ ,  $\langle I, \psi_1^1 \rangle$ ,  $\langle I, \psi_1^2 \rangle$  et  $\langle I, \psi_1^3 \rangle$ . L'application d'un second niveau de décomposition  $j = 2$  sur l'approximation de l'image  $\langle I, \phi_1^1 \rangle$  est illustré en (b) donnant 4 sous-images :  $\langle I, \phi_2^1 \rangle$ ,  $\langle I, \psi_2^1 \rangle$ ,  $\langle I, \psi_2^2 \rangle$  et  $\langle I, \psi_2^3 \rangle$ . Ainsi,  $j$  détermine le nombre de décompositions appliquées à notre image lors de sa transformée en ondelettes. De plus, la transformée en ondelettes permet d'obtenir une représentation parcimonieuse de l'image, illustrée par le nombre de coefficients de faible valeur, c'est-à-dire proches de zéro (couleur noire).

En considérant une image discrétisée  $\mathbf{I} \in \ell^2(\mathbb{Z}^2)$  comme l'approximation d'une fonction continue de carré intégrable  $I$  dépendant d'une fonction d'échelle  $\phi^{(1)}$  à une résolution donnée  $j \in \mathbb{Z}$ , la structure AMR permet de calculer les coefficients de la transformée en ondelettes de la fonction  $I$  (notre image). Cette image est définie de manière continue, en convoluant successivement l'image discrète  $\mathbf{I}$  (sous forme de matrice) avec les filtres  $\mathbf{h}$  et  $\mathbf{g}$ . Ce système fondé sur les filtres permet une implémentation plus intuitive des ondelettes discrètes exprimées à l'aide de la "transformée en ondelettes rapide" ou "Fast Wavelet Transform" (FWT) (Beylkin et al., 1991). Ce principe est schématisé sur la FIGURE 2.12(b). Nous pouvons noter que le nombre de coefficients reste à peu près constant (des petites variations sont rencontrées selon l'implémentation) et que la représentation multi-échelle contient à la fois une approximation de l'image d'origine en basse résolution ainsi que plusieurs niveaux de détails (horizontaux, verticaux et diagonaux).

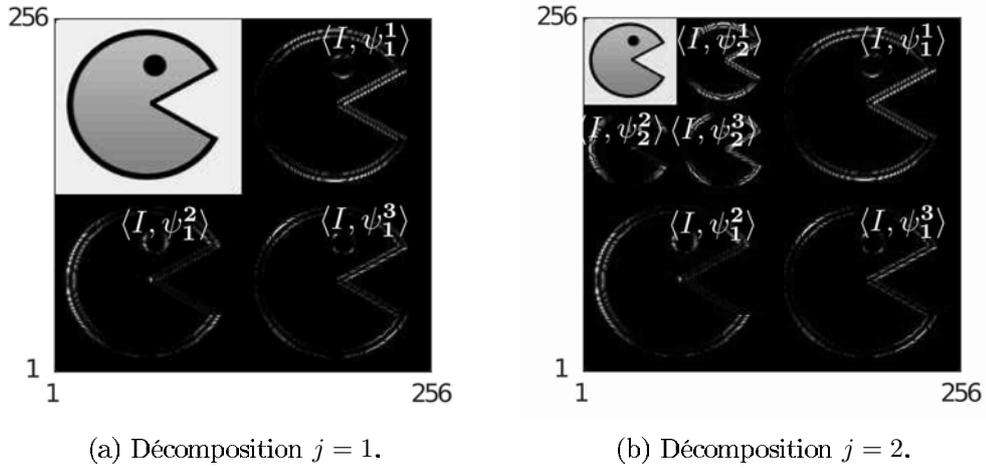


FIGURE 2.11 – Décomposition en ondelettes d’une image aux niveaux  $j = 1$  et  $j = 2$ .

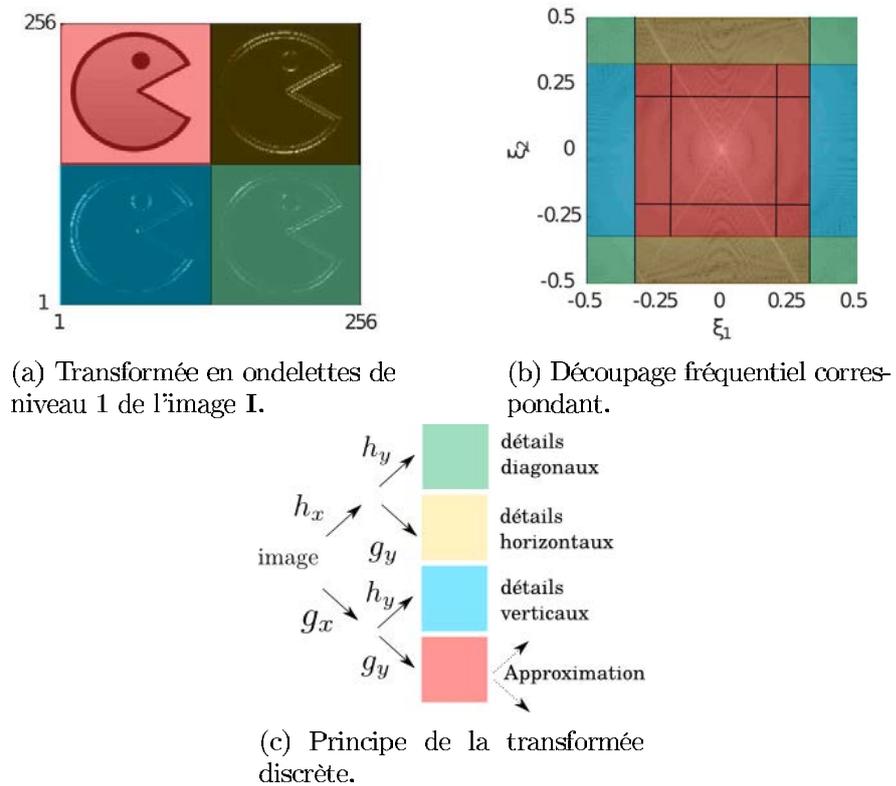


FIGURE 2.12 – Lien entre la représentation en ondelettes, les filtres discrets et le découpage dans le domaine fréquentiel.

### 2.2.3 Applications de la transformée en ondelettes

Les applications des ondelettes sont nombreuses et couvrent différents domaines. Nous nous limiterons dans cette sous-section aux applications de traitement du signal, et en particulier de traitement d'images.

- **Compression :**

- **Sons :** Les algorithmes de compression audio utilisent les capacités de l'audition humaine, optimales à certaines fréquences bien définies. L'allocation des fréquences utilisées pour réaliser une transformée en ondelettes est choisie dans le spectre d'audition humaine. Les fréquences contenant l'énergie la plus importante (c'est-à-dire des coefficients d'ondelettes élevés) sont détectées, et une enveloppe est calculée autour de ces fréquences permettant de mettre à zéro les coefficients d'ondelettes dans le voisinage (Rioul et Vetterli, 1991; Strang et Nguyen, 1996). La transformée en ondelettes est également utilisée en audionumérique pour réaliser des tâches de séparation de sources sonores (Plumbley et al., 2010).
- **Images :** Les images peuvent être analysées et synthétisées grâce à des banques de filtres 2D. Les ondelettes traduisant les basses fréquences représentent des zones uniformes tandis que celles de hautes fréquences représentent les régions plus texturées. Après une décomposition en ondelettes, les standards de compression réalisent des allocations mémoires différentes selon l'entropie des différents coefficients. En effet, une séquence de bits est produite pour chaque coefficient. Plus le coefficient est probable, plus sa taille mémoire est faible (Rieder et Nossek, 1997). Les images, décomposées en coefficients d'ondelettes, peuvent également être compressées en ignorant les coefficients situés sous un certain seuil (Grgic et al., 2001). Ce seuil est déterminé par l'estimation du ratio signal sur bruit souhaité. La compression utilisant les ondelettes a été rapportée dans la littérature notamment pour l'imagerie médicale (Erickson et al., 1998), en imagerie satellite (Delaunay et al., 2010) ou encore pour la représentation d'empreintes digitales dans la base de données du FBI (Bradley et Brislawn, 1994). Cependant, c'est l'utilisation des ondelettes dans le standard de compression JPEG2000 qui est la plus connue (Bradley et Brislawn, 1994). Grâce à une étape préparatoire visant à découper l'image en tuiles rectangulaires, compressées indépendamment et à la transformée en ondelettes, JPEG2000 (Christopoulos et al., 2000) permet de représenter l'image avec une taille mémoire réduite et des performances supérieures aux autres standards (Usevitch, 2001). La FIGURE 2.13 montre la compression de la même image par JPEG et JPEG2000, des artefacts apparaissent sur la compression JPEG, car les hautes fréquences sont filtrées rendant les contours ou le texte flous.

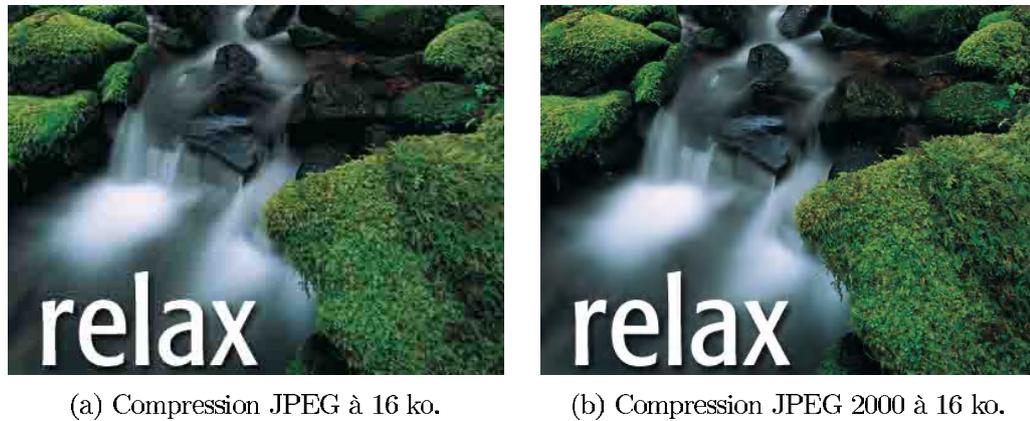


FIGURE 2.13 – Comparaison entre les compressions JPEG (a) et JPEG 2000 (b) <http://www.fnordware.com/j2k/jp2samples.html>.

- **Détection ou ajout d'informations :**

- **Segmentation et détection :** La détection de contours consiste à décomposer le signal en différents contenus fréquentiels. La transformée en ondelettes permet de réaliser ce découpage de manière efficace (Argoul et al., 1990; Shih et Tseng, 2005) (voir FIGURE 2.14). Les ondelettes permettent également de réaliser des détections de certaines cibles comme les visages (Garcia et Tzirittas, 1999; Zhou et Wei, 2006) ou les piétons (Oren et al., 1997; Mao et Chao, 2010) (voir FIGURE 2.15). Ce type d'applications est en plein développement pour les voitures autonomes.

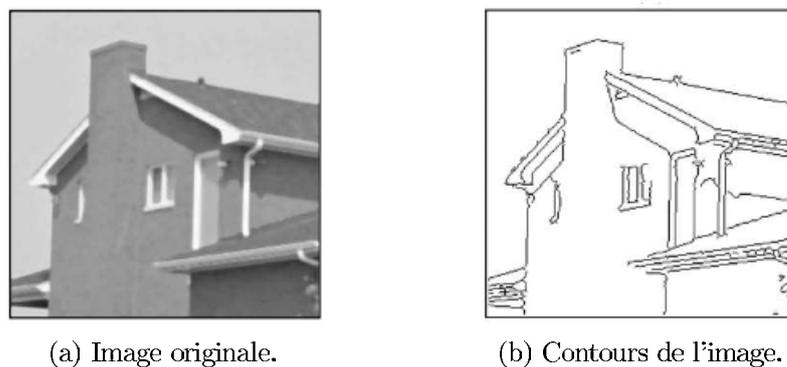


FIGURE 2.14 – Exemple de détection de contours utilisant la transformée en ondelettes (Shih et Tseng, 2005).

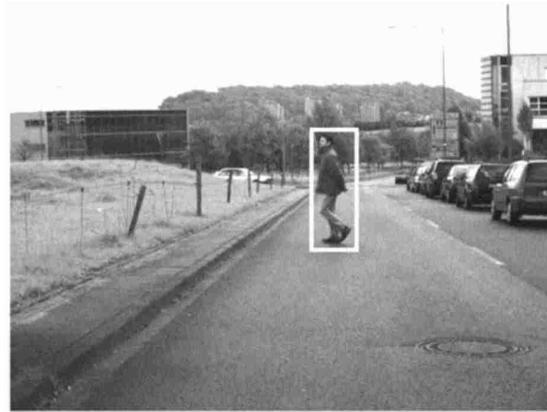


FIGURE 2.15 – Exemple de détection de piétons combinant les ondelettes et des machines à vecteur de support (Mao et Chao, 2010).

- **Tatouage numérique** : Cette technique permet d'ajouter des informations de copyright ou d'autres messages de vérification à une image ou à un autre document numérique (Martin et al., 2008). Le message inclus dans le signal hôte est constitué d'un ensemble de bits dont le contenu dépend de l'application. La transformée en ondelettes permet notamment de réaliser des tatouages invisibles, qui modifient le signal d'une manière imperceptible par l'utilisateur final (Kundur et Hatzinakos, 1998; Barni et al., 2001; Agreste et al., 2007). La localisation espace/fréquence des ondelettes et leur aspect multi-échelle leurs permettent d'être robustes aux attaques géométriques (consistant à retirer le tatouage de l'image). Enfin, elles rendent le tatouage compatible avec le standard JPEG2000.
  
- **Restauration d'image** :
  - **Fusion d'images** : Dans (Li et al., 1995; Nunez et al., 1999; Pajares et De La Cruz, 2004), la transformée en ondelettes est utilisée pour réaliser de la fusion d'images issues de modalités différentes. La transformée en ondelettes présente en effet de nombreux avantages pour ce type d'applications. Elle préserve l'information et permet de fusionner certains coefficients d'ondelettes, issues de différentes images. De plus, l'aspect multi-échelle permet de traiter des images de résolutions différentes. Cette application est particulièrement intéressante en imagerie médicale (voir FIGURE 2.16).

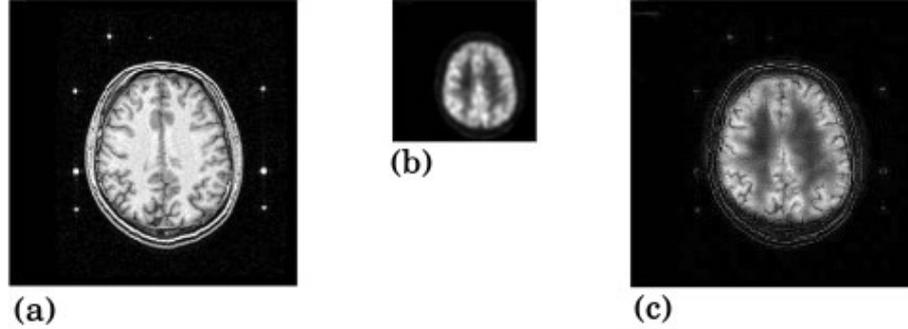


FIGURE 2.16 – Fusion (c) de deux images issues d’une IRM (a) et d’un PET-scan (b) grâce à la transformée en ondelettes (Pajares et De La Cruz, 2004).

- **Débruitage** : Les ondelettes sont également très utilisées en débruitage d’images par seuillage des coefficients d’ondelettes proches de zéro (Sardy et al., 2001; Xie et al., 2002; Chen et al., 2005). Ce seuillage s’appuie souvent sur la méthode de Dohono et Johnstone (Donoho et Johnstone, 1994), où le signal est décomposé en plusieurs niveaux d’ondelettes avant de subir un seuillage variant selon les niveaux. Deux types de seuillage existent, le seuillage brut et le seuillage doux. Ils sont définis par les relations suivantes :

$$y_{brut} = \begin{cases} x(t) & |x(t)| > a \\ 0 & |x(t)| \leq a \end{cases} \quad (2.20)$$

$$y_{doux} = \begin{cases} \text{signe}(x(t))(|x(t)| - a) & |x(t)| > a \\ 0 & |x(t)| \leq a \end{cases} \quad (2.21)$$

où  $a$  est la valeur de seuil. Le signal est ensuite reconstruit à partir des coefficients d’ondelettes restants après seuillage.

Il est toutefois important de signaler que la transformée en ondelettes n’est pas optimale pour modéliser les particularités *anisotropes* d’une image, c’est-à-dire les singularités curvilignes ou les zones fortement texturées. Ceci est dû au caractère *isotrope* de l’opérateur  $D_2$  défini en (2.16). Puisque  $D_2$  traite les deux axes de l’image de manière égale, les ondelettes 2D de hautes fréquences ne sont pas bien adaptées aux singularités autres que les singularités verticales, horizontales ou diagonales. Une décomposition récente fondée sur des opérateurs anisotropes a été proposée pour s’affranchir de la limitation de ne pas pouvoir traiter les informations anisotropes. Il s’agit de la transformée en *shearlets*, qui est décrite dans la sous-section suivante.

### 2.2.4 Transformée en shearlets

Contrairement aux transformées en ondelettes, les shearlets concernent uniquement les signaux 2D ou 3D mais ne sont pas adaptées aux signaux 1D. Nous verrons dans un premier temps la manière dont est constituée une base de shearlets 2D et ses applications en imagerie numérique. Nous aborderons également les différentes méthodes d'implémentation possibles des shearlets étudiées dans la littérature.

**Base de shearlets 2D :** Un système de shearlets peut être obtenu par dilatation anisotrope, translation et découpage d'un nombre fini de fonctions génératrices (Labate et al., 2005). Ces opérations sont réalisées dans  $L^2(\mathbb{R}^2)$  en appliquant un opérateur de translation défini en (2.17), un opérateur de dilatation anisotrope  $D_{\mathbf{A}}$  et un opérateur de découpage  $D_{\mathbf{S}_k}$ .

Soit  $|\det \mathbf{M}|$  la valeur absolue du déterminant d'une matrice  $\mathbf{M}$ , alors l'opérateur  $D_{\mathbf{M}}$  agit sur la fonction  $f \in L^2(\mathbb{R}^2)$  de variable  $\mathbf{u} = (u, v)$  correspondant aux coordonnées d'un pixel, de la façon suivante :

$$D_{\mathbf{M}}(f(\mathbf{u})) = |\det \mathbf{M}|^{1/2} f(\mathbf{M}\mathbf{u}), \text{ où } \mathbf{M} \in \mathbb{R}^{2 \times 2} \quad (2.22)$$

Le choix de la matrice  $\mathbf{M}$  détermine alors si l'opérateur  $D_{\mathbf{M}}$  est une dilatation ou un découpage.

**Définition 2.13. Opérateur de dilatation anisotrope :** En choisissant  $\mathbf{M}$  comme une matrice dite de dilatation anisotrope  $\mathbf{A}$  ou  $\tilde{\mathbf{A}}$ , on obtient les opérateurs de dilatation anisotrope  $D_{\mathbf{A}}$  et  $D_{\tilde{\mathbf{A}}}$ . Deux matrices sont nécessaires car la transformée en shearlets repose sur l'élaboration de deux cônes horizontal et vertical dans l'espace fréquentiel. Les matrices  $\mathbf{A}$  et  $\tilde{\mathbf{A}}$  les plus couramment utilisées sont :

$$\mathbf{A} = \begin{pmatrix} 2 & 0 \\ 0 & \sqrt{2} \end{pmatrix}, \text{ et } \tilde{\mathbf{A}} = \begin{pmatrix} \sqrt{2} & 0 \\ 0 & 2 \end{pmatrix}, \quad (2.23)$$

**Définition 2.14. Opérateur de découpage :** En choisissant  $\mathbf{M}$  comme une matrice de découpage ("shear matrice" en anglais)  $\mathbf{S}_k$ , on obtient l'opérateur de découpage  $D_{\mathbf{S}_k}$ .  $\mathbf{S}_k$  est définie par :

$$\mathbf{S}_k = \begin{pmatrix} 1 & k \\ 0 & 1 \end{pmatrix}, \text{ où } k \in \mathbb{Z}, \quad (2.24)$$

Comme nous pouvons le voir sur la FIGURE 2.17, grâce à la dilatation anisotrope, les coefficients de shearlets de hautes fréquences présents sur la dernière ligne de la FIGURE 2.17 (b) sont allongés et mieux adaptés pour représenter les informations pertinentes d'une image (par exemple, les contours) tandis que ceux des hautes fréquences des ondelettes, illustrées par la dernière ligne de la FIGURE 2.17 (a), sont beaucoup moins localisés.

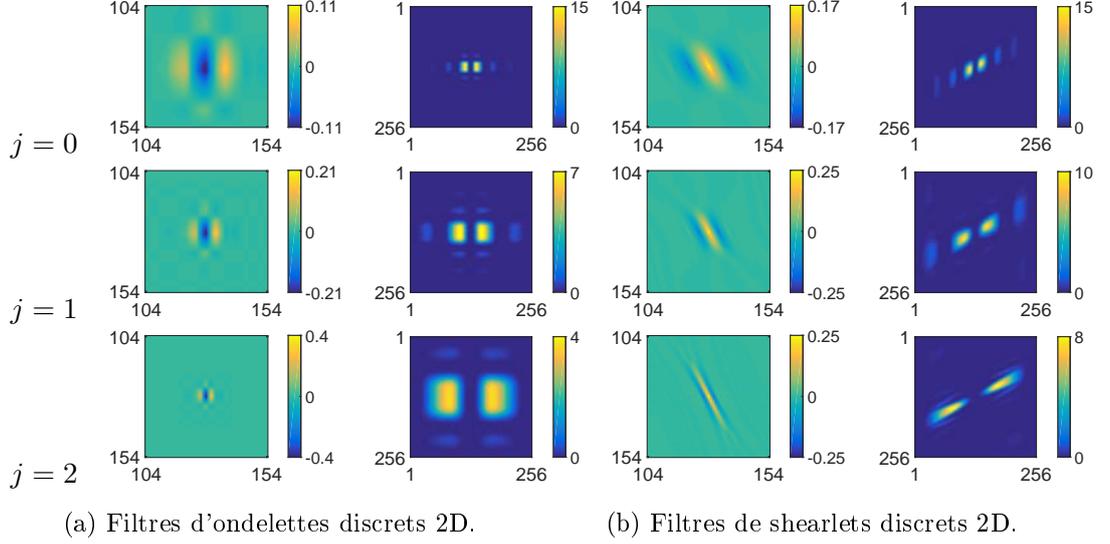


FIGURE 2.17 – Comparaison entre les filtres discretes 2D des ondelettes et des shearlets. Le domaine spatial est représenté sur les colonnes de gauche tandis que celui des fréquences est donné sur les colonnes de droite.

**Définition 2.15. Système conique de shearlets :** Pour une fonction d'échelle 2D  $\phi^{(1)} \in L^2(\mathbb{R}^2)$  et pour des fonctions génératrices de shearlets  $\psi^{(1)}, \psi^{(2)} \in L^2(\mathbb{R}^2)$ , un système conique de shearlets peut être défini de manière analogue à celui des ondelettes présenté en (2.18). Ce système est le suivant :

$$\begin{aligned}
 & \left\{ \phi_{\mathbf{m}}^{(1)} = T_{\mathbf{m}} \phi^{(1)} : \mathbf{m} \in \mathbb{Z}^2 \right\} \cup \\
 & \left\{ \psi_{j,k,\mathbf{m}}^{(1)} = D_{\mathbf{A}}^j D_{\mathbf{S}_k} T_{\mathbf{m}} \psi^{(1)} : j \in \mathbb{N}_0, |k| < \left\lceil 2^{\frac{j}{2}} \right\rceil, \mathbf{m} \in \mathbb{Z}^2 \right\} \cup \\
 & \left\{ \psi_{j,k,\mathbf{m}}^{(2)} = D_{\tilde{\mathbf{A}}}^j D_{\tilde{\mathbf{S}}_k} T_{\mathbf{m}} \psi^{(2)} : j \in \mathbb{N}_0, |k| < \left\lceil 2^{\frac{j}{2}} \right\rceil, \mathbf{m} \in \mathbb{Z}^2 \right\},
 \end{aligned} \tag{2.25}$$

avec  $\lceil \cdot \rceil$  la fonction de seuillage.

La condition  $|k| < \left\lceil 2^{\frac{j}{2}} \right\rceil$  empêche les orientations de l'opérateur de découpage de dépasser  $45^\circ$ . Cela force le support des shearlets, dans le domaine fréquentiel, à rester dans les cônes dans lesquels la fonction génératrice correspondante est localisée. Ainsi, pour couvrir tout le domaine fréquentiel, il est important de choisir des fonctions génératrices  $\psi^{(1)}$  et  $\psi^{(2)}$  portées par des cônes différents. Le découpage du cône de fréquences horizontal induit par le système (2.25) est représenté schématiquement sur la FIGURE 2.18.

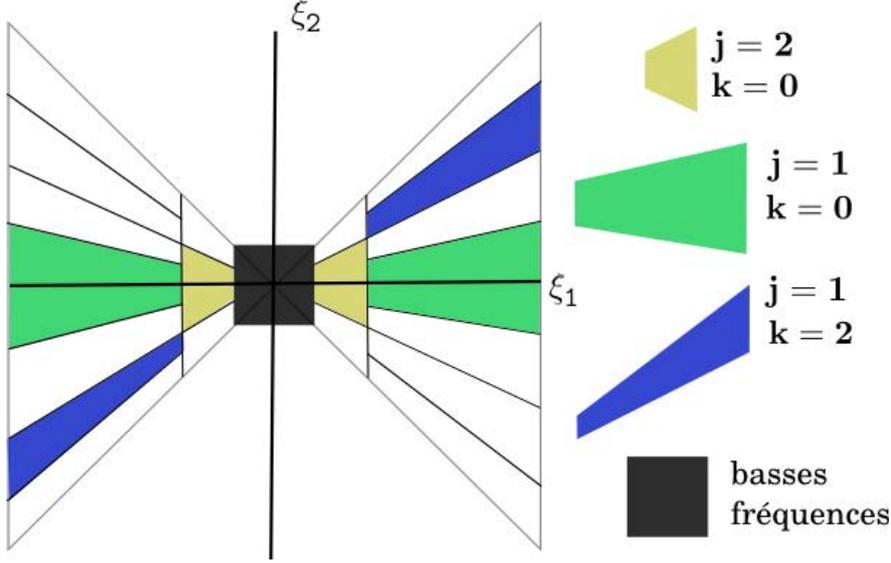


FIGURE 2.18 – Découpage du cône de fréquences horizontal induit par le système des shearlets (2.25).  $\xi_1$  et  $\xi_2$  représentent les fréquences selon chacun des axes  $x$  ou  $y$  d'une image.

**Définition 2.16. Transformée en shearlets 2D :** Comme dans le cas des ondelettes, la transformée en shearlets  $\mathcal{S}_\psi f$  d'une fonction  $f \in L^2(\mathbb{R}^2)$  selon une fonction génératrice  $\psi \in L^2(\mathbb{R}^2)$ , supportée essentiellement sur le cône des fréquences horizontal est donnée au niveau  $j \in \mathbb{Z}$ , avec un paramètre de découpage  $|k| < \lceil 2^{\frac{j}{2}} \rceil$  et un paramètre de translation  $\mathbf{m} \in \mathbb{Z}^2$  par le produit scalaire :

$$(\mathcal{S}_\psi f)(j, k, \mathbf{m}) = \langle f, \psi_{j,k,\mathbf{m}} \rangle = \iint_{\mathbb{R}^2} f(u, v) \overline{D_{\mathbf{A}}^j D_{\mathbf{S}_k} T_{\mathbf{m}} \psi(u, v)} du dv. \quad (2.26)$$

Pour les fonctions génératrices d'ondelettes qui sont dans le cône vertical, les matrices  $\mathbf{A}$  et  $\mathbf{S}_k$  de l'équation (2.26) sont remplacées par  $\tilde{\mathbf{A}}$  et  $\mathbf{S}_k^\top$  car les supports des opérateurs ne sont plus les mêmes. Pour plus de détails sur la théorie et l'application des shearlets, nous invitons le lecteur à se référer à (Kutyniok et Labate, 2012a).

## 2.2.5 Différentes implémentations des shearlets

Durant la dernière décennie, plusieurs bibliothèques ont été développées pour implémenter différents types de transformées en shearlets :

- "Fast Finite Shearlet Transform" (FFST) (Häuser et Steidl, 2012), qui implémente une transformée en shearlets non sous-échantillonnée dite "band-limited",
- ShearLab (Kutyniok et al., 2012), qui implémente une version sous-échantillonnée de la transformée en shearlets dite "compactly supported",

- ShearLab 3D (Kutyniok et al., 2016), qui contient une implémentation 2D mais aussi 3D de la transformée en shearlets non sous-échantillonnée dite "*compactly supported*".

La FIGURE 2.19 montre le niveau de décomposition  $j = 0$  de la transformée en shearlets tandis que les FIGURES 2.20 et 2.21 présentent les coefficients d'une transformée en shearlets non sous-échantillonnée d'une image de taille  $256 \times 256$  pixels résultant de deux niveaux de décompositions ( $j = 1, j = 2$ ). Le paramètre  $j_{max}$  correspond au nombre de niveaux utilisés dans la transformée en shearlets et le paramètre de découpage  $k$  contrôle son orientation. Toutes les matrices générées font la même taille que l'image de départ. La comparaison avec les FIGURES 2.23 et 2.24 permet de voir la différence avec une transformée sous-échantillonnée. L'avantage des shearlets sous-échantillonnées est que le nombre total de coefficients reste du même ordre de grandeur que le nombre de pixels de l'image de départ, tandis que les transformations non sous-échantillonnées ont beaucoup plus de coefficients. Par exemple, pour une image de  $256 \times 256$  pixels, la transformée en shearlets non sous-échantillonnée avec le paramètre  $j_{max} = 2$  possède 851968 coefficients contre 151552 pour sa version sous-échantillonnée. Il est important de noter qu'une transformée en ondelettes classique aurait eu 65536 coefficients indépendamment du nombre de décompositions.

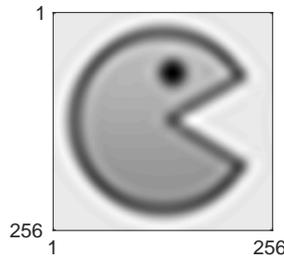


FIGURE 2.19 – Représentation basses fréquences ( $j = 0$ ) de l'image initiale obtenue par la fonction d'échelle de l'approche non sous-échantillonnée.

De manière générale, avec les ondelettes ou les shearlets, le niveau  $j = 1$  est toujours le plus détaillé (correspondant aux plus hautes fréquences), suivi par le niveau  $j = 2$ , etc. En revanche, nous utiliserons toujours, dans la suite, le niveau  $j = 0$  pour désigner l'approximation basses fréquences de l'image.

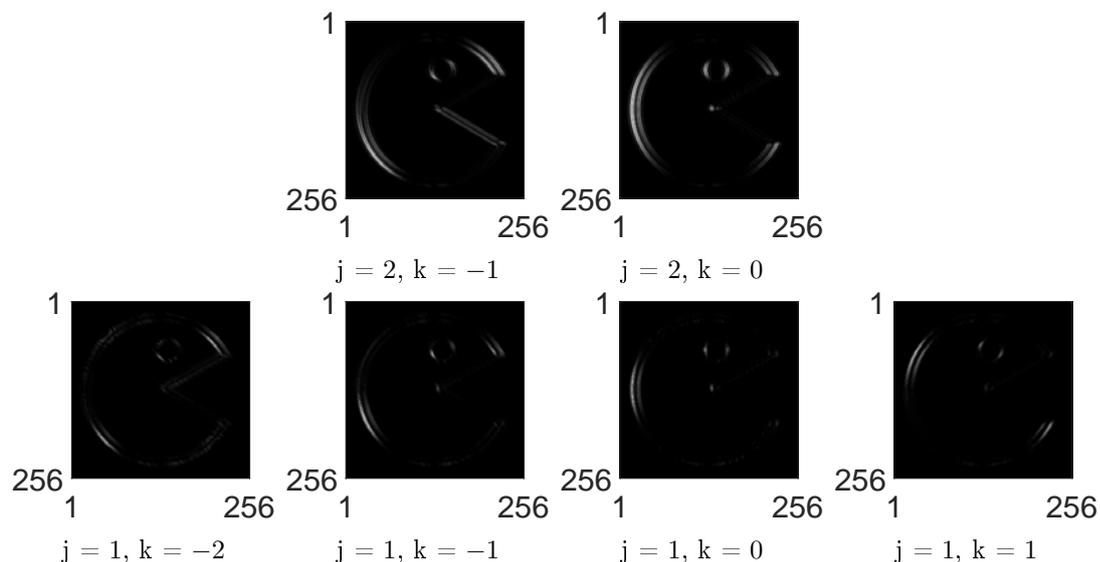


FIGURE 2.20 – Transformées en shearlets non sous-échantillonnées ( $j_{max} = 2$ ) issues du cône horizontal de fréquences avec des niveaux  $j$  et des découpages  $k$  différents.

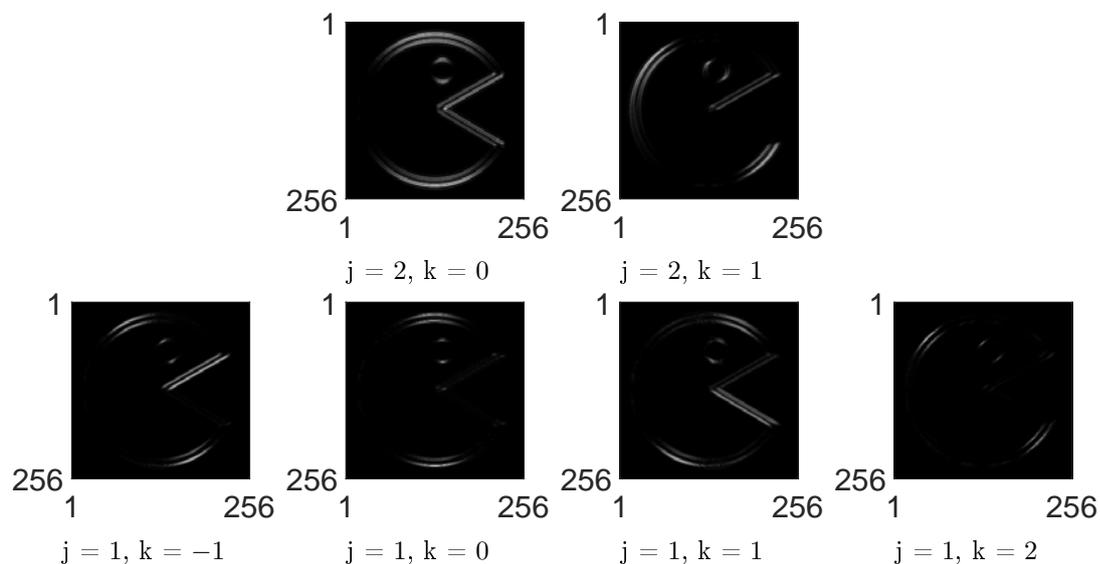


FIGURE 2.21 – Transformées en shearlets non sous-échantillonnées ( $j_{max} = 2$ ) issues du cône vertical de fréquences avec des niveaux  $j$  et des découpages  $k$  différents.

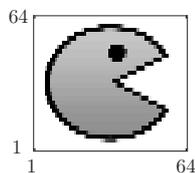


FIGURE 2.22 – Représentation basses fréquences ( $j = 0$ ) de l'image initiale, obtenue par la fonction d'échelle de l'approche sous-échantillonnée.

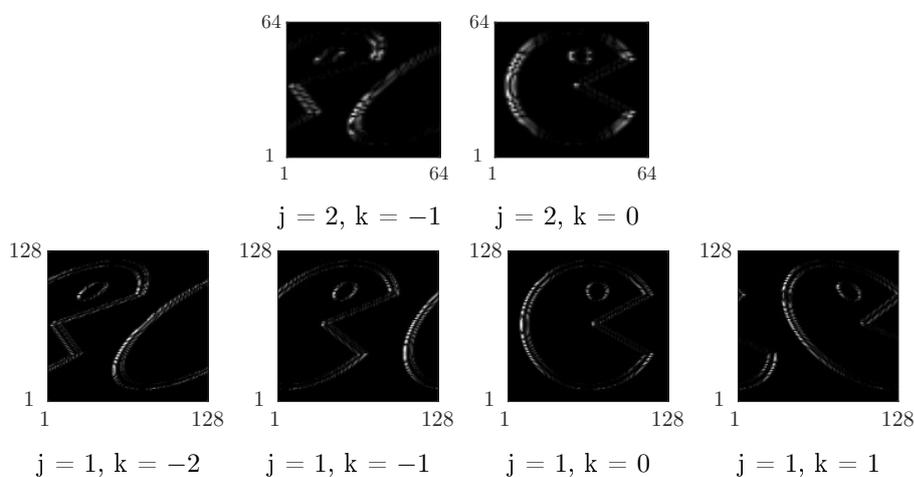


FIGURE 2.23 – Transformées en shearlets sous-échantillonnées ( $j_{max} = 2$ ) issues du cône horizontal de fréquences avec des niveaux  $j$  et des découpages  $k$  différents.

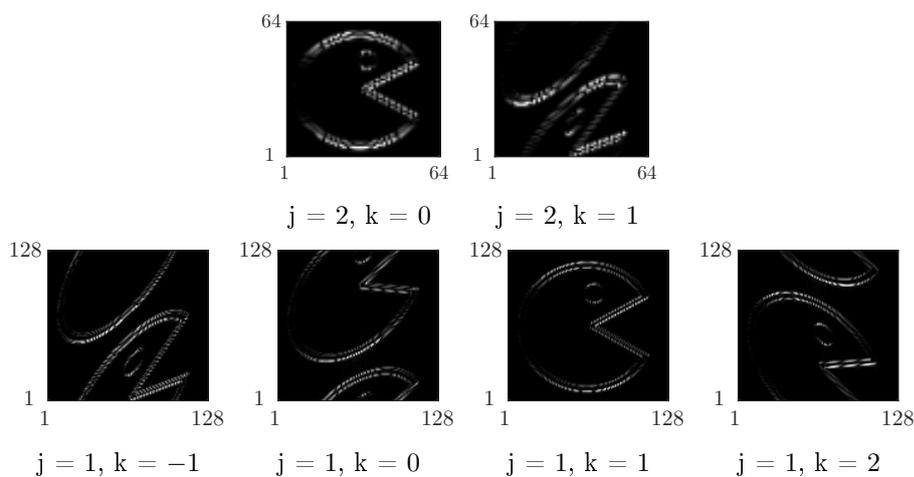


FIGURE 2.24 – Transformées en shearlets sous-échantillonnées ( $j_{max} = 2$ ) issues du cône vertical de fréquences avec des niveaux  $j$  et des découpages  $k$  différents.

### 2.2.6 Applications de la transformée en shearlets

Les applications des shearlets sont semblables à celles des ondelettes, hormis pour les signaux sonores. En effet, les shearlets sont spécialement conçues pour traiter les signaux 2D.

- **Détection ou ajout d'information :**

- **Segmentation et détection :** De nombreuses applications requièrent la détection automatique de certaines informations, en particulier les contours ou les points d'intérêt des images. L'orientation des contours peut être une information d'importance et la représentation d'une image en shearlets est spécialement conçue pour capturer l'information de structures anisotropes. La première approche permettant d'obtenir la localisation et l'orientation des contours d'une image à l'aide des shearlets a été développée par Yi et ses collègues (Yi et al., 2009). La FIGURE 2.25 montre que les performances de la segmentation de contour utilisant les shearlets surpassent celles de l'approche utilisant les ondelettes. La prise en compte des valeurs complexes dans la transformée en shearlets a également été considérée pour la détection de flammes (Reisenhofer et al., 2016).

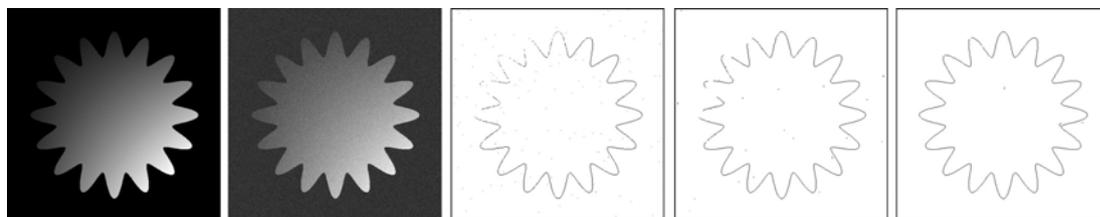


FIGURE 2.25 – Résultats de détection de contours. De gauche à droite : Image initiale, image bruitée, détection avec filtre de Sobel, avec les ondelettes, avec les shearlets (Yi et al., 2009).

- **Tatouage numérique :** Comme vu dans les applications des ondelettes, les images numériques peuvent être recopiées illégalement très facilement. Une méthode pour assurer leur authenticité est d'utiliser un tatouage numérique. Ce tatouage ne doit pas changer l'apparence de l'image. De plus, l'algorithme utilisé pour ce tatouage doit être réversible. Des approches utilisant les shearlets ont donc été développées à cette fin ces dernières années (Ahmaderaghi et al., 2014; Mardanpour et Chahooki, 2016).
- **Restauration d'images :**

- **Séparation d'images :** Les images contiennent parfois différentes classes de composants qui diffèrent par leur structure morphologique, telles que les points et les courbes. Dans de nombreuses applications, il est nécessaire de séparer ces caractéristiques. En imagerie médicale, les neurones sont, par

exemple, composés d'épines et de dendrites. Les méthodes récentes d'analyse harmonique ont abouti à des résultats convaincants. Les différents composants peuvent être séparés à l'aide de la parcimonie de l'image dans des représentations multi-échelles et d'algorithmes de minimisation en norme  $l_1$  (voir FIGURE 2.26). Les transformées en shearlets sont spécifiquement construites pour représenter efficacement les données anisotropes. Leur combinaison avec les systèmes d'ondelettes, qui révèlent bien les singularités ponctuelles, forme un choix naturel pour cette tâche (Kutyniok et Lim, 2010).

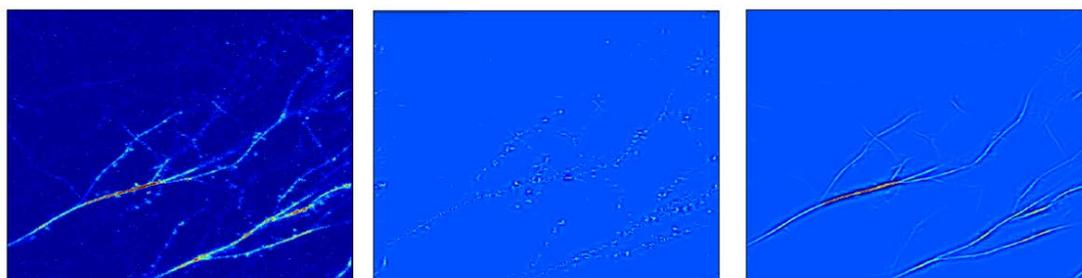


FIGURE 2.26 – Séparation d'une image représentant des neurones. De gauche à droite : images initiales, épines extraites de l'image, dendrites extraites de l'image (Kutyniok et Lim, 2010).

- **Débruitage :** Comme nous l'avons déjà mentionné, les images peuvent être débruitées à partir de leur transformées en ondelettes, en seuillant les coefficients des ondelettes puis en appliquant la transformée inverse. Cependant, cette méthode a tendance à produire des artefacts autour de discontinuités pointues. Pour surmonter ce problème, les shearlets ont été utilisées à la place des ondelettes dans des algorithmes plus sophistiqués (Easley et al., 2009; Chen et al., 2013). Dans (Kutyniok et al., 2016), de nombreuses applications des shearlets sont rapportées notamment pour le débruitage d'images avec diverses représentations non sous-échantillonnées (ondelettes, curvelets, shearlets,...). Les résultats d'une étude comparative entre l'utilisation des ondelettes et de deux implémentations différentes des shearlets : FFST et Shearlab3D y sont présentées (voir FIGURE 2.27). Le score *Peak Signal Noise Ratio* (PSNR) est utilisé pour effectuer cette comparaison. Plus ce score est élevé, plus la reconstruction obtenue est proche de l'image initiale. L'implémentation des shearlets joue un rôle dans les performances des algorithmes utilisés. Pour cette raison, nous avons fait le choix d'utiliser la librairie Shearlab3D dans le contexte de l'asservissement visuel. Ces travaux seront discutés dans les chapitres suivants.



(a) Image bruitée, PSNR=16.06



(b) Reconstruction avec des ondelettes, PSNR=23.55



(c) Reconstruction avec les shearlets de la librairie FFST, PSNR=25.49



(d) Reconstruction avec les shearlets de la librairie Shearlab3D, PSNR=26.28

FIGURE 2.27 – Comparaison de la reconstruction d’une image en utilisant les ondelettes ou les shearlets (Kutyniok et al., 2016).

## 2.3 Conclusion

Ce chapitre nous a permis de nous familiariser avec la transformée de Fourier standard et certaines autres méthodes inspirées de celle-ci comme les travaux de Gabor. Nous pouvons souligner que Gabor a apporté une première méthode fournissant une représentation temps-fréquence d'un signal. Ceci grâce à l'introduction du concept de la fenêtre glissante dans la transformée de Fourier standard. Cependant, l'utilisation d'une fenêtre glissante, dont la taille est fixée au préalable, pose le problème de la représentation des signaux transitoires. En effet, si la taille de la fenêtre est faible, elle ne permet pas de localiser les basses fréquences et inversement. En conséquence, le choix de la fenêtre s'est avéré être un dilemme important. Les ondelettes sont apparues par la suite pour y remédier. En effet, la transformée en ondelettes a l'avantage de fournir une représentation à la fois spatiale et fréquentielle du signal avec des fenêtres de tailles variables. La transformée en ondelettes nous donne ainsi une représentation encore plus hiérarchique des informations contenues dans un signal, notamment parce qu'elle forme une analyse multi-résolution. Son application la plus célèbre est le standard de compression JPEG2000. De nombreuses transformées, dérivées des ondelettes, sont ensuite apparues. L'une d'entre elles, la transformée en shearlets, a particulièrement retenu notre attention. Cette transformée, spécialement conçue pour détailler les singularités des signaux 2D (les images) a montré sa précision en détection de contours ou en débruitage d'image. Son utilisation dans une tâche de commande par asservissement visuel utilisant les images échographiques sera le thème du prochain chapitre.



## Chapitre 3

# Asservissement visuel utilisant les shearlets appliqué à l'imagerie échographique

Nous avons pu voir dans le chapitre précédent que les shearlets étaient très robustes au bruit. En général, l'imagerie médicale est caractérisée par un rapport signal/bruit défavorable. Ceci est d'autant plus vrai lorsqu'il s'agit de l'échographie ou encore de la tomographie par cohérence optique (OCT). Ces deux modalités seront présentées dans une première section. Ensuite, nous verrons dans une seconde section comment nous avons mis au point une commande fondée sur les shearlets avec une estimation numérique de la matrice d'interaction. Enfin, une troisième section présentera une méthode permettant l'élaboration analytique de la matrice d'interaction pour une commande d'asservissement visuel. Les deux méthodes seront évaluées en simulation et sur un système de validation expérimental utilisant l'imagerie échographique. Le sommaire ci-dessous présente la structure de ce chapitre :

### Sommaire

---

<b>3.1 Imagerie médicale</b> . . . . .	<b>70</b>
3.1.1 Imagerie échographique . . . . .	70
3.1.2 Tomographie par cohérence optique . . . . .	80
<b>3.2 Matrice d'interaction numérique</b> . . . . .	<b>85</b>
3.2.1 Concept et implémentation . . . . .	85
3.2.2 Résultats de simulation . . . . .	87
3.2.3 Résultats expérimentaux . . . . .	90
<b>3.3 Matrice d'interaction analytique</b> . . . . .	<b>95</b>
3.3.1 Concept et implémentation . . . . .	95
3.3.2 Résultats de simulation . . . . .	97
3.3.3 Résultats expérimentaux . . . . .	104
<b>3.4 Conclusion</b> . . . . .	<b>106</b>

---

### 3.1 Imagerie médicale

Voir à l'intérieur du corps humain a toujours été un enjeu majeur pour la médecine mais il aura fallu attendre 1895 pour que Wilhem Röntgen réalise la première radiographie après avoir étudié les rayons X. De nombreuses autres techniques ont été ensuite développées tout au long du 20<sup>ème</sup> siècle. La FIGURE 3.1 présente une frise chronologique indiquant les dates d'introduction de ces différentes techniques. Certaines sont invasives et donc à utiliser avec parcimonie, comme le scanner ou la tomographie par émission de positons (TEP-scan) tandis que d'autres sont non-invasives, en particulier l'imagerie par résonance magnétique (IRM), l'échographie ou encore la tomographie par cohérence optique pour ne citer que celles-ci.

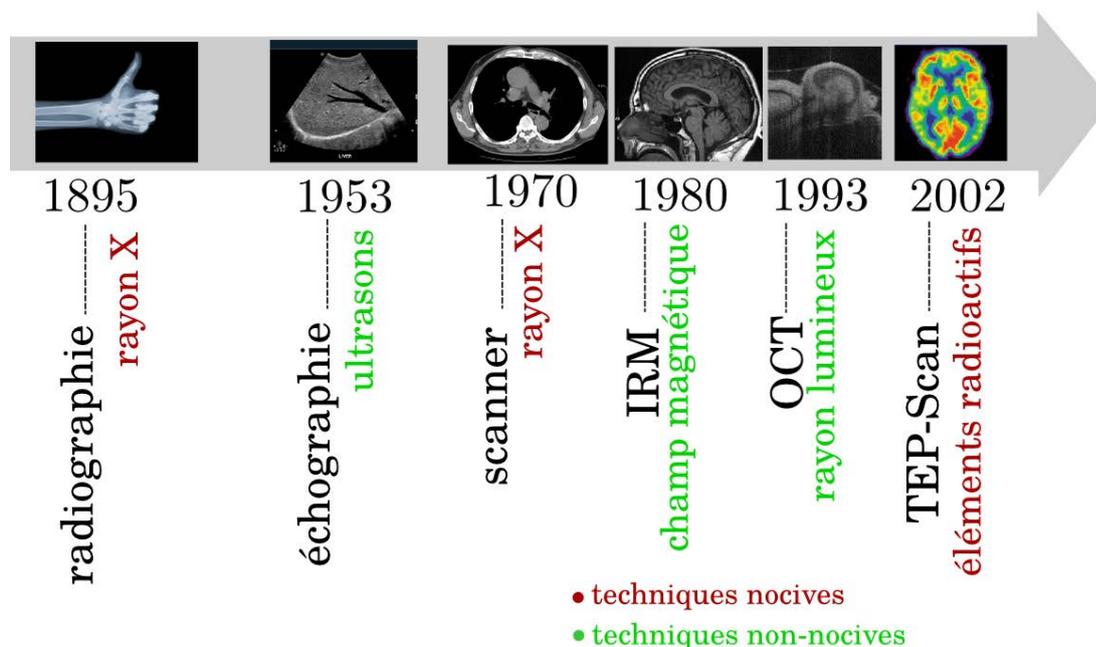


FIGURE 3.1 – Frise chronologique présentant les dates d'introduction des différentes techniques d'imagerie médicale.

Ces différentes techniques présentent de grandes disparités en termes de résolution et de profondeur de pénétration comme le montre la FIGURE 3.2. Nous allons particulièrement nous intéresser à l'imagerie échographique et à l'imagerie OCT qui ont été utilisées expérimentalement pour valider certains travaux décrits dans ce manuscrit.

#### 3.1.1 Imagerie échographique

L'imagerie échographique dite aussi ultrasonore est une technique d'imagerie non-invasive très utilisée dans plusieurs domaines de la médecine. Elle est obtenue à l'aide d'un échographe (voir FIGURE 3.3). Cet appareil est constitué d'une sonde effectuant l'émission et la réception d'ondes ultrasonores, un système informatique qui génère une

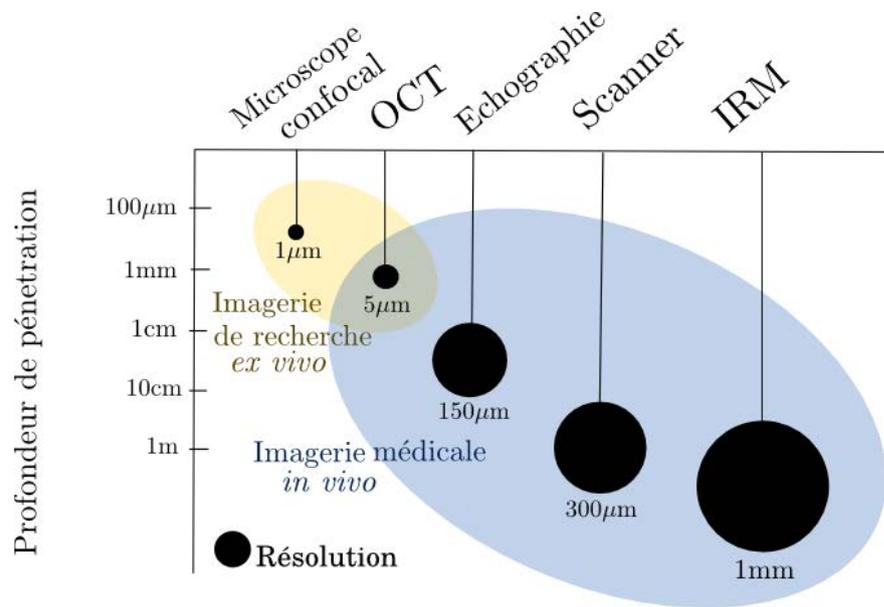


FIGURE 3.2 – Résolution et profondeur de pénétration des différentes techniques d'imageries médicales.

image à partir de la mesure du décalage temporel entre les ondes envoyées et leurs échos captés, une console de commande et un moniteur.

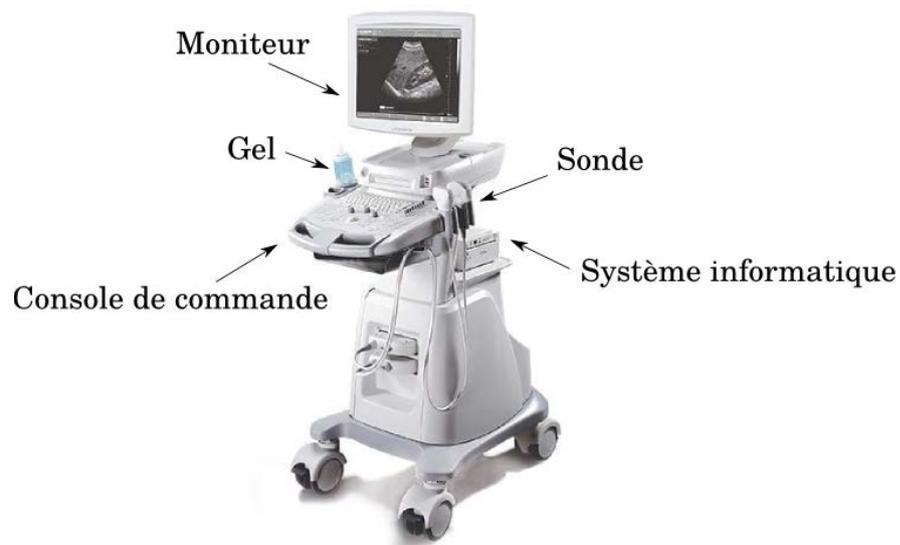


FIGURE 3.3 – Photographie d'une station échographique médicale.

Historiquement, l'étude des ultrasons a été largement développée durant la première guerre mondiale afin de détecter les sous-marins (Chilowsky et Langevin, 1916). Si quelques travaux ont tenté d'utiliser les ultrasons en imagerie médicale (Dussik, 1942; Ludwig et Struthers, 1949), ce n'est qu'en 1952 que le médecin Wild et l'électronicien Reid ont inventé l'échographe tel qu'on le connaît actuellement, en l'appliquant dans le cadre de la recherche de tumeurs cérébrales (Wild et Reid, 1952). Son usage courant en obstétrique est apparu dans les années 1970.

**Principe de formation de l'image ultrasonore :** L'élément de base d'une sonde échographique est le transducteur piézoélectrique qui, associé à un balayage électronique, permet d'obtenir une image 2D, techniquement appelée B-scan. La piézoélectricité est un phénomène permettant la production d'électricité à partir de certains matériaux qui se polarisent sous l'action de contraintes mécaniques et réciproquement se déforment sous l'application d'un champ électrique (Curie, 1880). Certains minéraux comme le quartz, comportant des agencements atomiques réguliers (isotropes), sont en effet susceptibles de produire une différence de potentiel sous l'effet d'une déformation mécanique et à l'inverse se déforment s'ils sont soumis à une différence de potentiel. Ce phénomène est utilisé en électronique pour capter et transformer les sons en signaux électriques (microphone) ou pour convertir des signaux électriques en vibrations sonores (écouteurs). Ce type de matériaux est également largement utilisé dans la conception de micro-actionneurs à haute résolution. Les transducteurs piézoélectriques positionnés sur l'extrémité de la sonde échographique jouent à la fois le rôle d'émetteur et de récepteur de l'onde ultrasonore. L'émission d'une onde ultrasonore se fait successivement sur chaque transducteur le long d'un ensemble de lignes de tir balayant l'image. Les échos enregistrés s'atténuent en fonction des obstacles rencontrés par l'onde incidente. Afin de supprimer la couche d'air qui sépare le tissu de la sonde, du gel échographique est utilisé.

La célérité  $c$  du son est reliée à la fréquence  $f$  et à la longueur d'onde  $\Lambda$  de l'onde sonore par la relation :

$$c = \Lambda f \quad (3.1)$$

Des fréquences entre 7 et 20 MHz sont utilisées généralement pour observer les structures superficielles et des fréquences comprises entre 1 et 6 MHz permettent d'atteindre les structures plus profondes, au détriment de la résolution de l'image. En pénétrant dans les différentes couches des tissus, les ondes ultrasonores subissent un phénomène de réflexion dû à la présence d'une interface entre deux tissus. En considérant que la longueur d'onde est petite par rapport à la structure rencontrée, l'interaction à la frontière de deux milieux possédant des propriétés acoustiques différentes peut être décrite par la loi de Snell-Descartes.

**Définition 3.1. Lois de Snell-Descartes :** Soit un milieu 1 avec une vitesse de propagation du son  $c_1$  et un milieu 2 avec une vitesse de propagation du son  $c_2$ . Soit une onde incidente arrivant sur l'interface entre les deux milieux avec un angle  $\theta_i$  par rapport à la normale au plan d'incidence (voir FIGURE 3.4).

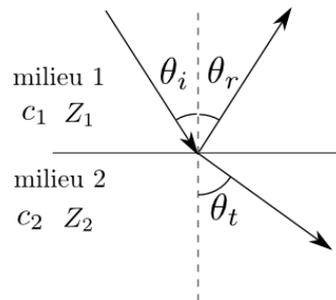


FIGURE 3.4 – Réflexion et transmission d'une onde sonore à l'interface entre deux milieux.

- *Première loi : l'onde incidente, l'onde réfléchie et l'onde transmise appartiennent au même plan.*
- *Deuxième loi : l'angle  $\theta_r$  de l'onde réfléchie par rapport à la normale est égale à l'angle  $\theta_i$ .*
- *Troisième loi : la relation entre l'angle  $\theta_t$  de l'onde transmise par rapport à la normale et  $\theta_i$  est :*

$$\frac{\sin \theta_i}{c_1} = \frac{\sin \theta_t}{c_2} \quad (3.2)$$

Lorsque l'onde incidente d'amplitude  $A_i$  rencontre une interface entre deux milieux d'impédance acoustique  $Z_1$  et  $Z_2$ , un écho d'intensité initiale  $A_r$  est renvoyé vers le capteur selon la formule suivante :

$$A_r = R A_i \quad \text{avec} \quad R = \left( \frac{Z_2 - Z_1}{Z_2 + Z_1} \right)^2 \quad (3.3)$$

L'interface tissus mous/air présente un coefficient de réflexion de 99 %. Avec un tel taux de réflexion, l'énergie acoustique transmise est alors trop faible pour permettre l'exploration des structures sous-jacentes. L'utilisation du gel échographique permet de résoudre ce problème. Dans le cas où l'onde incidente rencontre une interface tissus mous/os, le coefficient de réflexion peut atteindre une valeur importante de 30 % qui ne permet également pas de transmettre suffisamment d'énergie acoustique nécessaire à l'imagerie des tissus derrière l'os. Il en résulte la formation d'une zone d'ombre dans l'image. En revanche, lorsque le coefficient  $R$  est inférieur à 10%, l'intensité de l'onde transmise est suffisante pour permettre l'imagerie des structures plus profondes. Par exemple, l'interface rein/graisse présente un coefficient de réflexion d'environ 1 %.

Lorsque l'onde ultrasonore rencontre une interface de petite taille par rapport à sa longueur d'onde, un phénomène de diffusion d'énergie est engendré : l'onde incidente est renvoyée dans toutes les directions. Par conséquent, les échos réceptionnés par les transducteurs piézoélectriques proviennent à la fois de la réflexion des ondes ultrasonores et de ce phénomène de diffusion, qui provoque un aspect granulaire sur les images

échographiques, appelé *speckle*, visible sur la FIGURE 3.5. Cet aspect granulaire constitue un inconvénient majeur de l'utilisation des images échographiques en asservissement visuel.

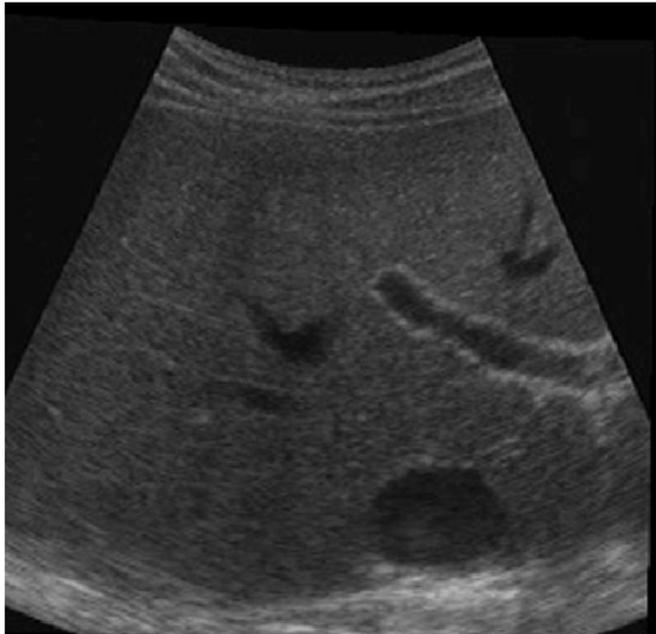


FIGURE 3.5 – Image échographique obtenue sur une maquette d'abdomen en gélatine.

**Asservissement visuel échographique :** Afin de pallier aux inégalités géographiques concernant l'accessibilité aux soins médicaux ou d'éviter à des patients à mobilité réduite de trop nombreux allers-retours dans un centre d'imagerie médicale, les moyens de télé-médecine sont de plus en plus plébiscités et notamment la télé-échographie. L'idée est de pouvoir faire un diagnostic à distance à l'aide d'un système robotique (Sublett et al., 1995). C'est en 2001 qu'une équipe de Vancouver a pour la première fois, utilisé l'imagerie échographique pour asservir un robot porteur d'une sonde ultrasonore (Abolmaesumi et al., 2001). Dans ces travaux, seuls les 3 ddl du plan de l'image échographique sont asservis, les trois autres étant commandés par téléopération. La tâche consistait à centrer la section de l'artère carotide dans l'image dans le but de réaliser une imagerie échographique 3D de cette dernière.

Nous pouvons distinguer plusieurs type de travaux, ceux portant sur le contrôle de la sonde (configuration embarquée) ou ceux portant sur le contrôle d'un instrument médical (configuration déporté). De plus, l'imagerie échographique avec une sonde 2D permet seulement d'obtenir une coupe dans un plan d'observation, les travaux d'asservissement échographique sur sonde 2D se distinguent également selon la nature du mouvement, uniquement dans le plan ou possédant des composantes hors plan.

Pour le guidage de la sonde dans le plan d'observation, les travaux présentés dans (Abolmaesumi et al., 2001) ont été élargis dans (Abolmaesumi et al., 2002). Plusieurs

algorithmes de détection  $y$  sont comparés. Le système d'imagerie échographique robotisée utilisé dans ce travail est présenté sur la FIGURE 3.6 (a) et le résultat de la détection de la section d'artère carotide avec la méthode Star-Kalman est présenté en FIGURE 3.6 (b). Des méthodes de détection de contours type "snake" ont également été utilisées dans (Krupa et al., 2016) pour maintenir la visibilité d'un élément anatomique d'intérêt par asservissement visuel.

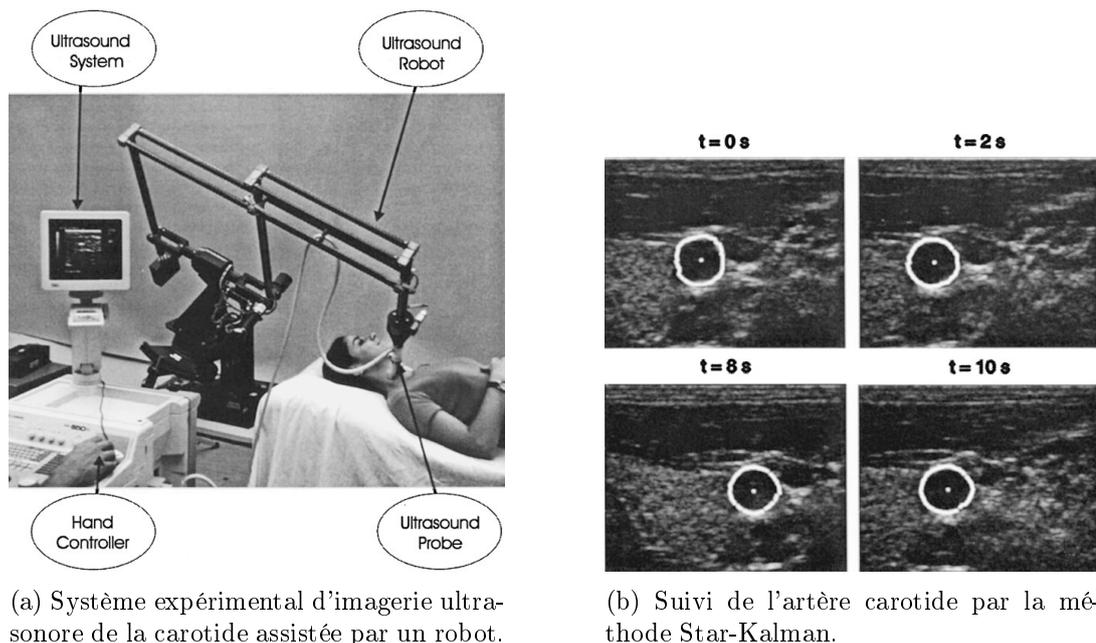


FIGURE 3.6 – Illustration d'une approche utilisant des primitives visuelles géométriques pour une tâche d'asservissement visuel échographique (Abolmaesumi et al., 2002).

Le guidage de la sonde a également été considéré avec des déplacements dans et en dehors du plan d'observation de la coupe échographique obtenue. Dans (Lee et al., 2007), l'application consiste à dissoudre un calcul rénal à l'aide d'un transducteur ultrasonique de forte intensité par une méthode de focalisation d'énergie ultrasonore. Un asservissement visuel échographique, utilisant deux coupes orthogonales dont l'intersection contient le point focal des ultrasons focalisés, est mis en œuvre pour une commande à 3 ddl. Le but étant de suivre les mouvements du calcul rénal, dus aux mouvements physiologiques du patient. Les informations visuelles utilisées sont les coordonnées 3D du centre du calcul rénal.

D'autres approches se servent de l'imagerie échographique pour faire du guidage d'outils. En effet, le diagnostic de nombreux cancers passe d'abord par une biopsie, notamment pour le cancer du sein (Mallapragada et al., 2011; Kobayashi et al., 2012) mais surtout dans le cas du cancer de la prostate, nécessitant l'utilisation d'une sonde trans-rectale (Han et al., 2012; Vitrani et al., 2016). Dans (Hong et al., 2004), un bras

robotisé est commandé par asservissement visuel échographique pour insérer une aiguille dans la vésicule biliaire afin de réaliser un drainage. Plus récemment, une autre approche d'asservissement visuel échographique a permis l'insertion d'une aiguille de ponction flexible dans le plan d'observation d'une sonde ultrasonore 2D (Neubach et Shoham, 2010). La tâche consistait à amener le bout de l'aiguille à un endroit prédéfini par l'utilisateur tout en contournant des obstacles en modifiant la courbure de l'aiguille. Ces méthodes utilisent également des techniques de prédiction de la trajectoire de l'aiguille à travers les tissus.

Les travaux de Vitrani (Vitrani et al., 2005, 2007) ont permis un asservissement visuel de 4 ddl d'un instrument de chirurgie afin d'assister la chirurgie sur cœur battant. Les primitives visuelles utilisées étaient de type géométriques. Cette méthode a été validée *in vivo* lors d'une opération où l'outil est inséré dans le ventricule gauche d'un cœur de cochon (voir FIGURE 3.7).



FIGURE 3.7 – Banc expérimental composé d'un robot et d'une sonde échographique utilisés pour insérer un outil dans le ventricule gauche d'un cœur de cochon (Vitrani et al., 2007).

De nombreuses primitives géométriques ont été utilisées en asservissement visuel échographique. Dans (Mebarki et al., 2010), une matrice d'interaction fondée sur l'approche des moments géométriques a été modélisée dans le cas d'une image échographique. Les primitives visuelles suivantes sont proposées pour contrôler les 6 ddl du robot tenant la sonde échographique :

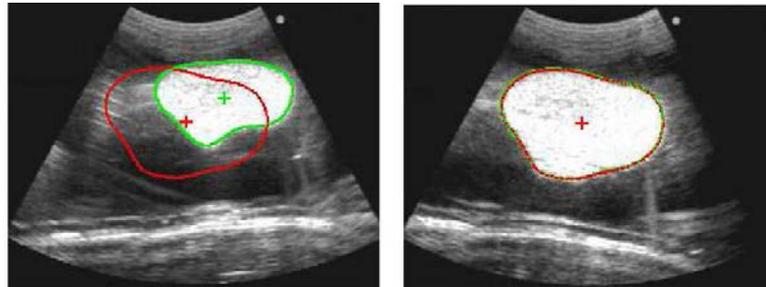
$$\mathbf{s} = (x_g, y_g, \alpha, \sqrt{a}, \mu_1, \mu_2) \quad (3.4)$$

où  $(x_g, y_g)$  représente le centre de gravité de l'objet,  $\alpha$  représente l'orientation principale de l'objet,  $a$  représente son aire,  $\mu_1$  est une valeur obtenue à partir des moments d'ordres

$\mu_2$  et  $\mu_3$  est obtenue à partir d'une combinaison des moments d'ordre 3 (voir paragraphe 1.3.1). La matrice d'interaction associée à cette méthode utilise une estimation du vecteur normal à la surface du contour observé. En pratique, ce vecteur est estimé en ligne à l'aide de coupes successives, utilisant des techniques d'odométrie. La FIGURE 3.8 (a) montre l'expérience réalisée pour valider une technique d'asservissement visuel à 6 ddl fondée sur les moments géométriques à l'aide d'un objet en gélatine immergé dans un bac d'eau, avec en FIGURE 3.8 (b) la segmentation de l'objet en question. Dans (Nadeau et al., 2016), cette méthode a été étendue à trois plans échographiques orthogonaux.



(a) Installation pour validation expérimentale d'asservissement visuel fondé sur les moments géométriques.



(a)

(b)

(b) Détection d'un objet sur une image échographique.

FIGURE 3.8 – Validation de la méthode des moments géométriques en asservissement visuel échographique 2D (Mebarki et al., 2010).

Pour pallier à l'inconvénient de devoir faire une segmentation de chaque image, une méthode directe utilisant l'intensité des pixels (voir paragraphe 1.3.2) appliquée à l'asservissement échographique a été mise en œuvre pour la première fois en 2011 avec une sonde échographique 2D (Nadeau et Krupa, 2011). Elle a été améliorée par la suite pour asservir les mouvements en dehors du plan d'observation et étendue à l'utilisation d'une sonde 3D (Nadeau et Krupa, 2013). Cette approche est fondamentale à la compréhens-

sion de notre étude assimilant les shearlets. Nous commencerons par la décrire dans le cas du contrôle d'une sonde échographique 2D dans le plan d'observation, puis nous verrons comment contrôler les mouvements hors de ce plan.

Soit  $\mathbf{I}$ , l'image échographique et  $\mathcal{R}$  une région d'intérêt de taille  $m \times n$  pixels dans l'image, la primitive visuelle utilisée  $\mathbf{s}$  correspond alors à un vecteur composé de toutes les intensités contenues dans la région d'intérêt de l'image échographique :

$$\mathbf{s} = (I_{1,1} \dots I_{u,v} \dots I_{m,n}) \quad (3.5)$$

La matrice d'interaction associée à  $\mathbf{s}$  nécessite alors de connaître la dérivée de  $\mathbf{s}$  par rapport au temps. Notons  $\mathbf{I}(\mathbf{x}, t)$  l'intensité du point située en  $\mathbf{x} = (x, y, z)$ , les coordonnées (en mètre) du point exprimées dans le repère de la sonde échographique à l'instant  $t$ . Nous avons alors la relation suivante :

$$\dot{\mathbf{I}}(\mathbf{x}, t) = \frac{\partial \mathbf{I}(\mathbf{x}, t)}{\partial \mathbf{x}} \dot{\mathbf{x}} + \frac{\partial \mathbf{I}(\mathbf{x}, t)}{\partial t} \quad (3.6)$$

avec

$$\frac{\partial \mathbf{I}(\mathbf{x}, t)}{\partial \mathbf{x}} = \nabla \mathbf{I} = (\nabla I_x \nabla I_y \nabla I_z)^\top \quad (3.7)$$

L'hypothèse que l'intensité d'un point reste constante dans une image échographique sur un faible intervalle de temps  $\Delta t$ , c'est-à-dire que  $\frac{\partial \mathbf{I}(\mathbf{x}, t)}{\partial t} = 0$ , est alors utilisée. La relation suivante est alors obtenue :

$$\dot{\mathbf{I}}(\mathbf{x}, t) = \nabla \mathbf{I} \dot{\mathbf{x}} \quad (3.8)$$

La variation temporelle de  $\mathbf{x}$  peut être obtenue à partir de la vitesse de la sonde  $\mathbf{v}$ , en utilisant la formule de composition des vitesses suivante :

$$\dot{\mathbf{x}} = \begin{bmatrix} 1 & 0 & 0 & 0 & z & -y \\ 0 & 1 & 0 & -z & 0 & x \\ 0 & 0 & 1 & y & -x & 0 \end{bmatrix} \mathbf{v} \quad (3.9)$$

Dans le cadre d'utilisation d'une sonde 2D, la section observée se trouve dans une coupe où  $z = 0$ . La matrice d'interaction associée à un point  $I_{u,v}$  de l'image de coordonnées pixelliques  $(u, v)$  est alors obtenue comme suit :

$$\mathbf{L}_{I_{u,v}} = [\nabla I_x \quad \nabla I_y \quad \nabla I_z \quad y \nabla I_z \quad -x \nabla I_z \quad x \nabla I_y - y \nabla I_x] \in \mathbb{R}^{1 \times 6} \quad (3.10)$$

$\Delta I_x$  et  $\Delta I_y$  peuvent être calculées depuis l'image tandis que le terme  $\Delta I_z$ , correspondant au gradient de l'intensité de l'image dans la direction orthogonale au plan image n'est pas calculable directement. Deux approches ont été proposées pour estimer ce gradient :

- en appliquant des petites translations à la sonde dans les directions hors du plan d'observation afin d'obtenir des coupes tangentes à ce plan et de calculer  $\Delta I_z$  à partir de ces coupes.

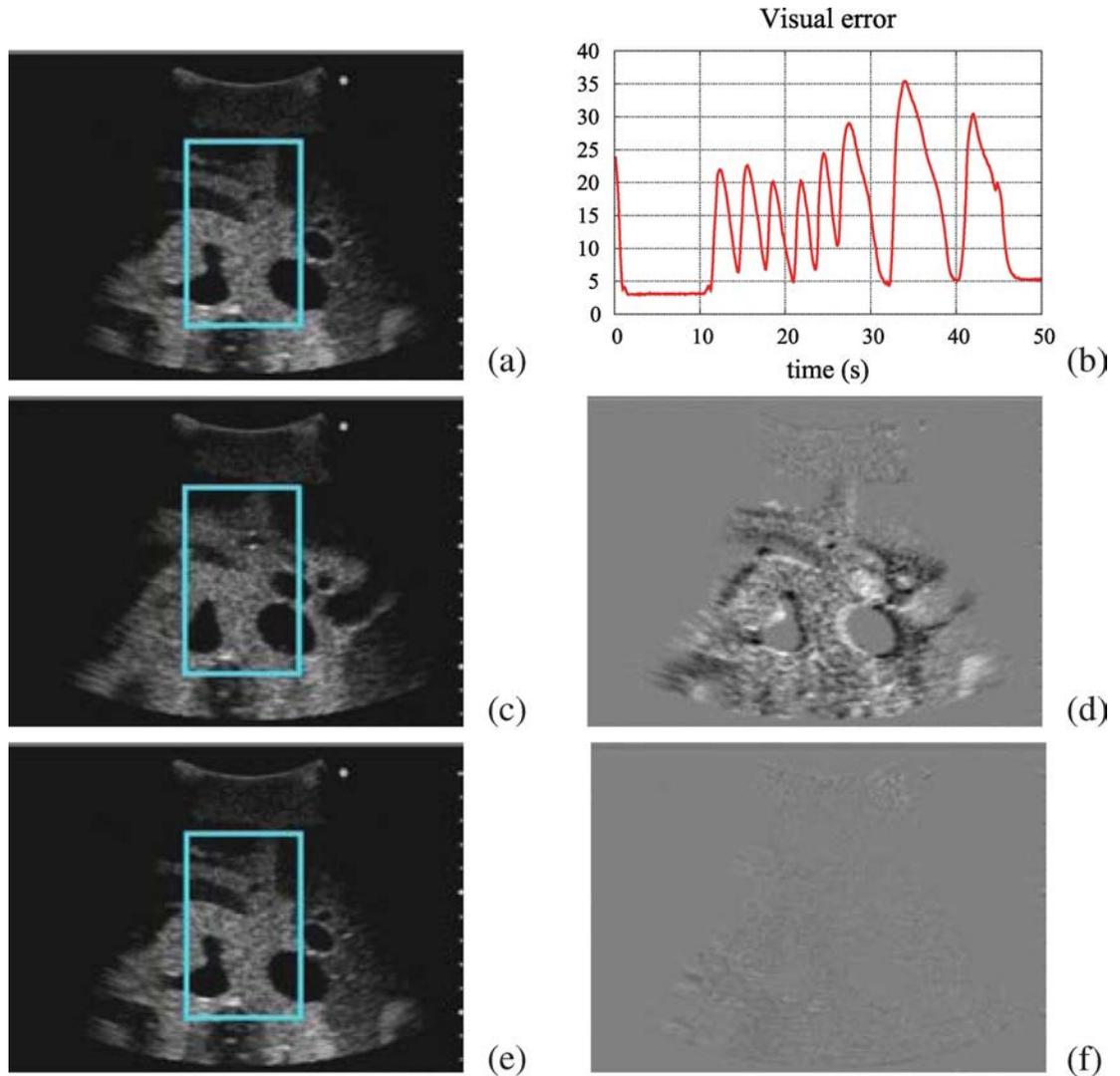


FIGURE 3.9 – Tâche de suivi avec un sonde échographique 2D : (a) image désirée, (b) courbe de la norme l'erreur visuelle entre l'intensité des pixels de l'image courante et ceux de l'image désirée, (c) image échographique courante obtenue à  $t = 33$ s, soit le moment où la norme de l'erreur visuelle est la plus élevée, (d) image de différence correspondante, (e) image échographique lors de la dernière itération de la tâche de suivi, (f) image de différence correspondante (Nadeau et Krupa, 2013).

- en estimant en ligne la différence d'intensité entre des coupes successives grâce à l'odométrie du robot.

La matrice d'interaction associée à  $\mathbf{s}$  est obtenue en concaténant les matrices  $\mathbf{L}_{I_{u,v}} \in \mathbb{R}^{1 \times 6}$  de tous les pixels contenus dans une région d'intérêt  $\mathcal{R}$ .

$$\mathbf{L}_{\mathbf{s}} = [\mathbf{L}_{I_{1,1}} \dots \mathbf{L}_{I_{m,n}}]^\top \quad (3.11)$$

La loi de commande suivante, de type Gauss-Newton, peut alors être utilisée :

$$\mathbf{v} = -\lambda \widehat{\mathbf{L}}_{\mathbf{s}}^+ (\mathbf{s}(\mathbf{t}) - \mathbf{s}^*) \quad (3.12)$$

La FIGURE 3.9 présente les résultats d'une tâche de suivi utilisant l'intensité des pixels de l'image échographique provenant d'une sonde 2D comme information visuelle. Nous pouvons également noter que dans le cas d'une sonde échographique 3D, la région d'intérêt  $\mathcal{R}$  est un volume et toutes les composantes du gradient sont directement disponibles pour calculer la matrice d'interaction.

La deuxième modalité d'imagerie médicale utilisée, dans ce travail, est la tomographie par cohérence optique (OCT). Elle sera présentée dans la suite de ce document.

### 3.1.2 Tomographie par cohérence optique

La tomographie par cohérence optique (OCT) est une des plus récentes innovations en matière d'imagerie médicale non-invasive. Les bases de cette technique ont été posées en 1991 par les équipes de Huang (Swanson et al., 1992) qui l'ont testée en condition *ex vivo* sur la rétine d'un œil et l'artère coronaire. Des tests en conditions *in vitro* ont ensuite été réalisés sur la rétine par les équipes de Fercher en 1993 (Fercher, 1993) et également par les équipes de Huang, la même année (Swanson et al., 1993). Si l'OCT a largement montré son efficacité en ophtalmologie par le diagnostic des glaucomes ou les maladies de la macula (Schuman et al., 1995; Puliafito et al., 1995), son utilisation en dermatologie (Pierce et al., 2004) ou par voie endoscopique (Tearney et al., 1997) est de plus en plus plébiscitée.

**Principe de formation de l'image OCT :** Le principe de fonctionnement d'un appareil OCT est très proche de celui d'un appareil échographique par ultrasons. Si des ultrasons sont utilisés dans le cas d'échographies, dans le cas de l'OCT, de la lumière dans l'infrarouge proche (entre 600 nm et 1300 nm) est utilisée. Cette lumière est absorbée par les tissus biologiques sur environ un millimètre de profondeur. Le temps de parcours du rayon lumineux nous informe sur la position des différentes couches du tissu et sur ses propriétés optiques (Fujimoto, 2003). La vitesse de la lumière ne permet pas d'étudier le délai entre l'émission du signal et la réception de son écho comme avec les ultrasons. C'est grâce à une technique d'interférométrie à faible cohérence de type Michelson qu'une image peut être reconstruite à partir des rayons lumineux envoyés sur l'échantillon (Michelson et Morley, 1887).

Ce dispositif inventé par Michelson et Morley permet de faire interférer deux faisceaux lumineux séparés par une lame, dite séparatrice et qui se réfléchissent sur deux miroirs. La FIGURE 3.10 représente un interféromètre utilisé en mode "lame d'air", c'est-à-dire dans lequel les deux miroirs, notés  $M1$  et  $M2$ , sont parfaitement perpendiculaires. On parle de mode "coin d'air" lorsque l'un des miroirs est très légèrement incliné. La lame séparatrice, semi-réfléchissante et inclinée à 45 degrés, permet de produire deux faisceaux de même intensité, le premier est réfléchi et le second est transmis. Le miroir  $M2$  se situe à une distance  $d$  par rapport à  $M1'$ , l'image de  $M1$  par la séparatrice. Cette distance  $d$  induit une *différence de marche* entre les deux rayons égale à  $2d$ . Cette différence des chemins optiques parcourus par ces deux rayons permet d'évaluer le retard, ou le déphasage d'un de ces rayons par rapport à l'autre.

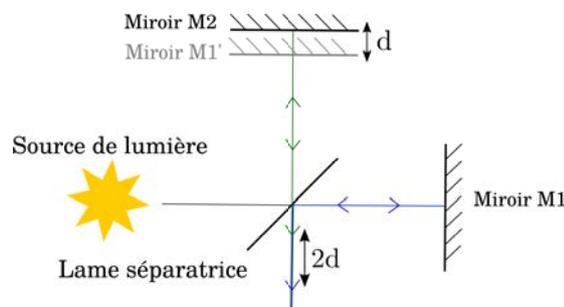


FIGURE 3.10 – Structure d'un interféromètre de Michelson.

Pour l'OCT, un rayon lumineux est envoyé sur une séparatrice, il se sépare alors en deux rayons, l'un envoyé vers un miroir tandis que l'autre est dirigé vers l'échantillon à analyser, qui fait office de deuxième miroir. La résolution axiale est alors proportionnelle à la largeur spectrale du rayon utilisé. L'amplitude du signal d'interférence correspond alors à une coupe donnée, et une frontière entre deux milieux d'indice de réfraction différents donne un fort signal d'interférence. C'est l'étude de ses interférences qui permet de retrouver la *différence de marche* entre le miroir de référence et les différentes couches de l'échantillon observé. Si l'analyse d'un rayon lumineux correspond au mode "A-scan" (carotte optique), l'envoi de plusieurs rayons sur une ligne permet d'obtenir un "B-scan" (image 2D) et donc une coupe de l'échantillon. Enfin, l'assemblage de plusieurs coupes permet d'obtenir un volume "C-scan" (image 3D), appelé également biopsie optique 3D.

Il existe deux grandes familles de systèmes d'OCT comme le montre la FIGURE 3.11. Ces systèmes sont :

- système OCT temporel (TD-OCT) : dans ce mode, le miroir de référence est translaté afin de faire varier la différence de marche. En effet, les deux faisceaux réfléchis se recombinaient seulement quand leur distance parcourue est la même.
- système OCT fréquentiel (FD-OCT) : dans ce mode, des détecteurs séparateurs de spectre sont utilisés et la transformée de Fourier reliant l'auto-corrélation et

la densité spectrale est utilisée afin d'en déduire directement la profondeur de l'échantillon. Cette technique permet une plus grande rapidité et également un meilleur rapport signal sur bruit (Leitgeb et al., 2003).

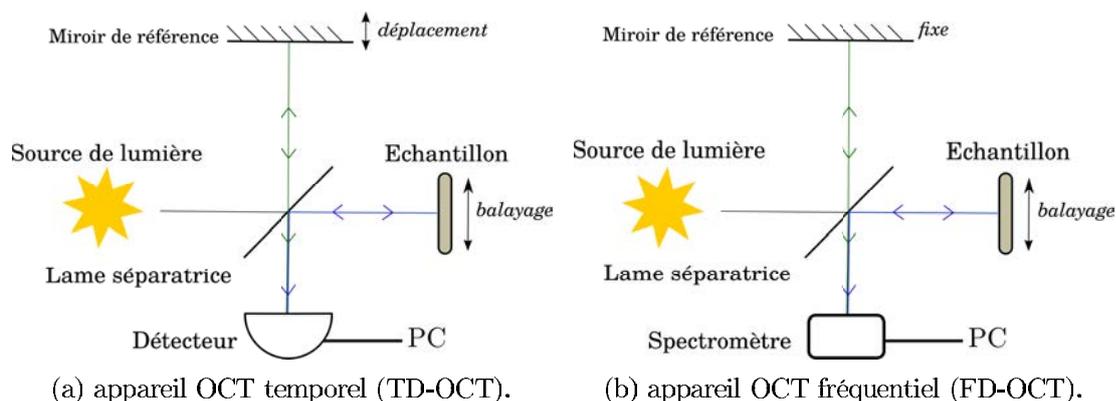


FIGURE 3.11 – Différence de structure entre un appareil OCT temporel et fréquentiel.

Pour la validation expérimentale de nos travaux sur l'asservissement visuel OCT, nous avons utilisé un appareil OCT fréquentiel Telesio II de Thorlabs ([www.thorlabs.com](http://www.thorlabs.com)) représenté sur la FIGURE 3.12. Par ailleurs, la FIGURE 3.13 (a) représente la photographie d'une mouche grâce à une caméra CCD classique. La ligne rouge représente la ligne utilisée pour obtenir la coupe "B-scan" correspondante visible sur la FIGURE 3.13 (b). La FIGURE 3.14 donne un exemple de reconstruction OCT 3D (biopsie optique) sur ce même échantillon.

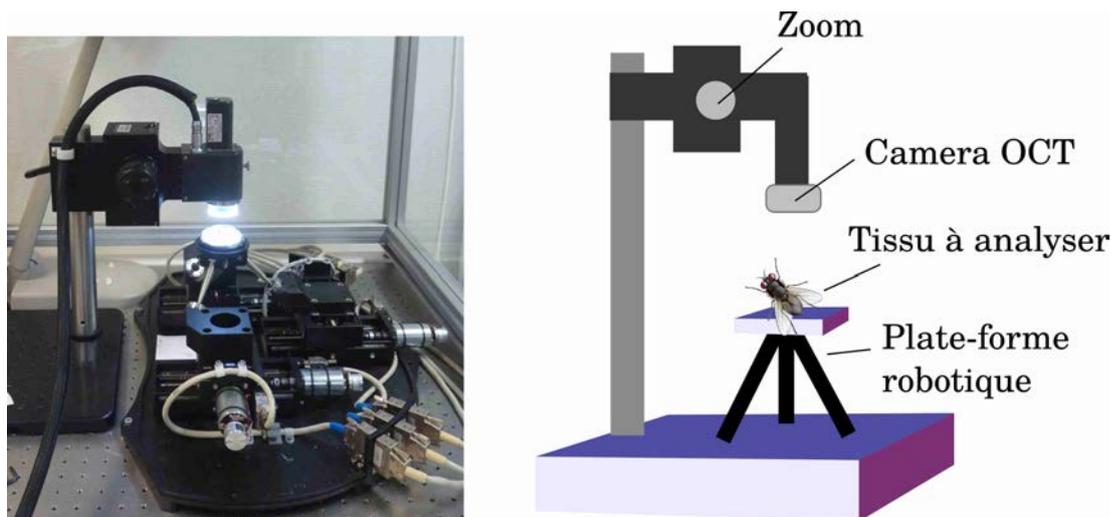
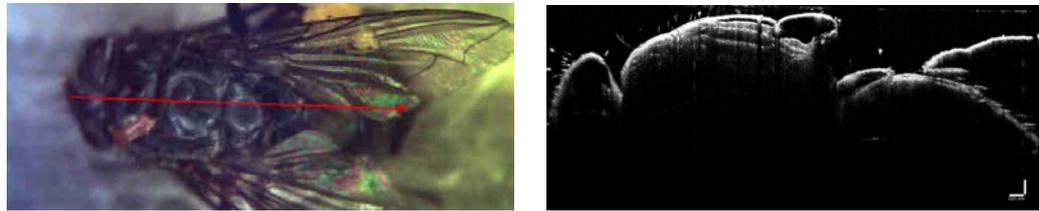


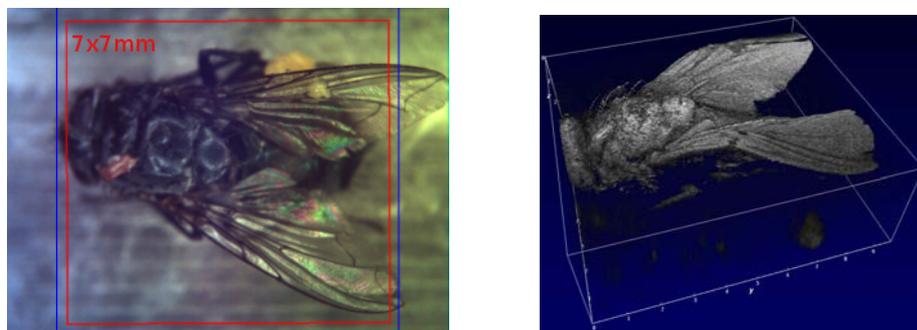
FIGURE 3.12 – Photographie du système d'acquisition OCT placé au dessus d'une plate-forme robotique.



(a) Image de la mouche obtenue à l'aide de la caméra CCD.

(b) Image 2D correspondant.

FIGURE 3.13 – Illustration de l'obtention d'une coupe de type "B-scan" (image OCT 2D).



(a) Image de la mouche obtenue à l'aide de la caméra CCD.

(b) Volume 3D correspondant.

FIGURE 3.14 – Illustration de l'obtention d'une coupe de type "C-scan" (image OCT 3D).

**Asservissement visuel OCT :** Les seuls travaux d'asservissement visuel OCT ont été réalisés par M. Ourak et ses collègues en 2016 (Ourak et al., 2016a,b). Dans ces travaux, une commande d'asservissement visuel pour réaliser des tâches répétées de biopsie optique a été mise au point. Ce travail utilise des informations du domaine fréquentiel pour calculer le déplacement relatif entre deux coupes OCT successives. Il permet le contrôle d'un système micro-robotique dans le plan de la coupe OCT (voir FIGURE 3.15).

Dans (Ourak et al., 2016b), cette loi de commande est étendue à 6 ddl en utilisant d'une part la coupe OCT pour contrôler les ddl du plan d'observation de la coupe OCT mais aussi l'image obtenue par la caméra CCD pour les 3 ddl restants. La disposition des degrés de liberté est représentée sur la FIGURE 3.16.

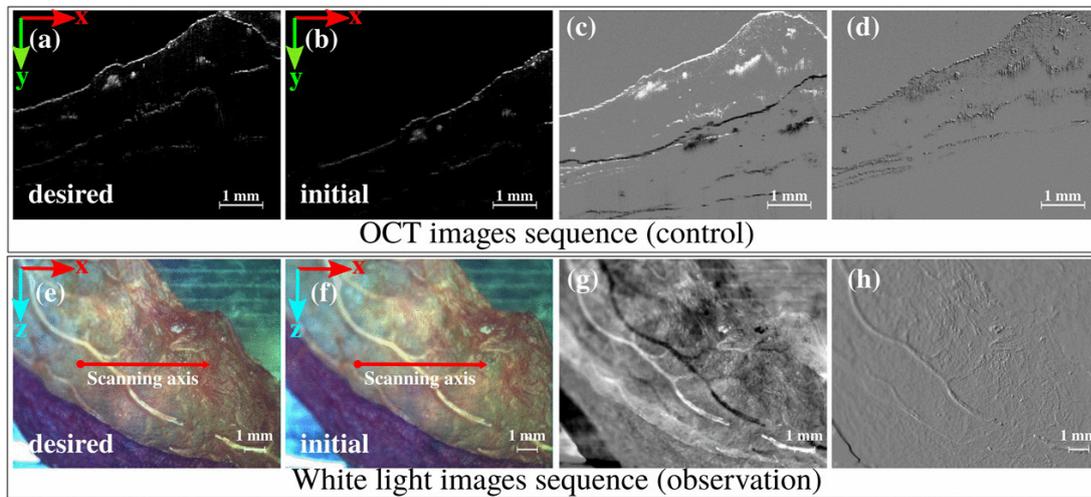


FIGURE 3.15 – Technique d’asservissement visuel OCT utilisant les ondelettes. Première ligne : images OCT lors d’une tâche d’asservissement visuel à 3 ddl ; deuxième ligne : les images correspondantes obtenues par la caméra CCD. (Ourak et al., 2016a).

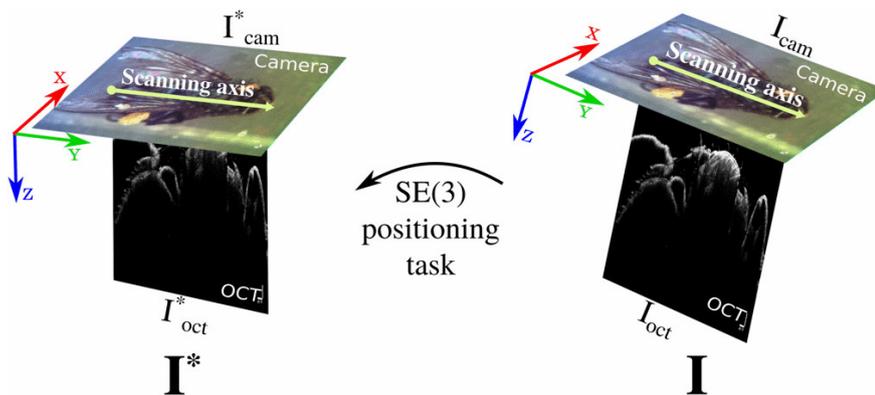


FIGURE 3.16 – Représentation de la tâche de positionnement réalisée par asservissement visuel OCT (Ourak et al., 2016b).

Après cette présentation des modalités d’imagerie médicale utilisées dans cette thèse, nous montrerons nos deux contributions en asservissement visuel appliqué à l’imagerie échographique.

## 3.2 Matrice d'interaction numérique

Cette première contribution consiste à étudier la pertinence du choix des coefficients en shearlets comme informations visuelles pour l'asservissement visuel d'un bras robotique à 6 ddl. Afin de simplifier l'étude, nous avons dans un premier temps estimé numériquement la matrice d'interaction. Pour ce faire, nous avons étudié la variation temporelle des coefficients d'une transformée en shearlets lorsqu'un déplacement est appliqué à la sonde échographique. Cette loi de commande a été testée en simulation puis expérimentalement à l'aide d'une sonde échographique 2D interagissant avec une maquette d'abdomen en gélatine (Dufflot et al., 2016c).

### 3.2.1 Concept et implémentation

Cette section s'intéresse au choix de l'implémentation des shearlets. Comme mentionné précédemment (voir section 2.2.5), la transformée en shearlets possède plusieurs manières d'être implémentées. Le choix de la variante guidera ainsi la méthode élaborée pour le calcul de la matrice d'interaction associée.

**Choix des shearlets à utiliser :** Comme présentées dans la section 2.2.4, les shearlets  $\mathcal{S}_\psi \mathbf{I}$  d'une image  $\mathbf{I}$  selon une fonction génératrice  $\psi$  sont données pour un niveau de détails  $j \in \mathbb{Z}$ , avec un paramètre de découpage  $|k| < \left\lceil 2^{\frac{j}{2}} \right\rceil$  et de translation  $\mathbf{m} \in \mathbb{Z}^2$  par le produit scalaire suivant :

$$(\mathcal{S}_\psi \mathbf{I})(j, k, \mathbf{m}) = \langle \mathbf{I}, \psi_{j,k,\mathbf{m}} \rangle \quad (3.13)$$

En faisant varier les paramètres  $j$  et  $k$  (en pratique  $\mathbf{m}$  dépend du nombre total de niveau de décomposition  $j_{max}$ ) des coefficients de shearlets sous-échantillonnées sont obtenus. Ils correspondent à deux cônes : le cône vertical ( $c = 1$ ) ou le cône horizontal ( $c = 2$ ). La FIGURE 3.17 montre la représentation des coefficients des shearlets d'une image composée de  $128 \times 128$  pixels avec 5 niveaux de décomposition ( $j_{max} = 5$ ). Chaque niveau contient un certain nombre de sous-images correspondant aux différents paramètres de découpage  $k$ , respectant la règle  $|k| < \left\lceil 2^{\frac{j}{2}} \right\rceil$ .

**Loi de commande :** Pour mettre en oeuvre notre première loi de commande, nous avons calculé les coefficients des shearlets d'une image donnée en utilisant le niveau  $j = 5$  avec  $k = [-1, 0, 1]$  du cône horizontal ( $c = 1$ ) et vertical ( $c = 2$ ) pour définir les primitives visuelles. Ainsi les primitives visuelles utilisées pour calculer la matrice d'interaction sont définies par la concaténation des coefficients des shearlets comme suit :

$$\mathbf{s} = \left\langle \mathbf{I}, \psi_{j,k,\mathbf{m}} \right\rangle_{(j_{max}=5)} \quad j = 5, \quad k = \{-1, 0, 1\}, \quad c = \{1, 2\} \quad (3.14)$$

Ce qui correspond aux 6 sous-images les moins détaillées de la transformée en shearlets de l'image originale (*Lena*) présentées sur la FIGURE 3.17.

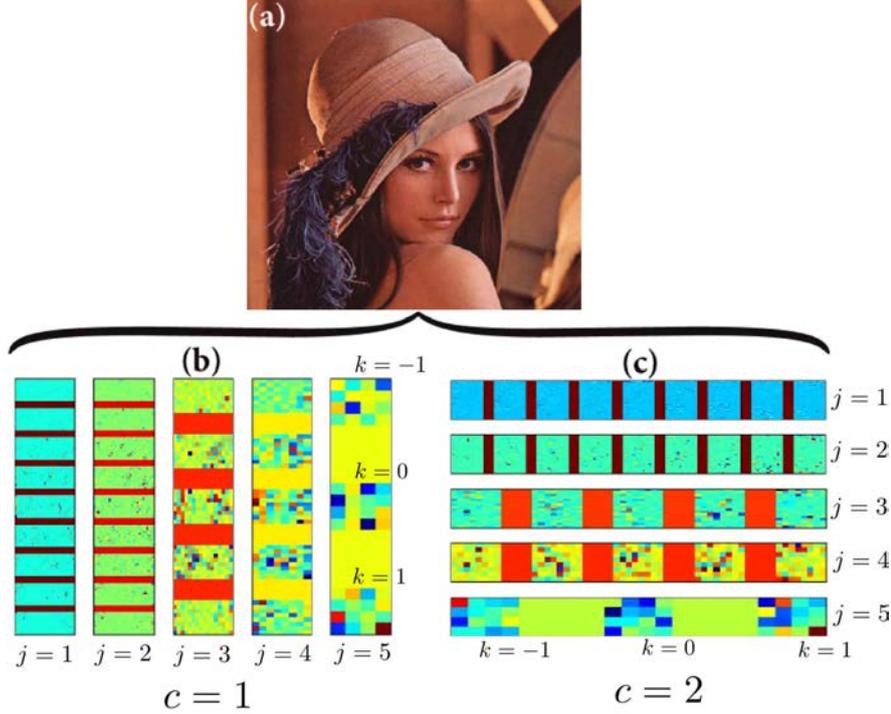


FIGURE 3.17 – Coefficients des shearlets ( $j_{max} = 5$ ) calculés à partir de la photographie *Lena*. (a) Image originale (b) coefficients du cône horizontal (c) coefficients du cône vertical.

Nous avons calculé la matrice d'interaction  $\mathbf{L}_s^*$  à la position désirée à l'aide d'une procédure hors ligne. Cette dernière consiste en application successive et indépendante de petits déplacements de la sonde échographique. Ces petits déplacements correspondent aux 3 translations  $\Delta x$ ,  $\Delta y$ ,  $\Delta z$  et aux 3 rotations  $\Delta r_x$ ,  $\Delta r_y$ ,  $\Delta r_z$  le long et autour de chaque axe du repère Cartésien attaché à la sonde (voir FIGURE 3.18).

Les différences mesurées sur chaque coefficient des shearlets  $\Delta s_{\Delta x}^l$ ,  $\Delta s_{\Delta y}^l$ ,  $\Delta s_{\Delta z}^l$ ,  $\Delta s_{\Delta r_x}^l$ ,  $s_{\Delta r_y}^l$ ,  $\Delta s_{\Delta r_z}^l$  induites par ces petits déplacements permettent de calculer la  $l^{\text{ème}}$  ligne de la matrice d'interaction  $\mathbf{L}_s^*$  de la manière suivante :

$$\mathbf{L}_s(l)^* = \left( \frac{\Delta s_{\Delta x}^l}{\Delta x} \quad \frac{\Delta s_{\Delta y}^l}{\Delta y} \quad \frac{\Delta s_{\Delta z}^l}{\Delta z} \quad \frac{\Delta s_{\Delta r_x}^l}{\Delta r_x} \quad \frac{\Delta s_{\Delta r_y}^l}{\Delta r_y} \quad \frac{\Delta s_{\Delta r_z}^l}{\Delta r_z} \right) \in \mathbb{R}^{1 \times 6} \quad (3.15)$$

Nous obtenons la matrice d'interaction en concaténant les  $N$  lignes de  $\mathbf{L}_s^*$  :

$$\mathbf{L}_s^* = \begin{bmatrix} \mathbf{L}_s(1)^* \\ \vdots \\ \mathbf{L}_s(N)^* \end{bmatrix} \quad (3.16)$$

La loi de commande est alors la suivante :

$$\mathbf{v} = -\lambda \mathbf{L}_s^{*+} (\mathbf{s}(t) - \mathbf{s}^*) \quad (3.17)$$

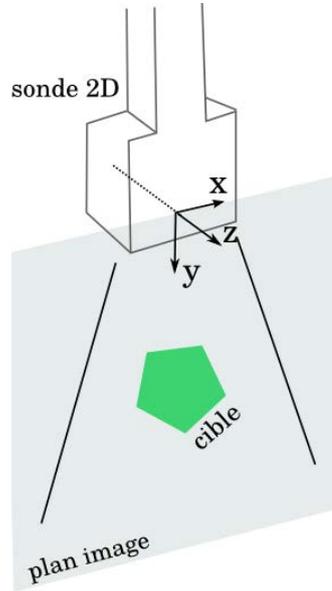


FIGURE 3.18 – Disposition des axes de la sonde échographique 2D.

où  $\lambda > 0$  est un gain positif et  $\mathbf{L}_s^{*+}$  est la pseudo-inverse de *Moore-Penrose* de  $\mathbf{L}_s^*$ . Comme dans (Nadeau et Krupa, 2013), nous avons utilisé une optimisation de type Gauss-Newton.

### 3.2.2 Résultats de simulation

Avant d'être validée expérimentalement, notre commande a été testée à l'aide d'un simulateur développé en C++. Cette étape nous a permis de calculer la zone de convergence du contrôleur proposé ainsi qu'effectuer une validation suivant plusieurs scénarios et dans différentes conditions, sans risquer d'endommager le matériel (notamment la sonde échographique).

**Présentation du simulateur :** Notre simulateur imite une sonde échographique 2D placé sur un abdomen. Il a été développé en s'appuyant sur les bibliothèques open-source ViSP, Visualization ToolKit library (VTK) et une version simplifiée et ré-implémentée en C++ de Shearlab (développée initialement en Matlab). Un exemple de volume échographique et d'une coupe 2D sont visibles sur la FIGURE 3.19. Le volume échographique est composé de 335 images 2D parallèles de taille  $250 \times 250$  pixels (taille de pixel de  $0.6\text{mm} \times 0.6\text{mm}$ ) et un intervalle de  $0.3\text{mm}$  entre deux images successives.

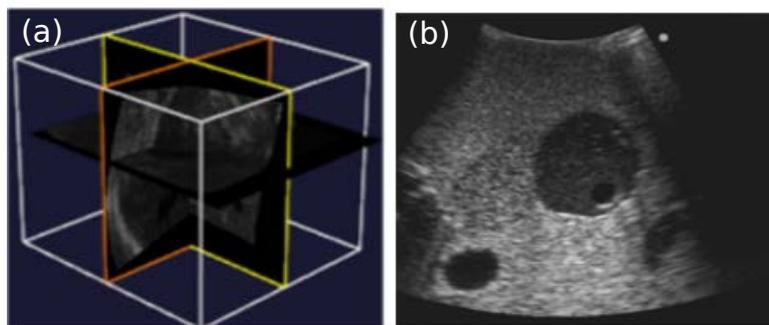
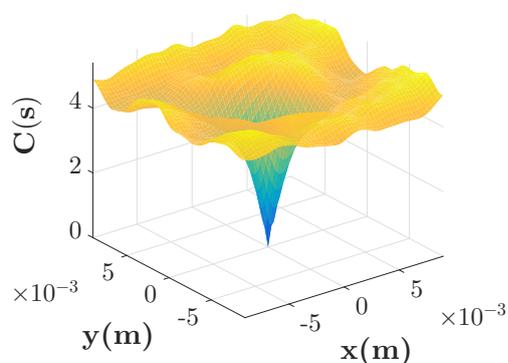


FIGURE 3.19 – (a) Volume échographique 3D. (b) Coupe 2D extraite du volume.

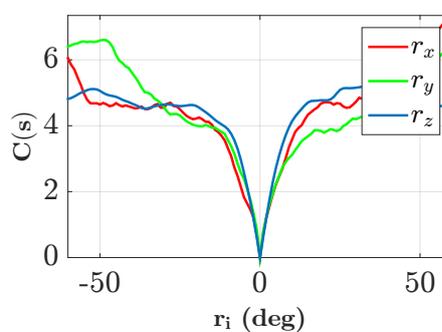
**Domaine de convergence :** Afin de juger du choix des primitives visuelles, nous avons calculé une fonction de coût  $C(s)$ , pour différentes positions de l'image, comme suit :

$$C(s) = \left( (s - s^*)^T (s - s^*) \right)^{\frac{1}{2}} \quad (3.18)$$

Comme le montre la FIGURE 3.20, la fonction de coût a une forme convexe et possède un minimum global bien défini. Notre choix de primitives visuelles est donc encourageant.



(a) Cône de convergence pour une translation en  $x$  et  $y$ .



(b) Cône de convergence pour les rotations autour de  $x$ ,  $y$  et  $z$ .

FIGURE 3.20 – Cône de convergence d'un asservissement visuel utilisant les shearlets.

**Tâche de positionnement :** Cette simulation consiste à réaliser une tâche de positionnement à 6 ddl en utilisant la loi de commande (3.17). Pour cela, l'objectif est d'atteindre une position désirée  $\mathbf{r}^* \in SE(3)$  (correspondant à une coupe échographique désirée) à partir d'une position initiale arbitraire  $\mathbf{r} \in SE(3)$  (correspondant à une coupe échographique initiale). A chaque nouvelle acquisition d'une image échographique, les coefficients des shearlets sont calculés dans une certaine région d'intérêt (entourée en

couleur cyan sur les images). Plus précisément, seuls les 96 coefficients correspondant aux paramètres  $j = 5$  et  $k = \{-1, 0, 1\}$  sont calculés. Les vitesses, générées par la loi de commande (3.17), sont appliquées à la sonde virtuelle avec un gain  $\lambda = 0.5$  afin d'atteindre la position désirée. La translation en  $y$  ne sera pas asservie visuellement car en pratique ce degré de liberté est contrôlé par une commande en effort pour maintenir le contact de la sonde avec l'abdomen. Dans la suite, le symbole  $\times$  sera utilisé pour indiquer que la direction  $y$  n'a pas été asservie visuellement. Les FIGURES 3.22(a)-(b) montrent respectivement l'image initiale observée avec une erreur de positionnement initiale  $\Delta \mathbf{e}_{init}$  (mm, deg) = (9,  $\times$ , 5, 2, 2, 5), et l'image désirée. La FIGURE 3.22(c) représente une image qui correspond à la différence entre l'image courante et l'image désirée selon une formule qui sera souvent utilisée par la suite :

$$\mathbf{I}_{diff} = \frac{(\mathbf{I} - \mathbf{I}^*) + 255}{2} \quad (3.19)$$

Grâce à l'ajout de la valeur 255, une convergence parfaite se traduit alors par une image  $\mathbf{I}_{diff}$  complètement grise. Sur la FIGURE 3.22(d), qui correspond à l'image de différence obtenue à la convergence de l'asservissement avec  $\Delta \mathbf{e}_{final}$  (mm, deg) = (0.1,  $\times$ , 0.2, 0.02, 0.18, 0.08), l'image contient essentiellement du bruit de type *speckle*, inhérent aux images échographiques.

Les FIGURES 3.21 et 3.23 illustrent la décroissance de l'erreur de positionnement sur chaque degré de liberté ainsi que l'évolution de la vitesse au cours du temps.

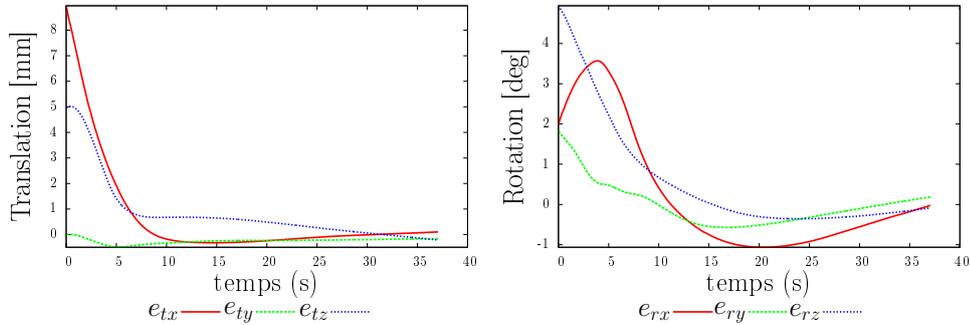


FIGURE 3.21 – [Simulation] Erreurs de positionnement sur chaque degré de liberté.

Ce premier résultat de simulation montre la faisabilité d'un asservissement visuel utilisant les coefficients des shearlets comme signal d'entrée de la boucle de commande. Son utilisation avec une vraie sonde échographique embarquée sur un bras robotique sera le thème de la prochaine sous-section.

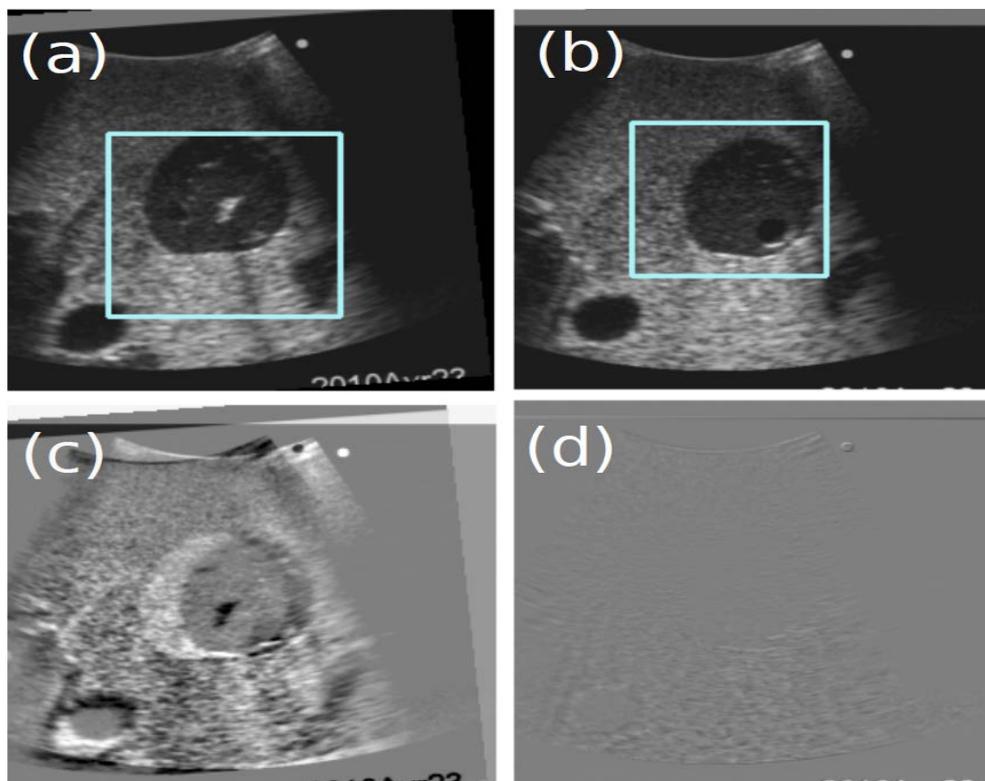


FIGURE 3.22 – [Simulation] Séquence d'images acquises pendant la tâche de positionnement : (a) image initiale  $\mathbf{I}_0$ , (b) image désirée  $\mathbf{I}(\mathbf{r}^*)$ , (c) image de différence  $\mathbf{I}_{diff}$  initiale, (d)  $\mathbf{I}_{diff}$  finale.

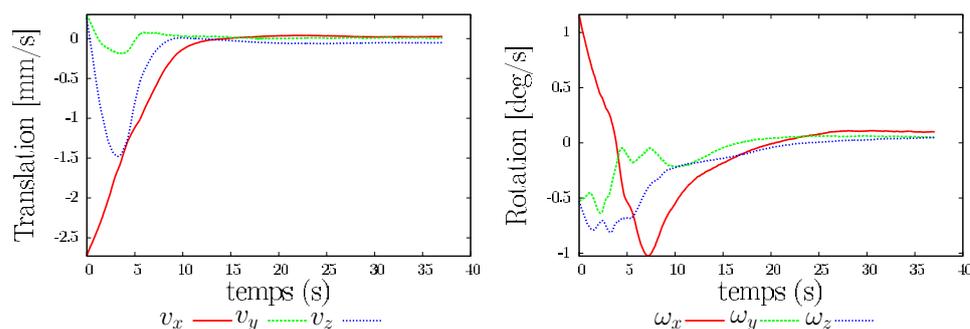


FIGURE 3.23 – [Simulation] Évolution de la vitesse sur chaque degré de liberté.

### 3.2.3 Résultats expérimentaux

L'objectif ici est de reproduire les résultats obtenus en simulation sur un système robotique. Notre dispositif de validation sera présenté ainsi que les performances de la loi de commande lors d'une tâche de positionnement et lors d'une tâche de suivi.

**Plate-forme expérimentale :** Nous avons testé et validé la loi de commande d'asservissement visuel avec les primitives visuelles constituées des coefficients shearlets de l'image dans des conditions expérimentales réalistes (voir FIGURE 3.24). Cette validation est réalisée sur une plate-forme expérimentale composée de :

- Un bras robotique anthropomorphique à 6 ddl de type *Adept Viper s850*;
- Une sonde échographique 2D fonctionnant à une fréquence comprise entre 2 et 5MHz (C60, Sonosite 180+) avec une profondeur de pénétration de 12cm produisant des images de  $480 \times 640$  pixels (taille du pixel :  $0.35 \times 0.36\text{mm}^2$ );
- Un capteur de force 6 axes ATI Gamma SI-65-5;
- Un ordinateur 3-GHz sous une distribution Linux où est implémenté en C++ notre contrôleur;
- Un abdomen en gélatine AB-41900-030 Kyoto Kagaku-ABDFAN.

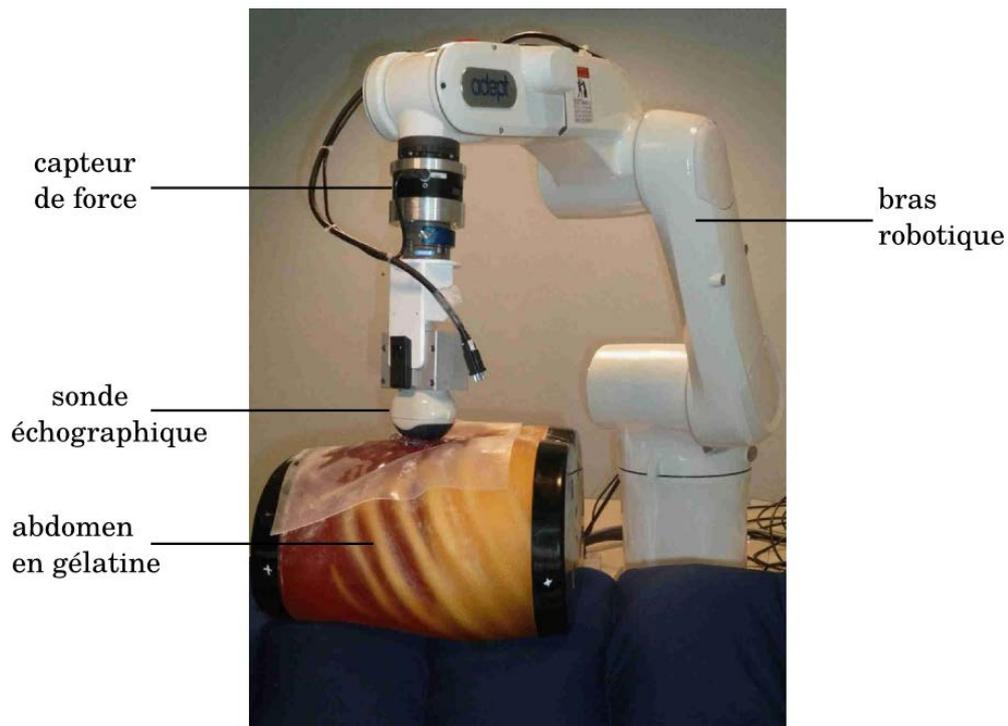


FIGURE 3.24 – Photographie la plate-forme robotique utilisée pour la validation expérimentale.

La sonde échographique, attachée à un robot Viper à 6 ddl, est placée à une position de référence où les coefficients des shearlets sont calculés et la matrice d'interaction numériquement estimée grâce à l'application lors d'une procédure initiale de petits déplacements suivants les 6 ddl. La sonde est alors déplacée sans que l'abdomen en gélatine

ne bouge. L'objectif est de retourner à la position désirée automatiquement en utilisant l'asservissement visuel proposé. Cette position désirée pourrait correspondre à une image acquise quelques jours plus tôt par un médecin qui souhaite se replacer au même endroit pour une ré-examenation. La direction  $y$  correspond à celle de contact entre la sonde et l'abdomen en gélatine. Il est très important d'appliquer des mouvements très délicats dans cette direction pour ne pas endommager le matériel. Dans ce but, une commande en force selon cette direction est effectuée, en maintenant une force désirée de contact de  $2N$  tandis que les 5 autres ddl sont contrôlés par asservissement visuel.

**Résultats obtenus sous conditions nominales** La FIGURE 3.25 représente les images obtenues avec une erreur de positionnement entre l'image initiale  $\mathbf{I}_0$  et l'image désirée  $\mathbf{I}^*$  de  $\Delta \mathbf{e}_{init}$  (mm, deg) = (7,  $\times$ , 8, 2,2, 2). L'erreur de positionnement finale obtenue à la convergence est de  $\Delta \mathbf{e}_{final}$  (mm, deg) = (0.16,  $\times$ , 0.09, 0.09, 0.22, 0.07), traduisant une précision importante de l'approche proposée.

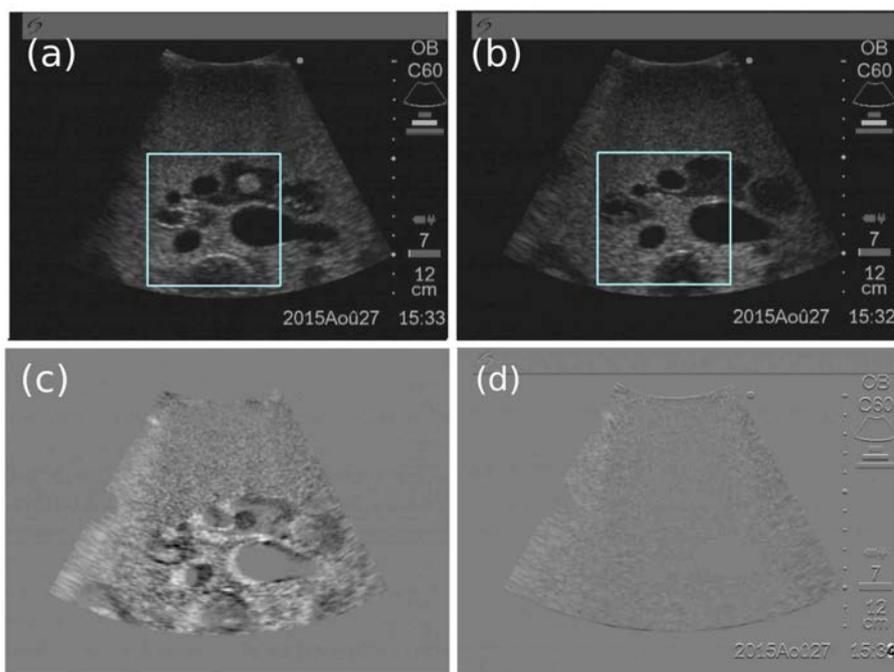


FIGURE 3.25 – [Validation expérimentale] Séquence d'images acquises pendant la tâche de positionnement : (a) image initiale  $\mathbf{I}(\mathbf{r}_0)$ , (b) image désirée  $\mathbf{I}(\mathbf{r}^*)$ , (c) image de différence  $\mathbf{I}_{diff}$  initiale, (d)  $\mathbf{I}_{diff}$  finale.

La FIGURE 3.26 présente l'évolution de l'erreur de positionnement obtenue durant la tâche. Nous pouvons voir que les erreurs de positionnement en translation suivent une trajectoire de type décroissance exponentielle. L'erreur sur les rotations hors-plan ( $e_{rx}, e_{ry}$ ) connaît, quant à elle, une phase de divergence avant de décroître également

vers zéro. Ce comportement peut être expliqué par l'estimation numérique de la matrice d'interaction avec de petits déplacements de la sonde qui est peu adapté aux rotations hors-plan. La FIGURE 3.27 présente l'évolution de la vitesse sur chaque ddl. Ces vitesses décroissent rapidement.

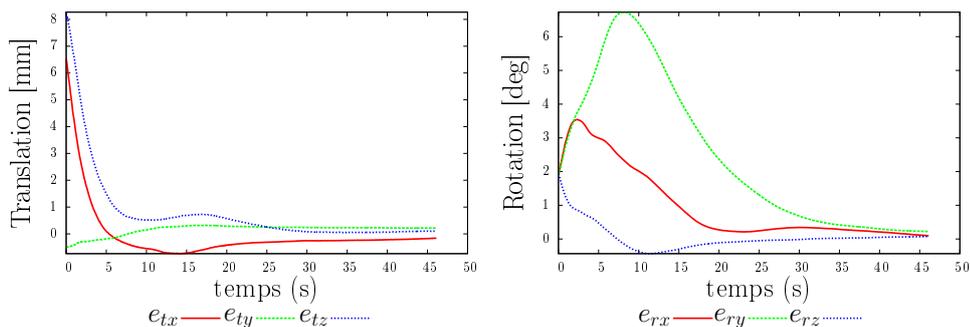


FIGURE 3.26 – [Validation expérimentale] Erreurs de positionnement sur chaque degré de liberté.

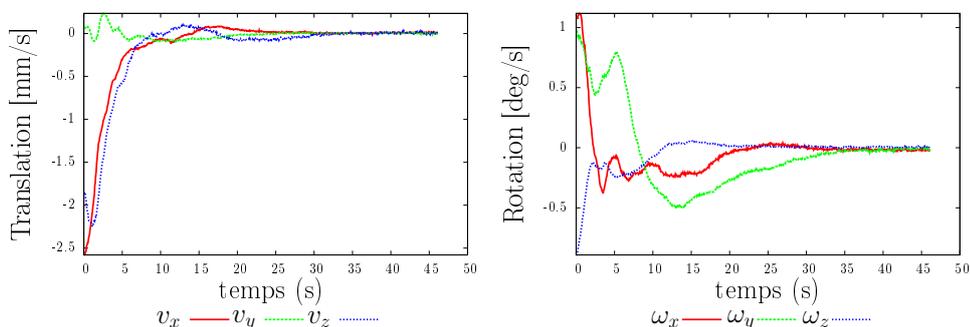


FIGURE 3.27 – [Validation expérimentale] Évolution de la vitesse sur chaque degré de liberté.

Des tâches de positionnement avec d'autres positions initiales ont été réalisées. Le TABLEAU 3.1 résume les erreurs de positionnement initiales et finales enregistrées pendant les différents tests effectués.

L'utilisation des coefficients de la transformée en shearlets de l'image présente de bons résultats pour des tâches d'asservissement visuel en termes de précision lorsque le déplacement initial de la sonde est peu important. Dans cette optique, nous avons testé notre loi de commande pour une tâche de suivi.

TABLE 3.1 – Étude de la robustesse du contrôleur ( $\Delta T_i$  (mm),  $\Delta R_i$  (deg),  $\mathbf{e}_0$  et  $\mathbf{e}_f$  représentent respectivement les erreurs initiales et finales acquises pendant l’accomplissement des différentes tâches de positionnement.)

$\mathbf{e}_f$ and $\mathbf{e}_0$	$\Delta T_x$	$\Delta T_y$	$\Delta T_z$	$\Delta R_x$	$\Delta R_y$	$\Delta R_z$
$\mathbf{e}_0$	1	×	2	5	5	5
$\mathbf{e}_f$	0,41	×	0,09	0,06	0,13	0,14
$\mathbf{e}_0$	10	×	5	5	5	5
$\mathbf{e}_f$	0,3	×	0,26	0,37	0,5	0,21
$\mathbf{e}_0$	7	×	8	2	2	2
$\mathbf{e}_f$	0,16	×	0,09	0,09	0,22	0,07
$\mathbf{e}_0$	8	×	5	2	2	5
$\mathbf{e}_f$	0,41	×	0,11	0,11	0,3	0,85
$\mathbf{e}_0$	4	×	5	5	5	5
$\mathbf{e}_f$	0,3	×	0,06	0,1	0,22	0,08
$\mathbf{e}_0$	2	×	4	5	5	4
$\mathbf{e}_f$	0,09	×	0,02	0,03	0,35	0,31
<b>Moyenne</b>	<b>0,28</b>	×	<b>0,10</b>	<b>0,13</b>	<b>0,29</b>	<b>0,28</b>
<b>Écart-type</b>	<b>0,13</b>	×	<b>0,08</b>	<b>0,12</b>	<b>0,13</b>	<b>0,29</b>

**Tâche de compensation des mouvements de la sonde :** Cette tâche consiste à déplacer automatiquement la sonde échographique pour suivre le mouvement de l’abdomen en gélatine déplacé manuellement. Dans ce cas, l’image désirée est également l’image initiale, le but de l’asservissement est de la maintenir invariante. L’idée est de compenser d’éventuels mouvements physiologiques du patient durant un examen échographique. Pour cette expérience, nous avons déplacé l’abdomen avec une amplitude de mouvement  $\Delta A$  (mm, deg) = (20, 2, 2, 3, 17, 3). Comme le montre la FIGURE 3.28, notre contrôleur fondé sur les shearlets compense efficacement les mouvements appliqués sur l’abdomen en gélatine.

Les premiers résultats de notre investigation ont montré que nous pouvions effectivement utiliser les coefficients des shearlets d’une image dans une loi de commande d’asservissement visuel échographique. Néanmoins, ce premier travail se base sur une estimation numérique de la matrice d’interaction. Nous avons donc essayé d’utiliser une autre implémentation des shearlets pour obtenir une formulation analytique de la matrice d’interaction. Cette étude présentée ci-après constitue une seconde contribution du travail de cette thèse.

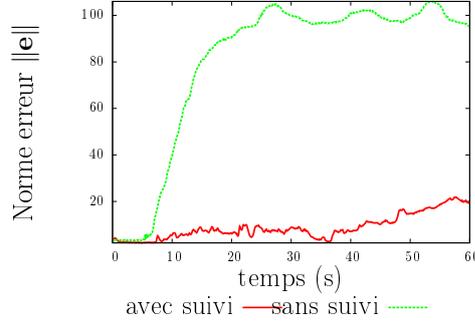


FIGURE 3.28 – Évolution de la norme de l'erreur des coefficients shearlets de l'image lors d'une tâche de suivi.

### 3.3 Matrice d'interaction analytique

Cette troisième partie présente la mise en place d'une commande d'asservissement visuel fondée sur les shearlets non sous-échantillonnées avec une estimation analytique de la matrice d'interaction associée. Cette méthode a été testée en simulation puis expérimentalement à l'aide d'une sonde échographique 2D et d'un abdomen en gélatine (DufLOT et al., 2016a).

#### 3.3.1 Concept et implémentation

Dans cette section, il est question du choix des primitives visuelles et du calcul analytique de la matrice d'interaction.

**Choix des shearlets à utiliser :** Dans la méthode présentée dans la section précédente, la commande visuelle avait pour signal d'entrée les coefficients de la transformée sous-échantillonnée des shearlets (Easley et al., 2008). Avec ce type de transformée, un nombre de coefficients du même ordre de grandeur que l'image initiale de taille  $n \times m$  pixels était obtenu (ce nombre variant selon le nombre de niveau  $j$  et d'orientations  $k$  considérés). Les ondelettes possèdent, quant à elles, un nombre d'orientations fixe (3) à chaque niveau. Nous proposons à présent d'utiliser une autre implémentation des shearlets : les shearlets non sous-échantillonnées (Kutyniok et al., 2016). Pour chaque couple  $(j, k)$ , nous avons désormais une image composée de  $n \times m$  coefficients. Le niveau  $j = 0$  correspond à une représentation grossière de l'image tandis que pour  $j = 4$ , nous avons 4 valeurs de  $k$  correspondant chacune à un niveau de détail focalisé sur une direction différente. L'idée est d'utiliser ces 5 imagettes (en réalité les coefficients de shearlets) comme signal d'entrée dans notre boucle de commande :

$$\mathbf{s} = \left\langle \mathbf{I}, \psi_{j,k,\mathbf{m}} \right\rangle_{(j_{max}=4)} \quad j = \{0, 4\}, \quad k = \{-1, 0, 1\}, \quad c = \{1, 2\} \quad (3.20)$$

où  $\mathbf{s}$  est le vecteur d'informations visuelles de taille  $5 \times m \times n$  correspondant à la concaténation des valeurs des coefficients des 5 imagettes représentées sur la FIGURE 3.29. Ces imagettes résultent des deux niveaux de décomposition focalisés sur les basses fréquences de l'image. Chacune des  $q \in [1, 2, 3, 4, 5]$  imagettes sera notées  $I^q$  dans la suite.

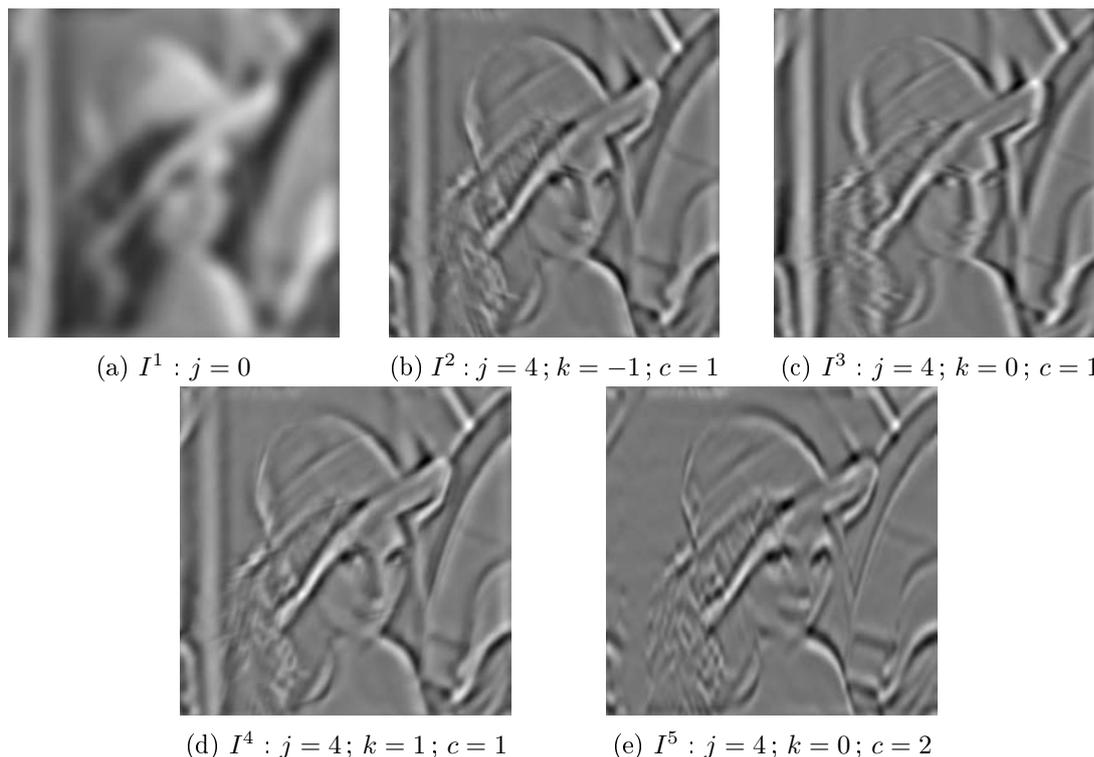


FIGURE 3.29 – Coefficients des shearlets non sous-échantillonnées calculés à partir de la photographie de *Lena*.

Une différence notable entre l'implémentation de la transformée non sous-échantillonnées et sa variante sous-échantillonnée est qu'au niveau  $j = j_{max}$ , c'est à dire la représentation la moins détaillée, quatre représentations seulement sont calculées contre six pour la méthode utilisant les shearlets sous-échantillonnées. En effet, dans l'implémentation non sous-échantillonnée des shearlets, les représentations  $\{ k = -1; c = 1 \}$  et  $\{ k = 1; c = 2 \}$  sont équivalentes ainsi que les représentations  $\{ k = 1; c = 1 \}$  et  $\{ k = -1; c = 2 \}$ . Pour cette raison, une seule des deux représentations est utilisée uniquement, ce qui réduit à quatre le nombre de représentations différentes.

**Loi de commande :** Nous avons considéré uniquement le contrôle des 3 ddl de la sonde qui correspondent aux mouvements dans le plan de l'image échographique  $(t_x, t_y, r_z)^T$  avec  $r_z$  la rotation autour de l'axe orthogonal au plan de l'image échographique. De manière similaire à l'approche d'asservissement visuel utilisant les intensités

des pixels de l'image échographique (Nadeau et Krupa, 2013), chaque ligne  $\mathbf{L}_{I_{i,j,q}} \in \mathbb{R}^{1 \times 3}$  de la matrice d'interaction  $\mathbf{L}_{\mathbf{s}} \in \mathbb{R}^{(5 \times m \times n) \times 3}$  qui relie la variation d'intensité du pixel  $(i, j)$  de l'imagette  $q$  au vecteur vitesse  $\mathbf{v} = (v_x, v_y, \omega_\theta)$  est donné par :

$$\mathbf{L}_{I_{i,j,q}} = [\nabla I_x^q(i, j) \quad \nabla I_y^q(i, j) \quad x \nabla I_y^q(i, j) - y \nabla I_x^q(i, j)] \in \mathbb{R}^{1 \times 3} \quad (3.21)$$

où  $x$  et  $y$  sont les coordonnées métriques 2D du pixel  $(i, j)$  exprimées dans le repère de la sonde 2D centré sur le milieu de l'image échographique :

$$\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} s_x(j - j_0) \\ s_y(i - i_0) \end{pmatrix} \quad (3.22)$$

avec  $(s_x, s_y)$  la taille en mètre d'un pixel et  $(i_0, j_0)$  les coordonnées en pixels du centre de l'image échographique.

Par ailleurs,  $\nabla I_{i,j}^q = [\nabla I_x^q(i, j) \quad \nabla I_y^q(i, j)]$  correspond au gradient spatial 2D associé au pixel  $(i, j)$  qui peut être calculé par l'application d'un filtre dérivatif sur l'image. La matrice d'interaction  $\mathbf{L}_{\mathbf{s}} \in \mathbb{R}^{(5.m.n) \times 3}$  relative aux primitives visuelles  $\mathbf{s}$  est alors obtenue en concaténant les  $\mathbf{L}_{I_{i,j,q}}$  relatifs à chaque pixel :

$$\mathbf{L}_{\mathbf{s}} = \begin{bmatrix} \mathbf{L}_{I_{1,1,1}} \\ \vdots \\ \mathbf{L}_{I_{m,n,5}} \end{bmatrix} \quad (3.23)$$

L'utilisation des shearlets non sous-échantillonnées permet d'être sensible aux singularités (points d'intérêt, zone texturée, contours, etc) de l'image, grâce aux imagettes  $\mathbf{I}^q$  focalisées sur des fréquences de l'image dans des directions distinctes (c'était déjà le cas avec la méthode sous-échantillonnée décrite précédemment). L'avantage, ici, est de pouvoir aussi calculer une estimation analytique de la matrice d'interaction. L'implémentation des shearlets non sous-échantillonnées, issue de SHEARLAB3D, une librairie sous MATLAB nous a imposé l'appel de fonctions MATLAB dans notre implémentation C++ ainsi que le transfert de données entre MATLAB et C++. L'inconvénient étant que le transfert de données entre les deux langages augmente le temps de calcul de la méthode.

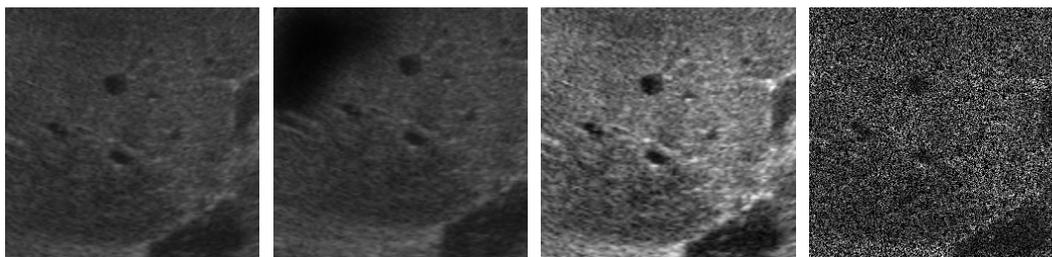
### 3.3.2 Résultats de simulation

Ces travaux de validation utilisent le même simulateur que celui décrit au paragraphe 3.2.2. Nous nous intéresserons particulièrement à l'étude du domaine de convergence de la nouvelle loi de commande utilisant une matrice d'interaction analytique. Nous discuterons, par la suite, les performances de la méthode en conditions nominales et défavorables (présence de bruits ou d'occultations partielles).

**Domaine de convergence :** Afin d'évaluer cette nouvelle information visuelle, nous avons calculé la fonction de coût, présentée dans l'équation (3.18), lors de déplacement de la sonde échographique en translation le long des axes  $x$  et  $y$  et en rotation autour

de l'axe  $z$  orthogonal au plan de l'image. Le domaine de convergence obtenu est alors représenté en évaluant cette fonction numériquement. Elle est rappelée ci-dessous :

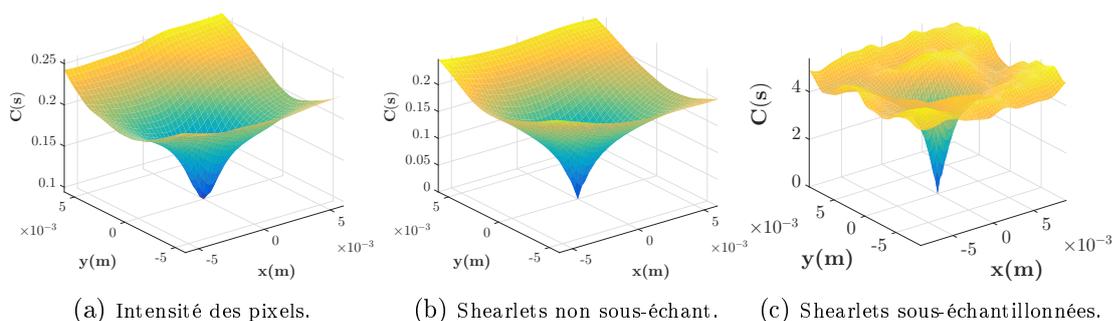
$$C(\mathbf{s}) = \left( (\mathbf{s} - \mathbf{s}^*)^\top (\mathbf{s} - \mathbf{s}^*) \right)^{\frac{1}{2}} \quad (3.24)$$



(a) Conditions nominales (b) Occultations partielles (c) Bruit Gaussien (d) Bruit de *speckle*

FIGURE 3.30 – Différentes images désirées utilisées lors de la validation de la loi de commande proposée.

Nous avons comparé les résultats d'un asservissement visuel utilisant l'intensité des pixels de l'image, de la méthode des shearlets sous-échantillonnées et des shearlets non sous-échantillonnées. Ces comparaisons ont été réalisées dans des conditions nominales (voir FIGURE 3.30 (a)) mais aussi en rajoutant du bruit ou des occultations partielles aux images pendant la tâche de positionnement (voir FIGURE 3.30 (b)-(c)-(d)).



(a) Intensité des pixels. (b) Shearlets non sous-échant. (c) Shearlets sous-échantillonnées.

FIGURE 3.31 – Domaine de convergence des différentes méthodes : intensité des pixels, shearlets sous-échantillonnées ou non.

La FIGURE 3.31 compare les cônes de convergence des trois méthodes suscitées. Chaque cône possède une forme convexe avec un minimum global bien défini. Aux alentours de la position désirée, le domaine de convergence reste convexe dans les approches utilisant les shearlets alors qu'il devient concave pour l'approche fondée sur l'intensité des pixels. A noter que la méthode fondée sur les shearlets sous-échantillonnées présente de nombreux minima locaux. Ceci provient du fait que dans cette approche, nous

n'avions pas pris en compte le niveau  $j = 0$  (voir FIGURE 3.29(a) pour son équivalent non sous-échantillonnées), c'est à dire les plus basses fréquences de l'image originale. Cette information, proche de l'information d'intensité des pixels de l'image est très riche.

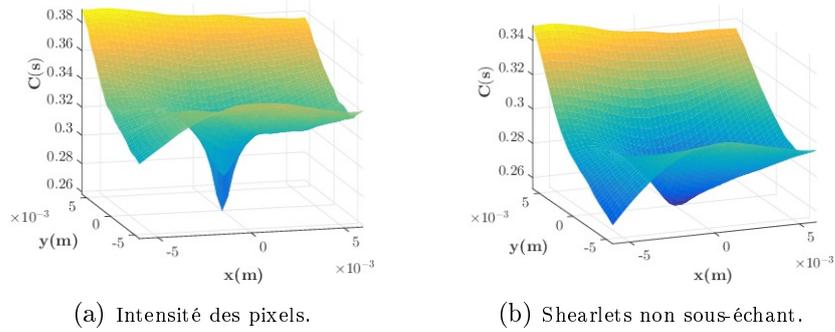


FIGURE 3.32 – Comparaison des cônes de convergence sous l'effet d'occultations partielles.

En présence d'une occultation partielle, le cône de convergence des shearlets non sous-échantillonnées est plus évasé que celui de l'approche utilisant les intensité des pixels mais présente deux minima, ce qui est un gros inconvénient, comme le montre la FIGURE 3.32. En présence d'un bruit Gaussien, les shearlet non sous-échantillonnées offrent un cône permettant une convergence plus rapide vers le minimum lorsque l'on est loin de la position désirée (voir FIGURE 3.33).

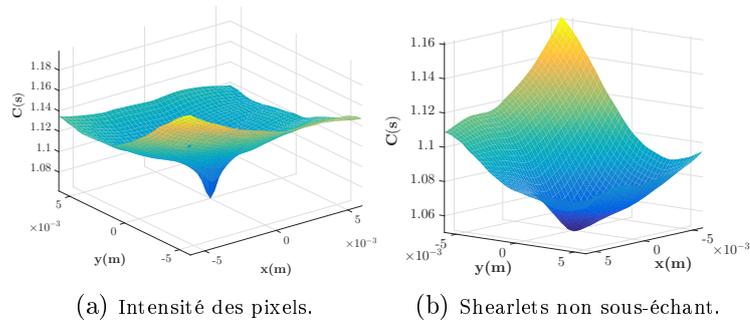


FIGURE 3.33 – Comparaison des cônes de convergence en présence de bruit Gaussien.

En présence de bruit de type *speckle*, la FIGURE 3.34 montre que le cône obtenu par les shearlets non sous-échantillonnées est beaucoup plus lisse que celui de l'approche fondée sur l'intensité des pixels. Les images échographiques étant naturellement affectées par ce type de bruit, cette méthode est un bon candidat pour l'asservissement visuel échographique.

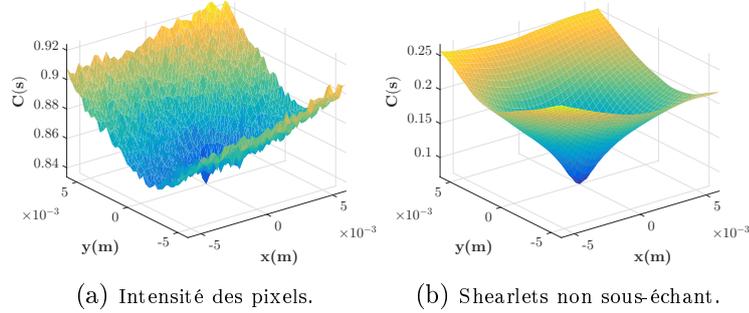


FIGURE 3.34 – Comparaison des cônes de convergence en présence de bruit de *speckle*.

**Tâche de positionnement :** Le test consiste à réaliser une tâche de positionnement automatique à 3 ddl en utilisant l’approche analytique avec les shearlets non sous-échantillonnées et à les comparer avec la méthode fondée sur l’intensité des pixels. La tâche consiste en l’atteinte d’une position désirée à partir d’une position initiale arbitraire. A chaque acquisition d’image, les coefficients des shearlets sont calculés et les vitesses  $\mathbf{v} = (v_x, v_y, \omega_z)^\top$  sont appliquées au bras robotique qui porte la sonde échographique.

Nous avons effectué six tests différents pour comparer la méthode analytique fondée sur les shearlets non sous-échantillonnées et l’approche intensité. Cest tests ont été réalisés suivant trois scénarios différents : conditions nominales, présence de bruit Gaussien et sous l’effet d’un bruit de *speckle*. La FIGURE 3.35 montre l’image initiale (a) et l’image de différence  $\mathbf{I}_{diff}$  (b) observées lors du démarrage de ces six tests.

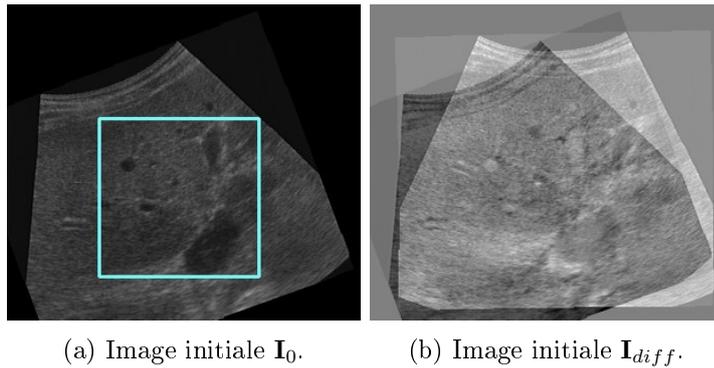


FIGURE 3.35 – Images de référence pour tous les tests.

Pour chaque test, l’erreur de positionnement initiale est de  $\Delta \mathbf{e}_{init} = (10, 5, 20)$  (mm,mm,deg) et le gain  $\lambda$  a été réglé empiriquement afin d’avoir la convergence la plus rapide tout en évitant l’apparition de comportements oscillatoires du robot. La région d’intérêt est entourée en couleur cyan (voir FIGURE 3.36).

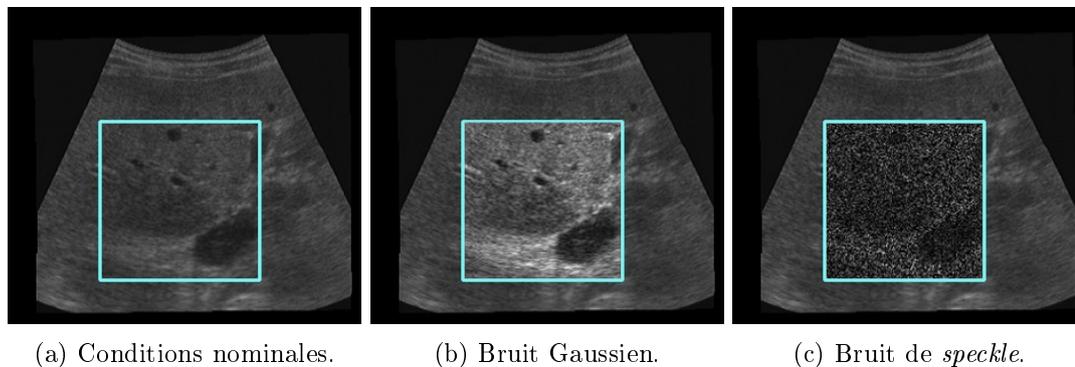


FIGURE 3.36 – Représentation des images désirées pour la comparaison entre l'approche intensité et les shearlets.

Le temps d'exécution d'une itération a été fixée à 40ms, ce qui correspond à une cadence d'acquisition de la sonde échographique virtuelle de 25 images par seconde. Pour les deux approches, les meilleurs performances ont été obtenues avec des gains de  $\lambda = 0.08$  dans les conditions nominales,  $\lambda = 0.03$  avec du bruit Gaussien et  $\lambda = 0.05$  pour du bruit de *speckle*.

**Conditions nominales :** Dans ce scénario, les erreurs finales obtenues par l'approche shearlets et celle utilisant l'intensité convergent toutes vers zéro et les images de différences obtenues, en fin de tâche, sont similaires (voir FIGURE 3.37).

Les FIGURES 3.38 et 3.39 montrent l'évolution de l'erreur de positionnement et de la vitesse appliquée sur chaque ddl pour les deux méthodes suscitées. La méthode des shearlets est plus rapide que celle fondée sur l'intensité des pixels (environ 15 itérations pour les shearlets contre 140 pour la méthode fondée sur l'intensité des pixels).

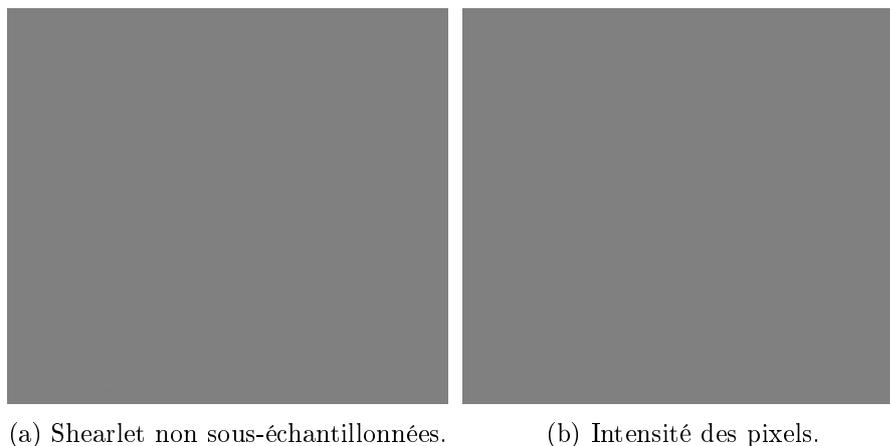


FIGURE 3.37 – Images de différence obtenues après convergence.

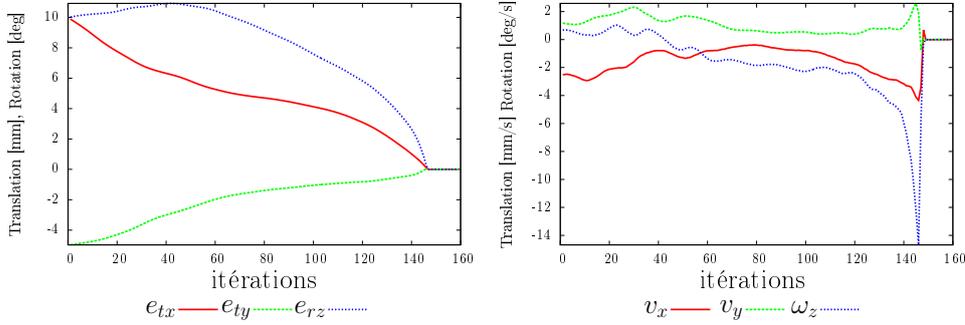


FIGURE 3.38 – [Simulation] Courbes d’erreurs de positionnement et de vitesses obtenues avec l’approche intensité en conditions nominales.

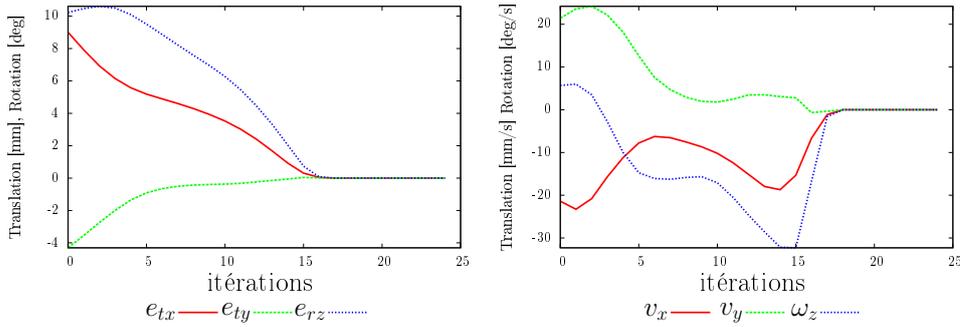
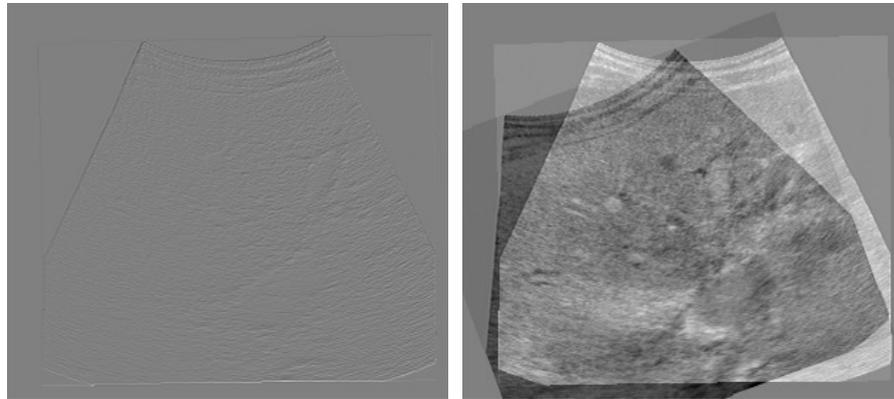


FIGURE 3.39 – [Simulation] Courbes d’erreurs de positionnement et de vitesses obtenues avec les shearlets en conditions nominales.

**Bruit gaussien :** Lorsque du bruit Gaussien est ajouté sur l’image désirée, les deux approches convergent également vers zéro. Les évolutions des erreurs des vitesses sont similaires à celles du premier test.

**Bruit speckle :** Lorsque du bruit de *speckle* est ajouté, l’image de différence finale obtenue avec les shearlets est quasiment grise, signe d’une bonne convergence (voir FIGURE 3.40 (a)) alors que l’autre approche échoue à ce test (voir FIGURE 3.40 (b)). Cela confirme les résultats obtenus dans le paragraphe 3.3.2, montrant que l’approche fondée sur l’intensité est moins adaptée au bruit de type *speckle* que celle des shearlets.

L’erreur de positionnement et l’évolution de la vitesse pour les deux approches sont présentées sur les FIGURES 3.41 et 3.42. Le résultat des shearlets est comparable à celui obtenu en condition nominale, cette méthode n’est pas affectée par le bruit de type *speckle*. En revanche, l’approche intensité échoue à converger dans ce cas et tombe dans un minimum local.



(a) Shearlet non sous-échantillonnées. (b) Intensité des pixels.

FIGURE 3.40 – Erreur finale dans le cas d'un bruit de type *speckle*.

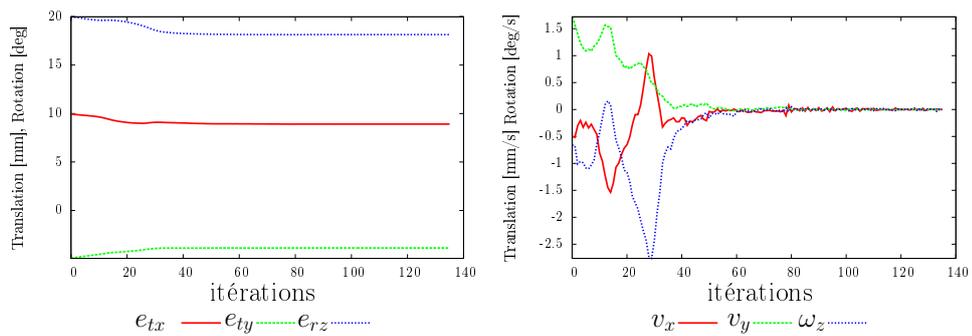


FIGURE 3.41 – [Simulation] Courbes d'erreurs de positionnement et de vitesses obtenues de l'approche intensité avec du bruit de type *speckle*.

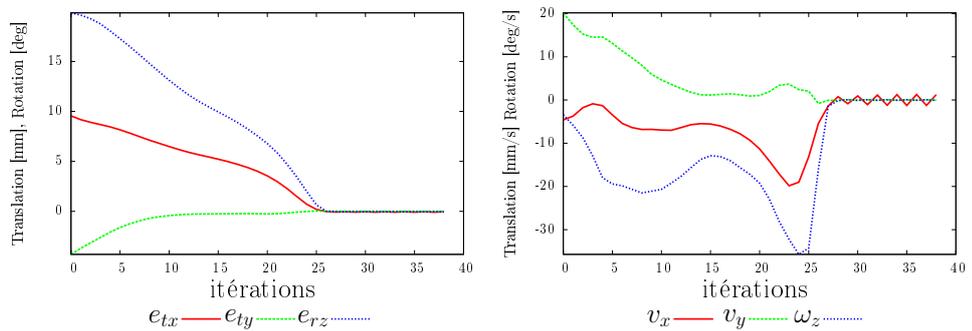


FIGURE 3.42 – [Simulation] Courbes d'erreurs de positionnement et de vitesses obtenues des shearlets avec du bruit de type *speckle*.

Les résultats de simulation sont en adéquation avec les cônes de convergence obtenus dans le paragraphe 3.3.2. En effet, la forme convexe du cône des shearlets au voisinage de la position désirée présageait une convergence rapide de la loi de commande. Cette rapidité est confirmée par le nombre d'itérations nécessaires (moins de 20) pour converger. L'étude de ces cônes indiquait également que la méthode fondée sur les shearlets n'était pas affectée par le bruit de type *speckle*. La simulation a confirmé ces résultats. Ce *speckle* est un inconvénient majeur des images échographiques, même s'il contient de l'information permettant de caractériser certains tissus par leur texture (Krupa et al., 2007).

### 3.3.3 Résultats expérimentaux

Les résultats expérimentaux ont été obtenus avec le même système que celui présenté en section 3.2.3. Certaines images captées lors d'une tâche de positionnement contenant beaucoup d'occultations sont présentées sur la FIGURE 3.43. Les courbes de positionnement et d'évolution de la vitesse correspondantes (voir FIGURE 3.44) montrent que l'approche des shearlets non sous-échantillonnées est robuste aux occultations partielles.

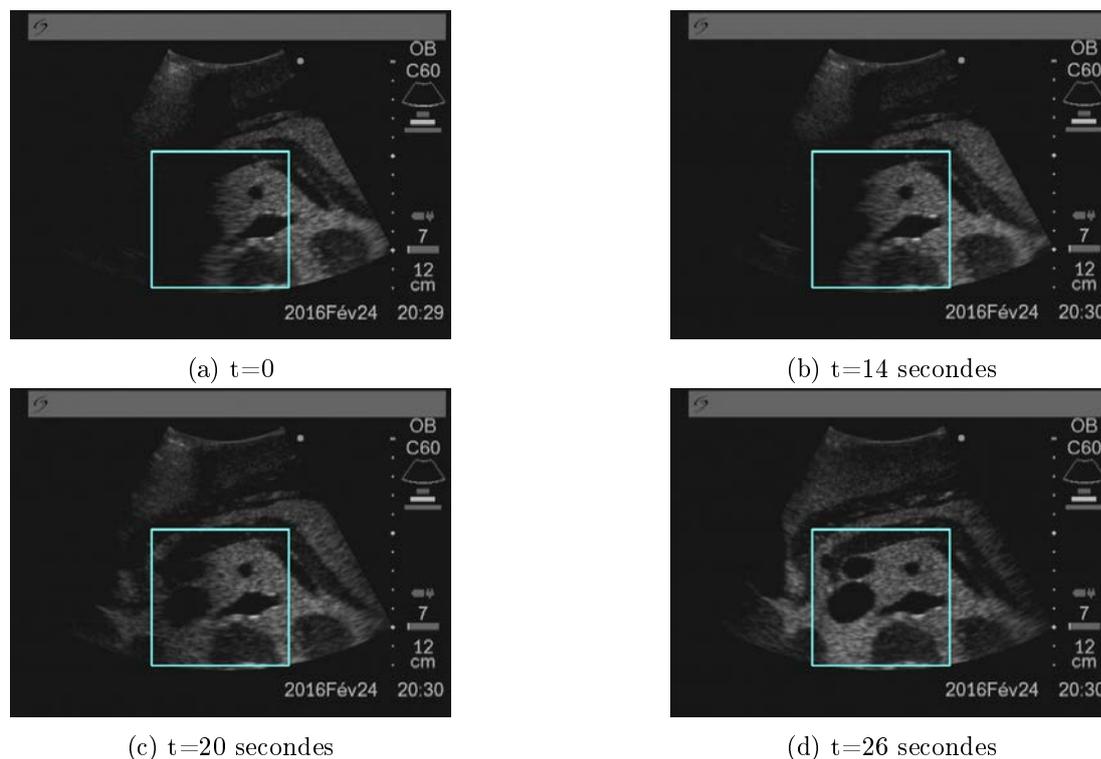


FIGURE 3.43 – Séquence d'images lors d'une tâche de positionnement expérimentale à 3 ddl.

La courbe de la FIGURE 3.44 (a) montre une erreur de positionnement avec une décroissance exponentielle après une courte divergence en  $e_{tx}$  tandis que la FIGURE 3.44 (b) montre l'évolution du vecteur de vitesse. Le comportement du contrôleur reste très similaire à celui observé en simulation.

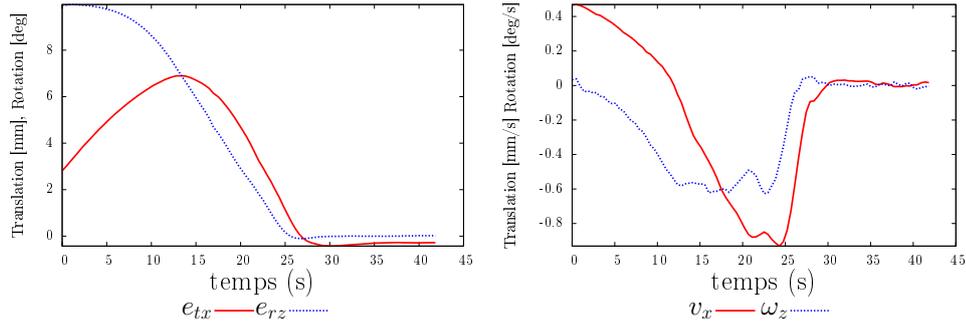


FIGURE 3.44 – [Simulation] Courbes d'erreurs de positionnement et de vitesses obtenues par l'approche des shearlets en conditions nominales.

Le test expérimental a été reproduit plusieurs fois en utilisant : la méthode intensité (I), la méthode des shearlets sous-échantillonnées (SE) présentée en section 3.2 et la méthode des shearlets non sous-échantillonnées (SNE). Trois scénarios ont été considérés : conditions nominales, bruit Gaussien et bruit *speckle*. Pour chaque scénario, trois positions initiales différentes ont été testées. Les gains optimaux :  $\lambda = 0.008$  pour la méthode I,  $\lambda = 0.007$  pour la méthode SE and  $\lambda = 1$  pour la méthode SNE ont été choisis. De plus, la boucle de commande n'a pas la même durée pour chaque méthode avec 40ms pour I, 60ms pour SE et 200ms pour NSE.

Le TABLEAU 3.2 résume les résultats obtenus dans chaque cas. La méthode SNE fonctionne pour chaque test, la méthode I échoue deux fois lors de l'ajout du bruit de *speckle*. Le cône de cette méthode intensité était très impacté par la présence de *speckle*, cet échec est donc en adéquation avec les résultats de simulation. La méthode SE échoue pour une seule position initial sans "explications".

L'inconvénient majeur de la méthode SNE est le temps de calcul nécessaire pour obtenir les coefficients des shearlets. Nous avons alors un temps de boucle de 200ms, mais le temps de convergence demeure du même ordre que pour les deux autres méthodes car cette approche nécessite peu d'itérations pour converger.

TABLE 3.2 – Étude de la robustesse des lois de commande pour les méthodes intensité (I), shearlets sous-échantillonnées (SE) et shearlets non sous-échantillonnées (SNE) ( $e_x$  (mm),  $e_z$  (deg),  $\Delta \mathbf{e}_{init}$  and  $\mathbf{e}_f$  représentent les erreurs de positionnements initiales et finales.)

$\mathbf{e}_f$ and $\mathbf{e}_0$	SNE		I		SE	
	$e_x$	$e_z$	$e_x$	$e_z$	$e_x$	$e_z$
$\Delta \mathbf{e}_{init}$	10	8	10	8	10	8
$\mathbf{e}_f$ Nominal	0.3	0.063	0.3	0.057	0.4	0.5
$\mathbf{e}_f$ Gaussian	0.3	0.106	0.5	0.017	1.3	1.2
$\mathbf{e}_f$ Speckle	0.2	0.487	0.4	0.0744	0.02	0.12
$\Delta \mathbf{e}_{init}$	15	5	15	5	15	5
$\mathbf{e}_f$ Nominal	0.2	0.109	0.4	0.080	0.1	0.44
$\mathbf{e}_f$ Gaussian	0.7	0.091	0.4	0.022	0.07	0.09
$\mathbf{e}_f$ Speckle	0.3	0.206	échec	échec	1.7	2.1
$\Delta \mathbf{e}_{init}$	5	10	5	10	5	10
$\mathbf{e}_f$ Nominal	0.3	0.051	0.6	0.114	échec	échec
$\mathbf{e}_f$ Gaussian	0.9	0.143	0.3	0.051	échec	échec
$\mathbf{e}_f$ Speckle	0.3	0.0171	échec	échec	échec	échec
temps moyen	14.6	secondes	17	secondes	14	secondes

### 3.4 Conclusion

Nous avons vu dans ce chapitre les particularités de certaines modalités d'imagerie médicale (échographie et OCT) qui sont caractérisées par un rapport signal/bruit généralement défavorable. Deux méthodes d'asservissement visuel fondées sur les shearlets (sous-échantillonnées et non sous-échantillonnées) ont été présentées utilisant deux implémentations différentes et démontrent expérimentalement une robustesse au bruit de type *speckle*. Si la première méthode, utilisant des shearlets dites sous-échantillonnées, nécessite peu de coefficients pour le calcul de la matrice d'interaction, elle se base sur une estimation numérique de cette dernière et possède un cône de convergence réduit. Les primitives visuelles considérées n'incluent pas les basses-fréquences de l'image (niveau  $j = 0$ ), ce qui explique l'étroitesse de ce cône de convergence. La seconde méthode proposée afin de réaliser un asservissement visuel sur imagerie échographique robuste au bruit de *speckle*, repose sur l'utilisation de shearlets dites non sous-échantillonnées, et inclut les basses fréquences de l'image. Cette variante d'implémentation des shearlets permet aisément de définir une formulation analytique de la matrice d'interaction mais engendre le transfert de nombreuses données et un temps de calcul important. Le chapitre suivant permettra d'établir les fondements d'un asservissement visuel générique utilisant différentes représentations multi-échelles. Nous exposerons le cas des ondelettes et des shearlets (sous-échantillonnées ou non).

## Chapitre 4

# Généralisation à l'asservissement visuel fondé sur les transformées multi-échelles

Dans le chapitre précédent, nous avons vu comment la transformée en shearlets sous-échantillonnée avait été validée pour une tâche d'asservissement visuel sur l'imagerie échographique avec une estimation numérique de la matrice d'interaction. Une première formulation analytique de cette matrice a ensuite été présentée. Même si elle a montré un plus grand domaine de convergence et une meilleure robustesse aux bruits et occultations, cette nouvelle formulation est applicable uniquement avec les coefficients des shearlets non-échantillonnés. Elle possède également l'inconvénient d'être coûteuse en temps de calcul. Dans ce chapitre, nous présenterons une méthode générique d'asservissement visuel fondée sur les représentations multi-échelles et testée sur des images fournies par une caméra conventionnelle. Quatre transformées multi-échelles seront étudiées et comparées, à savoir les ondelettes et les shearlets dans leurs versions sous-échantillonnées ou non. L'approche utilisant les ondelettes sous-échantillonnées sera également validée sur une autre modalité d'imagerie médicale qui est la tomographie par cohérence optique (OCT). Le sommaire ci-dessous présente la disposition de ce chapitre :

### Sommaire

---

<b>4.1 Principe de l'asservissement visuel fondé sur les transformées multi-échelles . . . . .</b>	<b>108</b>
4.1.1 Approche utilisant la transformée en ondelettes . . . . .	108
4.1.2 Approche utilisant la transformée en shearlets . . . . .	114
<b>4.2 Simulation et validation expérimentale sur caméra conventionnelle . . . . .</b>	<b>115</b>
4.2.1 Détails d'implémentation . . . . .	116
4.2.2 Domaine de convergence . . . . .	119

4.2.3	Simulation . . . . .	120
4.2.4	Validation expérimentale . . . . .	126
<b>4.3</b>	<b>Validation expérimentale en utilisant l'OCT . . . . .</b>	<b>138</b>
4.3.1	Description du dispositif robotique . . . . .	138
4.3.2	Détails d'implémentation . . . . .	139
4.3.3	Validation expérimentale . . . . .	141
<b>4.4</b>	<b>Conclusion . . . . .</b>	<b>147</b>

---

## 4.1 Principe de l'asservissement visuel fondé sur les transformées multi-échelles

Nous avons voulu établir une méthode permettant d'utiliser des transformées multi-échelles en asservissement visuel avec une approche analytique la plus générique possible. Nous distinguerons, dans la suite, le cas des ondelettes et des shearlets, mais le principe qui sera développé pourrait être étendu à d'autres transformées multi-échelles. L'ensemble de ces résultats théoriques ainsi que la validation sur caméra conventionnelle a fait l'objet d'une publication (Dufflot et al., 2018).

### 4.1.1 Approche utilisant la transformée en ondelettes

Au lieu d'utiliser l'intensité de chaque pixel de l'image comme préconisé dans la méthode d'asservissement visuel photométrique (voir section 1.3.2), nous proposons d'utiliser dans cette nouvelle approche, les coefficients d'ondelettes de l'image. En considérant un ensemble de  $L \in \mathbb{N}$  fonctions génératrices d'ondelettes  $(\psi^{(l)})_{l \leq L} \subset L^2(\mathbb{R}^2)$  comme introduites en (2.15), nous pouvons définir les primitives visuelles comme étant la concaténation du produit scalaire de l'image à l'instant  $t$ , notée  $I_t$ , avec les fonctions génératrices d'ondelettes  $\psi_{j,\mathbf{m}}^{(l)}$  où  $j$  correspond à l'échelle de la transformée,  $l \in \{0, 1, 2, 3\}$  le type de fonction ondelette (générant une approximation du signal ou des détails verticaux, horizontaux ou diagonaux) et  $\mathbf{m} = (m_1, m_2)$  un facteur de translation dépendant de  $j$ .

En utilisant la notation suivante :

$$(\mathcal{W}_{\psi^{(l)}} f)(j, \mathbf{m}) = \langle I_t, \psi_{j,\mathbf{m}}^{(l)} \rangle = \iint_{\mathbb{R}^2} I(x, y, t) \psi_{j,\mathbf{m}}^{(l)} dx dy, \quad (4.1)$$

nous pouvons définir le vecteur des primitives visuelles sous la forme suivante :

$$\mathbf{s}_w(t) = \left( (\mathcal{W}_{\psi^{(l_1)}} I_t)(j_1, \mathbf{m}_1), (\mathcal{W}_{\psi^{(l_2)}} I_t)(j_2, \mathbf{m}_2), \dots, (\mathcal{W}_{\psi^{(l_N)}} I_t)(j_N, \mathbf{m}_N) \right)^\top \quad (4.2)$$

L'erreur visuelle est donnée, en utilisant les coefficients d'ondelettes de l'image, par :

$$\mathbf{e}_w(t) = \mathbf{s}_w(t) - \mathbf{s}_w^* \quad (4.3)$$

Afin de minimiser  $\mathbf{e}_w$ , une solution consiste à mettre en œuvre une loi de commande fondée sur une optimisation de type *Levenberg-Marquardt* :

$$\mathbf{v}(t) = -\lambda \left( \mathbf{H}(t) + \mu \text{diag}(\mathbf{H}(t)) \right)^{-1} \left( \mathbf{L}_{s_w}(t) \right)^\top \mathbf{e}_w(t), \quad (4.4)$$

où  $\lambda > 0$  est un gain,  $\mu > 0$  un poids et  $\mathbf{H}(t) = \left( \mathbf{L}_{s_w}(t) \right)^\top \mathbf{L}_{s_w}(t)$  est la matrice Hessianne associée à  $\mathbf{s}_w$ . Une haute valeur pour  $\mu$  ( $\mu = 1$ ) correspond à une optimisation de type de descente de gradient tandis qu'une valeur faible de  $\mu$  ( $\mu = 10^{-3}$ ) revient à utiliser la technique de *Gauss-Newton* présentée en (1.12).

La difficulté est alors de définir la matrice d'interaction  $\mathbf{L}_{s_w} \in \mathbb{R}^{N \times 6}$  qui relie le vecteur de vitesse de la caméra  $\mathbf{v}$  à la variation temporelle des coefficients d'ondelettes :

$$\frac{d\mathbf{s}_w(t)}{dt} = \mathbf{L}_{s_w}(t)\mathbf{v}(t) \quad (4.5)$$

La formulation analytique de cette nouvelle matrice d'interaction, sera établie de deux manières, dans la suite de ce document. Premièrement nous présenterons la formulation exacte de cette matrice et dans un second temps, une approximation qui a l'avantage d'être moins coûteuse en temps de calcul par rapport à la version exacte. A noter que notre première intuition était de générer à partir d'une base d'ondelettes  $\psi^{(l)}$ , une base d'ondelettes  $\Psi^{(l)}$  dont la transformée donnerait directement les coefficients de la transformée utilisant les ondelettes  $\psi^{(l)}$  appliquée sur la dérivée de l'image initiale. Cette approche n'a toutefois pas porté ses fruits.

**Matrice d'interaction fondée sur les coefficients d'ondelettes :** En supposant que les fonctions génératrices d'ondelettes  $\psi_{j,\mathbf{m}}^{(l)}$  ainsi que les intensités de l'image  $I_t$  sont différentiables et en utilisant la définition du produit scalaire pour les fonctions de  $\mathcal{L}^2(\mathbb{R}^2)$  (voir équation (4.1)), nous pouvons écrire la relation suivante :

$$\frac{d(\mathcal{W}_{\psi^{(l)}} I_t)(j, \mathbf{m})}{dt} = \frac{d \langle I_t, \psi_{j,\mathbf{m}}^{(l)} \rangle}{dt} = \frac{d \iint_{\mathbb{R}^2} I(x, y, t) \psi_{j,\mathbf{m}}^{(l)}(x, y) dx dy}{dt} \quad (4.6)$$

**Définition 4.1. Théorème de transport de Reynolds dans  $\mathbb{R}^2$  :** Soit  $f$  une fonction de  $\mathcal{L}^2(\mathbb{R}^2)$  alors, le théorème de transport de Reynolds nous permet d'écrire la relation suivante :

$$\frac{d}{dt} \int_{\mathbb{R}^2} f dV = \int_{\mathbb{R}^2} \frac{\partial f}{\partial t} dV \quad (4.7)$$

Ce théorème peut être appliqué à l'équation (4.6), nous permettant d'écrire la relation suivante :

$$\frac{d(\mathcal{W}_{\psi^{(l)}} I_t)(j, \mathbf{m})}{dt} = \frac{d \langle I_t, \psi_{j,\mathbf{m}}^{(l)} \rangle}{dt} = \iint_{\mathbb{R}^2} \frac{\partial I(x, y, t)}{\partial t} \psi_{j,\mathbf{m}}^{(l)}(x, y) dx dy \quad (4.8)$$

Or nous avons vu dans le paragraphe sur l'asservissement visuel photométrique (voir section 1.3.2) que la continuité du flux optique nous permet d'écrire la relation (1.32), rappelée ci-dessous :

$$\frac{\partial I(x, y, t)}{\partial t} = -\left(\nabla I_t(x, y)\right)^\top \mathbf{L}_p(x, y) \mathbf{v}(t), \quad (4.9)$$

où  $\nabla I_t(x, y)$  est le gradient de l'image  $I_t$  et où  $\mathbf{L}_p$  est la matrice d'interaction associée à un point de la scène de coordonnées  $(x, y)$  dans le plan image et de profondeur  $Z$  par rapport au repère caméra, définie par la relation suivante :

$$\mathbf{L}_p(x, y) = \begin{bmatrix} -1/Z & 0 & x/Z & xy & -(1+x^2) & y \\ 0 & -1/Z & y/Z & 1+y^2 & -xy & -x \end{bmatrix} \quad (4.10)$$

Grâce à l'égalité (4.9), l'équation (4.8) peut alors s'écrire de la manière suivante :

$$\frac{d \langle I_t, \psi_{j, \mathbf{m}}^{(l)} \rangle}{dt} = - \iint_{\mathbb{R}^2} (\nabla I_t(x, y))^\top \mathbf{L}_p(x, y) \mathbf{v}(t) \psi_{j, \mathbf{m}}^{(l)}(x, y) dx dy \quad (4.11)$$

Afin de simplifier les notations, nous utiliserons dans la suite l'égalité suivante :

$$I_t^{(i)} = (\nabla I_t(x, y))^\top \mathbf{L}_p(x, y) \mathbf{e}_i, \quad (4.12)$$

où  $i \in \{1, \dots, 6\}$  et  $\mathbf{e}_i \in \mathbb{R}^6$  est le  $i^{\text{ème}}$  vecteur unité canonique de  $\mathbb{R}^6$ . En d'autres termes,  $I_t^{(i)}$  correspond à la  $i^{\text{ème}}$  colonne d'une matrice d'interaction de l'approche d'asservissement visuel photométrique.

Nous obtenons alors la relation suivante :

$$\frac{d \langle I_t, \psi_{j, \mathbf{m}}^{(l)} \rangle}{dt} = - \iint_{\mathbb{R}^2} \left[ I_t^{(1)}, I_t^{(2)}, I_t^{(3)}, I_t^{(4)}, I_t^{(5)}, I_t^{(6)} \right] \psi_{j, \mathbf{m}}^{(l)}(x, y) \mathbf{v}(t) dx dy \quad (4.13)$$

Or, par définition, nous savons que :

$$\iint_{\mathbb{R}^2} I_t^{(i)} \psi_{j, \mathbf{m}}^{(l)}(x, y) dx dy = \langle I_t^{(i)}, \psi_{j, \mathbf{m}}^{(l)}(x, y) \rangle = (\mathcal{W}_{\psi^{(l)}} I_t^{(i)})(j, \mathbf{m}) \quad (4.14)$$

Finalement, pour répondre à la définition (4.5), la matrice d'interaction  $\mathbf{L}_{s_w}$  peut s'écrire comme suit :

$$\mathbf{L}_{s_w}(t) = - \begin{bmatrix} (\mathcal{W}_{\psi^{(l_1)}} I_t^{(1)})(j_1, \mathbf{m}_1) & \cdots & (\mathcal{W}_{\psi^{(l_1)}} I_t^{(6)})(j_1, \mathbf{m}_1) \\ \vdots & \ddots & \vdots \\ (\mathcal{W}_{\psi^{(l_N)}} I_t^{(1)})(j_N, \mathbf{m}_N) & \cdots & (\mathcal{W}_{\psi^{(l_N)}} I_t^{(6)})(j_N, \mathbf{m}_N) \end{bmatrix} \in \mathbb{R}^{N \times 6} \quad (4.15)$$

La FIGURE 4.1 illustre le principe utilisé pour obtenir la matrice d'interaction  $\mathbf{L}_{s_w}$  en schématisant les équations présentées précédemment. Nous pouvons remarquer qu'en retirant du schéma les transformées en ondelette, nous obtenons le principe d'élaboration de la matrice d'interaction associée à l'information photométrique.

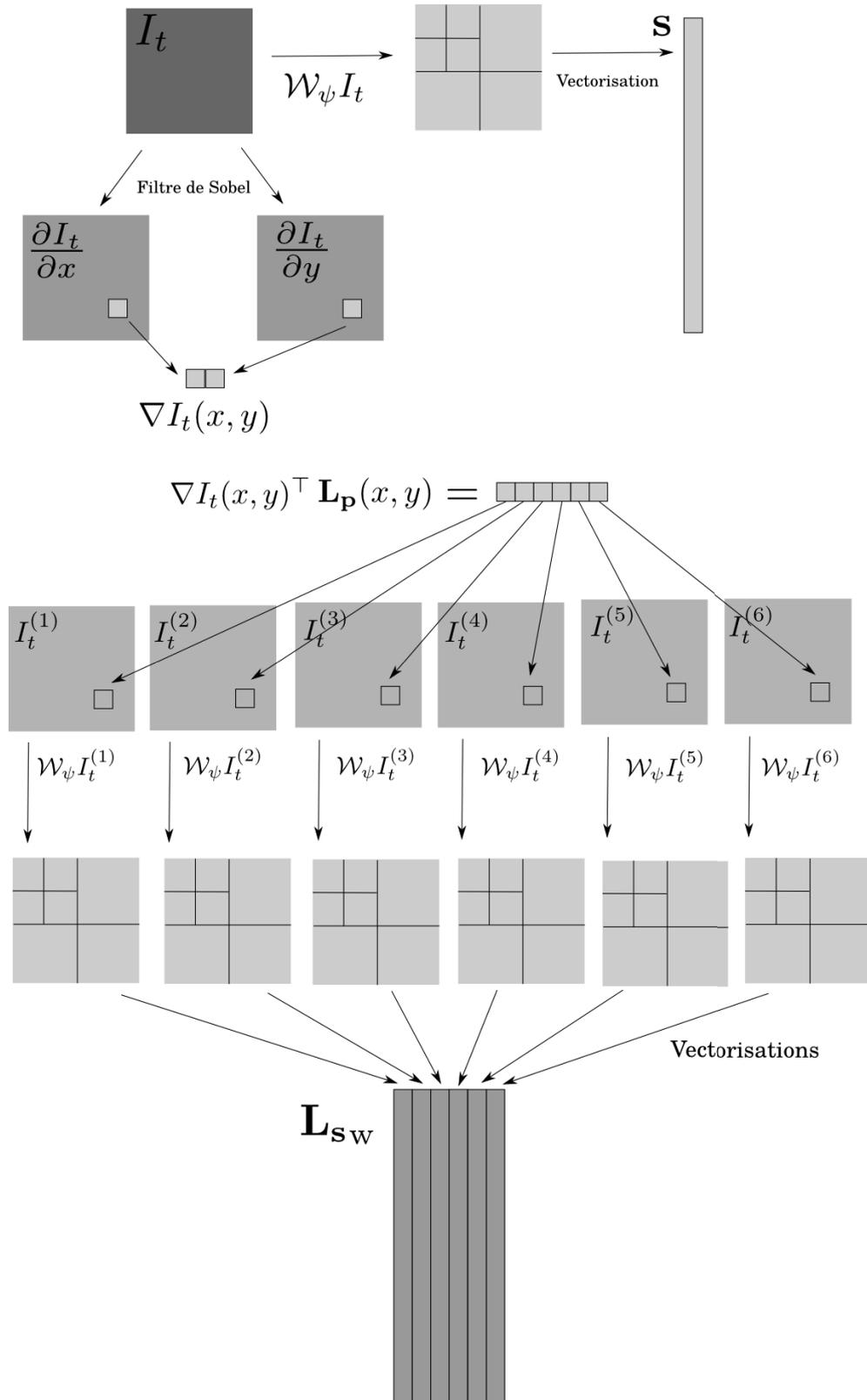


FIGURE 4.1 – Schématisation du calcul de la matrice d'interaction associée aux coefficients des ondelettes.

**Approximation de la matrice d'interaction fondée sur les coefficients d'ondelettes :** Bien que la transformée en ondelettes rapide (FWT) soit un algorithme de complexité ( $O(n)$ ), calculer la matrice d'interaction à chaque itération revient à effectuer six transformées en ondelettes en plus de celle nécessaire pour obtenir le vecteur d'informations visuelles. Pour limiter le temps de calcul, nous pouvons, soit réduire le nombre de niveaux de décompositions de la transformée, ou bien construire une approximation de la matrice d'interaction  $\mathbf{L}_{\mathbf{s}_w}$  qui utilise uniquement deux transformées en ondelettes.

**Définition 4.2. Support d'une fonction :** Le support d'une fonction ou d'une application est la partie de son ensemble de définition sur laquelle se concentre l'information utile de cette fonction. Pour une fonction numérique, c'est la partie du domaine où elle n'est pas nulle.

Nous supposons dans la suite que toutes les fonctions génératrices  $\psi^{(l)}$  sont compactes sur leur support dans le domaine spatial. Il en résulte que les vitesses  $\left(\frac{dx}{dt}, \frac{dy}{dt}\right)$  de l'équation du flux optique, présentée en (1.31), restent constantes sur le support des ondelettes  $\psi_j^{(l)}$ . Cette supposition reste inchangée en remplaçant  $\mathbf{L}_p(x, y)$  dans l'intégrale (4.11) par une matrice  $\mathbf{L}_{pw}(j, \mathbf{m})$  qui dépend uniquement du centre du support des ondelettes  $\psi_j^{(l)}$  et cela pour chaque niveau  $j$ . Ainsi, la matrice  $\mathbf{L}_{pw}(j, \mathbf{m})$  peut être simplement calculée en évaluant  $\mathbf{L}_p$  au niveau du centre du support des ondelettes :

$$\mathbf{L}_{pw}(j, \mathbf{m}) = \mathbf{L}_p(2^{-j}m_1, 2^{-j}m_2). \quad (4.16)$$

En utilisant la relation (4.16), une approximation  $\tilde{\mathbf{L}}_{\mathbf{s}_w}(t) \approx \mathbf{L}_{\mathbf{s}_w}(t)$  de la matrice d'interaction associée aux ondelettes à l'instant  $t$  peut être obtenue en effectuant uniquement deux transformées en ondelettes, une sur la dérivée partielle en  $x$  de l'image, l'autre sur celle en  $y$  :

$$\tilde{\mathbf{L}}_{\mathbf{s}_w}(t) = - \begin{bmatrix} \left( (\mathcal{W}_{\psi^{(l_1)}} \frac{\partial I_t}{\partial x})(j_1, \mathbf{m}_1), (\mathcal{W}_{\psi^{(l_1)}} \frac{\partial I_t}{\partial y})(j_1, \mathbf{m}_1) \right) \mathbf{L}_{pw}(j_1, \mathbf{m}_1) \\ \vdots \\ \left( (\mathcal{W}_{\psi^{(l_N)}} \frac{\partial I_t}{\partial x})(j_N, \mathbf{m}_N), (\mathcal{W}_{\psi^{(l_N)}} \frac{\partial I_t}{\partial y})(j_N, \mathbf{m}_N) \right) \mathbf{L}_{pw}(j_N, \mathbf{m}_N) \end{bmatrix} \in \mathbb{R}^{N \times 6} \quad (4.17)$$

La FIGURE 4.2 résume le principe utilisé pour obtenir la matrice d'interaction approchée  $\tilde{\mathbf{L}}_{\mathbf{s}_w}$  à l'aide d'un schéma. En le comparant à la FIGURE 4.1, nous pouvons constater que cette manière d'obtenir la matrice d'interaction demande de réaliser trois fois moins de transformées en ondelettes que la précédente.

Ainsi, les paragraphes précédents ont montré comment utiliser la transformée en ondelettes du gradient de l'image pour calculer notre matrice d'interaction. La section suivante montrera comment étendre ce résultat aux shearlets.

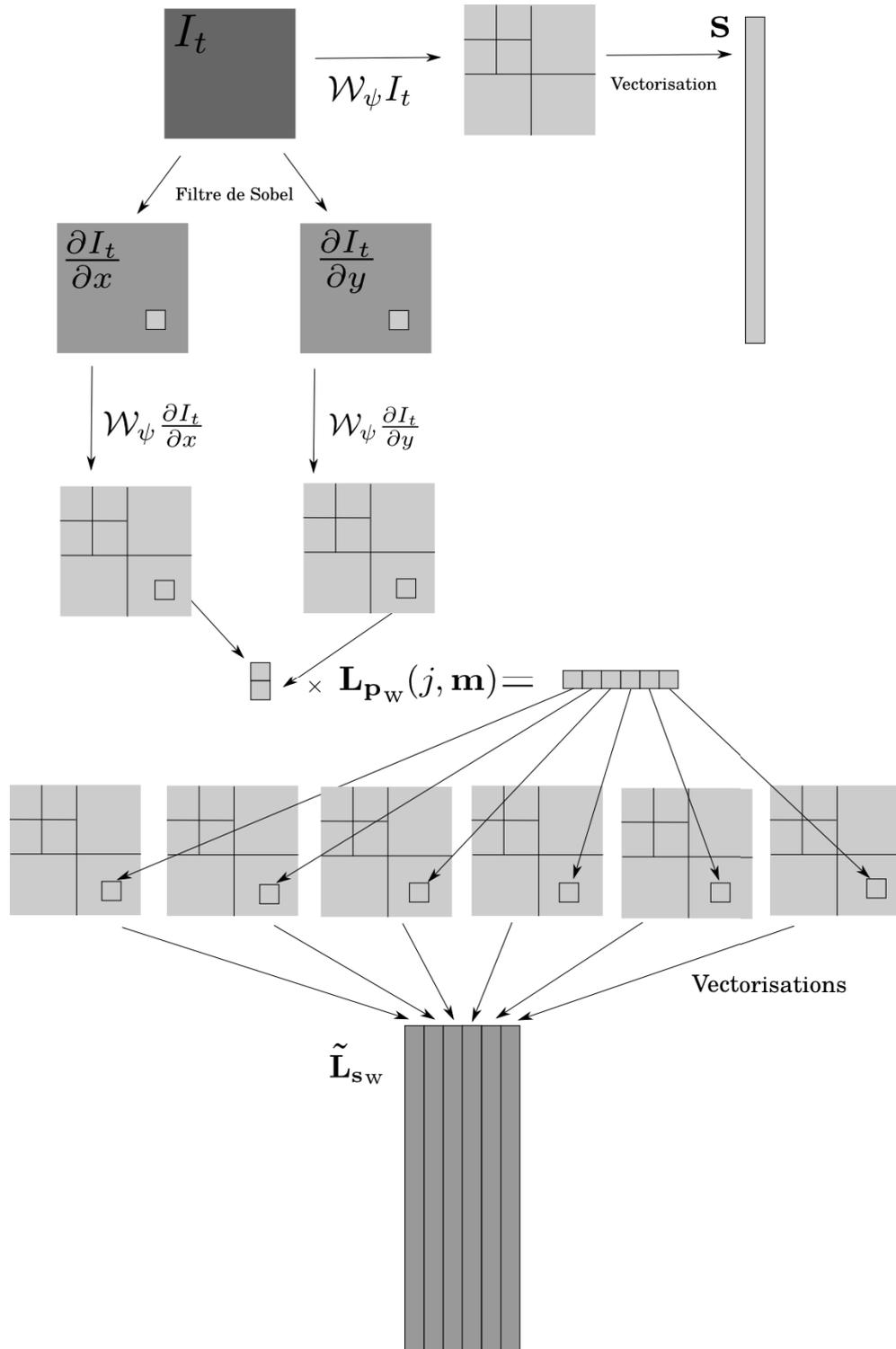


FIGURE 4.2 – Schématisation du calcul de l'approximation de la matrice d'interaction associée aux coefficients des ondelettes.

### 4.1.2 Approche utilisant la transformée en shearlets

Les représentations fondées sur les shearlets peuvent être également utilisées de la même manière que celles des ondelettes pour définir la matrice d'interaction. Pour un nombre fini  $L \in \mathbb{N}$  de fonctions génératrices de shearlets  $(\psi^{(l)})_{l \leq L} \subset L^2(\mathbb{R}^2)$ , les informations visuelles correspondent aux coefficients des shearlets définies avec  $l_n, j_n, \mathbf{m}_n, k_n \subset \{1, \dots, L\} \times \mathbb{N} \times \mathbb{Z}^2 \times \mathbb{Z}$ .  $l_n$  représente la fonction génératrice utilisée,  $j_n$  correspond à l'échelle,  $\mathbf{m}_k$  à l'opérateur de translation et  $k_n$  à celui de découpage (*shearing* en anglais). Ces primitives visuelles sont données à l'instant  $t$  par :

$$\mathbf{s}_{\text{sh}}(t) = \left( (\mathcal{S}_{\psi^{(l_1)}} I_t)(j_1, k_1, \mathbf{m}_1), \dots, (\mathcal{S}_{\psi^{(l_N)}} I_t)(j_N, k_N, \mathbf{m}_N) \right)^\top \quad (4.18)$$

Le terme  $(\mathcal{S}_{\psi^{(l)}} I_t)(j, k, \mathbf{m})$  est défini de deux manières différentes selon le cône dans lequel la transformée en sheartlet est calculée.

Si la fonction génératrice de shearlets  $\psi^{(l)}$  appartient au cône fréquentiel horizontal,  $(\mathcal{S}_{\psi^{(l)}} I_t)(j, k, \mathbf{m})$  correspond à :

$$(\mathcal{S}_{\psi^{(l)}} I_t)(j, k, \mathbf{m}) = \iint_{\mathbb{R}^2} I_t(x, y) D_{\mathbf{A}}^j D_{\mathbf{S}_k} T_{\mathbf{m}} \psi^{(l)}(x, y) dx dy \quad (4.19)$$

Si la fonction génératrice de shearlets  $\psi^{(l)}$  appartient au cône fréquentiel vertical, il est donné par la relation :

$$(\mathcal{S}_{\psi^{(l)}} I_t)(j, k, \mathbf{m}) = \iint_{\mathbb{R}^2} I_t(x, y) D_{\mathbf{A}}^j D_{\mathbf{S}_k^\top} T_{\mathbf{m}} \psi^{(l)}(x, y) dx dy \quad (4.20)$$

Nous définissons l'erreur visuelle à minimiser comme suit :

$$\mathbf{e}_{\text{sh}}(t) = \mathbf{s}_{\text{sh}}(t) - \mathbf{s}_{\text{sh}}^* \quad (4.21)$$

Nous proposons comme dans le cas des ondelettes d'utiliser une loi de commande avec optimisation de type *Levenberg-Marquardt* pour calculer le vecteur vitesse de commande à appliquer au robot :

$$\mathbf{v}(t) = -\lambda \left( \mathbf{H}(t) + \mu \text{diag}(\mathbf{H}(t)) \right)^{-1} \left( \mathbf{L}_{\mathbf{s}_{\text{sh}}}(t) \right)^\top \mathbf{e}_{\text{sh}}(t), \quad (4.22)$$

où  $\lambda > 0$  est un gain,  $\mu > 0$  un facteur de poids et  $\mathbf{H}(t) = \left( \mathbf{L}_{\mathbf{s}_{\text{sh}}}(t) \right)^\top \mathbf{L}_{\mathbf{s}_{\text{sh}}}(t)$ .

**Matrice d'interaction fondée sur les coefficients des shearlets :** De manière similaire aux équations (4.8) à (4.12), la matrice d'interaction associée au vecteur  $\mathbf{s}_{\text{sh}}$  à l'instant  $t$  peut être écrite sous la forme suivante :

$$\mathbf{L}_{\mathbf{s}_{\text{sh}}}(t) = - \begin{bmatrix} (\mathcal{S}_{\psi^{(l_1)}} I_t^{(1)})(j_1, k_1, \mathbf{m}_1) & \cdots & (\mathcal{S}_{\psi^{(l_1)}} I_t^{(6)})(j_1, k_1, \mathbf{m}_1) \\ \vdots & \ddots & \vdots \\ (\mathcal{S}_{\psi^{(l_N)}} I_t^{(1)})(j_N, k_N, \mathbf{m}_N) & \cdots & (\mathcal{S}_{\psi^{(l_N)}} I_t^{(6)})(j_N, k_N, \mathbf{m}_N) \end{bmatrix} \in \mathbb{R}^{N \times 6} \quad (4.23)$$

Son obtention consiste à appliquer, de manière analogue au cas des ondelettes, une transformée en shearlets sur chaque colonne de la matrice d'interaction définie dans la méthode photométrique.

**Approximation de la matrice d'interaction fondée sur les coefficients des shearlets :** La différence entre la méthode des ondelettes et celle des shearlets se trouve dans la définition de la matrice d'interaction approchée  $\tilde{\mathbf{L}}_{\text{s sh}}$ . Ceci est dû au fait que le centre du support des shearlets n'est pas seulement déterminé par le paramètre  $j$  mais dépend également du découpage  $k$ . En faisant l'hypothèse que le centre du support des shearlets est situé autour de l'origine du domaine spatial, nous obtenons deux formulations de  $\mathbf{L}_{\text{psh}}(j, k, \mathbf{m})$ . Dans le cas où la fonction génératrice de shearlets appartient au cône fréquentiel horizontal, nous pouvons écrire :

$$\mathbf{L}_{\text{psh}}(j, k, \mathbf{m}) = \mathbf{L}_{\text{p}}\left(2^{-j}(m_1 - km_2), 2^{-j/2}m_2\right), \quad (4.24)$$

Dans le cas où la fonction génératrice de shearlets appartient au cône fréquentiel vertical, nous utilisons la relation suivante :

$$\mathbf{L}_{\text{psh}}(j, k, \mathbf{m}) = \mathbf{L}_{\text{p}}\left(2^{-j/2}m_1, 2^{-j}(m_2 - km_1)\right), \quad (4.25)$$

L'approximation de la matrice d'interaction  $\tilde{\mathbf{L}}_{\text{s sh}}(t) \approx \mathbf{L}_{\text{s sh}}(t)$  est donnée, de manière analogue à celle des ondelettes, par :

$$\tilde{\mathbf{L}}_{\text{s sh}}(t) = - \left[ \begin{array}{c} \left( (\mathcal{S}_{\psi^{(l_1)}} \frac{\partial I_t}{\partial x})(j_1, k_1, \mathbf{m}_1), (\mathcal{S}_{\psi^{(l_1)}} \frac{\partial I_t}{\partial y})(j_1, \mathbf{m}_1) \right) \mathbf{L}_{\text{psh}}(j_1, k_1, \mathbf{m}_1) \\ \vdots \\ \left( (\mathcal{S}_{\psi^{(l_N)}} \frac{\partial I_t}{\partial x})(j_N, k_N, \mathbf{m}_N), (\mathcal{S}_{\psi^{(l_N)}} \frac{\partial I_t}{\partial y})(j_N, k_N, \mathbf{m}_N) \right) \mathbf{L}_{\text{psh}}(j_N, \mathbf{m}_N) \end{array} \right] \in \mathbb{R}^{N \times 6} \quad (4.26)$$

Dans cette section, nous avons exposé la méthodologie pour calculer la matrice d'interaction lorsque les informations visuelles sont constituées des coefficients de la transformée en ondelettes ou en shearlets de l'image. Contrairement à l'approche de la section 3.3, celle-ci permet également d'utiliser des transformées sous-échantillonnées réduisant le temps de calcul par rapport à celles non sous-échantillonnées.

## 4.2 Simulation et validation expérimentale sur caméra conventionnelle

Afin de comparer les approches d'asservissement visuel utilisant les ondelettes et les shearlets, nous avons considéré différents scénarios. Il est à noter que pour chaque méthode, nous avons implémenté une version sous-échantillonnée et sa variante non sous-échantillonnée, ce qui nous amène à comparer quatre méthodes au total. Ces notions de transformées sous-échantillonnées ou non ont été présentées dans le chapitre 2. Dans un premier temps, des simulations ont permis de comparer nos quatre méthodes

à celle de la photométrie provenant de la librairie ViSP, ceci dans des conditions favorables et dégradées. Dans un deuxième temps, les deux méthodes sous-échantillonnées ont été validées expérimentalement (les méthodes non sous-échantillonnées présentant des résultats moins pertinents).

#### 4.2.1 Détails d'implémentation

Nous avons développé un simulateur en C++ permettant de tester chaque méthode dans différentes conditions de travail. Le calcul des coefficients d'ondelettes (ondelettes sous-échantillonnées : *e-ondelette* et non sous-échantillonnées : *ne-ondelette*) et ceux des shearlets (shearlets sous-échantillonnées : *e-shearlet* et non sous-échantillonnées : *ne-shearlet*) est effectué sous l'environnement MATLAB en utilisant MATLAB WAVELET TOOLBOX<sup>1</sup> et SHEARLAB 3D<sup>2</sup>. Des fonctions permettant de calculer les ondelettes ou les shearlets ont été appelées dans l'algorithme principal codé en C++. Le résultat des transformées est retourné par MATLAB vers cet algorithme. Ce choix de ne pas se focaliser sur la réécriture des toolboxes des ondelettes et des shearlets permet de consacrer davantage de temps aux aspects théoriques de ce travail de thèse qu'aux détails d'implémentation ou d'optimisation de code.

méthodes	<i>e-ondelette</i>	<i>e-shearlet</i>	<i>ne-ondelette</i>	<i>ne-shearlet</i>
matrice d'interaction	$\mathbf{L}_{s_w}$	$\mathbf{L}_{s_{sh}}$	$\mathbf{L}_{s_w}$	$\mathbf{L}_{s_{sh}}$
$t_{\mathbf{L}_s}$	208.18 ms	498.57 ms	392.93 ms	220.65 ms
$t_{loop}$	100 ms	140 ms	700 ms	800 ms
$n_s$	20 059	24 576	262 144	327 680
$n_{level}$	2	2	1	1

TABLE 4.1 – Tableau récapitulatif des différentes méthodes.  $t_{\mathbf{L}_s}$  : temps de calcul de la matrice d'interaction,  $t_{loop}$  durée d'une itération,  $n_s$  : taille du vecteur d'informations visuelles,  $n_{level}$  : nombre de niveaux de décompositions utilisé dans la transformée en ondelettes ou en shearlets.

Le TABLEAU 4.1 montre le temps de boucle requis pour les différentes étapes de la commande (c'est-à-dire le temps de calcul du vecteur des informations visuelles et celui de la matrice d'interaction). La matrice d'interaction a été calculée une seule fois à la position désirée et maintenue constante durant toute la tâche d'asservissement visuel. Le nombre de niveaux de décomposition  $n_{level}$  des différentes méthodes varie selon le cas d'une approche sous-échantillonnée ou non sous-échantillonnée. En effet, plus ce nombre est grand, plus le vecteur  $\mathbf{s}_i$  ( $i = [w, sh]$ ) est de taille importante. Ainsi, dans le cas des méthodes non sous-échantillonnées, un seul niveau de décomposition est considéré (correspondant à plus de 250000 coefficients) engendrant déjà un temps de calcul non négligeable. De plus, pour chaque cas, les coefficients très haute-fréquences qui correspondent à du bruit dans l'image n'ont pas été pris en compte.

1. <https://fr.mathworks.com/products/wavelet.html>

2. <http://www.shearlab.org/software>

Afin d'obtenir le meilleur comportement possible avec toutes les lois de commande, nous avons choisi d'utiliser des gains adaptatifs pour  $\lambda$  et  $\mu$  :

$$\lambda = 10^{\log_{10}(\|\mathbf{I}_{diff}\|) - 7,5}$$

$$\mu = 10^{2 * \log_{10}(\|\mathbf{I}_{diff}\|) - 17}$$

où le gain  $\lambda$  et le facteur de poids  $\mu$  décroissent lorsque l'image courante<sup>3</sup>  $\mathbf{I}$  se rapproche de l'image désirée  $\mathbf{I}^*$ . Cela permet d'avoir une vitesse  $\mathbf{v}$  élevée au début de la tâche d'asservissement tout en évitant les oscillations en fin de tâche. Les valeurs 7,5 et 17 ont été choisies empiriquement.

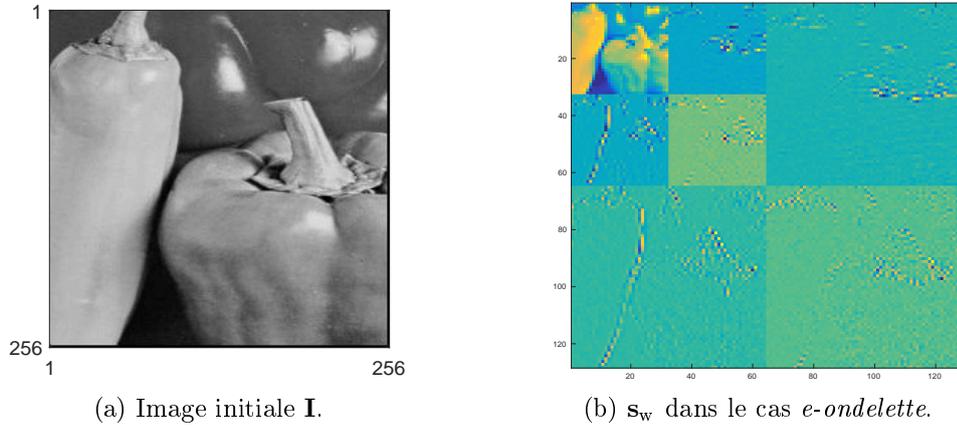
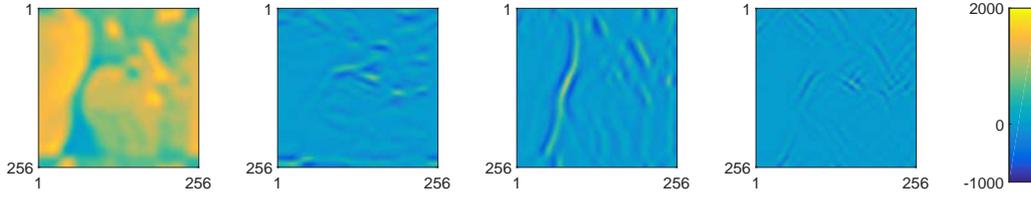


FIGURE 4.3 – Image initiale et sa transformée en ondelettes sous-échantillonnées.

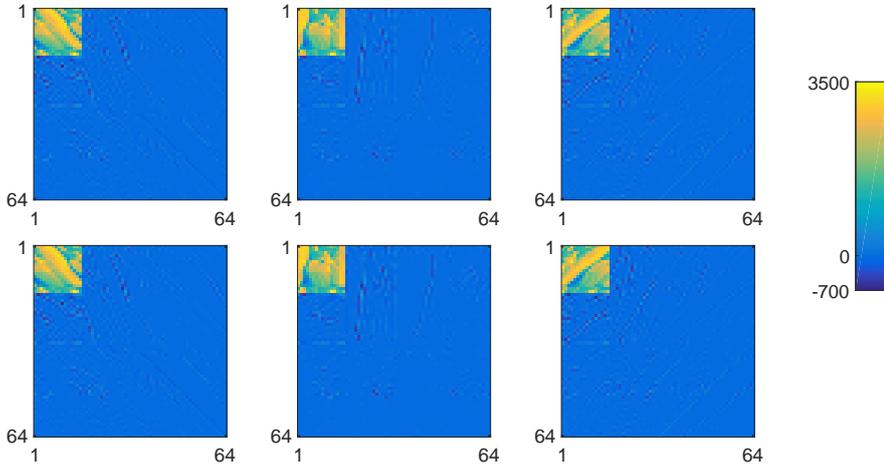
Les FIGURES 4.3, 4.4, 4.5 et 4.6 présentent les décompositions en ondelettes et en shearlets sous-échantillonnées ou non, utilisées dans la loi de commande, obtenues à partir de l'image 4.3(a) :

- Le vecteur d'informations visuelles  $\mathbf{s}_w$  est obtenu en appliquant une transformée en ondelettes de niveau 3 (voir FIGURE 4.3(b)). Le coin en haut à gauche contient l'approximation basses fréquences de taille  $32 \times 32$  pixels de l'image originale. Les autres coefficients correspondent aux détails horizontaux, verticaux et diagonaux de l'image aux niveaux de décomposition  $j = 2$  et  $j = 3$ . Le premier niveau (contenant essentiellement du bruit) n'est pas utilisé dans la commande et n'est donc pas représenté sur la figure.
- La concaténation des quatre matrices représentées sur la FIGURE 4.4 nous donne le vecteur des informations visuelles  $\mathbf{s}_w$  de la méthode *ne-ondelette* de niveau  $j = 4$ .

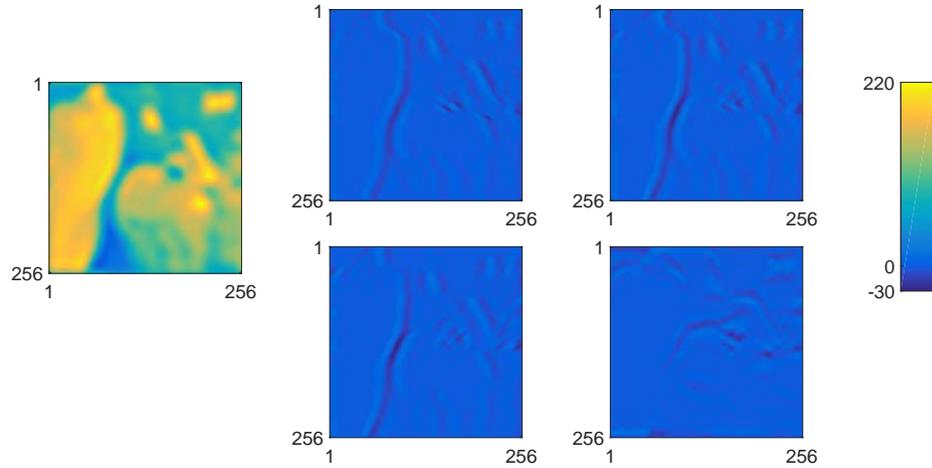
3. Dans la section 4.1, l'image considérée comme une fonction continue de l'espace  $L^2(\mathbb{R}^2)$  était notée  $I_t$ , alors que dans la suite de ce chapitre, nous noterons l'image courante  $\mathbf{I}$ , car elle est considérée cette fois-ci comme une matrice.

FIGURE 4.4 –  $\mathbf{s}_w$  dans le cas *ne-ondelette*.

La première matrice correspond à une approximation de l'image initiale tandis que les trois autres correspondent aux détails horizontaux, verticaux et diagonaux de cette image. Chaque matrice possède la même taille que l'image initiale.

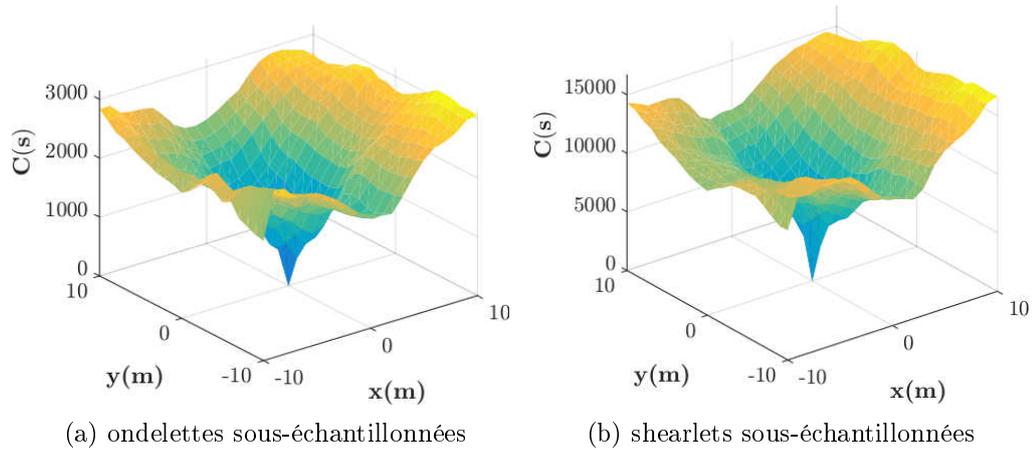
FIGURE 4.5 –  $\mathbf{s}_{sh}$  dans le cas *e-shearlet*.

- La concaténation des six matrices représentées sur la FIGURE 4.5 forme le vecteur d'informations visuelles  $\mathbf{s}_{sh}$ . Ces matrices sont obtenues en appliquant une transformée en shearlets sous-échantillonnées à l'image initiale, avec quatre niveaux de décompositions. Chaque coin supérieur gauche correspond à une approximation de taille  $16 \times 16$  pixels de l'image initiale mettant l'accent sur les fréquences de six directions distinctes. Les autres coefficients représentent les détails aux niveaux  $j = 3$  et  $j = 4$  de la transformée. Les coefficients de niveaux  $j = 1$  et  $j = 2$  n'ont pas été représentés puisque nous les considérons comme étant associés à du bruit.
- La concaténation des cinq matrices représentées sur la FIGURE 4.6 forme le vecteur des informations visuelles  $\mathbf{s}_{sh}$  obtenue avec la méthode des shearlets non sous-échantillonnées avec 4 niveaux de décomposition. La première matrice correspond à l'approximation basse fréquence de l'image initiale tandis que les quatre autres correspondent aux détails dans différentes directions. Comme pour les ondelettes non sous-échantillonnées, seuls les coefficients de niveau  $j = 4$  ont été conservés.

FIGURE 4.6 –  $s_{sh}$  dans le cas *ne-shearlet*.

#### 4.2.2 Domaine de convergence

Afin de juger de l'efficacité des approches fondées sur les ondelettes et les shearlets précédemment présentées, nous avons utilisé le simulateur développé afin de représenter les cônes de convergence de chacune des approches. Ils sont calculés en faisant varier les degrés de liberté du robot virtuel de la même manière que dans le chapitre 3.

FIGURE 4.7 – Cône de convergence le long des axes  $x$  et  $y$  des approches sous-échantillonnées.

Nous pouvons constater sur les FIGURES 4.7 et 4.8 qu'un minimum global est bien défini pour chaque méthode. Ceci montre qu'elles permettent toutes d'obtenir une convergence vers la position désirée sans "grande difficulté".

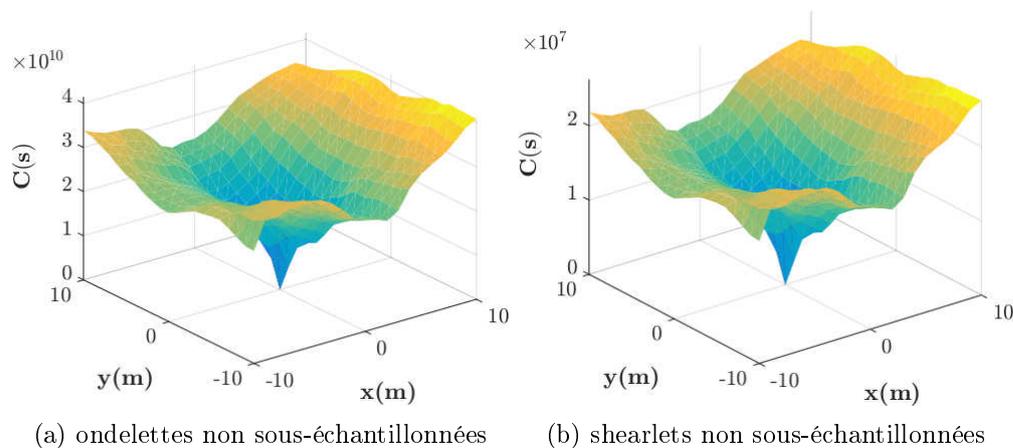


FIGURE 4.8 – Cône de convergence le long des axes  $x$  et  $y$  des approches non sous-échantillonnées.

### 4.2.3 Simulation

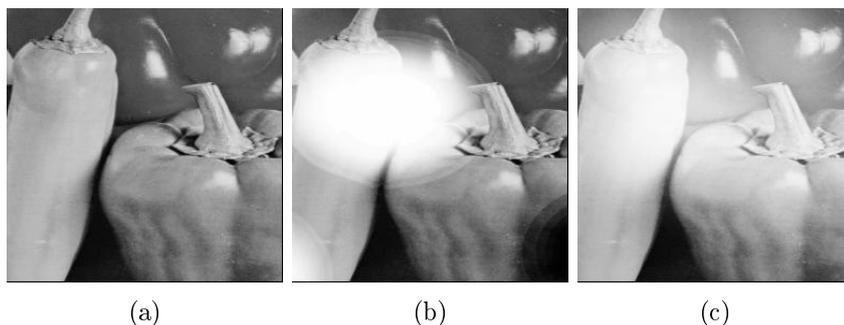


FIGURE 4.9 – Image désirée (a) en conditions nominales, (b) avec une occultation partielle (c) avec une variation d’illumination.

Afin d’analyser les performances des lois de commande présentées dans la section précédente, nous avons réalisé plusieurs simulations sous différentes conditions de travail. Chaque loi de commande a été testée dans des conditions nominales (voir FIGURE 4.9(a)) ou défavorables : occultations partielles (voir FIGURE 4.9(b)) et variations d’illumination (voir FIGURE 4.9(c)). L’erreur initiale de positionnement entre l’image initiale et l’image désirée est représentée sur la FIGURE 4.10.

Le TABLEAU 4.2 présente les trois positions initiales considérées, notées *test 1*, *test 2* et *test 3*. Chacune de ces positions sera testée dans chaque scénario mentionné auparavant.

Nous utilisons la différence d’image, notée  $\mathbf{I}_{diff}$  entre l’image courante et l’image désirée pour une meilleure visualisation de la qualité de convergence ( $\mathbf{I}_{diff}$  est complètement grise lorsque l’image finale est similaire à l’image désirée) des différentes lois de

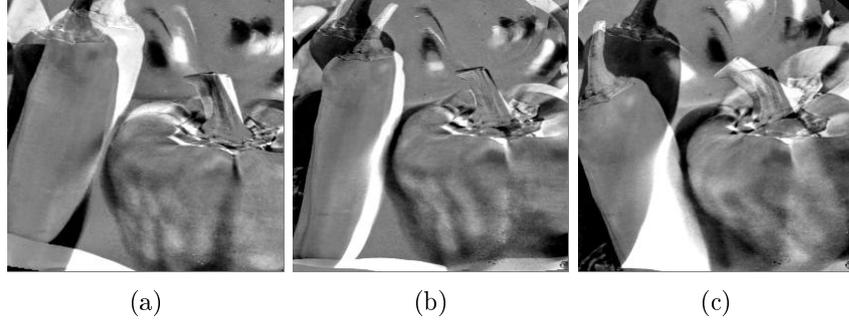


FIGURE 4.10 – Différence  $\mathbf{I}_{diff}$  à  $t = 0$  : (a) pour le *test 1*, (b) pour le *test 2*, (c) pour le *test 3*.

test #	$\Delta T_x$	$\Delta T_y$	$\Delta T_z$	$\Delta R_x$	$\Delta R_y$	$\Delta R_z$	figure
<i>test 1</i>	10	-10	100	5	-5	10	Fig. 4.10(a)
<i>test 2</i>	5	-5	100	-20	10	-5	Fig. 4.10(b)
<i>test 3</i>	-2	2	-50	5	10	-30	Fig. 4.10(c)

TABLE 4.2 – Erreur de positionnement initiale  $\mathbf{e}_0$  entre la pose de la caméra à l'état initial et désiré ( $\Delta T$  en *mm* et  $\Delta R$  en degrés).

commandes (voir l'équation (3.19)). La taille d'un côté d'un pixel de l'image planaire simulée est fixée à  $\eta = 1,2\text{mm}$ .

**Conditions nominales [*scenario 1*] :** Considérons le premier scénario visant à étudier le comportement des quatre lois de commande dans des conditions nominales d'utilisation. En pratique, nous réalisons trois tests dans lesquels la caméra virtuelle observe une scène planaire et se trouve à une position donnée où l'image désirée est enregistrée. La caméra rejoint alors une position de départ, entraînant une erreur de positionnement initiale. Cette erreur  $\mathbf{e} = \mathbf{e}_0$  ( $\mathbf{e} \in SE(3)$ ) pour chaque test est indiquée dans le TABLEAU 4.2. L'erreur de positionnement finale  $\mathbf{e}_f$  est mesurée à la convergence de la commande.

Le TABLEAU 4.3 rapporte les résultats obtenus par les quatre méthodes proposées ainsi que ceux dans le cas de la méthode photométrique. A la convergence, l'erreur finale en translation  $\Delta T_i$  ( $i = [x, y, z]$ ) a une valeur d'environ  $10^{-4}\text{mm}$  et celle en rotation  $\Delta R_i$  ( $i = [x, y, z]$ ) d'environ  $10^{-4}$  degré pour chaque méthode. Il en résulte que tous les contrôleurs présentent les mêmes performances en termes de précision en conditions nominales.

**Occultations partielles [*scenario 2*] :** Le *scenario 2* traite de la validation des quatre méthodes et de leur comparaison avec l'approche photométrique dans le cas d'occultations partielles. Il permet d'étudier le comportement de chaque méthode lors-

	méthode / test	$\Delta T_x$	$\Delta T_y$	$\Delta T_z$	$\Delta R_x$	$\Delta R_y$	$\Delta R_z$
$\mathbf{e}_0$	<i>test 1</i>	<b>10</b>	<b>-10</b>	<b>100</b>	<b>5</b>	<b>-5</b>	<b>10</b>
$\mathbf{e}_f$	photométrie	$10^{-4}$	$10^{-4}$	$10^{-4}$	$10^{-4}$	$-10^{-4}$	$10^{-4}$
$\mathbf{e}_f$	<i>e-ondelette</i>	$10^{-4}$	$10^{-4}$	$10^{-4}$	$10^{-4}$	$-10^{-4}$	$10^{-4}$
$\mathbf{e}_f$	<i>e-shearlet</i>	$10^{-4}$	$10^{-4}$	$10^{-4}$	$10^{-4}$	$-10^{-4}$	$10^{-4}$
$\mathbf{e}_f$	<i>ne-ondelette</i>	$10^{-4}$	$10^{-4}$	$10^{-3}$	$10^{-4}$	$-10^{-3}$	$10^{-4}$
$\mathbf{e}_f$	<i>ne-shearlet</i>	$10^{-4}$	$10^{-4}$	$10^{-3}$	$10^{-3}$	$-10^{-4}$	$10^{-4}$
$\mathbf{e}_0$	<i>test 2</i>	<b>5</b>	<b>-5</b>	<b>100</b>	<b>-20</b>	<b>10</b>	<b>-5</b>
$\mathbf{e}_f$	photométrie	$10^{-4}$	$10^{-4}$	$10^{-4}$	$10^{-4}$	$-10^{-4}$	$10^{-4}$
$\mathbf{e}_f$	<i>e-ondelette</i>	$10^{-4}$	$10^{-4}$	$10^{-4}$	$10^{-4}$	$10^{-4}$	$10^{-4}$
$\mathbf{e}_f$	<i>e-shearlet</i>	$10^{-4}$	$10^{-4}$	$10^{-4}$	$10^{-4}$	$10^{-4}$	$10^{-4}$
$\mathbf{e}_f$	<i>ne-ondelette</i>	$10^{-4}$	$10^{-4}$	$10^{-4}$	$10^{-3}$	$10^{-4}$	$-10^{-4}$
$\mathbf{e}_f$	<i>ne-shearlet</i>	$10^{-4}$	$10^{-4}$	$10^{-4}$	$10^{-4}$	$10^{-4}$	$10^{-4}$
$\mathbf{e}_0$	<i>test 3</i>	<b>-2</b>	<b>2</b>	<b>-50</b>	<b>5</b>	<b>10</b>	<b>-30</b>
$\mathbf{e}_f$	photométrie	$10^{-4}$	$10^{-4}$	$10^{-4}$	$10^{-4}$	$10^{-4}$	$10^{-4}$
$\mathbf{e}_f$	<i>e-ondelette</i>	$10^{-4}$	$10^{-4}$	$10^{-4}$	$10^{-4}$	$10^{-4}$	$10^{-4}$
$\mathbf{e}_f$	<i>e-shearlet</i>	$10^{-4}$	$10^{-4}$	$10^{-4}$	$10^{-4}$	$10^{-4}$	$10^{-4}$
$\mathbf{e}_f$	<i>ne-ondelette</i>	$10^{-4}$	$10^{-4}$	$10^{-4}$	$10^{-4}$	$10^{-4}$	$10^{-4}$
$\mathbf{e}_f$	<i>ne-shearlet</i>	$10^{-4}$	$10^{-4}$	$10^{-4}$	$10^{-4}$	$10^{-4}$	$10^{-4}$

TABLE 4.3 – [*scenario 1*] Comparaison des différentes lois de commande dans des conditions nominales.  $\Delta T_i$  (mm),  $\Delta R_i$  (deg),  $\mathbf{e}_0$  et  $\mathbf{e}_f$  représentent respectivement les erreurs de positionnement initiales et finales.

qu'une partie de l'image désirée est dissimulée par la présence d'une zone de saturation (ou tout autre type d'occultations) comme le montre la FIGURE 4.9(b). Les mêmes erreurs initiales de positionnement que celles utilisées pour le *scenario 1* sont considérées ici. Comme présenté sur le TABLEAU 4.4, lors du *test 1*, les quatre contrôleurs ainsi que la méthode photométrique convergent vers la position désirée malgré une erreur statique pour les rotations hors plan ( $R_x$  et  $R_y$ ) ainsi que pour la translation  $T_z$ . Cependant, l'approche *e-ondelette* réagit mieux, notamment pour la translation le long de l'axe  $z$ . De manière générale, les méthodes utilisant les ondelettes et les shearlets présentent de meilleures performances que la photométrie dans ce scénario. Enfin, pour le *test 3*, tous les contrôleurs échouent (✗) en raison de l'erreur de positionnement initiale  $\mathbf{e}_0$  trop grande combinée à l'occultation partielle de l'image désirée.

	test / méthode	$\Delta T_x$	$\Delta T_y$	$\Delta T_z$	$\Delta R_x$	$\Delta R_y$	$\Delta R_z$
$\mathbf{e}_0$	<i>test 1</i>	<b>10</b>	<b>-10</b>	<b>100</b>	<b>5</b>	<b>-5</b>	<b>10</b>
$\mathbf{e}_f$	photométrie	0.69	0.01	-4.61	4.30	-5.64	0.05
$\mathbf{e}_f$	<i>e-ondelette</i>	0.56	-0.16	0.52	4.43	-5.60	0.17
$\mathbf{e}_f$	<i>e-shearlet</i>	0.56	-0.16	-1.94	4.31	-5.51	0.08
$\mathbf{e}_f$	<i>ne-ondelette</i>	0.78	-0.14	-2.79	4.30	-5.60	0.15
$\mathbf{e}_f$	<i>ne-shearlet</i>	0.74	-0.14	-2.74	4.33	-5.54	0.13
$\mathbf{e}_0$	<i>test 2</i>	<b>5</b>	<b>-5</b>	<b>100</b>	<b>-20</b>	<b>10</b>	<b>-5</b>
$\mathbf{e}_f$	photométrie	1.18	-0.24	-42.84	-20.29	9.14	-1.50
$\mathbf{e}_f$	<i>e-ondelette</i>	0.55	2.06	-11.06	-20.30	9.17	-1.27
$\mathbf{e}_f$	<i>e-shearlet</i>	0.52	1.18	-28.98	-20.40	9.18	-1.50
$\mathbf{e}_f$	<i>ne-ondelette</i>	1.10	0.10	-31.24	-20.37	9.18	-1.43
$\mathbf{e}_f$	<i>ne-shearlet</i>	1.04	0.16	-31.04	-20.36	9.22	1.47
$\mathbf{e}_0$	<i>test 3</i>	<b>-2</b>	<b>2</b>	<b>-50</b>	<b>5</b>	<b>10</b>	<b>-30</b>
$\mathbf{e}_f$	photométrie	<b>×</b>	<b>×</b>	<b>×</b>	<b>×</b>	<b>×</b>	<b>×</b>
$\mathbf{e}_f$	<i>e-ondelette</i>	<b>×</b>	<b>×</b>	<b>×</b>	<b>×</b>	<b>×</b>	<b>×</b>
$\mathbf{e}_f$	<i>e-shearlet</i>	<b>×</b>	<b>×</b>	<b>×</b>	<b>×</b>	<b>×</b>	<b>×</b>
$\mathbf{e}_f$	<i>ne-ondelette</i>	<b>×</b>	<b>×</b>	<b>×</b>	<b>×</b>	<b>×</b>	<b>×</b>
$\mathbf{e}_f$	<i>ne-shearlet</i>	<b>×</b>	<b>×</b>	<b>×</b>	<b>×</b>	<b>×</b>	<b>×</b>

TABLE 4.4 – [*scenario 2*] Comparaison des différents contrôleurs dans le cas d’occultations partielles.  $\Delta T_i$  (mm),  $\Delta R_i$  (deg),  $\mathbf{e}_0$  et  $\mathbf{e}_f$  représentent respectivement les erreurs de positionnement initiales et finales. **×** = échec.

**Variations d’illumination [*scenario 3*] :** Le *scenario 3* traite de la validation des quatre méthodes et de leur comparaison avec la photométrie dans le cas de variations d’illumination appliquées sur l’image désirée. Ainsi, l’image désirée  $\mathbf{I}^*$  (voir FIGURE 4.9(c)) est acquise avec une saturation due à une source de lumière tandis que les images courantes sont capturées dans des conditions normales pendant la tâche de positionnement. Le TABLEAU 4.5 montre que tous les contrôleurs convergent pour le *test 1*. Cependant, en raison des variations d’illumination, une erreur plus importante sur les rotations hors plan ( $R_x$  and  $R_y$ ) peut être observée (voir FIGURE 4.11). Lors du *test 2* les méthodes utilisant les shearlets et les ondelettes convergent près de la position désirée contrairement à l’approche photométrique. Les erreurs statiques observées au niveau de la translation le long de l’axe  $z$  et des rotations  $R_x$  et  $R_y$  peuvent être expliquées par le fait que même si les approches présentées sont robustes aux perturbations extérieures, la loi de commande utilise des informations globales qui sont fortement impactées par ces perturbations. Ainsi, l’information commune entre l’image initiale et l’image désirée est faible, entraînant des erreurs statiques non-négligeables. Enfin, lors du *test 3*, seules les approches de type sous-échantillonnées ne divergent pas. Cela est dû au fait que le vecteur des informations visuelles contient moins de coefficients intégrant des informations peu pertinentes que dans les approches non sous-échantillonnées.

	test / méthode	$\Delta T_x$	$\Delta T_y$	$\Delta T_z$	$\Delta R_x$	$\Delta R_y$	$\Delta R_z$
$\mathbf{e}_0$	<i>test 1</i>	<b>10</b>	<b>-10</b>	<b>100</b>	<b>5</b>	<b>-5</b>	<b>10</b>
$\mathbf{e}_f$	photométrie	0.84	0.57	-31.49	4.27	-5.61	-0.26
$\mathbf{e}_f$	<i>e-ondelette</i>	0.10	0.69	-23.48	4.43	-5.65	-0.10
$\mathbf{e}_f$	<i>e-shearlet</i>	0.52	0.49	-13.57	4.26	-5.60	-0.07
$\mathbf{e}_f$	<i>ne-ondelette</i>	0.83	0.60	-23.66	4.26	-5.65	-0.22
$\mathbf{e}_f$	<i>ne-shearlet</i>	0.78	0.59	-22.76	4.27	-5.62	-0.21
$\mathbf{e}_0$	<i>test 2</i>	<b>5</b>	<b>-5</b>	<b>100</b>	<b>-20</b>	<b>10</b>	<b>-5</b>
$\mathbf{e}_f$	photométrie	<b>X</b>	<b>X</b>	<b>X</b>	<b>X</b>	<b>X</b>	<b>X</b>
$\mathbf{e}_f$	<i>e-ondelette</i>	0.34	0.66	-59.68	-20.27	9.25	-2.40
$\mathbf{e}_f$	<i>e-shearlet</i>	0.80	0.89	-52.28	-20.31	9.28	-2.12
$\mathbf{e}_f$	<i>ne-ondelette</i>	1.34	0.456	-60.99	-20.32	9.33	-2.20
$\mathbf{e}_f$	<i>ne-shearlet</i>	1.20	0.52	-59.19	-20.28	9.35	-2.22
$\mathbf{e}_0$	<i>test 3</i>	<b>-2</b>	<b>2</b>	<b>-50</b>	<b>5</b>	<b>10</b>	<b>-30</b>
$\mathbf{e}_f$	photométrie	<b>X</b>	<b>X</b>	<b>X</b>	<b>X</b>	<b>X</b>	<b>X</b>
$\mathbf{e}_f$	<i>e-ondelette</i>	-5.26	-20.07	-176.26	2.91	10.28	-5.21
$\mathbf{e}_f$	<i>e-shearlet</i>	26.50	-22.76	-317.09	2.26	10.73	-0.04
$\mathbf{e}_f$	<i>ne-ondelette</i>	<b>X</b>	<b>X</b>	<b>X</b>	<b>X</b>	<b>X</b>	<b>X</b>
$\mathbf{e}_f$	<i>ne-shearlet</i>	<b>X</b>	<b>X</b>	<b>X</b>	<b>X</b>	<b>X</b>	<b>X</b>

TABLE 4.5 – [*scenario 3*] Comparaison des différents contrôleurs dans le cas de variations d’illumination.  $\Delta T_i$  (mm),  $\Delta R_i$  (deg),  $\mathbf{e}_0$  et  $\mathbf{e}_f$  représentent respectivement les erreurs de positionnement initiales et finales. **X** = échec.

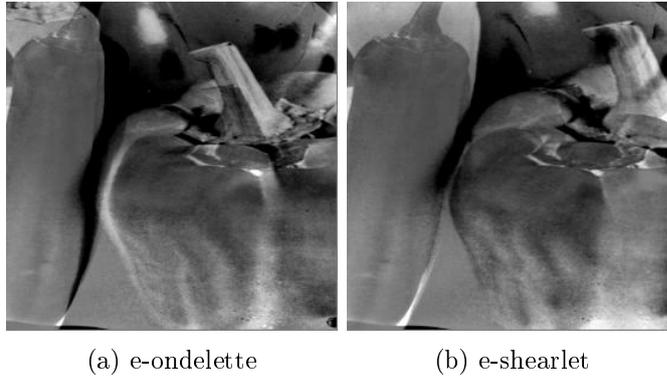


FIGURE 4.11 – [*scenario 3, test 3*] Images de différences  $\mathbf{I}_{diff}$  à la convergence lors de variations d’illumination. Même s’il y a une erreur statique, la méthode ne diverge pas.

L’objectif de ces simulations était l’étude de la capacité des quatre contrôleurs développés à réussir des tâches d’asservissement visuel dans différentes conditions d’utilisation ainsi que leur comparaison avec la méthode photométrique. Nous avons pu constater que les approches proposées permettent d’atteindre les objectifs en conditions nominales comme le laissait pressentir l’étude des cônes de convergence. Lorsqu’il y a

la présence de perturbations externes, les lois de commande convergent toujours malgré parfois une importante erreur statique sur les rotations hors plan ( $R_x$  et  $R_y$ ) ou sur la translation  $T_z$ . De plus, la méthode photométrique échoue sur certains cas comme dans le scénario comportant des variations d'illumination, à cause notamment de son modèle reposant sur l'hypothèse forte de continuité du flux optique. Même si les approches utilisant les ondelettes et les shearlets utilisent la même hypothèse, elles y sont plus robuste car seule la transformées de niveau  $j = 0$  s'appuie fortement sur cette hypothèse. Les autres niveaux comporte des coefficients élevés pour des contours de l'image selon certaines directions. Ces informations changent peu selon le type d'illumination. Nous pouvons également remarquer que les versions sous-échantillonnées fournissent de meilleurs résultats lors de conditions défavorables que celles non sous-échantillonnées, en plus d'être moins coûteuses en temps de calcul ( $t_{loop} = 120$  ms pour les méthodes sous-échantillonnées contre  $t_{loop} = 700$  ms pour les méthodes non sous-échantillonnées). Cela est dû au fait qu'elles comportent davantage de niveaux de détails contenus dans un nombre restreint de coefficients tandis que les approches non sous-échantillonnées contiennent beaucoup plus de coefficients, mais dont la plupart sont très proches de zéro, et contiennent, de ce fait, peu d'information utile. De plus, les détails de certains niveaux, assimilable à du bruit, ne sont pas conservés dans le vecteur  $\mathbf{s}_i$  ( $i = [w, sh]$ ). Dans la suite, nous retiendrons uniquement les méthodes *e-ondelette* et *e-shearlet*.

**Robustesse aux erreurs d'étalonnage [scenario 4] :** La simulation permet aussi de tester facilement la robustesse des contrôleurs proposés aux erreurs d'étalonnage de la caméra. Nous avons ici étudié la robustesse aux erreurs sur les paramètres intrinsèques de la caméra. Ainsi, les deux méthodes sous-échantillonnées sont comparées sur ce point. Deux paramètres ont été volontairement entachés d'erreurs : les coordonnées du point principal et la distance focale (voir TABLEAU 4.6). En effet, les vraies coordonnées du point principal de l'image sont  $[u_0, v_0] = (160, 120)$  pixels tandis que la vraie distance focale est  $p_x = p_y = 870$  pixels. Lorsque l'erreur est en faveur d'une surestimation de la distance focale, les deux méthodes convergent mais nécessitent davantage d'itérations. Cependant, à l'inverse, lorsque l'erreur est en faveur d'une sous-estimation de la distance focale, la méthode d'ondelettes sous-échantillonnées atteint un minimum local tandis que la méthode des shearlets sous-échantillonnées converge vers le minimum global.

TABLE 4.6 – Étude de la robustesse des contrôleurs aux erreurs sur la distance focale avec l'erreur de positionnement initiale correspondant au *test 2*.

Longueur focale erronée	Erreur	<i>e-ondelette</i>	<i>e-shearlet</i>
en pixels	(en %)	en itérations	en itérations
435	-50	✗	✓ 190 itérations
609	-30	✗	✓ 440 itérations
696	-20	✓ 700	✓ 580 itérations
870	0	✓ 950	✓ 1030 itérations
1305	+50	✓ 2970	✓ 3720 itérations

Les deux méthodes convergent lorsque la longueur focale erronée est 20% plus faible que sa valeur exacte ( $p_x = p_y = 696$  pixels). Cette erreur a donc été conservée pour le test suivant où nous avons également volontairement modifié les coordonnées du point principal. Les résultats sont présentés dans le TABLEAU 4.7 et ne montrent aucune différence entre les deux méthodes, qui peuvent être considérées comme robustes à ce type d'erreur.

TABLE 4.7 – Étude de la robustesse des contrôleurs aux erreurs d'étalonnage sur le point principal de l'image avec l'erreur de positionnement initiale correspondant au *test 1*.

Coordonnées erronées	Distance	<i>e-ondelette</i>	<i>e-shearlet</i>
(160,120)	0	✓ (160 itérations)	✓ (70 itérations)
(320,240)	200	✓ (360 itérations)	✓ (180 itérations)
(160,-100)	220	✓ (540 itérations)	✓ (170 itérations)
(-200,120)	360	✓ (310 itérations)	✓ (330 itérations)
(500,-200)	467	✗	✗

Après avoir démontré en simulation l'intérêt des approches sous-échantillonnées par rapport à celles non sous-échantillonnées, nous allons étudier le comportement de ces lois de commande dans des conditions expérimentales.

#### 4.2.4 Validation expérimentale

Afin de valider expérimentalement les méthodes d'asservissement visuel proposées dans ce chapitre, nous utilisons un système robotique cartésien à 6 ddl (voir FIGURE 4.12). Une caméra de type CCD est montée sur l'effecteur du robot dans une configuration *caméra embarquée*. Cette caméra permet d'obtenir des images de taille  $450 \times 450$  pixels à une cadence de 25 images par seconde. Ces images sont ensuite redimensionnées à  $256 \times 256$  pixels. Notre boucle de commande est gérée par une station de travail PC équipée d'un processeur 2,4 GHz sous la distribution Linux. Comme pour la simulation, les transformées en ondelettes et en shearlets sont réalisées sous MATLAB avec la MATLAB WAVELET TOOLBOX et SHEARLAB 1.1. Enfin, tous les calculs sont effectués en considérant une profondeur de la scène constante de  $Z = 0.8\text{m}$ .

Pour cette validation expérimentale, nous avons comparé les méthodes sous-échantillonnées *e-ondelette* et *e-shearlet*. Comme pour la partie simulation, nous avons réalisé des tests en conditions nominales et défavorables afin de déterminer et de comparer les performances de ces deux approches. Le TABLEAU 4.8 présente les deux erreurs de positionnement initiales considérées pour chacun de ces tests. La scène planaire est remplacée par un objet 3D dans les derniers tests de cette section.

**Conditions nominales et scène planaire [*scenario 1*] :** La première validation expérimentale consiste à tester de deux contrôleurs *e-ondelette* et *e-shearlet* dans des conditions nominales d'utilisation avec une profondeur de scène fixe. La FIGURE 4.13



FIGURE 4.12 – Photographie du système expérimental avec une scène planeaire.

Test	$\Delta T_x$ (mm)	$\Delta T_y$ (mm)	$\Delta T_z$ (mm)	$\Delta R_x$ (°)	$\Delta R_y$ (°)	$\Delta R_z$ (°)
<i>test 1</i>	5	50	100	5	-5	-4
<i>test 2</i>	-20	-30	-20	10	2	2

TABLE 4.8 – Erreur initiale  $\mathbf{e}_0$  entre l'image initiale  $\mathbf{I}$  et l'image désirée  $\mathbf{I}^*$  dans le cas d'une scène planeaire.

nous montre l'image obtenue par la caméra à différents moments de la tâche de positionnement : la FIGURE 4.13(a) montre l'image désirée  $\mathbf{I}^*$ , la FIGURE 4.13(b) l'image de différence  $\mathbf{I}_{diff}$  entre l'image désirée et l'image initiale, tandis que les FIGURES 4.13(c)-(d) illustrent respectivement  $\mathbf{I}_{diff}$  finale (à la convergence) pour les approches *e-ondelette* et *e-shearlet*. Les deux méthodes convergent de manière précise vers la position désirée.

L'erreur de positionnement  $\mathbf{e}$  durant la tâche, pour chaque ddl, est visible sur les FIGURES 4.14 et 4.15. Ces valeurs sont obtenues à partir des encodeurs très précis du robot. La FIGURE 4.14 représente l'évolution de l'erreur pour la méthode *e-ondelette* tandis que la FIGURE 4.15 présente celle de la méthode *e-shearlet*. Nous pouvons voir que tous les ddl du robot convergent vers leur position désirée. Néanmoins, la décroissance de l'erreur n'est pas exponentielle en raison de notre loi de commande qui utilise un algorithme de type *Levenberg-Marquardt* avec un gain adaptatif.

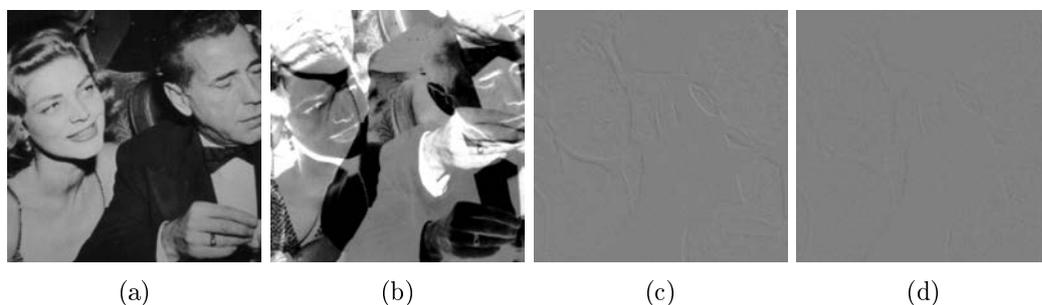


FIGURE 4.13 – [*scenario 1, test 1*] (a) image désirée  $\mathbf{I}^*$  (b) image de différence initiale  $\mathbf{I}_{diff}$ , (c) image de différence finale  $\mathbf{I}_{diff}$  dans le cas *e-ondelette*, et (d) dans le cas *e-shearlet*.

Les valeurs de l'erreur de positionnement finale pour chacune des deux méthodes sont :

- *e-ondelette*

- translation :  $moyenne(\Delta T_i) = 0,30$  mm
- rotation :  $moyenne(\Delta R_i) = 0,029$  deg

- *e-shearlet*

- translation :  $moyenne(\Delta T_i) = 0,08$  mm
- rotation :  $moyenne(\Delta R_i) = 0,014$  deg

Le contrôleur *e-ondelette* atteint la position désirée en 500 itérations tandis que celui *e-shearlet* converge en seulement 400 itérations. De plus, la méthode *e-shearlet* présente une meilleure précision en translation et en rotation.

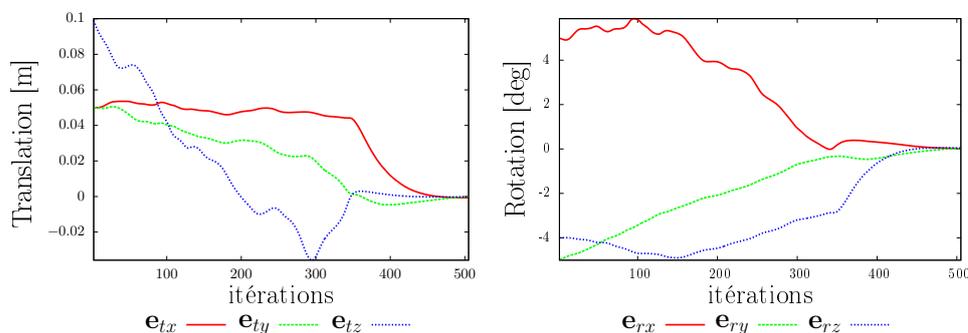
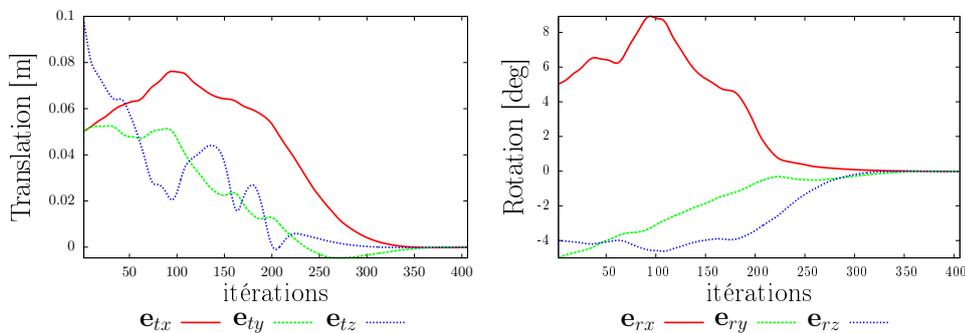
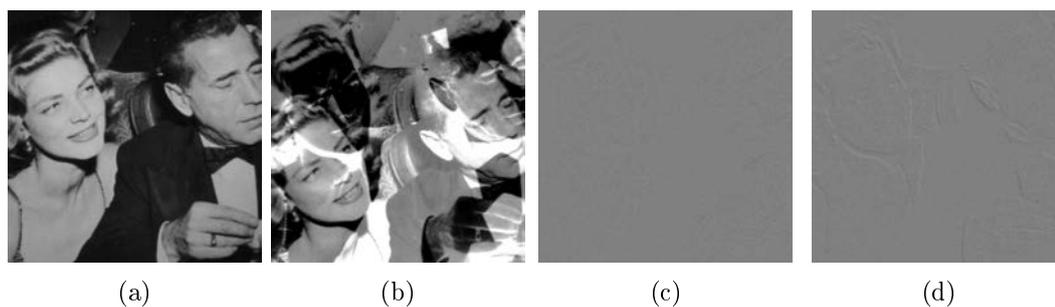
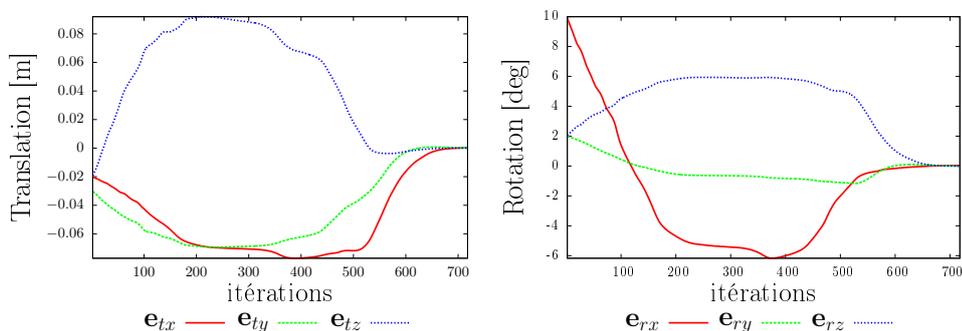


FIGURE 4.14 – [*scenario 1, test 1*] Erreur de positionnement  $\mathbf{e}$  avec la méthode *e-ondelette*.

FIGURE 4.15 – [scenario 1, test 1] Erreur de positionnement  $\mathbf{e}$  avec la méthode *e-shearlet*.

Les résultats pour le *test 2* sont présentés sur la FIGURE 4.16. Nous pouvons constater que les deux contrôleurs convergent avec précision comme le montrent les images de différence  $\mathbf{I}_{diff}$  finales complètement grise. (FIGURE 4.16(c) et (d)).

FIGURE 4.16 – [scenario 1, test 2] (a) image désirée  $\mathbf{I}^*$  (b) image de différence initiale  $\mathbf{I}_{diff}$ , (c)  $\mathbf{I}_{diff}$  finale dans le cas *e-ondelette*, et (d) dans le cas *e-shearlet*.FIGURE 4.17 – [scenario 1, test 2] Erreur de positionnement  $\mathbf{e}$  avec la méthode *e-ondelette*.

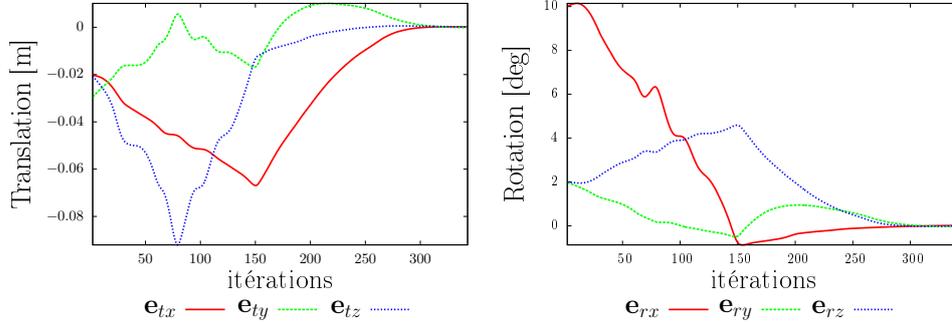


FIGURE 4.18 – [scenario 1, test 2] Erreur de positionnement  $\mathbf{e}$  avec la méthode *e-shearlet*.

La décroissance de l'erreur  $\mathbf{e}$  sur chaque ddl pour le *test 2* a été représentée sur la FIGURE 4.17 pour la méthode *e-ondelette* et sur la FIGURE 4.18 pour celle *e-shearlet*. Les résultats sont très comparables à ceux obtenus pour le *test 1* même si la commande a un comportement davantage non-linéaire en début de tâche. Cette fois-ci, la méthode *e-ondelette* met 700 itérations pour converger contre 300 pour la méthode *e-shearlet*. Finalement, les valeurs numériques d'erreurs de positionnement finales en translation et en rotation à la fin de la tâche sont :

- *e-ondelette*

- translation :  $moyenne(\Delta T_i) = 0,15$  mm
- rotation :  $moyenne(\Delta R_i) = 0,16$  deg

- *e-shearlet*

- translation :  $moyenne(\Delta T_i) = 0,13$  mm
- rotation :  $moyenne(\Delta R_i) = 0,02$  deg

En conditions nominales, la sensibilité de la méthode *e-shearlet* à des fréquences selon différentes directions, adaptées aux informations visuelles anisotropes, semble lui conférer un avantage en termes de rapidité de convergence ainsi qu'une meilleure précision que la méthode *e-ondelette*. Néanmoins, les courbes d'erreur de positionnement pour les deux tests montrent un comportement plus instable pour l'approche des shearlets que celle des ondelettes. Nous pouvons également noter qu'en fin de tâche, les courbes d'erreur de positionnement prennent une allure de décroissance exponentielle. Cela est dû au fait qu'à partir d'un moment, l'image courante devient si proche de l'image désirée que  $\|\mathbf{I}_{diff}\|$  est faible. Or cette valeur est utilisée pour adapter en ligne la valeur de  $\mu$  dans notre optimisation de type *Levenberg-Marquardt* (voir équation (4.27)). Lorsque  $\mu$  a une faible valeur, la loi de commande revient à utiliser une optimisation de type Gauss-Newton.

**Occultations partielles et scène planaire [scenario 2] :** Dans ce scénario, nous avons étudié le comportement des deux contrôleurs en présence d'occultations partielles sur l'image désirée. A cette fin, l'image désirée a été cachée par un objet 3D qui en plus de couvrir une partie de l'image, induit des effets d'ombres non négligeables. La FIGURE 4.19(a) montre l'image désirée  $\mathbf{I}^*$  sur laquelle l'objet 3D a été posé (puis retiré lors de la tâche de positionnement). L'image  $\mathbf{I}_{diff}$  initiale est montrée à la FIGURE 4.19(b). Enfin, les FIGURES 4.19(c)-(d) montrent que les deux approches atteignent la position désirée avec la même précision.

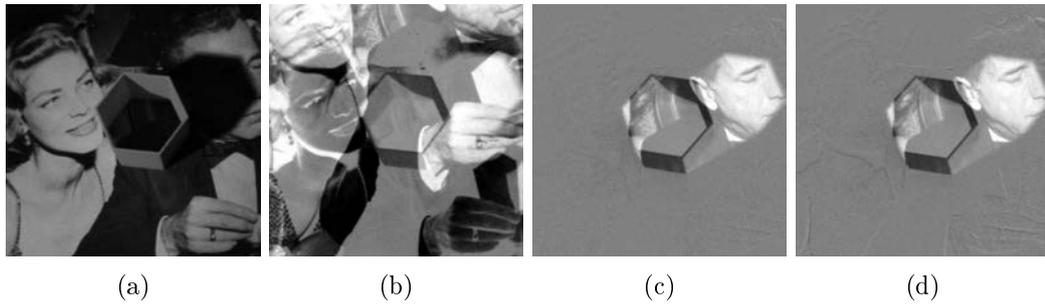


FIGURE 4.19 – [scenario 2, test 1] (a) image désirée  $\mathbf{I}^*$  (b) image de différence initiale  $\mathbf{I}_{diff}$ , (c)  $\mathbf{I}_{diff}$  finale dans le cas *e-ondelette*, et (d) dans le cas *e-shearlet*.

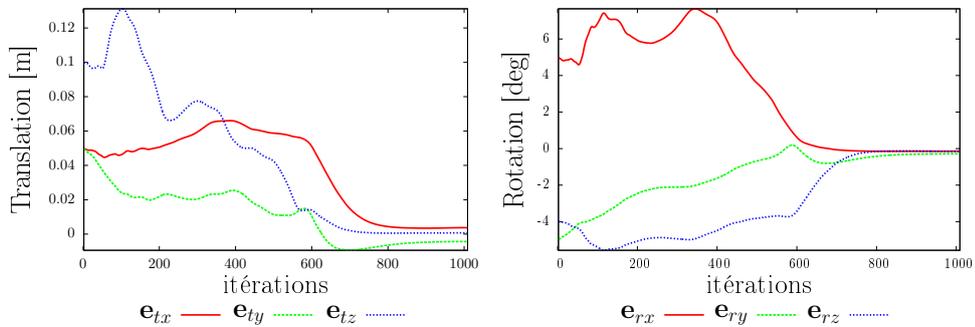


FIGURE 4.20 – [scenario 2, test 1] Erreur de positionnement  $\mathbf{e}$  avec la méthode *e-ondelette*.

La décroissance de l'erreur de positionnement pour chaque méthode est présentée à la FIGURE 4.20 pour la méthode *e-ondelette* et à la FIGURE 4.21 pour celle *e-shearlet*. Les deux méthodes convergent mais l'erreur de positionnement finale n'est pas aussi faible qu'en conditions nominales.

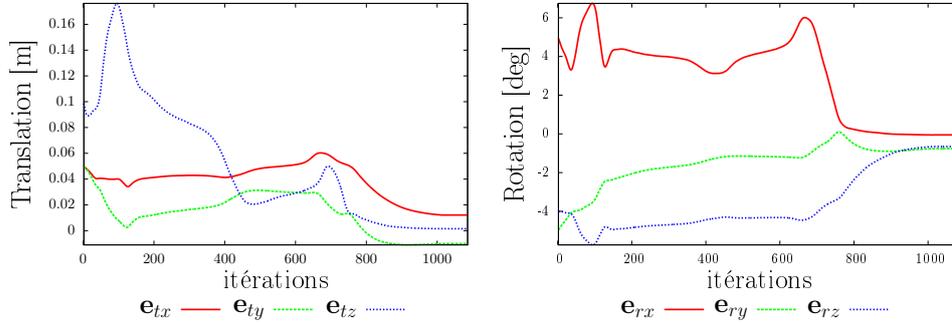


FIGURE 4.21 – [scenario 2, test 1] Erreur de positionnement  $\mathbf{e}$  avec la méthode *e-shearlet*.

Les valeurs numériques des erreurs de positionnement finales en translation et en rotation sont :

- ***e-ondelette***

- translation :  $moyenne(\Delta T_i) = 2.91$  mm
- rotation :  $moyenne(\Delta R_i) = 0,19$  deg

- ***e-shearlet***

- translation :  $moyenne(\Delta T_i) = 7.91$  mm
- rotation :  $moyenne(\Delta R_i) = 0,49$  deg

De manière générale, l'occultation générée par l'objet 3D a davantage perturbé la convergence en translation qu'en rotation. Cependant, l'approche *e-ondelette* est plus précise pour tous les ddl en présence d'occultations.

Ce scénario a également été réalisé avec l'erreur initiale de positionnement du *test 2* ainsi qu'un autre objet 3D. La FIGURE 4.22(a) nous présente l'image désirée avec l'occultation partielle, la FIGURE 4.22(b) l'image de différence initiale et les FIGURES 4.22(c)-(d) montrent les images de différence à la fin de la tâche de positionnement. Nous pouvons voir que cette fois-ci la méthode *e-ondelette* converge avec précision vers la position désirée (FIGURE 4.22 (c)) tandis que la méthode *e-shearlet* échoue (FIGURE 4.22(d)). Finalement les erreurs de positionnement à la convergence sont :

- ***e-ondelette***

- translation :  $moyenne(\Delta T_i) = 3.57$  mm
- rotation :  $moyenne(\Delta R_i) = 0.28$  deg

- *e-shearlet*

- translation : ✗
- rotation : ✗

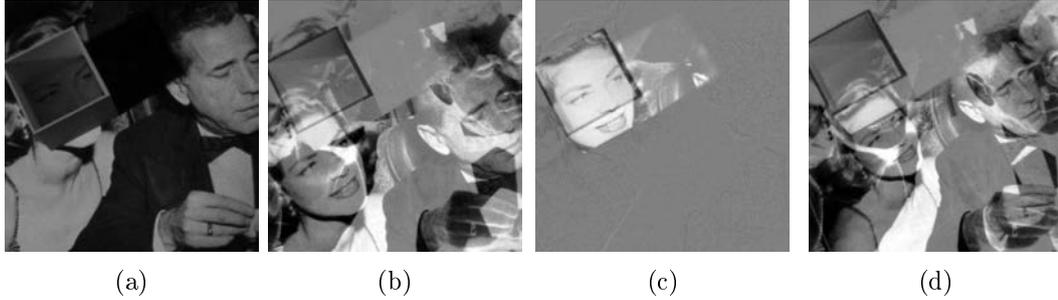


FIGURE 4.22 – [*scenario 2, test 2*] (a) image désirée  $\mathbf{I}^*$  (b) image de différence initiale  $\mathbf{I}_{diff}$ , (c)  $\mathbf{I}_{diff}$  finale dans le cas *e-ondelette*, et (d)  $\mathbf{I}_{diff}$  dans le cas *e-shearlet*.

Ce deuxième test expérimental est cohérent avec le premier et permet d'affirmer que la méthode *e-ondelette* est plus robuste aux occultations que la méthode *e-shearlet*.

**Variations d'illumination et scène planaire [*scenario 3*] :** La FIGURE 4.23(a) montre l'image désirée exposée à un spot lumineux, la FIGURE 4.23(b) celle de différence initiale et les FIGURE 4.23(c)-(d) illustrent respectivement l'image de différence finale pour l'approche *e-ondelette* et *e-shearlet*. Pour ce premier test, la méthode *e-shearlet* converge avec une erreur statique et des oscillations tandis que l'approche *e-ondelette* tombe dans un minimum local après 400 itérations.



FIGURE 4.23 – [*scenario 3, test 1*] (a) image désirée  $\mathbf{I}^*$  (b) image de différence initiale  $\mathbf{I}_{diff}$ , (c)  $\mathbf{I}_{diff}$  finale dans le cas *e-ondelette*, et (d) dans le cas *e-shearlet*.

Pour le *test 2*, dont les résultats sont présentés sur la FIGURE 4.24, les deux méthodes atteignent la position désirée avec une erreur statique plus importante. Cela peut s'expliquer par le fait que, comme pour l'approche photométrique, le modèle utilisé pour établir les matrices d'interaction de ces approches repose l'hypothèse de la continuité du flux optique. Cette hypothèse n'est pas validée en cas de variations d'illuminations sur

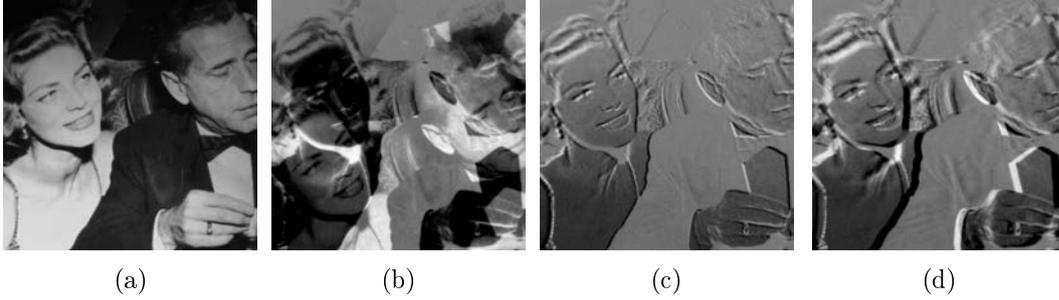


FIGURE 4.24 – [*scenario 3, test 2*] (a) image désirée  $\mathbf{I}^*$  (b) image de différence initiale  $\mathbf{I}_{diff}$ , (c)  $\mathbf{I}_{diff}$  finales dans le cas *e-ondelette*, et (d) dans le cas *e-shearlet*.

l'image. Cependant, les méthodes ne divergent pas complètement, ce qui leur confère une meilleure robustesse à cette hypothèse que la méthode photométrique. Cela peut être expliqué par le fait que les transformées en ondelettes et en shearlets autres que celle de niveau  $j = 0$ , sont parcimonieuses, avec des hauts coefficients pour des contours de l'image selon certaines directions et une majorité de coefficients proches de zéros ailleurs. Ces transformées sont donc peu impacté par les changements d'illumination.

Les valeurs de l'erreur de positionnement finale sont :

- *e-ondelette*

- translation :  $moyenne(\Delta T_i) = 23,5$  mm
- rotation :  $moyenne(\Delta R_i) = 1,3$  deg

- *e-shearlet*

- translation :  $moyenne(\Delta T_i) = 21,3$  mm
- rotation :  $moyenne(\Delta R_i) = 1,3$  deg

**Conditions nominales et scène 3D [scenario 4] :** Jusqu'à présent, les scénarios présentés utilisaient essentiellement des scènes planaires. Dans ce dernier scénario de validation, nous avons remplacé la photographie utilisée précédemment par un objet 3D tout en augmentant l'angle d'incidence entre la caméra et cet objet (voir FIGURE 4.25). Bien que les contrôleurs aient été élaborés dans le cas général, notre caméra monoculaire ne nous permet pas d'obtenir une valeur exacte de la profondeur  $Z$  pour chaque point de l'image. Le but de ce scénario est donc d'évaluer la robustesse de nos contrôleurs aux erreurs sur la profondeur de la scène. Les deux erreurs initiales considérées dans ce scénario entre l'image initiale  $\mathbf{I}$  et l'image désirée  $\mathbf{I}^*$  sont présentées dans le TABLEAU 4.9.

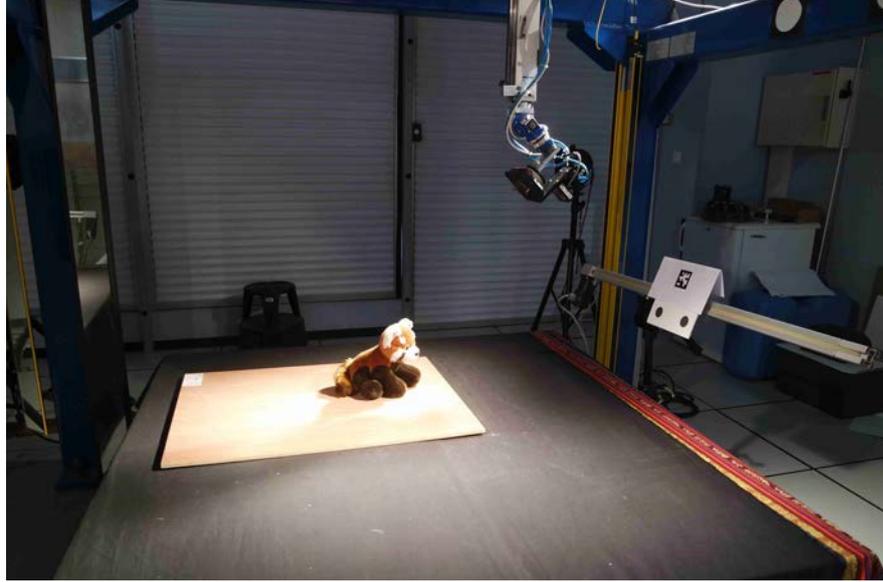
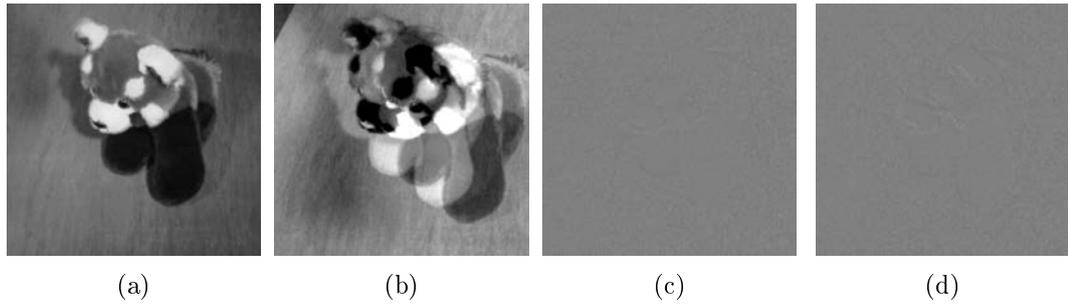


FIGURE 4.25 – Photographie du banc expérimental avec une scène 3D.

Test	$\Delta T_x$ (mm)	$\Delta T_y$ (mm)	$\Delta T_z$ (mm)	$\Delta R_x$ (°)	$\Delta R_y$ (°)	$\Delta R_z$ (°)
<i>test 3</i>	40	30	50	-5	2	-3
<i>test 4</i>	-100	-40	-200	3	-3	4

TABLE 4.9 – Erreur de positionnement initiale  $\mathbf{e}_0$  entre l'image initiale  $\mathbf{I}$  et l'image désirée  $\mathbf{I}^*$  pour une scène 3D.FIGURE 4.26 – [scenario 4, test 3] (a) image désirée  $\mathbf{I}^*$  (b) image de différence initiale  $\mathbf{I}_{diff}$ , (c)  $\mathbf{I}_{diff}$  finale dans le cas *e-ondelette*, et (d) dans le cas *e-shearlet*.

Les valeurs numériques des erreurs de positionnement finales pour le *test 3* sont :

- *e-ondelette*

- translation :  $moyenne(\Delta T_i) = 0.43$  mm
- rotation :  $moyenne(\Delta R_i) = 0.03$  deg

- *e-shearlet*

- translation :  $moyenne(\Delta T_i) = 0.15$  mm
- rotation :  $moyenne(\Delta R_i) = 0.02$  deg

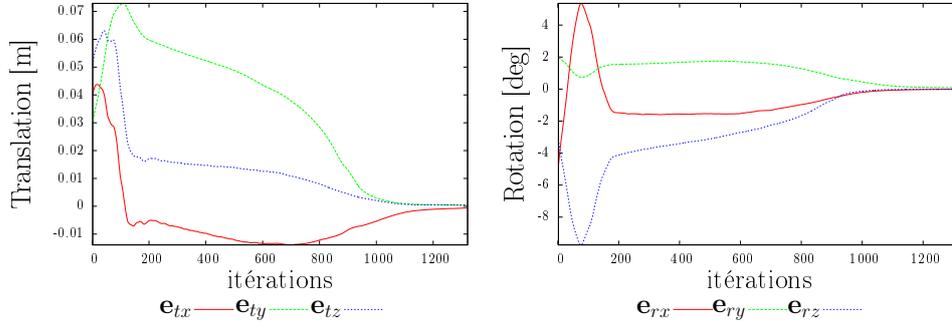


FIGURE 4.27 – [scenario 4, test 3] Erreur de positionnement  $\mathbf{e}$  avec la méthode *e-ondelette*.

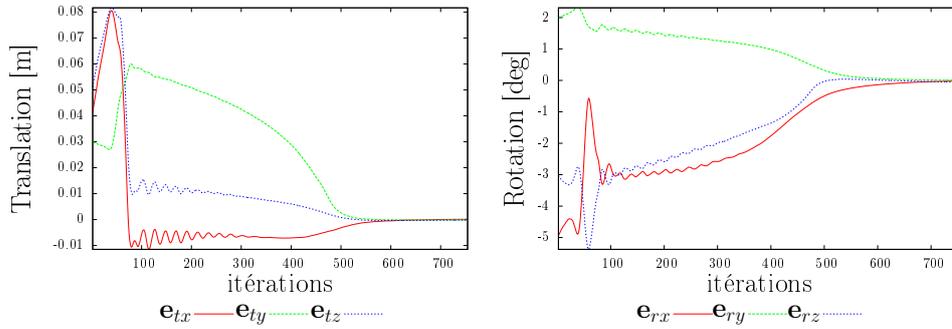


FIGURE 4.28 – [scenario 4, test 3] Erreur de positionnement  $\mathbf{e}$  avec la méthode *e-shearlet*.

Les deux méthodes sont très précises (moins d'un millimètre d'erreur en translation et moins de 0.05 degrés en rotation) pour le *test 3*. Les deux lois de commande convergent comme dans le cas d'une scène plane (voir FIGURE 4.26), mais la méthode *e-shearlet* présente une phase oscillatoire (voir FIGURE 4.27 et 4.28).

Pour le *test 4*, l'erreur initiale en translation est plus importante que pour le *test 3*. La FIGURE 4.29 nous montre que la méthode *e-ondelette* converge avec une bonne précision tandis que celle *e-shearlet* échoue. Les valeurs numériques des erreurs de positionnement finales sont :

- ***e-ondelette***

- translation :  $\text{moyenne}(\Delta T_i) = 0.53 \text{ mm}$
- rotation :  $\text{moyenne}(\Delta R_i) = 0.11 \text{ deg}$

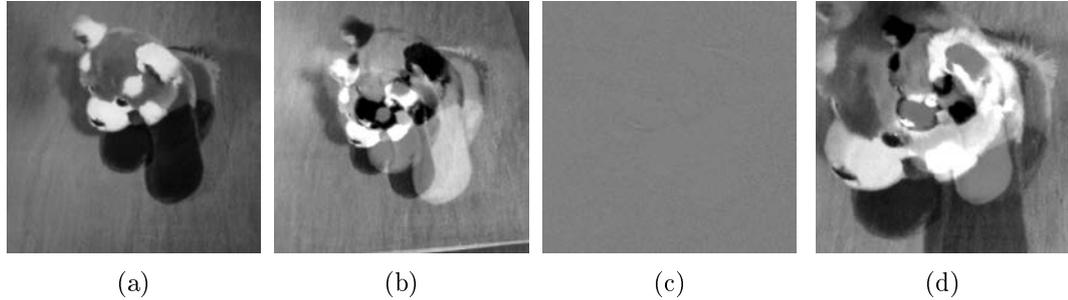


FIGURE 4.29 – [scenario 4, test4](a) image désirée  $\mathbf{I}^*$  (b) image de différence initiale  $\mathbf{I}_{diff}$ , (c)  $\mathbf{I}_{diff}$  finale dans le cas *e-ondelette*, et (d) dans le cas *e-shearlet*.

- *e-shearlet*

- translation :  $moyenne(\Delta T_i) = \mathbf{X}$
- rotation :  $moyenne(\Delta R_i) = \mathbf{X}$

Ce dernier scenario a montré expérimentalement que la méthode *e-ondelette* surpasse les performances de celle *e-shearlet*. Nous pouvons souligner que les rotations appliquées dans le cas d'une scène 3D sont assez faibles, mais une rotation de plus de 10 degrés faisait sortir l'objet 3D du champ de la caméra, à moins d'y associer un important mouvement de translation. De tels déplacements ne permettent pas la convergence des méthodes présentées dans ce chapitre, car elles ne sont pas assez robustes aux approximations sur la profondeur  $Z$  de la scène. Nous pouvons par conséquent conclure qu'en présence d'une scène 3D, la convergence peut être obtenue à condition d'avoir une erreur de positionnement initiale faible et ceci malgré la forte hypothèse d'une profondeur  $Z$  constante en chaque point de l'image.

Ainsi, cette validation expérimentale de l'asservissement visuel sur caméra CCD a permis de comparer les deux approches sous-échantillonnées. La méthode *e-shearlet* est légèrement plus précise en conditions nominales et plus robuste aux erreurs d'étalonnage et aux variations d'illumination mais la méthode *e-ondelette* surpasse la méthode *e-shearlet* lorsqu'il y a présence d'occultations ou une scène 3D. En conclusion, la méthode à privilégier dépendra du type de conditions défavorables susceptibles d'apparaître dans l'application visée. Nous avons choisi de tester la méthode *e-ondelette* sur des images de type OCT. Ce sujet sera développé dans la prochaine section.

## 4.3 Validation expérimentale en utilisant l'OCT

Cette section s'intéresse à la validation de l'approche *e-ondelette* avec une sonde OCT (voir sous-section 3.1.2) comme capteur visuel et une plate-forme robotique constituée d'un robot parallèle à 6 ddl. La vitesse générée par la loi de commande sera limitée aux composantes linéaires en raison de difficultés pratiques à commander le robot en rotations. La première sous-section présente le système robotique utilisé, la deuxième détaille le calcul de la matrice d'interaction adaptée à l'OCT et enfin la dernière expose les résultats expérimentaux obtenus.

### 4.3.1 Description du dispositif robotique

La méthode *e-ondelette* a été testée expérimentalement en utilisant une plate-forme robotique sur laquelle sont placés des échantillons biologiques (morceau de crevette) ou non-biologiques (LEGO en plastique) en vue d'une caractérisation optique à l'aide d'un système OCT.

La plate-forme robotique est un robot parallèle à 6 ddl (Space-FAB SF-3000 BS de l'entreprise Micros), caractérisée par une envergure de déplacement de  $(T_x, T_y, T_z)_{max} = (50, 100, 12.7)$  (mm) en translation et  $(R_x, R_y, R_z)_{max} = (10, 10, 10)$  (°) en rotation ainsi que par une résolution linéaire de  $0.2\mu\text{m}$  et angulaire de  $0.0005^\circ$ . Nous y avons relié deux ordinateurs communiquant avec un protocole TCP/IP. L'ordinateur commandant la partie vision est doté d'un processeur core Intel de 3.2 GHz et d'une distribution Linux tandis que l'ordinateur contrôlant le robot est doté d'un processeur Xeon Intel 2.33 GHz sous une configuration Windows.

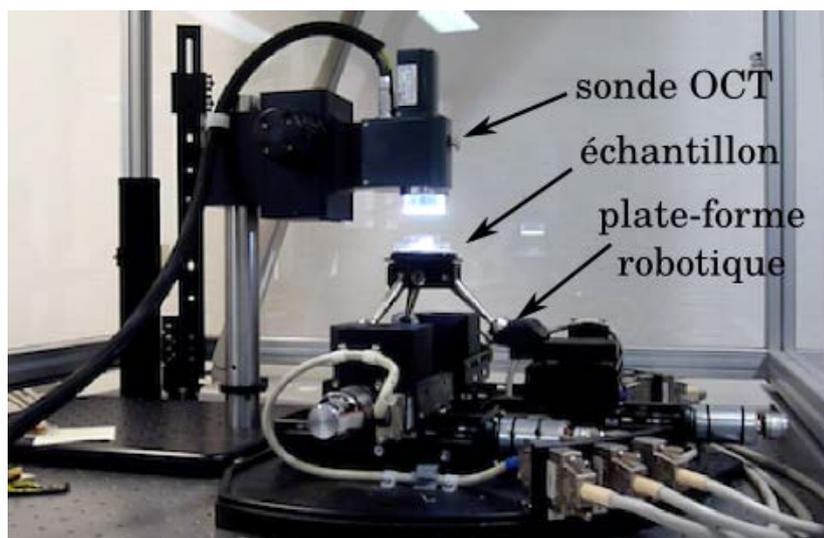


FIGURE 4.30 – Photographie du système expérimental OCT utilisé.

Le système d'imagerie OCT utilisé est présenté en FIGURE 4.30. Il s'agit d'un appareil OCT fréquentiel Telesto II de l'entreprise THORLABS caractérisé par des résolutions axiale et latérale de  $5.5\mu\text{m}$  and  $7\mu\text{m}$  avec une profondeur de pénétration de  $3.5\text{mm}$ . Dans notre étude, nous asservissons la plate-forme robotique placée sous le système OCT, sur lequel se trouve un échantillon à analyser. Nous sommes donc dans une configuration de type *caméra déportée*.

Pour des raisons techniques, il n'a pas été possible d'asservir la plate-forme robotique en rotation. Par conséquent, l'asservissement visuel, proposé ici, traite uniquement les 3 ddl correspondant aux trois translations. Comme pour notre validation avec une caméra CCD, le niveau de détail contenant les informations de plus hautes fréquences n'a pas été conservé afin de filtrer le bruit de l'image, particulièrement présent avec cette modalité d'imagerie.

### 4.3.2 Détails d'implémentation

Les informations utilisées dans l'asservissement visuel sont composées des coefficients de la transformée en ondelettes de l'image  $\mathbf{I}$ . Plus précisément, il s'agit des niveaux de détails  $j = 2$  et  $j = 3$  d'une transformée en ondelettes à 3 niveaux de décompositions :

$$\mathbf{s}_w(t) = \left( (\mathcal{W}_{\psi^{(l_1)}} I_t)(j_1, \mathbf{m}_1), (\mathcal{W}_{\psi^{(l_2)}} I_t)(j_2, \mathbf{m}_2), \dots, (\mathcal{W}_{\psi^{(l_N)}} I_t)(j_N, \mathbf{m}_N) \right)^\top, \quad (4.27)$$

avec  $j_i \in \{2, 3\}$  et  $l_i \in \{0, 1, 2, 3\}$ , (voir section 4.1.1) et  $\mathbf{m}_i$ , facteur de translation, dépendant de la valeur de  $j$ .

Comme précédemment, l'erreur entre les positions initiale et désirée est donnée comme suit :

$$\mathbf{e}_w(t) = \mathbf{s}_w(t) - \mathbf{s}_w^* \quad (4.28)$$

La FIGURE 4.31 permet de se représenter les informations visuelles utilisées dans la loi de commande.

La matrice d'interaction associée à  $\mathbf{s}_w$  nécessite alors de connaître la dérivée par rapport au temps de  $\mathbf{s}_w$ . Soit  $\mathbf{I}(\mathbf{x}, t)$  l'intensité du point située en  $\mathbf{x} = (x, y, z)$ , les coordonnées (en mètre) du point exprimées dans le repère de la coupe OCT à l'instant  $t$ . Nous avons alors la relation suivante :

$$\dot{\mathbf{I}}(\mathbf{x}, t) = \frac{\partial \mathbf{I}(\mathbf{x}, t)}{\partial \mathbf{x}}^\top \dot{\mathbf{x}} + \frac{\partial \mathbf{I}(\mathbf{x}, t)}{\partial t} \quad (4.29)$$

avec

$$\frac{\partial \mathbf{I}(\mathbf{x}, t)}{\partial \mathbf{x}} = \nabla \mathbf{I} = [\nabla \mathbf{I}_x \nabla \mathbf{I}_y \nabla \mathbf{I}_z]^\top \quad (4.30)$$

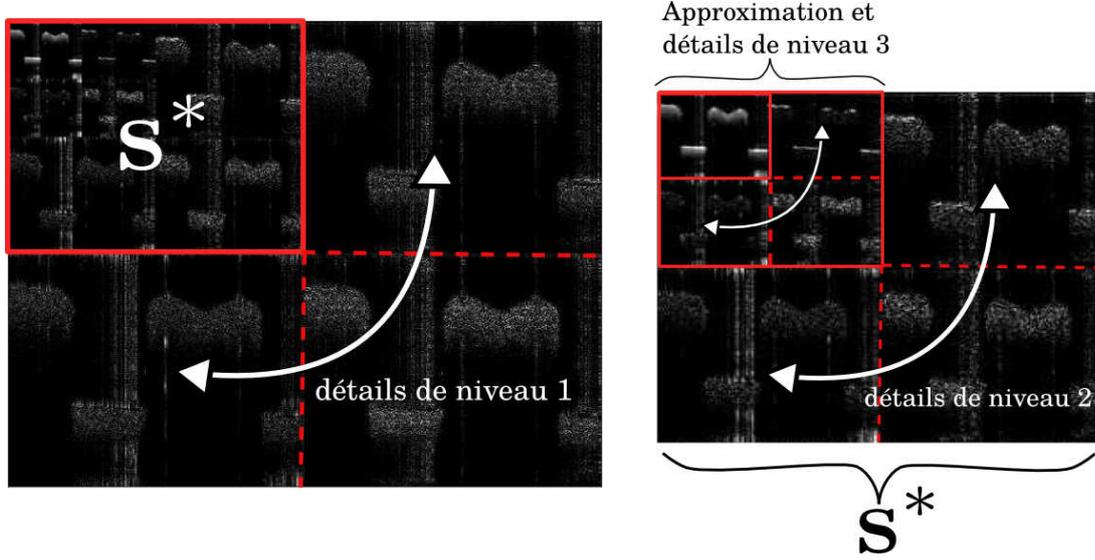


FIGURE 4.31 – Transformée en ondelettes d'une image OCT acquise sur un objet artificiel "LEGO".

Comme dans les travaux de (Nadeau et Krupa, 2013), en imagerie échographique, l'hypothèse que l'intensité d'un point reste constante dans une image OCT sur un faible intervalle de temps  $\Delta t$ , c'est-à-dire que  $\frac{\partial \mathbf{I}(\mathbf{x}, t)}{\partial t} = 0$  est alors utilisée. Ici encore, il s'agit d'une hypothèse forte compte tenu du bruit (*speckle*) contenu dans ce type d'image. La relation suivante est alors obtenue :

$$\dot{\mathbf{I}}(\mathbf{x}, t) = \nabla \mathbf{I} \dot{\mathbf{x}} \quad (4.31)$$

Pour calculer les composantes de  $\nabla \mathbf{I}$  liées aux translations selon la direction  $y$ , nous devons acquérir deux coupes parallèles de chaque côté de notre position désirée, comme illustré sur la FIGURE 4.32, afin de permettre une estimation numérique du gradient 3D de l'image désirée (Nadeau et Krupa, 2013).

La variation temporelle de  $\mathbf{x}$  peut être obtenue à partir de la vitesse de la sonde  $\mathbf{v}$ , en utilisant la formule de composition des vitesses suivante :

$$\dot{\mathbf{x}} = \begin{bmatrix} 1 & 0 & 0 & 0 & z & -y \\ 0 & 1 & 0 & -z & 0 & x \\ 0 & 0 & 1 & y & -x & 0 \end{bmatrix} \mathbf{v} \quad (4.32)$$

La section observée se trouve dans une coupe où  $y = 0$ . La matrice d'interaction associée à un point  $I_{u,v}$  de l'image de coordonnées pixelliques  $(u, v)$  est alors obtenue comme suit :

$$\mathbf{L}_{I_{u,v}} = [\nabla I_x \quad \nabla I_y \quad \nabla I_z \quad -z \nabla I_y \quad z \nabla I_x - x \nabla I_z \quad x \nabla I_y] \in \mathbb{R}^{1 \times 6} \quad (4.33)$$

Dans notre cadre de validation, seuls les translations sont asservies donc seule une projection de  $\mathbf{L}_{I_{u,v}}$  sur ses trois premières composantes est utilisée. La matrice d'interaction relative à  $\mathbf{s}_w$  est alors donnée par la relation suivante :

$$\mathbf{L}_{\text{sw}}(t) = - \begin{bmatrix} (\mathcal{W}_{\psi(t_1)} \nabla I_x)(j_1, \mathbf{m}_1) & (\mathcal{W}_{\psi(t_1)} \nabla I_y)(j_1, \mathbf{m}_1) & (\mathcal{W}_{\psi(t_1)} \nabla I_z)(j_1, \mathbf{m}_1) \\ \vdots & \vdots & \vdots \\ (\mathcal{W}_{\psi(t_N)} \nabla I_x)(j_N, \mathbf{m}_N) & (\mathcal{W}_{\psi(t_N)} \nabla I_y)(j_N, \mathbf{m}_N) & (\mathcal{W}_{\psi(t_N)} \nabla I_z)(j_N, \mathbf{m}_N) \end{bmatrix} \in \mathbb{R}^{N \times 3} \quad (4.34)$$

La loi de commande utilisée est la suivante :

$$\mathbf{v}(t) = -\lambda \left( \mathbf{H}(t) + \mu \text{diag}(\mathbf{H}(t)) \right)^{-1} \left( \mathbf{L}_{\text{sw}}(t) \right)^{\top} \mathbf{e}_{\text{w}}(t), \quad (4.35)$$

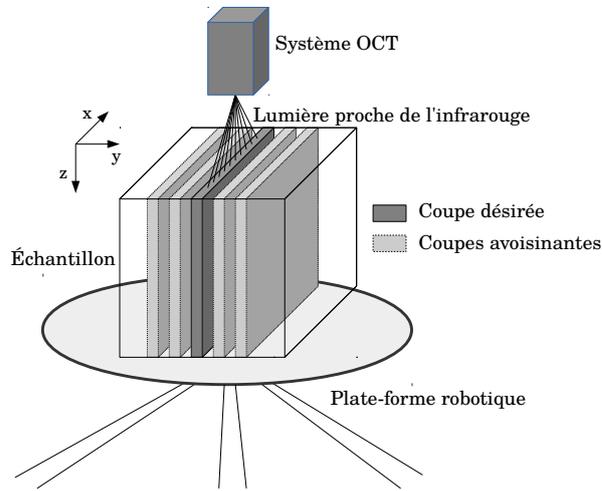


FIGURE 4.32 – Schéma des coupes utilisées pour calculer la matrice d'interaction.

### 4.3.3 Validation expérimentale

Afin de valider la loi de commande en considérant le système d'imagerie OCT comme capteur, nous l'avons testée sur un échantillon artificiel (une pièce de LEGO) et un échantillon biologique (un morceau de crevette). Ces échantillons sont présentés sur la FIGURE 4.33.

Le TABLEAU 4.10 expose les erreurs de positionnement initiales et finales pour l'ensemble des tests menés. Deux tests ont été réalisés en utilisant la pièce de LEGO et trois tests en utilisant le morceau de crevette. La moyenne de l'erreur de positionnement finale obtenue, en considérant tous les axes, est de 0.01mm et l'écart-type de 0.02mm. Ainsi, la méthode s'avère précise, même selon l'axe  $y$  alors que seules 4 coupes adjacentes sont utilisées pour estimer la deuxième colonne de la matrice d'interaction.

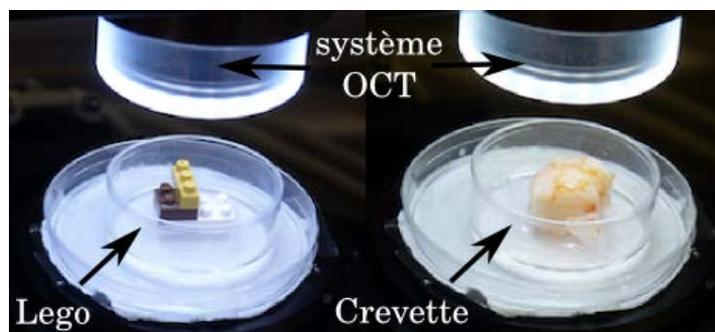


FIGURE 4.33 – Échantillons utilisés pour la validation expérimentale sur caméra OCT.

TABLE 4.10 – Tests réalisés pour valider la méthode *e-ondelette* sur un capteur OCT.

Cas	erreur	$\Delta T_x$ (mm)	$\Delta T_y$ (mm)	$\Delta T_z$ (mm)
LEGO test 1	$e_{initial}$	<b>-0.5</b>	<b>-0.5</b>	<b>0.3</b>
	$e_{final}$	-0.0118	-0.0149	0.0048
LEGO test 2	$e_{initial}$	<b>-0.3</b>	<b>-0.6</b>	<b>0.5</b>
	$e_{final}$	-0.0002	-0.0020	-0.0111
crevette test 1	$e_{initial}$	<b>-0.5</b>	<b>-0.5</b>	<b>0.3</b>
	$e_{final}$	-0.0074	0.0035	-0.0091
crevette test 2	$e_{initial}$	<b>-0.5</b>	<b>-0.2</b>	<b>0.3</b>
	$e_{final}$	-0.0209	0.0677	-0.0087
crevette test 3	$e_{initial}$	<b>0.2</b>	<b>-0.2</b>	<b>0.2</b>
	$e_{final}$	-0.0058	0.0054	-0.0070
	moyenne ( $ e_{final} $ )	0.0092	0.0187	0.0081
	écart-type ( $ e_{final} $ )	0.0077	0.0278	0.0024

Dans la suite, deux tests sont détaillés, le premier avec la pièce de LEGO et le deuxième avec le morceau de crevette.

**Expériences sur une pièce de LEGO :** La première batterie de tests a été réalisée avec une pièce de LEGO de taille  $1.6 \times 0.8$  cm<sup>2</sup>. Tout d'abord, l'échantillon a été placé à la position désirée  $\mathbf{r}^*$  et ensuite déplacé à une position initiale arbitraire  $\mathbf{r}$ . L'erreur de positionnement initiale est alors estimée par les encodeurs du robot à  $\Delta T_{initial} = (-0.5, -0.5, 0.3)$  (mm). L'image de différence entre l'image OCT initiale  $\mathbf{I}_{oct}$  et l'image OCT désirée  $\mathbf{I}_{oct}^*$  est visible sur la FIGURE 4.34(b). Le déplacement le long des axes  $x$  et  $z$  est visible sur la coupe OCT, contrairement au déplacement selon l'axe  $y$ . Ceci est dû à la structure presque invariante du LEGO le long de l'axe  $y$  (FIGURE 4.34(b)). Cependant, la différence d'image le long de cet axe est visible sur la FIGURE 4.35.(b). En effet, durant la tâche de positionnement, nous pouvons obtenir une image de la surface de l'échantillon à l'aide d'une caméra CCD, attachée à l'appareil OCT. Ainsi, même si elles ne sont pas utilisées dans la commande, ces images peuvent

être affichées pour illustrer visuellement la convergence de la loi de commande.

Les résultats de la première expérience avec la pièce de LEGO sont présentés sur la FIGURE 4.34. La FIGURE 4.34(a) représente l'image OCT initiale, la FIGURE 4.34(b) la différence initiale et la FIGURE 4.34(c) la différence finale, à la convergence. Cette dernière contient essentiellement du bruit de type *speckle*, très fréquent en imagerie OCT. L'erreur de positionnement finale obtenue est de  $\Delta T_{final} = (-0.0118, -0.0149, 0.0048)$  (mm), témoignant de la précision de la méthode (voir FIGURE 4.35).

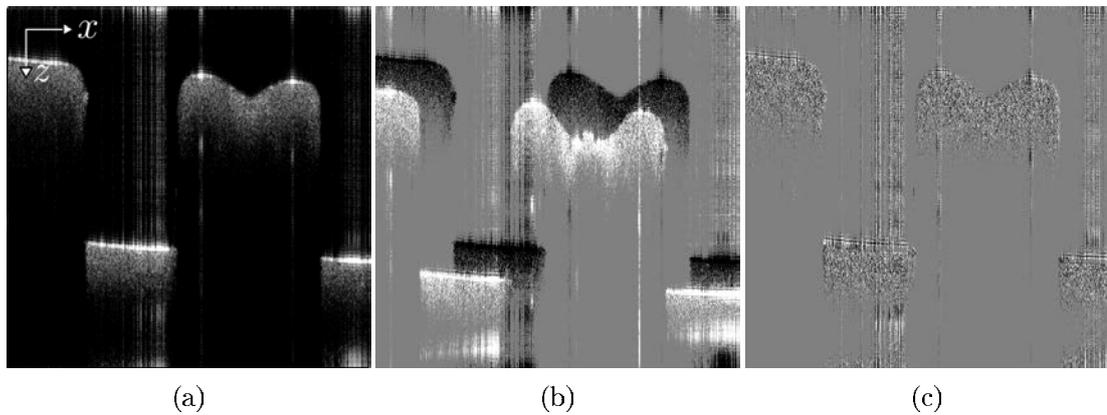


FIGURE 4.34 – Test 1 (OCT) : (a) image OCT désirée  $\mathbf{I}_{oct}^*$ ; (b)  $\mathbf{I}_{diff}^*$  initiale correspondante, (c)  $\mathbf{I}_{diff}^*$  finale correspondante.

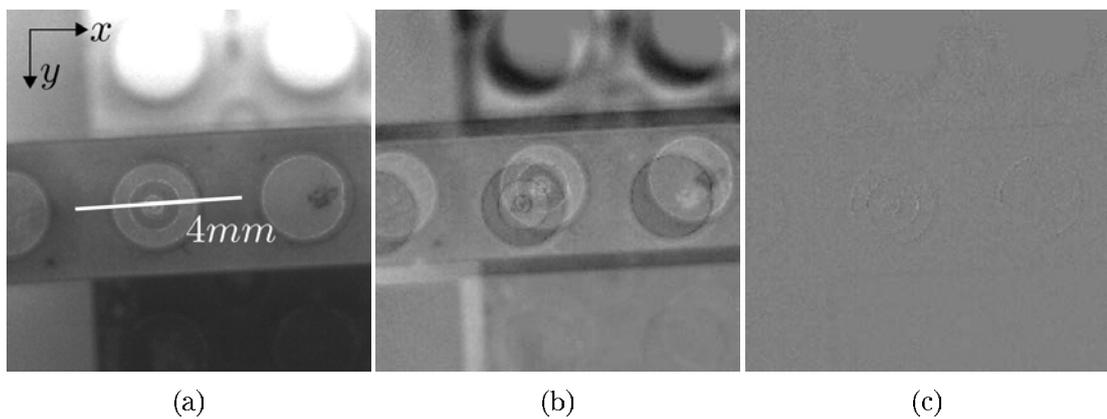


FIGURE 4.35 – Test 1 (camera CCD) : (a) image CCD désirée  $\mathbf{I}_{cam}^*$ ; (b)  $\mathbf{I}_{diff}^*$  initiale correspondante, (c)  $\mathbf{I}_{diff}^*$  finale correspondante.

L'évolution de l'erreur de positionnement sur chaque ddl est présentée sur la FIGURE 4.36 et la norme de l'erreur  $\|e\| = \|s_w - s_w^*\|$  sur la FIGURE 4.37. L'ensemble des erreurs converge vers 0 en environ 80 itérations. L'évolution des erreurs de positionnement est linéaire témoignant d'une commande stable. L'évolution de la norme de l'erreur n'est, quant à elle, pas lisse. Nous pouvons supposer que cet aspect est probablement dû à la présence de bruit *speckle* dans les images OCT mais cette hypothèse demande à être vérifiée.

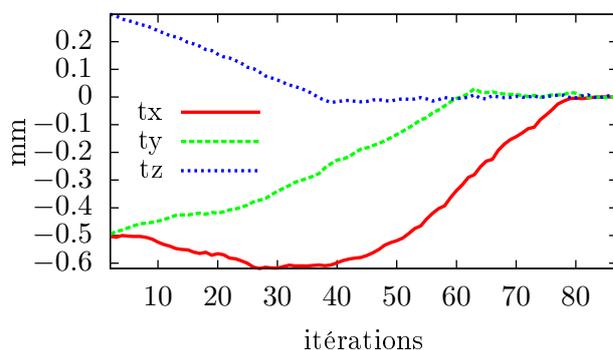


FIGURE 4.36 – LEGO : erreurs de positionnement.

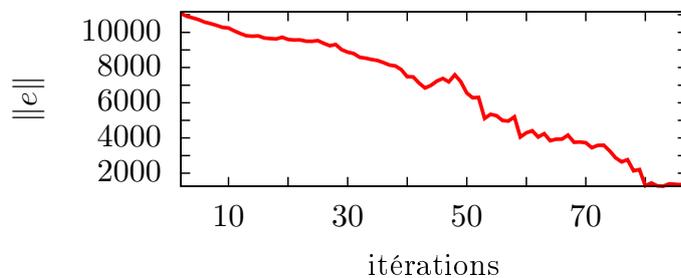


FIGURE 4.37 – LEGO : norme de l'erreur entre coefficients ondelettes de l'image courante et ceux de l'image désirée.

**Expérience sur un morceau de crevette :** Pour notre seconde expérience, la pièce de LEGO a été remplacée par un échantillon biologique : un morceau de crevette d'une taille d'environ  $2 \times 1 \text{ cm}^2$ . Ce scénario reste le même que celui du LEGO, l'image à la position désirée est enregistrée puis le robot déplace l'échantillon à une position initiale arbitraire. Notre objectif est de simuler, par exemple, une biopsie optique réalisée sur un même emplacement à plusieurs jours d'intervalle. L'erreur de positionnement initiale pour cette expérience est de  $\Delta T_{initial} = (-0.5, -0.5, 0.3) \text{ (mm)}$ . L'image de différence entre l'image initiale et l'image désirée est représentée sur la FIGURE 4.38.(b). Contrairement à l'expérience sur le LEGO, le déplacement le long de l'axe  $y$  est visible sur l'image OCT (voir FIGURE 4.38(b)). Comme le montre la FIGURE 4.38.(c), qui est complètement grise, la loi de commande converge vers la position désirée avec une erreur de positionnement finale  $\Delta T_{final} = (-0.0074, 0.0035, -0.0091) \text{ (mm)}$ . La FIGURE 4.39 montre également la convergence du point de vue de la caméra CCD.

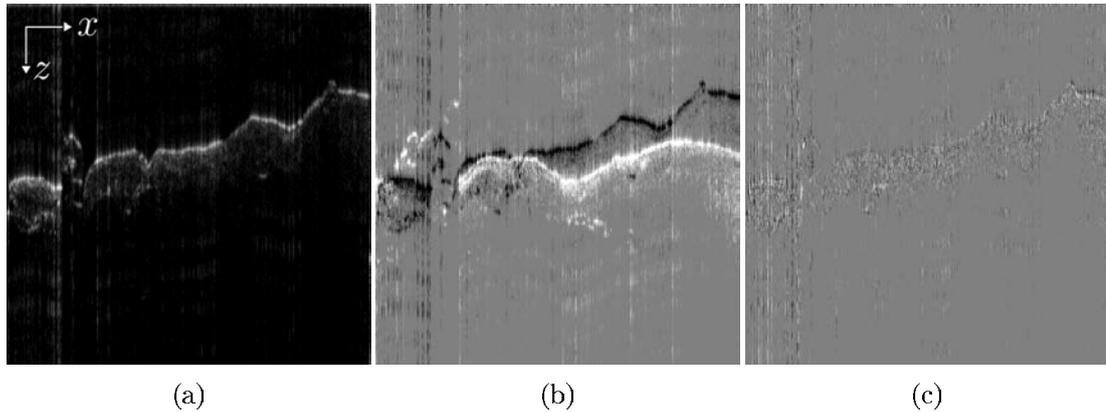


FIGURE 4.38 – Test 2 (OCT) : (a) image OCT désirée  $\mathbf{I}_{oct}^*$ , (b)  $\mathbf{I}_{diff}^*$  initiale correspondante, (c)  $\mathbf{I}_{diff}^*$  finale correspondante.

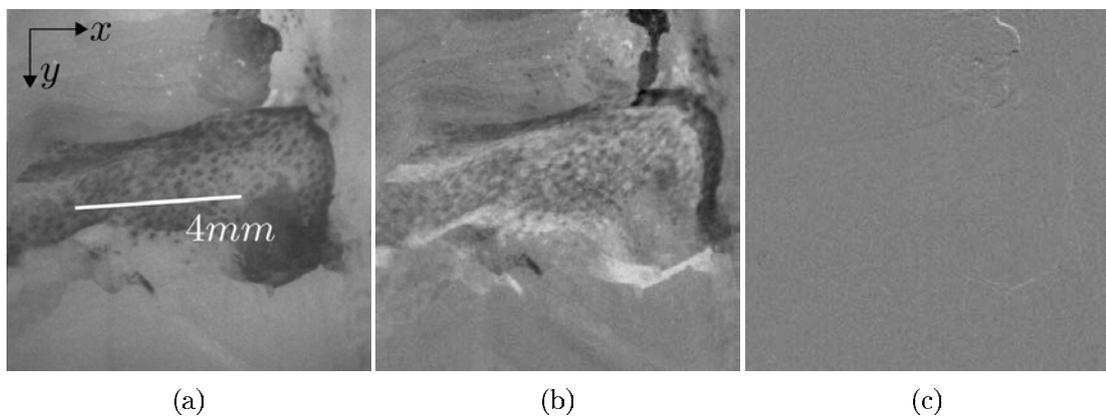


FIGURE 4.39 – Test 2 (camera CCD) : (a) image CCD désirée  $\mathbf{I}_{cam}^*$ , (b)  $\mathbf{I}_{diff}^*$  initiale correspondante, (c)  $\mathbf{I}_{diff}^*$  finale correspondante.

L'évolution de l'erreur sur chaque degré de liberté est présentée sur la FIGURE 4.40. Nous observons le même comportement que pour le test sur la pièce de LEGO, avec une évolution très linéaire de l'erreur de positionnement. De même, l'évolution de la norme des coefficients en ondelettes de l'image  $\|\mathbf{e}\| = \|\mathbf{s}_w - \mathbf{s}_w^*\|$  est visible sur la FIGURE 4.41 et sa décroissance vers 0 confirme la bonne convergence de la loi de commande proposée, avec une convergence très rapide en fin de tâche.

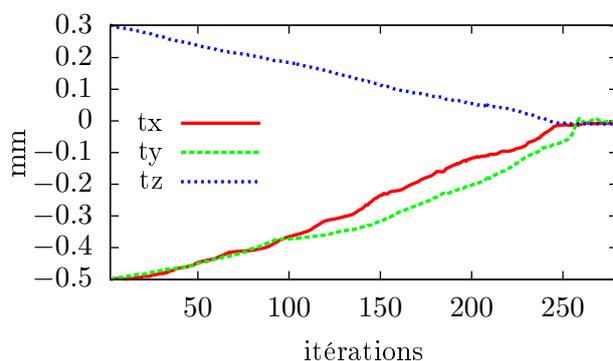


FIGURE 4.40 – Morceau de crevette : erreurs de positionnement.

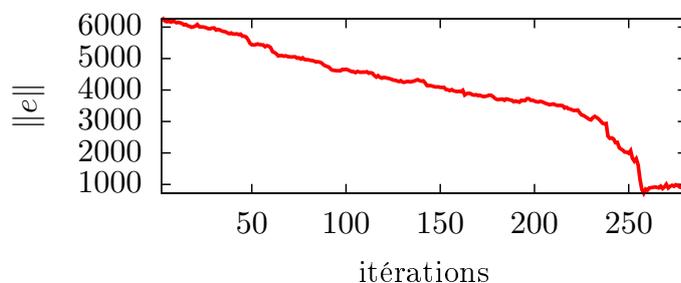


FIGURE 4.41 – Morceau de crevette : norme de l'erreur entre coefficients ondelettes de l'image courante et ceux de l'image désirée.

L'objectif de cette section était l'étude de notre approche fondée sur les ondelettes en imagerie OCT. A l'aide des deux types d'échantillon, le premier non-biologique et le second biologique, nous avons montré que la commande permettait d'obtenir des résultats satisfaisants en termes de précision et de robustesse au bruit, très présent en OCT.

## 4.4 Conclusion

Nous avons proposé quatre lois de commandes fondées sur les représentations multi-échelles (ondelettes et shearlets) afin de réaliser une tâche d’asservissement visuel à 6 ddl. Pour chacune des méthodes, nous avons mis au point une matrice d’interaction analytique reliant la variation temporelle des coefficients d’ondelettes ou de shearlets à celles des vitesses du capteur visuel (caméra conventionnelle ou système d’imagerie OCT). Les approches sous-échantillonnées et non sous-échantillonnées ont été prises en compte et comparées en simulation. Elles ont toutes montré une importante précision lors de conditions nominales d’utilisation. Cependant, les différents scénarios de simulation ont mis en évidence que les approches sous-échantillonnées étaient plus intéressantes que celles non sous-échantillonnées. En effet, elles sont d’une part plus robustes aux conditions défavorables (occultations partielles, variations d’illumination) et d’autre part moins coûteuses en temps de calcul. La suppression des plus hautes fréquences du signal dans l’élaboration du vecteur d’informations visuelles permet de filtrer le bruit de l’image.

Les approches sous-échantillonnées ont été également validées expérimentalement avec des conditions nominales ou dégradées. Elles ont montré une précision de moins d’un millimètre en translation et de moins d’un demi-degré en rotation en conditions nominales. Si la comparaison entre les approches sous-échantillonnées et non sous-échantillonnées a mis en évidence la supériorité des approches sous-échantillonnées, en terme de robustesse aux conditions défavorables, la comparaison entre les méthodes *e-ondelette* et *e-shearlet* est plus nuancée. Ainsi, nous avons pu constater que la méthode *e-shearlet* est plus robuste aux erreurs d’étalonnage du capteur de vision et aux variations d’illumination tandis que l’approche *e-ondelette* est plus robuste aux occultations partielles et à une mauvaise estimation de la profondeur de la scène (objets 3D). De plus, la méthode *e-ondelette* permet un calcul plus rapide de la matrice d’interaction. La rapidité de la commande avec cette méthode pourrait être améliorée en utilisant une implémentation en C++ lors du calcul des coefficients en *e-ondelette* par des bibliothèques préexistantes, ce qui serait impossible dans l’immédiat avec l’approche *e-shearlet*, puisqu’aucune implémentation C++ de la transformée en shearlets n’existe pour le moment.

Enfin, nous avons étudié la faisabilité d’une commande utilisant la méthode *e-ondelette* sur un dispositif OCT. Malgré la nature très bruitée des images OCT, la méthode reste fonctionnelle avec un comportement tout à fait satisfaisant. En effet, l’erreur finale moyenne obtenue à la convergence est de 0.01mm (soit environ 3% de la moyenne des déplacements initiaux).



## Chapitre 5

# Shearlets en acquisition comprimée : application à l'OCT

Dans le cadre d'un ambitieux projet traitant du diagnostic des maladies neurodégénératives (Alzheimer, Parkinson) par l'étude du dysfonctionnement du système olfactif (muqueuse olfactive), nous avons mené des travaux sur l'acquisition comprimée. Dans ce projet, un micro-robot flexible endonasal équipé, en sa partie distale, d'un système d'imagerie de type OCT pour l'acquisition de biopsies optiques 3D est utilisé. L'OCT est une modalité pour laquelle l'acquisition peut être très coûteuse en temps en raison de la manipulation et du transfert d'un grand nombre de données. Le but de nos recherches est la simulation d'une acquisition comprimée permettant un balayage intelligent de la zone à analyser et permettant de réduire le temps d'acquisition sans perte de qualité d'image. Nous avons donc simulé une méthode d'acquisition partielle de l'image avec un masque en spirale suivie d'une reconstruction utilisant différentes représentations multi-échelles dans lesquelles le signal est parcimonieux. Ces reconstructions obtenues avec une adaptation de l'algorithme *Fast Iterative Soft-Thresholding* (FISTA) ont été comparées à l'aide de deux métriques : "Peak Signal Noise Ratio" (PSNR) et "Structural Similarity" (SSIM) afin de déterminer quelle transformée multi-échelle donne la reconstruction la plus proche de l'image de départ. Le sommaire ci-dessous présente la disposition de ce chapitre :

### Sommaire

---

<b>5.1</b>	<b>Bases de l'acquisition comprimée</b>	<b>150</b>
5.1.1	Problème d'optimisation	150
5.1.2	Reconstruction d'un signal	153
5.1.3	Acquisition comprimée en imagerie	156
<b>5.2</b>	<b>Application à l'imagerie OCT</b>	<b>157</b>
5.2.1	Nouvelle formulation de l'algorithme FISTA	158
5.2.2	Validation numérique	159
<b>5.3</b>	<b>Conclusion</b>	<b>166</b>

---

## 5.1 Bases de l'acquisition comprimée

Le concept d'acquisition comprimée permet d'améliorer l'acquisition et la reconstruction d'un signal grâce à la résolution d'un système linéaire sous-déterminé. Cette technique fut introduite par Candès, Donoho, Romberg et Tao (Candes et Wakin, 2008) qui démontrèrent qu'il est possible de dépasser la limite imposée par le théorème de Shannon (voir ci-dessous), longtemps considéré comme fondamental en théorie de l'information.

**Théorème 5.1. Théorème de Shannon-Nyquist :** *La représentation discrète d'un signal nécessite des échantillons régulièrement espacés à une fréquence d'échantillonnage supérieure au double de la fréquence maximale présente dans ce signal.*

*Réciproquement, l'échantillonnage avec des échantillons régulièrement espacés peut décrire un signal à condition qu'il ne contienne aucune fréquence supérieure à la moitié de celle de l'échantillonnage, dite fréquence de Nyquist.*

Pour obtenir une bonne reconstruction du signal sans respecter le théorème de Shannon, la théorie de l'acquisition comprimée consiste à définir une base dans laquelle le signal à reconstruire est parcimonieux ("sparse" en anglais), c'est-à-dire une base dans laquelle le signal contient un nombre important de coefficients nuls. C'est ensuite grâce à une minimisation prenant en compte la parcimonie du signal dans cette base que nous pouvons l'approximer. Depuis les années 2000, l'utilisation de la parcimonie (c'est-à-dire des matrices creuses) est un domaine largement étudié en traitement du signal. L'acquisition comprimée est désormais en plein essor et compte des milliers d'articles, couvrant un nombre important de domaines d'applications : filtrage des signaux, manipulations de grands volumes de données, etc. La parcimonie des représentations en ondelettes a notamment été très utilisée dans les standards de compression comme MPEG (Le Gall, 1991) ou JPEG2000 (Christopoulos et al., 2000). Une nouvelle tendance consiste à utiliser la parcimonie directement dans l'élaboration des capteurs d'imagerie afin de prendre cette information en compte dès l'étape d'acquisition des données et d'éviter de stocker toutes les mesures et de les compresser ensuite. Les laboratoires *Nokia Bell labs* ont notamment créé des caméras sans lentille fonctionnant sur le principe de l'acquisition comprimée (Huang et al., 2013). La parcimonie peut également être utilisée pour la reconstruction de données sismiques (Wang et al., 2011), pour la séparation de sources sonores (Yang et al., 2015) ou pour la restauration d'images dégradées (Xu et Wang, 2009).

### 5.1.1 Problème d'optimisation

Depuis le milieu des années 2000, Dohono, Candès et Romberg ont démontré qu'il est possible de reconstruire un signal et *a fortiori* une image de taille  $n$  avec seulement  $m$  mesures ( $m \ll n$ ) sous réserve de respecter certaines conditions, développées dans la suite de cette section (Candès et Romberg, 2005; Candès et al., 2006a; Romberg, 2008;

(Donoho et al., 2010). La FIGURE 5.1 illustre le principe d'une méthode d'acquisition comprimée. En disposant d'un vecteur d'observations  $\mathbf{b}$ , nous cherchons à reconstruire un signal  $\boldsymbol{\varsigma}$  avec un minimum de mesures. Le problème est que le nombre d'observations disponibles est largement inférieur à la taille de  $\boldsymbol{\varsigma}$ . Or, si l'inconnue  $\boldsymbol{\varsigma}$  est parcimonieuse dans une certaine base et que nos observations  $\mathbf{b}$  fournissent des informations qui permettent de retrouver les coefficients non nuls de l'image de  $\boldsymbol{\varsigma}$  dans la base parcimonieuse alors une bonne estimation de  $\boldsymbol{\varsigma}$ , appelée  $\tilde{\boldsymbol{\varsigma}}$  peut être obtenue.

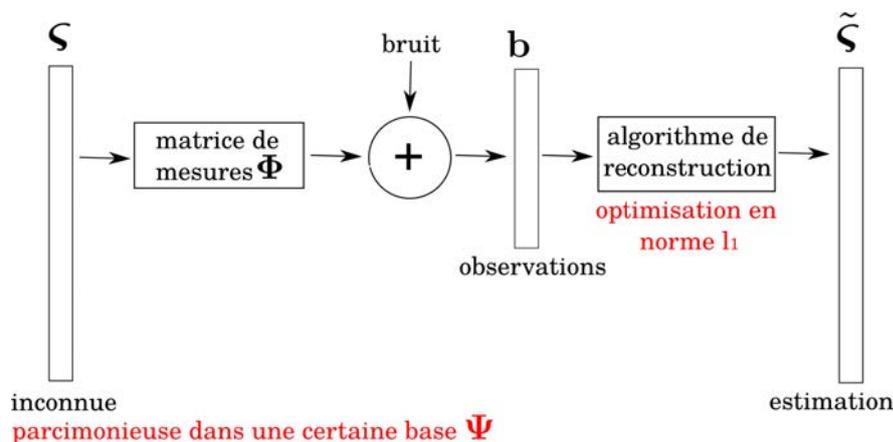


FIGURE 5.1 – L'acquisition comprimée.

**Définition 5.1. Principe de cohérence :** Le principe de cohérence mesure la corrélation maximum entre chaque paire d'éléments de deux matrices  $\Phi$  et  $\Psi$ . Ces dernières représentent deux bases différentes,  $\Phi$  est de taille  $m \times n$  et  $\Psi$  de taille  $n \times n$ .  $\Phi_j$  représente la  $j^{\text{ème}}$  colonne de la matrice  $\Phi$  et  $\Psi_k$  représente la  $k^{\text{ème}}$  colonne de la matrice  $\Psi$ . Alors la cohérence  $c$  entre ces deux bases est définie comme suit :

$$c(\Phi, \Psi) = \sqrt{n} \cdot \max_{1 \leq j \leq m, 1 \leq k \leq n} \langle \Phi_j, \Psi_k \rangle \quad (5.1)$$

Formellement, pour utiliser la théorie de l'acquisition comprimée, les conditions à respecter sont les suivantes :

**Parcimonie :** Il existe une base  $\Psi$  (appelée base parcimonieuse), dans laquelle le signal est parcimonieux, c'est-à-dire une base dans laquelle l'expression du signal est constituée d'un grand nombre de coefficients nuls. Le signal est alors représenté par  $k$  coefficients non nuls avec  $k \ll n$ . Ainsi, ce signal est dit  $k$ -sparse dans la base  $\Psi$ . En effet la projection d'un signal sur une base bien choisie permet d'obtenir de nombreux coefficients assez proches de zéro pour être négligeables. De telles bases peuvent être les ondelettes pour les images, les sinusoides localisées pour des signaux sonores ou encore les curvelets pour des données sismiques.

**Incohérence** En acquisition comprimée, la cohérence entre, la base de départ  $\Phi$  et la base où notre signal est  $k$ -sparse  $\Psi$  doit être la plus faible possible. Cela signifie

que la corrélation maximale entre deux éléments, le premier pris dans la base  $\Psi$  et le second pris dans la base  $\Phi$  doit être très faible. De ce fait, chacun des  $k$  éléments non nuls de la base  $\Psi$  doit dépendre d'un grand nombre d'éléments dans la base de mesures  $\Phi$  où le signal a été initialement défini. En raison de notre connaissance d'un nombre de coefficients limité dans la base  $\Phi$ , ce principe garantit qu'aucun coefficient indispensable à une bonne reconstruction dans la base  $\Psi$  ne pourra être omis. En pratique, la matrice  $\Phi$  correspond souvent à une distribution d'acquisition aléatoire, que l'on assimilera à une matrice appelée masque d'acquisition dans la suite.

En considérant les éléments suivants :

- $\zeta \in \mathbb{R}^n$  le signal à reconstruire.
- $\mathbf{b} \in \mathbb{R}^m$  avec  $m \ll n$  le vecteur de mesures.
- $\Phi \in \mathbb{R}^{m \times n}$  la matrice des mesures, telle que

$$\mathbf{b} = \Phi \zeta \quad (5.2)$$

- $\Psi \in \mathbb{R}^{n \times n}$  la base parcimonieuse telle que

$$\zeta = \Psi \sigma \quad (5.3)$$

avec  $\sigma$  un vecteur  $k$ -sparse,

Le problème consiste alors à minimiser la norme de  $\zeta$  en respectant les équations (5.2) et (5.3). L'enjeu est de définir la norme à utiliser pour minimiser la taille du vecteur  $\sigma$  sachant qu'il est constitué de nombreux coefficients nuls.

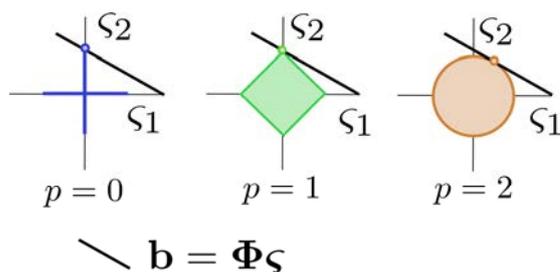


FIGURE 5.2 – Minimisation en norme  $l_p$

**Définition 5.2. Norme d'un vecteur :** La norme  $l_p$ ,  $p \in \{0, 1, 2\}$  d'un vecteur  $\zeta$  de taille  $n$  est égale à :

$$\|\zeta\|_p = \sqrt[p]{\sum_{i=1}^n |\zeta_i|^p} \quad (5.4)$$

Même si cette définition reste valable pour la norme  $l_0$ , la définition suivante est souvent utilisée :

$$\|\boldsymbol{\varsigma}\|_0 = \#\{i | \varsigma_i \neq 0\} \quad (5.5)$$

Cela signifie que la norme  $l_0$  d'un vecteur  $\boldsymbol{\varsigma}$  correspond au nombre de coefficients non nuls de ce vecteur.

Pour des raisons topologiques, c'est-à-dire de la forme des sphères unités pour les différentes normes considérées (voir FIGURE 5.2), la norme  $l_2$  ne permet pas d'obtenir un vecteur parcimonieux comme solution (Romberg, 2008; Davenport et al., 2011). En effet, sur cet exemple en dimension 2, nous pouvons voir une droite correspondant aux vecteurs respectant l'égalité  $\mathbf{b} = \Phi\boldsymbol{\varsigma}$  et leur intersection avec les sphères unités en norme  $l_p$  avec  $p \in \{0, 1, 2\}$ . La solution obtenue pour  $p = 2$  possède ses deux coordonnées  $\varsigma_1$  et  $\varsigma_2$  non nulles alors que les solutions pour  $p = 0$  et  $p = 1$  présente une coordonnée nulle. Étendue à  $\mathbb{R}^n$ , l'intersection d'un hyperplan avec la boule unité des normes  $l_0$  et  $l_1$  aura toujours davantage de coefficients nuls que l'intersection avec celle de la norme  $l_2$ . En pratique, la norme  $l_0$  n'est pas utilisée car cette technique s'avère trop coûteuse en temps de calcul, revenant à essayer  $\binom{n}{k}$  combinaisons pour un signal  $k$ -sparse de taille  $n$ . En effet, la minimisation en norme  $l_0$  est un problème NP complet<sup>1</sup>. De ce fait, la norme  $l_1$  est la plus adaptée. Le problème peut alors s'écrire :

$$\min \|\boldsymbol{\sigma}\|_1 \text{ sachant que } \mathbf{b} = \Phi\Psi\boldsymbol{\sigma} = \Omega\boldsymbol{\sigma} \quad (5.6)$$

**Définition 5.3. Propriété de restriction isométrique (RIP) :** Pour chaque  $k \in \mathbb{N}$ , définir l'isométrie de constante  $\delta_k \in ]0, 1[$  de la matrice  $\Omega$  consiste à trouver le plus petit nombre vérifiant l'inégalité suivante pour tout vecteur  $\boldsymbol{\sigma}$   $k$ -sparse :

$$(1 - \delta_k)\|\boldsymbol{\sigma}\|_2^2 \leq \|\Omega\boldsymbol{\sigma}\|_2^2 \leq (1 + \delta_k)\|\boldsymbol{\sigma}\|_2^2 \quad (5.7)$$

Candes et ses collègues ont montré que si la matrice  $\Omega \in \mathbb{R}^{m \times n}$  est RIP( $2k, \delta_{2k}$ ) avec  $\boldsymbol{\sigma} \in \mathbb{R}^n$  un vecteur  $k$ -sparse alors  $\boldsymbol{\varsigma}$  est la seule solution au problème (5.6) (Candes et al., 2006a).

Le nombre de mesures  $m$  nécessaires pour obtenir une parfaite reconstruction du signal dépend de la parcimonie du signal dans la base parcimonieuse. Des vérifications numériques suggèrent, qu'en pratique, la plupart des signaux  $k$ -sparses peuvent être parfaitement reconstruits lorsque  $m \geq 4k$  (Lustig et al., 2007). Si un signal n'est pas  $k$ -sparse, alors la qualité de la reconstruction obtenue sera aussi précise que celle obtenue en gardant les  $k$  plus grands coefficients du signal (Candes et al., 2006b).

### 5.1.2 Reconstruction d'un signal

Il existe de nombreux algorithmes de reconstruction en acquisition comprimée. Le travail présenté dans (Qaisar et al., 2013) nous permet de les diviser en six catégories :

---

1. Un problème NP complet est un problème pour lequel toute solution proposée peut être rapidement vérifiée mais pour lequel aucune solution ne peut être trouvée efficacement.

**Relaxation convexe** : ce type d'algorithme utilise la programmation linéaire pour reconstruire le signal. Les variantes les plus connues sont "*Basic Pursuit*" (BP) (Chen et al., 2001) ou "*Least Angle Regression*" (LARS) (Candès et Recht, 2009). Ces méthodes ont l'inconvénient de donner un résultat précis uniquement quand la solution est très parcimonieuse (Cai et Zhang, 2014).

**Algorithmes gloutons** : Ces algorithmes résolvent le problème de reconstruction de manière itérative par des sélections aléatoires de colonnes dans la matrice  $\Omega$ . L'algorithme le plus connu pour reconstruire petit à petit le support de la solution est "*Orthogonal Matching Pursuit*" (OMP) (Blumensath et al., 2012). Néanmoins, comme leur nom l'indique, ces algorithmes sont coûteux en ressources.

**Algorithmes de seuillages itératifs** : Ces techniques utilisent des seuillages bruts ou doux en partant de mesures bruitées du signal, obtenues grâce à la transposée de  $\Omega$  (Daubechies et al., 2004; Blumensath et Davies, 2009). Les méthodes "*Fast iterative soft thresholding*" (ISTA) (Daubechies et al., 2004) et la méthode "*Approximate message passing*" (AMP) (Donoho et al., 2010) sont les plus courantes.

**Algorithmes combinatoires** : Gilbert et ses collègues ont développé ces techniques efficaces et très rapides mais qui nécessitent que la matrice de mesures  $\Phi$  soit creuse, c'est-à-dire qu'elle soit constituée de nombreux coefficients nuls (Gilbert et al., 2007).

**Algorithmes de Bregman** : Ces algorithmes découpent le problème d'optimisation de départ en sous-problèmes qui sont résolus de manière itérative (Yin et al., 2008). Ils sont très bien adaptés aux implémentations GPU (Graphics Processing Unit).

Nous avons choisi d'éliminer les algorithmes gloutons en raison de leur coût en temps de calcul. Ne connaissant pas le degré de parcimonie de la solution de notre problème et n'ayant pas de matrice de mesures creuse, nous avons également écarté les méthodes de relaxation convexe et les algorithmes combinatoires. Enfin, les méthodes de minimisation non convexes et les algorithmes de Bregman étant moins documentés dans la littérature, nous avons choisi d'utiliser un algorithme de seuillage itératif dans nos travaux. Par ailleurs, ces algorithmes sont très utilisés en imagerie IRM, qui présente des similitudes avec l'OCT.

**De ISTA à FISTA** : Le problème d'optimisation rencontré est le suivant

$$\mathbf{A}\boldsymbol{\varsigma} = \mathbf{b} \quad (5.8)$$

avec  $\mathbf{A}$ , une matrice de mesures,  $\mathbf{b}$ , les mesures connues et  $\boldsymbol{\varsigma}$  le vecteur à reconstruire. Le modèle standard d'optimisation pour estimer  $\boldsymbol{\varsigma}$  est de la forme :

$$\min_{\boldsymbol{\varsigma}} (f(\boldsymbol{\varsigma}) + g(\boldsymbol{\varsigma})) \quad (5.9)$$

où  $f$  et  $g$  sont des fonctions qui dépendent de la méthode de résolution utilisée. Par exemple ce modèle appliqué aux moindres carrés donne :

$$\boldsymbol{\varsigma}_{LS} = \min_{\boldsymbol{\varsigma}} \|\mathbf{A}\boldsymbol{\varsigma} - \mathbf{b}\|_2^2 \quad (5.10)$$

$$\text{avec } f(\boldsymbol{\varsigma}) = \|\mathbf{A}\boldsymbol{\varsigma} - \mathbf{b}\|_2^2 \text{ et } g(\boldsymbol{\varsigma}) = 0 \quad (5.11)$$

En ajoutant une régularisation de Tikhonov pour contraindre la norme de  $\boldsymbol{\varsigma}$ , le modèle devient :

$$\boldsymbol{\varsigma}_{TIK} = \min_{\boldsymbol{\varsigma}} \|\mathbf{A}\boldsymbol{\varsigma} - \mathbf{b}\|_2^2 + \|\mathbf{\Gamma}\boldsymbol{\varsigma}\|_2^2 \quad (5.12)$$

$$\text{avec } f(\boldsymbol{\varsigma}) = \|\mathbf{A}\boldsymbol{\varsigma} - \mathbf{b}\|_2^2 \text{ et } g(\boldsymbol{\varsigma}) = \|\mathbf{\Gamma}\boldsymbol{\varsigma}\|_2^2 \quad (5.13)$$

où  $\mathbf{\Gamma}$ , appelée matrice de Tikhonov, est souvent égale à la matrice identité.

En acquisition comprimée, la norme de  $\boldsymbol{\varsigma}$  est également régularisée en utilisant cette fois la norme  $l_1$ . Le modèle devient :

$$\boldsymbol{\varsigma}_{l1} = \min_{\boldsymbol{\varsigma}} \|\mathbf{A}\boldsymbol{\varsigma} - \mathbf{b}\|_2^2 + \|\boldsymbol{\varsigma}\|_1 \quad (5.14)$$

$$\text{avec } f(\boldsymbol{\varsigma}) = \|\mathbf{A}\boldsymbol{\varsigma} - \mathbf{b}\|_2^2 \text{ et } g(\boldsymbol{\varsigma}) = l\|\boldsymbol{\varsigma}\|_1 \quad (5.15)$$

Formellement, les conditions nécessaires pour appliquer un algorithme de type "*Iterative Shrinkage-Thresholding*" (ISTA) (Daubechies et al., 2004) sont les suivantes :

- $f + g$  a un minimum  $\boldsymbol{\varsigma}^*$  dans  $\mathbb{R}^n$ .
- $f$  et  $g$  sont convexes.
- $f$  est lisse et différentiable.
- $g$  est connue et  $f$  peut être approximée par une fonction du 1<sup>er</sup> ordre.

ISTA utilise alors un processus itératif dont l'étape principale est :

$$\boldsymbol{\varsigma}_{k+1} = \text{prox}_g(\boldsymbol{\varsigma}_k - t_k \nabla f(\boldsymbol{\varsigma}_k)) \quad (5.16)$$

Dans le cas d'une régularisation en norme  $l_1$  définie en (5.14) :

- $\text{prox}_g$  est une fonction de seuillage de type  $T_a(\boldsymbol{\varsigma}) = (|\boldsymbol{\varsigma}_i| - a) * \text{signe}(\boldsymbol{\varsigma}_i)$
- $t_k \in [0, 1/\|\mathbf{A}^\top \mathbf{A}\|]$  est un pas de gradient.
- $\nabla f(\boldsymbol{\varsigma}) = \mathbf{A}^\top (\mathbf{A}\boldsymbol{\varsigma} - \mathbf{b})$ .

La méthode de gradient accélérée de Nesterov (Nesterov et al., 2007) permet de rendre l'algorithme ISTA plus rapide. L'algorithme obtenu est appelé "*Fast Iterative Shrinkage-Thresholding Algorithm*" (FISTA) (voir **Algorithme 1**). Alors que ISTA a une complexité de  $O(1/k)$ , FISTA possède une complexité de  $O(1/k^2)$ . Cette accélération de Nesterov introduit un terme  $\boldsymbol{\vartheta}$  qui prend en compte les gradients de l'étape précédente dans le calcul de la solution courante, permettant une convergence plus rapide vers notre solution  $k$ -sparse désirée.

**Données :**  $f, prox_g, \varsigma_0$   
**Résultat :**  $\varsigma$   
 $\vartheta_1 = \varsigma_0, t_1 = 1;$   
**Tant que**  $\varsigma_{k+1} - \varsigma_k > \epsilon$   
     $\varsigma_{k+1} = prox_g(\vartheta_k - t_k \nabla f(\vartheta_k))$   
     $t_{k+1} = \frac{1 + \sqrt{1 + 4t_k^2}}{2}$   
     $\vartheta_{k+1} = \varsigma_k + \frac{t_k - 1}{t_{k+1}}(\varsigma_k - \varsigma_{k-1})$   
**Fin**

Algorithme 1 : FISTA

### 5.1.3 Acquisition comprimée en imagerie

L'acquisition comprimée a été utilisée pour les systèmes radar (Baraniuk et Steeghs, 2007), ou la séparation de sources sonores (Xu et Wang, 2009) mais c'est en imagerie qu'elle trouve le plus d'applications notamment en imagerie médicale 3D. Une des contributions les plus impressionnantes a été réalisée à l'Université de Rice par Richard Baraniuk et ses collègues (Duarte et al., 2008). Il s'agit d'une "Single pixel caméra" (SPC). Cette caméra, composée d'un seul détecteur de photons permet de reconstruire des images avec très peu de mesures. La FIGURE 5.3 (a) montre l'image obtenue en photographiant un ballon de football avec une caméra traditionnelle dont la résolution est de  $64 \times 64$  pixels (soit 4096 pixels) et la FIGURE 5.3 (b) montre l'image reconstruite par acquisition comprimée à partir de seulement 1600 mesures aléatoires prises avec la caméra SPC. Cette technique a été améliorée récemment pour aboutir à l'acquisition comprimée ultrarapide (Satat et al., 2017).

L'acquisition comprimée a également été utilisée pour le traitement d'images sismiques. Ces images sont en effet parcimonieuses dès qu'une transformée en curvelets leur est appliquée. Ce type de données étant de hautes dimensions (deux pour la source, deux pour le récepteur et une dimension pour le temps) et souvent incomplètes, l'acquisition comprimée est par conséquent très largement utilisée dans ce domaine (Herrmann et Hennenfent, 2008; Herrmann et al., 2012).

Enfin, la troisième modalité d'image ayant largement bénéficié de la théorie de l'acquisition comprimée est l'imagerie médicale, notamment l'IRM grâce au travail de Lustig, Dohono et Pauly dès 2007 (Lustig et al., 2007). Le travail de Vasanawala et ses collègues témoigne de 2 ans d'expérimentations de l'acquisition comprimée en pédiatrie (Vasanawala et al., 2011). De nombreuses méthodes de reconstructions ont été explorées par la suite, comme des algorithmes de minimisation non convexes (Qu et al., 2010a), des techniques de Bregman (Smith et al., 2012) ou de seuillages itératifs (Qu et al., 2010b). Certains travaux, plus récents, utilisent l'algorithme FISTA en imagerie IRM (Pejoski et al., 2015; Huang et al., 2017). Des recherches sur l'acquisition comprimée avec l'imagerie ultrasonore ont également été menées, notamment par les équipes de l'IRIT à Toulouse (Quinsac et al., 2012) et du Creatis à Lyon (Liebgott et al., 2012).



(a) Image d'un ballon de football, prise par une caméra conventionnelle avec une résolution de  $64 \times 64$  pixels.



(b) Le même ballon, reconstruit à partir de 1600 mesures aléatoires prises par une caméra SPC.

FIGURE 5.3 – Comparaison entre l'image fournie par une caméra conventionnelle et celle obtenue par une caméra SPC (Baraniuk, 2007).

C'est en 2010 que les premiers travaux utilisant l'acquisition comprimée pour l'OCT sont apparus. Lebed et ses collègues ont utilisé l'algorithme de type seuillage itératif développé par Ingrid Daubechies (Daubechies et al., 2004) avec des masques d'acquisition en quadrillage pour la prise des mesures et la base des ondelettes pour la représentation parcimonieuse des images (Lebed et al., 2010). La même année, Liu et ses collègues ont utilisé un masque aléatoire et une méthode de gradient conjugué (Liu et Kang, 2010). Les travaux de Fang et ses collègues ont pris une toute autre direction en construisant une base parcimonieuse à partir d'un jeu de données et en utilisant également un système de segmentation d'image (Fang et al., 2013, 2017).

## 5.2 Application à l'imagerie OCT

Comme évoqué précédemment, l'OCT est une technique d'imagerie médicale non-invasive suscitant un intérêt clinique croissant, notamment pour la réalisation de diagnostics in-situ. Cependant un des inconvénients de l'OCT est le temps d'acquisition trop lent, en particulier pour les volumes OCT 3D. Des améliorations dans la fréquence d'acquisition des images ainsi que dans le transfert des données seraient une véritable avancée en OCT (une image 3D OCT pouvant avoir un poids de plusieurs mégabytes). La théorie de l'acquisition comprimée est par conséquent une solution très intéressante pour résoudre cette problématique.

L'utilisation de l'acquisition comprimée permettrait de réaliser des biopsies 3D beaucoup plus rapidement et également de réduire les éventuels artefacts dus au temps d'acquisition. Pour analyser un point  $P_1$  puis un point  $P_2$ , la conception de l'appareil OCT impose la nécessité de balayer tous les points situés sur la trajectoire entre  $P_1$  et  $P_2$  par les faisceaux laser du dispositif OCT. L'idée de faire des acquisitions partielles de données selon une trajectoire du faisceau laser en spirale nous est apparue comme la manière la plus rapide pour couvrir une surface donnée tout en respectant les contraintes cinématiques du système de balayage de l'appareil OCT, qui repose sur l'actionnement de miroirs galvanométriques. Néanmoins, la théorie de l'acquisition comprimée préconise l'utilisation de masques aléatoires pour l'acquisition des données. Le travail de Vasanawala a montré qu'un masque respectant une distribution de Poisson permettait une reconstruction précise des données manquantes à l'acquisition (Vasanawala et al., 2011). La question que nous nous sommes posée ici était de savoir si un masque en spirale pouvait donner des résultats aussi précis qu'un masque aléatoire. Pour cela, nous avons implémenté un algorithme dérivé de FISTA et comparé ces deux types de masques sur des images numériques d'échantillons typiquement obtenues par l'OCT. Nous avons appliqué notre algorithme avec trois bases dans lesquelles le signal est parcimonieux : les ondelettes, les curvelets et les shearlets. Ces recherches ont fait l'objet d'une publication (Dufflot et al., 2016b).

### 5.2.1 Nouvelle formulation de l'algorithme FISTA

Considérons  $\mathbf{b}$  comme étant le vecteur de mesures OCT,  $\Phi$  la matrice des coordonnées spatiales de  $\mathbf{b}$ ,  $\Psi$  la base dans laquelle le signal est parcimonieux. Le problème d'optimisation à résoudre est alors de la forme suivante :

$$\varsigma_{l1} = \min_{\varsigma} \|\Phi\varsigma - \mathbf{b}\|_2^2 + \|\Psi\varsigma\|_1 \quad (5.17)$$

avec  $\|\Phi\varsigma - \mathbf{b}\|_2^2$  représentant la correspondance entre l'inconnue  $\varsigma$  et les mesures  $\mathbf{b}$  et  $\|\Psi\varsigma\|_1$  représentant la parcimonie de  $\varsigma$  dans la base  $\Psi$ . Selon la formulation de FISTA,  $f(\varsigma)$  est pris comme étant  $\|\Phi\varsigma - \mathbf{b}\|_2^2$  et  $g(\varsigma)$  est définie comme  $\|\Psi\varsigma\|_1$ . En suivant l'Algorithme 1,  $prox_g$  et  $\nabla f$  sont définis comme ci-dessous :

- $prox_g = T_a(\varsigma) = (|\varsigma_i| - a) \times \text{signe}(\varsigma_i)$ , une fonction seuil de paramètre  $a$ .
- $\nabla f(\varsigma) = (\Phi\Psi^{-1})^T(\Phi\Psi^{-1}\varsigma - \mathbf{b})$  où  $\Psi^{-1}$  est la transformée inverse de  $\Psi$ .

Dans notre cas, pour accélérer l'algorithme et pour tester différentes bases où le signal est parcimonieux (puisque la matrice  $\Psi$  est implémentée sous forme de fonction pour les transformées en curvelets et les shearlets,  $\nabla f$  n'est pas calculable en tant que matrice), nous avons modifié l'algorithme initial FISTA en effectuant un changement sur  $prox_g$  et  $\nabla f$ . Ces changements sont les suivants :

- $prox_g = \Psi^{-1}T_a(\Psi(\varsigma))$

- $\nabla f(\boldsymbol{\varsigma}) = \boldsymbol{\Phi}^T(\boldsymbol{\Phi}\boldsymbol{\varsigma} - \mathbf{b})$

Notre algorithme, appelé  $\mu$ -FISTA est présenté ci-dessous (voir **Algorithme 2**).

**Données :**  $\mathbf{A}$ ,  $a$ ,  $\boldsymbol{\varsigma}_0$

**Résultat :**  $\boldsymbol{\varsigma}$

$\boldsymbol{\vartheta}_1 = \boldsymbol{\varsigma}_0, t_1 = 1;$

**Tant que**  $\boldsymbol{\varsigma}_{k+1} - \boldsymbol{\varsigma}_k > \epsilon$

$$\left| \begin{array}{l} \boldsymbol{\varsigma}_k = \boldsymbol{\Psi}^{-1}T_a(\boldsymbol{\Psi}(\boldsymbol{\vartheta}_k - t_k\boldsymbol{\Phi}^T(\boldsymbol{\Phi}\boldsymbol{\vartheta}_k - \mathbf{b}))) ; \\ t_{k+1} = \frac{1+\sqrt{1+4t_k^2}}{2} ; \\ \boldsymbol{y}_{k+1} = \boldsymbol{\varsigma}_k + \frac{t_k-1}{t_{k+1}}(\boldsymbol{\varsigma}_k - \boldsymbol{\varsigma}_{k-1}); \end{array} \right.$$

**Fin**

**Algorithme 2 :**  $\mu$ -FISTA

### 5.2.2 Validation numérique

Afin de comparer quantitativement les différentes bases parcimonieuses, nous utiliserons deux métriques pour calculer la similarité entre l'image de référence  $\mathbf{I}_r$  et l'image reconstruite  $\mathbf{I}_c$  avec  $\mu$ -FISTA. Le premier est le PSNR.

**Définition 5.4. Peak Signal Noise Ratio (PSNR) :** Soient deux images  $\mathbf{I}_r$  et  $\mathbf{I}_c$  de longueur  $N$ , alors le PSNR entre ces deux signaux est donné par :

$$PSNR = 10 * \log_{10}\left(\frac{255^2}{MSE}\right) \quad \text{avec} \quad MSE = \frac{(\mathbf{I}_c - \mathbf{I}_r)^2}{N} \quad (5.18)$$

Comparer les deux images  $\mathbf{I}_c$  et  $\mathbf{I}_r$  simplement avec leur signal ne donne pas toujours une bonne visualisation de leur similarité morphologique. C'est pourquoi, nous avons également considéré une métrique appelée "*Structural SIMilarity*" (SSIM) dont la valeur est située entre -1 et 1.

**Définition 5.5. Structural SIMilarity (SSIM) :** Soient deux signaux  $\mathbf{I}_r$  et  $\mathbf{I}_c$  de longueur  $N$ , alors le SSIM entre ces deux signaux est donné par :

$$SSIM = [\text{ss}(\mathbf{I}_c, \mathbf{I}_r)] \cdot [\text{ls}(\mathbf{I}_c, \mathbf{I}_r)] \cdot [\text{cs}(\mathbf{I}_c, \mathbf{I}_r)] \quad (5.19)$$

avec ss, ls et cs, la similarité de structure, de luminance et de contraste entre les deux images. Ainsi, (5.19) peut être reformulée comme suit :

$$SSIM = \frac{(2\mathbf{m}_{\mathbf{I}_c}\mathbf{m}_{\mathbf{I}_r} + \mathbf{c}_1)(2\mathbf{v}_{\mathbf{I}_c}\mathbf{v}_{\mathbf{I}_r} + \mathbf{c}_2)(2\text{cov}_{\mathbf{I}_c\mathbf{I}_r} + \mathbf{c}_3)}{(\mathbf{m}_{\mathbf{I}_c}^2 + \mathbf{m}_{\mathbf{I}_r}^2 + \mathbf{c}_1)(\mathbf{v}_{\mathbf{I}_c}^2 + \mathbf{v}_{\mathbf{I}_r}^2 + \mathbf{c}_2)(\mathbf{v}_{\mathbf{I}_c}\mathbf{v}_{\mathbf{I}_r} + \mathbf{c}_3)} \quad (5.20)$$

avec

- $\mathbf{m}_{\mathbf{I}}$  la moyenne de  $\mathbf{I}$ .

- $v_{\mathbf{I}}^2$  la variance de  $\mathbf{I}$ .
- $\text{cov}_{\mathbf{I}_1\mathbf{I}_2}$  la covariance de  $\mathbf{I}_1$  et  $\mathbf{I}_2$
- $c_1, c_2$  et  $c_3$  des variables servant à stabiliser la division.

Une comparaison des résultats fournis par l'algorithme  $\mu$ -FISTA avec différentes bases parcimonieuses et l'image reconstruite avec une simple interpolation bilinéaire a été réalisée. Ainsi, six images de taille  $256 \times 256$  pixels obtenues respectivement par l'imagerie OCT d'un morceau de raisin, de kiwi, de banane, de poulet et de deux échantillons de tissus biologiques (voir FIGURE 5.4) ont été reconstruites à partir des données acquises. Pour évaluer notre procédure de reconstruction, un masque a été appliqué à l'image OCT de référence, afin de simuler une acquisition partielle des données. La qualité de la reconstruction a ensuite été mesurée par les deux métriques précédemment présentées, à savoir le PSNR et le SSIM.

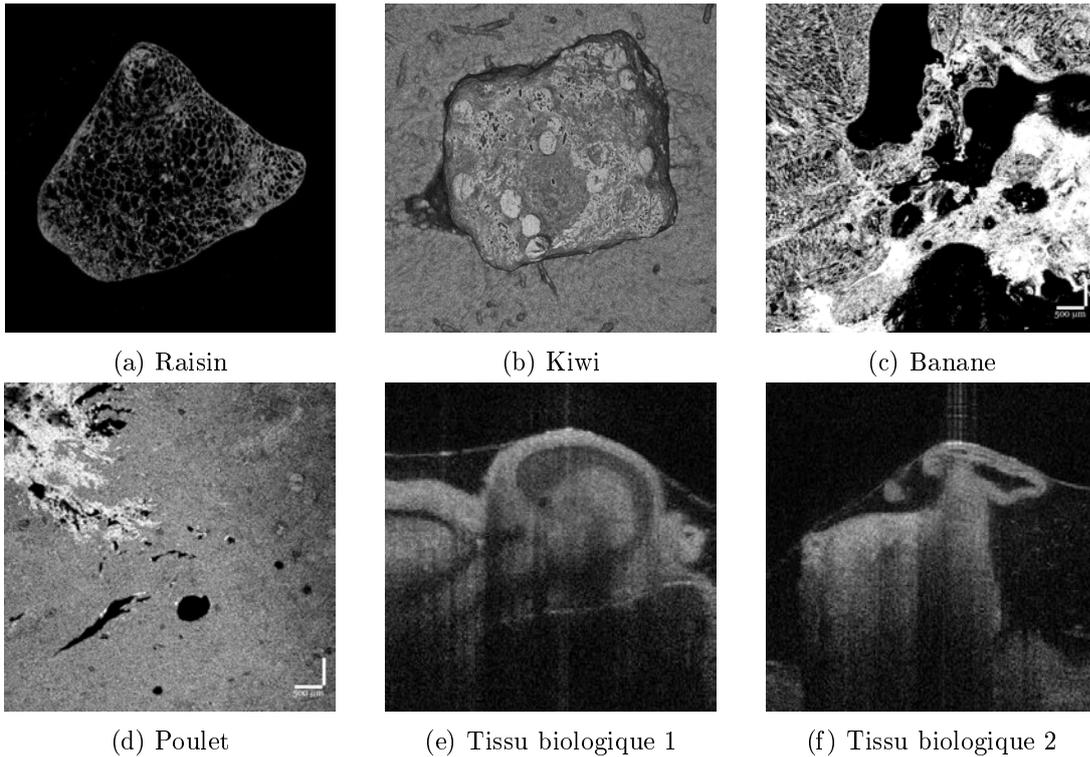


FIGURE 5.4 – Séquence d'images OCT de différents échantillons utilisés pour la validation de la méthode d'optimisation  $\mu$ -FISTA.

La théorie de l'acquisition comprimée nous suggère d'utiliser un masque aléatoire pour obtenir les données (voir FIGURE 5.5(a)). Cependant, ce choix s'avère difficile à réaliser dans le cas de l'OCT en raison de la cinématique des miroirs de balayage. De ce fait, nous avons comparé le masque aléatoire avec un masque en spirale (voir FIGURE 5.5

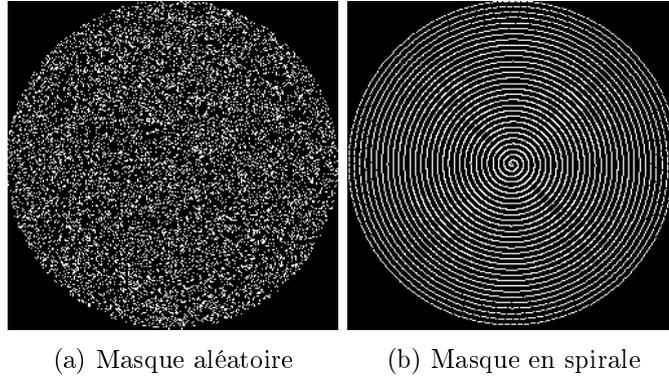


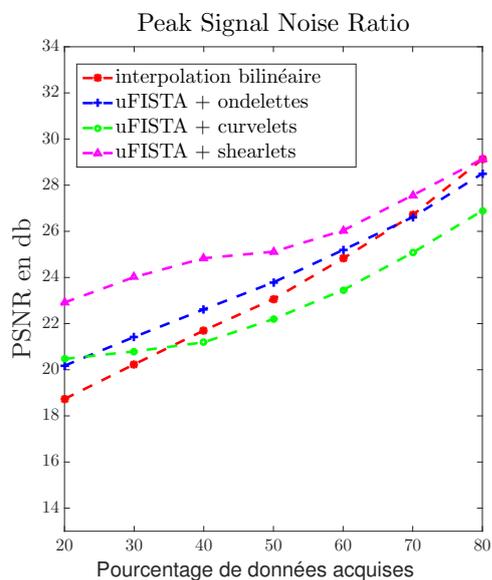
FIGURE 5.5 – Masques utilisés pour simuler l’acquisition partielle de données (20% d’échantillons sont utilisés pour la reconstruction).

(b)). Ce type de masque est compatible avec l’appareil OCT et permet d’acquérir de nombreuses données dans une zone définie dans un temps minimum.

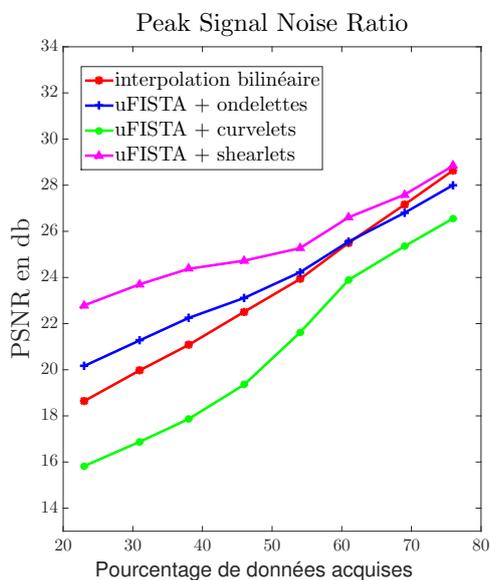
Les FIGURES 5.6 et 5.7 montrent respectivement la moyenne des PSNR et la moyenne des SSIM pour nos échantillons qui sont obtenus pour plusieurs types de reconstruction (interpolation bilinéaire ou bien  $\mu$ -FISTA avec ondelettes, curvelets ou shearlets comme base parcimonieuse) avec un masque aléatoire (lignes en pointillées) ou un masque en spirale (lignes continues). Différents pourcentages d’échantillonnage ont été testés et comparés.

Un masque en spirale réduit la qualité de notre reconstruction avec les curvelets, mais avec les deux autres bases, à savoir les ondelettes et les shearlets, ce masque montre les mêmes performances que le masque aléatoire. La reconstruction obtenue avec les shearlets permet d’obtenir le meilleur PSNR quels que soient le masque utilisé et le pourcentage de données acquises. A contrario, les curvelets présentent le PSNR le plus faible pour les deux types de masques utilisés, mais l’écart avec les autres méthodes est d’autant plus important lorsqu’un masque en spirale est utilisé. La FIGURE 5.7 permet de distinguer plus aisément les résultats des différentes méthodes. En effet, l’utilisation de l’algorithme  $\mu$ -FISTA avec les shearlets donne les meilleurs résultats dès que l’on dispose de moins de 50 % de données. A partir de 60 % de données acquises, l’interpolation bilinéaire donne les meilleurs scores au SSIM. Ces scores légèrement plus élevés que ceux de l’approche utilisant les shearlets sont parfaitement compréhensibles, car avec peu de données manquantes, une interpolation bilinéaire permet une bonne reconstruction de l’image.

Ainsi, l’étude des courbes PSNR et SSIM préconise de manière non-équivoque l’utilisation de l’algorithme  $\mu$ -FISTA avec la transformée en shearlets quel que soit le type de masque d’acquisition, mais plus encore, elle montre qu’avec cette transformée l’utilisation d’un masque en spirale n’entraîne aucune dégradation de la qualité de la reconstruction de l’image.

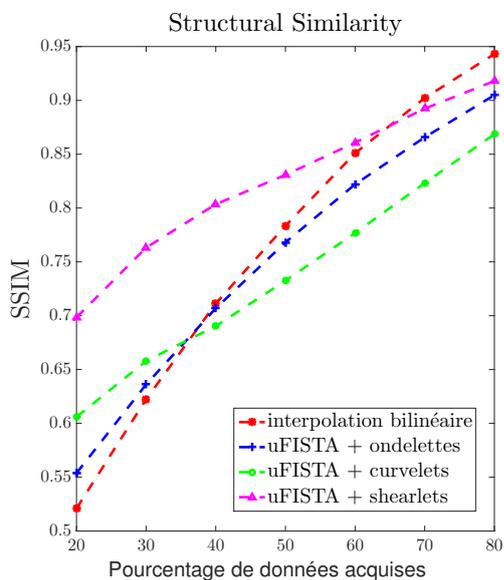


(a) Masque aléatoire

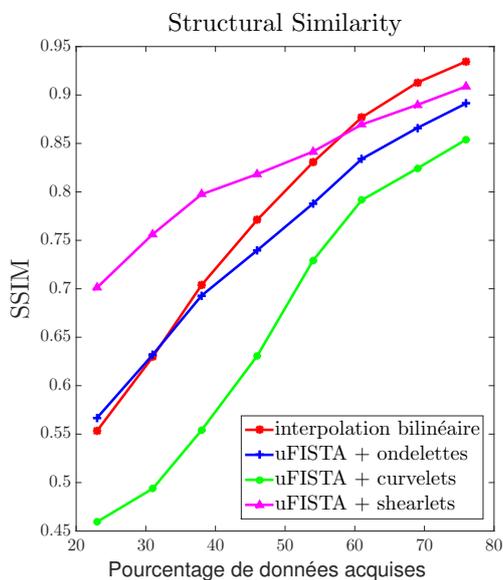


(b) Masque en spirale

FIGURE 5.6 – PSNR pour différents pourcentages de données acquises, avec des masques aléatoires et en spirale.



(a) Masque aléatoire



(b) Masque en spirale

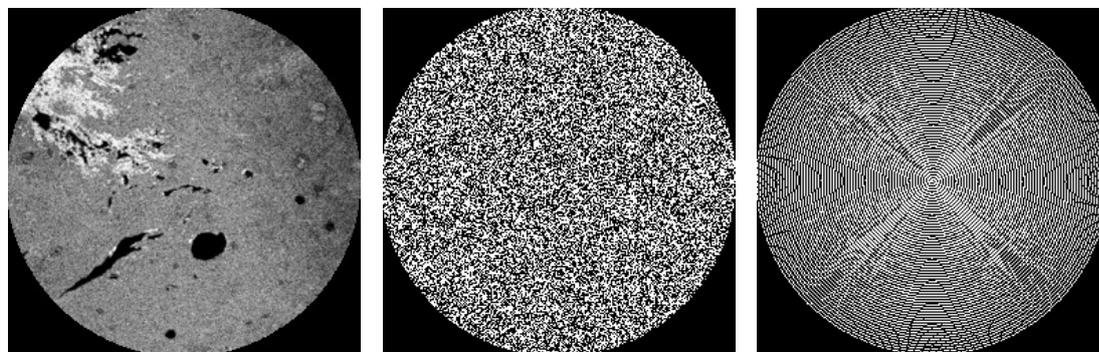
FIGURE 5.7 – SSIM pour différents pourcentages de données acquises, avec des masques aléatoires et en spirale.

Les FIGURES 5.9, 5.10, 5.12 et 5.13 présentent quelques exemples de reconstruction pour les deux masques (aléatoire et en spirale) pour différentes valeurs d'échantillonnage et différentes images. Avec les deux masques, la reconstruction utilisant une interpolation bilinéaire induit des artefacts structurels particulièrement visibles sur les FIGURES 5.12(a) et 5.13(a).

Avec l'algorithme  $\mu$ -FISTA, le choix de la transformée permettant de rendre notre signal parcimonieux affecte beaucoup la qualité de reconstruction de l'image :

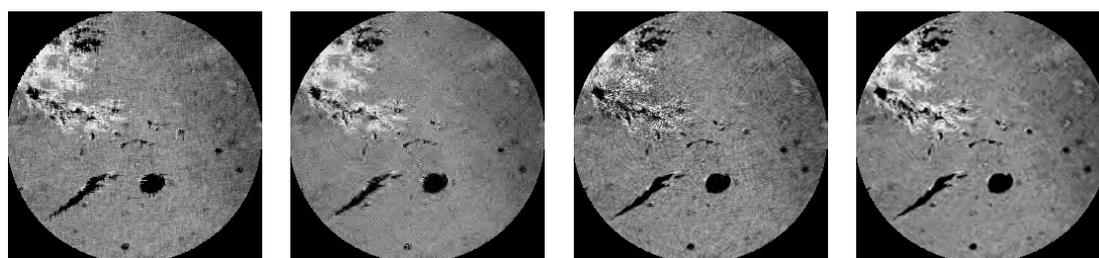
- Avec les ondelettes, certains détails importants de l'image reconstruite sont manquants, ce qui est particulièrement visible dans l'exemple du kiwi où certaines tâches sur l'échantillon n'apparaissent plus (voir FIGURES 5.12(b) et 5.13(b)).
- Avec les curvelets,  $\mu$ -FISTA fournit une reconstruction proche de l'image de départ avec un masque aléatoire. Cette qualité de reconstruction est particulièrement visible sur la FIGURE 5.9(c). Cependant, cette base n'est pas adaptée au masque en spirale. En effet, sur les FIGURES 5.10(c) et 5.13(c), nous pouvons voir apparaître les spirales du masque au sein de l'image reconstruite. Cela est probablement dû au fait que la transformée en curvelets intègre des rotations. Le masque en spirale peut perturber le passage du signal dans cette base.
- Avec les shearlets, que ce soit avec l'échantillon de poulet ou celui de kiwi, toute l'information est conservée dans la reconstruction. Les FIGURES 5.12(d) et 5.13(d) ne présentent pas de tâches manquantes comme cela était le cas pour les reconstructions utilisant les ondelettes ou les curvelets. En revanche, de faibles différences peuvent être remarquées entre le masque aléatoire et le masque en spirale dans le cas de l'échantillon de poulet. En effet, de légers artefacts sont présents sur parties les plus texturées de la FIGURE 5.10(d).

Ces observations obtenues à l'œil nu sont bien en adéquation avec les résultats des courbes de PSNR (voir FIGURE 5.6) et de SSIM (voir FIGURE 5.7).



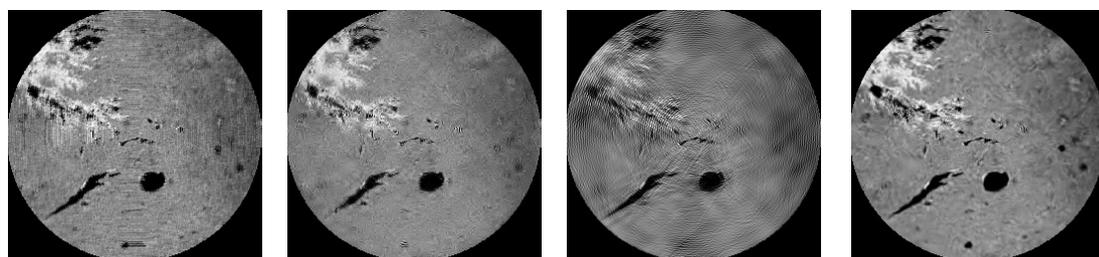
(a) Image de référence. (b) Masque aléatoire avec 50% de données. (c) Masque en spirale avec 46% de données.

FIGURE 5.8 – Image de référence et masques utilisés pour échantillon du poulet.



(a) Interpolation bilinéaire (b)  $\mu$ -FISTA avec ondelettes (c)  $\mu$ -FISTA avec curvelets (d)  $\mu$ -FISTA avec shearlets

FIGURE 5.9 – Résultats des reconstructions sur l'échantillon de poulet avec un masque aléatoire et 50% de données.



(a) Interpolation bilinéaire (b)  $\mu$ -FISTA avec ondelettes (c)  $\mu$ -FISTA avec curvelets (d)  $\mu$ -FISTA avec shearlets

FIGURE 5.10 – Résultats des reconstructions sur l'échantillon de poulet avec un masque spirale et 46% de données.

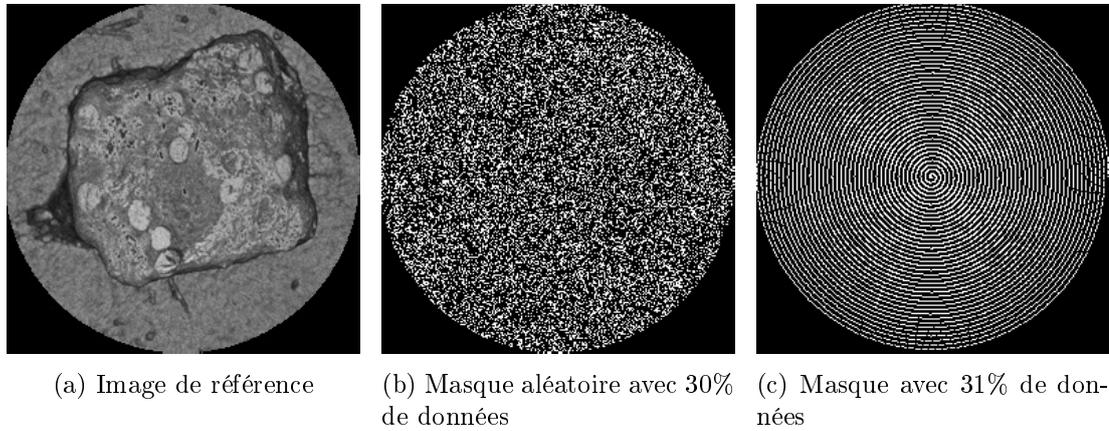


FIGURE 5.11 – Image de référence et masques utilisés pour l'échantillon du kiwi.

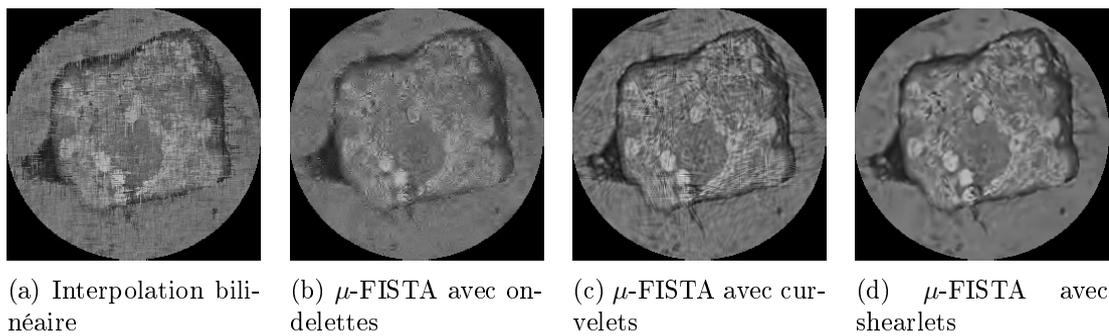


FIGURE 5.12 – Résultats des reconstructions sur l'échantillon de kiwi avec un masque aléatoire et 30% de données.

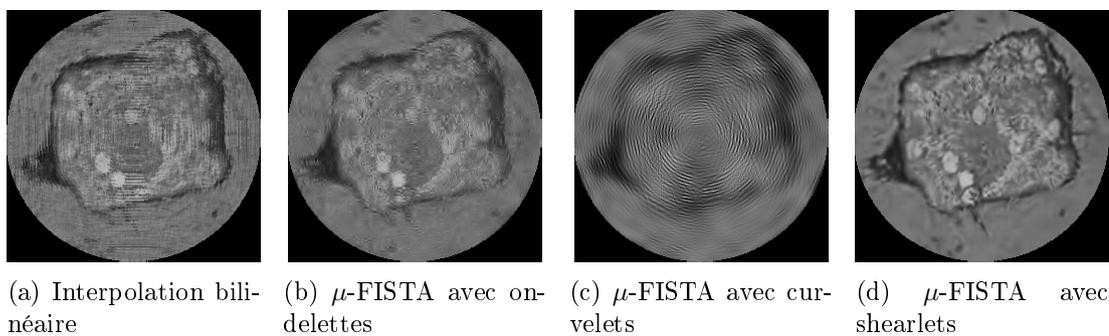


FIGURE 5.13 – Résultats des reconstructions sur l'échantillon de kiwi avec un masque spirale et 31% de données.

### 5.3 Conclusion

Dans ce chapitre, nous avons présenté une étude préliminaire de l'utilisation de l'acquisition comprimée en imagerie OCT. Nous avons modifié l'algorithme FISTA pour l'adapter à notre problématique. Notre algorithme, appelé  $\mu$ -FISTA, a permis de comparer qualitativement et quantitativement plusieurs bases dans lesquelles notre image de départ est parcimonieuse (constituée d'un grand nombre de zéros). Ces comparaisons ont été réalisées à l'aide de deux métriques : le PSNR et le SSIM. Elles ont montré que la base des shearlets offrait les meilleures performances en terme de qualité de reconstruction. L'observation des différentes reconstructions à l'œil nu confirme les résultats obtenus à l'aide des courbes de PSNR et de SSIM de chaque méthode. L'appareil OCT ne permettant pas de réaliser de masque aléatoire d'acquisition de mesures, préconisé dans la littérature, notre méthode a permis également de simuler l'utilisation d'un masque en spirale. En effet, ce dernier est davantage en adéquation avec la cinématique de l'appareil. Ainsi, l'utilisation d'une trajectoire en spirale pour effectuer l'acquisition des données OCT est compatible avec l'utilisation de l'algorithme  $\mu$ -FISTA et n'affecte que très peu la qualité de reconstruction de l'image en utilisant la transformée en shearlets. Une perspective intéressante à ce travail serait l'utilisation simultanée de l'acquisition comprimée et de l'asservissement visuel avec cette modalité d'imagerie.

# Conclusion

Depuis une dizaine d'années, l'utilisation de robots s'est développée dans de nombreux domaines. Le monde de l'industrie utilise des robots depuis les années 1960 pour effectuer des tâches élémentaires mais difficiles pour l'Homme (soulever des objets très lourds, travailler dans un environnement hostile, agir à haute cadence). Cependant, les nombreux progrès effectués dans le domaine de l'informatique avec des capacités de calculs toujours plus puissantes, et ce même pour des technologies embarquées, tendent à faire réaliser aux robots des tâches de plus en plus complexes. Ainsi, on retrouve des robots dans de nouveaux domaines comme la robotique domestique (exemple : Roomba de iRobot), la robotique militaire (exemple : Drone predator de l'US Air Force) ou les voitures autonomes (exemple : Google Car). La robotique médicale s'est également largement développée avec des robots interventionnels (exemple : DaVinci de Intuitive Surgical) ou des robots pour la visualisation comme les porte-endoscopes servant à assister le médecin durant ses interventions grâce à un retour visuel (qui peut être multimodal) de la zone à opérer. Quel que soit le domaine d'application, la commande de tels robots nécessite d'interpréter les signaux fournis par les différents capteurs (capteurs de lumière, de son, d'ultrasons ou de force). Le travail présenté dans ce manuscrit s'est intéressé à l'utilisation d'informations visuelles afin de contrôler un système robotique, en considérant plusieurs types de capteurs : une caméra standard, une sonde échographique et un système d'imagerie OCT.

L'objectif principal de cette thèse était l'élaboration de nouvelles approches d'asservissement visuel fondées sur les représentations multi-échelles telles que les ondelettes et les shearlets. Le chapitre 1 a rappelé les principes de la vision par ordinateur et a dressé un bref inventaire des différentes méthodes d'asservissement visuel existantes. Le chapitre 2 a introduit les représentations multi-échelles, en particulier la transformée en ondelettes et une de ses extensions : la transformée en shearlets. Dans les deux cas, la distinction entre la transformée sous-échantillonnée (possédant autant de coefficients que de pixels dans l'image initiale) et sa variante non sous-échantillonnée (possédant jusqu'à 10 fois plus de coefficients que de pixels dans l'image initiale) a été soulignée. Les chapitres 3 et 4 ont discuté les contributions scientifiques de cette thèse, en l'occurrence le développement de plusieurs lois de commande dont le signal est le résultat d'une transformée d'ondelettes ou de shearlets (plusieurs variantes ont été proposées). Enfin le chapitre 5 s'est intéressé à l'utilisation des représentations multi-échelles en acquisition comprimée dans le but de pouvoir réaliser un asservissement visuel sur des

données incomplètes obtenues par des méthodes d'acquisition comprimée, même si cela reste une perspective à long terme.

## Contributions

Le chapitre 3 a montré les premiers résultats prometteurs de l'utilisation des shearlets en asservissement visuel échographique. Dans un premier temps, une loi de commande utilisant une matrice d'interaction estimée numériquement est présentée. Cette matrice relie la variation temporelle des coefficients des shearlets et la vitesse spatiale de la sonde échographique. Ce travail, a ensuite été étendu par le calcul d'une nouvelle matrice d'interaction qui est cette fois-ci analytique. Néanmoins, cette dernière n'est applicable qu'avec une version non sous-échantillonnée de la transformée en shearlets. Cette première formulation a permis de montrer des performances intéressantes en termes de précision et de robustesse de la loi de commande qui en découle, notamment par rapport aux conditions de travail non-favorables qui caractérisent l'imagerie ultrasonore. Cependant, le fait de ne pas pouvoir utiliser de représentations sous-échantillonnées constituait un inconvénient majeur de cette méthode.

Dans le chapitre 4, l'attention a été portée sur le formalisme mathématique des nouvelles primitives visuelles étudiées et par conséquent, la mise en œuvre de matrices d'interaction analytique associées. Ces matrices d'interaction ont été formalisées en utilisant quatre types d'informations visuelles multi-échelles différentes, à savoir les coefficients des ondelettes sous-échantillonnées ou non et ceux des shearlets, également sous-échantillonnées ou non. Cette loi de commande est un asservissement visuel direct évitant les étapes de détection, d'extraction, de suivi et de mise en correspondance de points d'intérêt qui incombent aux approches géométriques. Néanmoins, la décomposition du signal dans l'espace des ondelettes ou celui des shearlets permet une représentation de l'image qui met l'accent sur ses singularités (comme les contours, les points d'intérêt ou les zones texturées), en les associant à des coefficients de valeur élevée tandis que les zones de l'image plus uniformes sont caractérisées par des coefficients proches de 0. Plusieurs scénarios de simulation ont pu mettre en valeur la supériorité des approches sous-échantillonnées sur celles non sous-échantillonnées en termes de précision, de rapidité et de robustesse. Enfin, des résultats expérimentaux ont permis de valider les lois de commande d'asservissement visuel proposées pour la tâche de positionnement d'une caméra CCD embarquée sur un robot à 6 ddl. Si la méthode utilisant la transformée en ondelettes est plus robuste aux occultations, celle utilisant les shearlets semble meilleure en terme de robustesse aux changements d'illumination. Enfin, la méthode utilisant les ondelettes sous-échantillonnées a été validée sur une autre modalité d'imagerie, l'OCT, montrant des résultats intéressants pour ce type d'images très bruitées.

Dans le chapitre 5, nous avons testé en simulation un algorithme d'acquisition comprimée appliqué à l'imagerie OCT. L'acquisition parcimonieuse des données (environ 30% des données sont considérées pour la reconstruction d'une image entière) est effectuée suivant des trajectoires de balayage spécifiques compatibles avec la cinématique d'une structure robotique, par exemple, une trajectoire sous forme d'une spirale. Différentes représentations multi-échelles ont été testées : ondelettes, curvelets, shearlets. Il s'est avéré que la transformée en shearlets offre la meilleure reconstruction comparées aux autres méthodes. Ce travail ouvre la voie vers l'association de l'acquisition comprimée à l'asservissement visuel dans la même boucle de commande. En d'autres termes, les tâches de perception et d'action peuvent être combinées de manière originale dans la même boucle.

## Perspectives

### A court terme

L'inconvénient majeur des méthodes présentées dans ce manuscrit demeure le temps requis pour le calcul des coefficients en ondelettes et en shearlets. Dans ce travail, nous avons utilisé des bibliothèques développées sous MATLAB, puisqu'il n'existe pas d'implémentation de la transformée en shearlets en C++. Nous espérons pouvoir obtenir un temps de calcul 10 fois plus rapide en utilisant une des nombreuses implémentations C++ de la transformée en ondelettes. Pour la transformée en shearlets, nous avons commencé la mise en œuvre d'une bibliothèque C++ de la version sous-échantillonnée, mais ce développement est à poursuivre. Les performances gagnées en temps de calcul pourraient nous permettre d'utiliser une matrice d'interaction, calculée à chaque itération à la position courante et non plus seulement à la position désirée.

De plus, nous avons mis en place dans le chapitre 4, une matrice d'interaction approximée, nécessitant moins de transformées en ondelettes ou en shearlets pour sa construction que celle que nous avons testée expérimentalement. L'étude des performances de cette matrice d'interaction est un autre axe de recherche à court terme. L'utilisation de cette matrice réduirait encore le temps d'exécution d'une boucle de commande dans le cadre d'une matrice d'interaction estimée en ligne à chaque itération.

Une autre perspective à court terme serait d'étendre notre commande utilisant la transformée en ondelettes d'une image OCT à 6 ddl en contrôlant également les rotations du robot parallèle supportant l'échantillon à analyser.

### A moyen terme

De futurs travaux pourraient s'intéresser à l'utilisation des valeurs imaginaires de la transformée en ondelettes ou en shearlets afin d'étudier si l'amplitude des valeurs complexes de ces transformées permettrait d'augmenter davantage la robustesse de notre

approche (ce qui est notre intuition). L'étude théorique de la stabilité de la commande est également un autre axe de recherche à mener.

Une autre perspective à moyen terme serait de faire varier les niveaux de décompositions utilisés lors de la commande en commençant par considérer uniquement les niveaux les plus grossiers puis de plus en plus détaillés en se rapprochant de la position désirée. En d'autres termes, utiliser les avantages de la matrice d'interaction multi-échelles pour développer des lois de commandes de plus en plus précises. Ceci permettrait d'augmenter le domaine de convergence sans affecter la précision de la commande.

L'utilisation d'une caméra RGB-D afin de connaître la profondeur des points de l'image est également à envisager. En effet, la méthode proposée dans le chapitre 4 permettait la convergence sur des scènes 3D uniquement dans le cas de déplacements assez faibles.

Enfin, nous pouvons poursuivre ces travaux afin de valider notre méthode pour des tâches de suivi où l'objectif serait de compenser les mouvements physiologiques du patient durant un examen échographique ou l'acquisition d'une biopsie optique à l'aide d'un système de tomographie par cohérence optique. Une étude de la robustesse de la loi de commande utilisant les ondelettes et/ou les shearlets aux déformations observées dans les images échographiques semble également être une perspective prometteuse. L'utilisation du simulateur présenté dans le chapitre 3 associé à une séquence d'images échographiques avec un organe se déformant permettrait de mettre en œuvre cette étude.

## A long terme

La perspective la plus ambitieuse de ce travail serait d'effectuer un asservissement visuel à partir de la transformée en shearlets de données partiellement acquises, sans passer par une reconstruction au préalable de l'image afin d'exploiter au maximum la parcimonie du signal obtenu à l'aide des transformées en ondelettes ou en shearlets. Un travail préliminaire a été fourni sur ce sujet dans le dernier chapitre de cette thèse. Cette perspective est particulièrement intéressante dans le domaine de l'imagerie médicale où les capteurs visuels utilisés nécessitent de réaliser un balayage de la scène pour acquérir l'image. En effet, une diminution du temps d'acquisition d'une image échographique 3D ou OCT 3D permettrait de réduire les artefacts sur l'image dus à d'éventuels mouvements du patient.

Enfin, un asservissement visuel multi-modal est envisageable à long terme. En effet, nous avons vu que les transformées multi-échelles étaient particulièrement utilisées dans le domaine de la fusion d'images. Nous pouvons alors imaginer une tâche d'asservissement visuel échographique dans lequel l'image à la position désirée est donnée par une modalité d'imagerie invasive (exemple : scanner à rayon X) ou peu rapide (exemple : IRM).

# Bibliographie

- Abdelkader, H. H., Y. Mezouar, N. Andreff, et P. Martinet  
2005. 2 1/2 d visual servoing with central catadioptric cameras. In *Intelligent Robots and Systems, 2005.(IROS 2005). 2005 IEEE/RSJ International Conference on*, Pp. 3572–3577. IEEE.
- Abolmaesumi, P., S. Salcudean, W.-H. Zhu, S. P. DiMaio, et M. R. Sirouspour  
2001. A user interface for robot-assisted diagnostic ultrasound. In *Proceedings 2001 ICRA. IEEE International Conference on Robotics and Automation (Cat. No.01CH37164)*, volume 2, Pp. 1549–1554 vol.2.
- Abolmaesumi, P., S. E. Salcudean, W.-H. Zhu, M. R. Sirouspour, et S. P. DiMaio  
2002. Image-guided control of a robot for medical ultrasound. *IEEE Transactions on Robotics and Automation*, 18(1) :11–23.
- Agreste, S., G. Andaloro, D. Prestipino, et L. Puccio  
2007. An image adaptive, wavelet-based watermarking of digital images. *Journal of Computational and Applied Mathematics*, 210(1-2) :13–21.
- Ahmaderaghi, B., J. M. Del Rincon, F. Kurugollu, et A. Bouridane  
2014. Perceptual watermarking for discrete shearlet transform. In *Visual Information Processing (EUVIP), 2014 5th European Workshop on*, Pp. 1–6. IEEE.
- Alkhalil, F. et C. Dognon  
2012. Stereo visual servoing with decoupling control. In *Intelligent Robots and Systems (IROS), 2012 IEEE/RSJ International Conference on*, Pp. 1671–1676. IEEE.
- Allen, J.  
1977. Short term spectral analysis, synthesis, and modification by discrete fourier transform. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 25(3) :235–238.
- Allen, P. K., A. Timcenko, B. Yoshimi, et P. Michelman  
1993. Automated tracking and grasping of a moving object with a robotic hand-eye system. *IEEE Transactions on Robotics and Automation*, 9(2) :152–165.
- Andreff, N., B. Espiau, et R. Horaud  
2002. Visual servoing from lines. *The International Journal of Robotics Research*, 21(8) :679–699.

- Argoul, F., A. Arneodo, J. Elezgaray, G. Grasseau, et R. Mureni  
1990. Wavelet analysis of the self-similarity of diffusion-limited aggregates and electrodeposition clusters. *Physical Review A*, 41(10) :5537.
- Bakthavatchalam, M., F. Chaumette, et E. Marchand  
2013. Photometric moments : New promising candidates for visual servoing. In *Robotics and Automation (ICRA), 2013 IEEE International Conference on*, Pp. 5241–5246. IEEE.
- Barajas, M., J. P. Dávalos-Viveros, S. Garcia-Lumbreras, et J. L. Gordillo  
2013. Visual servoing of uav using cuboid model with simultaneous tracking of multiple planar faces. In *2013 IEEE/RSJ International Conference on Intelligent Robots and Systems*, Pp. 596–601.
- Baraniuk, R. et P. Steeghs  
2007. Compressive radar imaging. In *2007 IEEE Radar Conference*, Pp. 128–133.
- Baraniuk, R. G.  
2007. Compressive sensing [lecture notes]. *IEEE signal processing magazine*, 24(4) :118–121.
- Barni, M., F. Bartolini, et A. Piva  
2001. Improved wavelet-based watermarking through pixel-wise masking. *IEEE transactions on image processing*, 10(5) :783–791.
- Bateux, Q. et E. Marchand  
2017. Histograms-based visual servoing. *IEEE Robotics and Automation Letters*, 2(1) :80–87.
- Bateux, Q., E. Marchand, J. Leitner, et F. Chaumette  
2017. Visual servoing from deep neural networks. *arXiv preprint arXiv :1705.08940*.
- Benhimane, S. et E. Malis  
2007. Homography-based 2d visual tracking and servoing. *The International Journal of Robotics Research*, 26(7) :661–676.
- Beylkin, G., R. Coifman, et V. Rokhlin  
1991. Fast wavelet transforms and numerical algorithms i. *Communications on pure and applied mathematics*, 44(2) :141–183.
- Blumensath, T. et M. E. Davies  
2009. Iterative hard thresholding for compressed sensing. *Applied and computational harmonic analysis*, 27(3) :265–274.
- Blumensath, T., M. E. Davies, et G. Rilling  
2012. Greedy algorithms for compressed sensing. *YC Eldar and G. Kutyniok*, Pp. 348–393.

- Bradley, J. N. et C. M. Brislawn  
1994. The wavelet/scalar quantization compression standard for digital fingerprint images. In *Circuits and Systems, 1994. ISCAS'94., 1994 IEEE International Symposium on*, volume 3, Pp. 205–208. IEEE.
- Burt, P. et E. Adelson  
1983. The laplacian pyramid as a compact image code. *IEEE Transactions on communications*, 31(4) :532–540.
- Cai, T. T. et A. Zhang  
2014. Sparse representation of a polytope and recovery of sparse signals and low-rank matrices. *IEEE Transactions on Information Theory*, 60(1) :122–132.
- Candes, E. et D. Donoho  
2000. Curvelets : A surprisingly effective nonadaptive representation for objects with edges. Technical report, Stanford Univ Ca Dept of Statistics.
- Candes, E., J. Romberg, et T. Tao  
2006a. Robust uncertainty principles : exact signal reconstruction from highly incomplete frequency information. *Information Theory, IEEE Transactions on*, 52(2) :489–509.
- Candes, E. et M. Wakin  
2008. An introduction to compressive sampling. *Signal Processing Magazine, IEEE*, 25(2) :21–30.
- Candès, E. J. et B. Recht  
2009. Exact matrix completion via convex optimization. *Foundations of Computational mathematics*, 9(6) :717.
- Candes, E. J., J. K. Romberg, et T. Tao  
2006b. Stable signal recovery from incomplete and inaccurate measurements. *Communications on pure and applied mathematics*, 59(8) :1207–1223.
- Candès, E. et J. Romberg  
2005. *11-magic : Recovery of Sparse Signals via Convex Programming*.
- Chaumette, F.  
2004. Image moments : a general and useful set of features for visual servoing. *IEEE Transactions on Robotics*, 20(4) :713–723.
- Chaumette, F. et S. Hutchinson  
2006. Visual servo control. i. basic approaches. *IEEE Robotics & Automation Magazine*, 13(4) :82–90.
- Chaumette, F. et S. Hutchinson  
2007. Visual servo control. ii. advanced approaches [tutorial]. *IEEE Robotics & Automation Magazine*, 14(1) :109–118.

- Chaumette, F., P. Rives, et B. Espiau  
1993. Classification and realization of the different vision-based tasks. In *Visual Servoing : Real-Time Control of Robot Manipulators Based on Visual Sensory Feedback*, Pp. 199–228. World Scientific.
- Chen, G., T. D. Bui, et A. Krzyzak  
2005. Image denoising using neighbouring wavelet coefficients. *Integrated Computer-Aided Engineering*, 12(1) :99–107.
- Chen, S. S., D. L. Donoho, et M. A. Saunders  
2001. Atomic decomposition by basis pursuit. *SIAM review*, 43(1) :129–159.
- Chen, Z., X. Hao, et Z. Sun  
2013. Image denoising in shearlet domain by adaptive thresholding. *JOURNAL OF INFORMATION & COMPUTATIONAL SCIENCE*, 10(12) :3741–3749.
- Cherubini, A. et F. Chaumette  
2013. Visual navigation of a mobile robot with laser-based collision avoidance. *The International Journal of Robotics Research*, 32(2) :189–205.
- Chesi, G., E. Malis, et R. Cipolla  
2000. Automatic segmentation and matching of planar contours for visual servoing. In *Robotics and Automation, 2000. Proceedings. ICRA '00. IEEE International Conference on*, volume 3, Pp. 2753–2758. IEEE.
- Chilowsky, C. et P. Langevin  
1916. Procédés et appareils pour la production de signaux sous-marins dirigés et pour la localisation à distance d'obstacles sous-marins. *French patent*, 502913.
- Christopoulos, C., A. Skodras, et T. Ebrahimi  
2000. The jpeg2000 still image coding system : an overview. *IEEE transactions on consumer electronics*, 46(4) :1103–1127.
- Collewet, C.  
1999. *Contributions à l'élargissement du champ applicatif des asservissements visuels 2D*. PhD thesis, Rennes 1.
- Collewet, C. et E. Marchand  
2011. Photometric visual servoing. *IEEE Transactions on Robotics*, 27(4) :828–834.
- Collewet, C., E. Marchand, et F. Chaumette  
2008. Visual servoing set free from image processing. In *2008 IEEE International Conference on Robotics and Automation*, Pp. 81–86.
- Cooley, J. W. et J. W. Tukey  
1965. An algorithm for the machine calculation of complex fourier series. *Mathematics of computation*, 19(90) :297–301.

- Crombez, N., G. Caron, et E. M. Mouaddib  
2015. Photometric gaussian mixtures based visual servoing. In *Intelligent Robots and Systems (IROS), 2015 IEEE/RSJ International Conference on*, Pp. 5486–5491. IEEE.
- Curie, J.  
1880. Développement par compression de l'électricité polaire dans les cristaux hémihédres à faces inclinées. *Bull. Soc. Fr. Mineral.*, 3 :90.
- Dame, A. et E. Marchand  
2011. Mutual information-based visual servoing. *IEEE Transactions on Robotics*, 27(5) :958–969.
- Daubechies, I.  
1988. Orthonormal bases of compactly supported wavelets. *Communications on pure and applied mathematics*, 41(7) :909–996.
- Daubechies, I., M. Defrise, et C. De Mol  
2004. An iterative thresholding algorithm for linear inverse problems with a sparsity constraint. *Communications on pure and applied mathematics*, 57(11) :1413–1457.
- Davenport, M. A., M. F. Duarte, Y. C. Eldar, et G. Kutyniok  
2011. Introduction to compressed sensing. *preprint*, 93(1) :2.
- Deguchi, K. et T. Noguchi  
1996. Visual servoing using eigenspace method and dynamic calculation of interaction matrices. In *Proceedings of 13th International Conference on Pattern Recognition*, volume 1, Pp. 302–306 vol.1.
- Delabarre, B. et E. Marchand  
2012. Visual servoing using the sum of conditional variance. In *Intelligent Robots and Systems (IROS), 2012 IEEE/RSJ International Conference on*, Pp. 1689–1694. IEEE.
- Delaunay, X., M. Chabert, V. Charvillat, et G. Morin  
2010. Satellite image compression by post-transforms in the wavelet domain. *Signal Processing*, 90(2) :599 – 610.
- Do, M. et M. Vetterli  
2003. Contourlets. In *Beyond wavelets*, G. V. Welland, ed., volume 10 of *Studies in Computational Mathematics*, Pp. 1–27. Academic Press/Elsevier Science, San Diego, CA.
- Doignon, C.  
2015. A single 3-d feature-based visual servoing. In *Methods and Models in Automation and Robotics (MMAR), 2015 20th International Conference on*, Pp. 41–46. IEEE.

- Donoho, D.  
2001. Sparse components of images and optimal atomic decompositions. *Constructive Approximation*, 17(3) :353–382.
- Donoho, D., A. Maleki, et A. Montanari  
2010. Message passing algorithms for compressed sensing : I. motivation and construction. In *Information Theory Workshop (ITW), 2010 IEEE*, Pp. 1–5.
- Donoho, D. L. et I. M. Johnstone  
1994. Threshold selection for wavelet shrinkage of noisy data. In *Engineering in Medicine and Biology Society, 1994. Engineering Advances : New Opportunities for Biomedical Engineers. Proceedings of the 16th Annual International Conference of the IEEE*, volume 1, Pp. A24–A25. IEEE.
- Drummond, T. et R. Cipolla  
1999. Real-time tracking of complex structures for visual servoing. In *International Workshop on Vision Algorithms*, Pp. 69–84. Springer.
- Duarte, M. F., M. A. Davenport, D. Takhar, J. N. Laska, T. Sun, K. F. Kelly, et R. G. Baraniuk  
2008. Single-pixel imaging via compressive sampling. *IEEE Signal Processing Magazine*, 25(2) :83–91.
- Dufлот, L.-A., A. Krupa, B. Tamadazte, et N. Andreff  
2016a. Shearlet-based vs. photometric-based visual servoing for robot-assisted medical applications. In *IEEE/RSJ Int. Conf. on Intelligent Robots and Systems, IROS'16*.
- Dufлот, L. A., A. Krupa, B. Tamadazte, et N. Andreff  
2016b. Shearlet transform : A good candidate for compressed sensing in optical coherence tomography. In *2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, Pp. 435–438.
- Dufлот, L.-A., A. Krupa, B. Tamadazte, et N. Andreff  
2016c. Toward ultrasound-based visual servoing using shearlet coefficients. In *IEEE Int. Conf. on Robotics and Automation, ICRA'16*.
- Dufлот, L. A., R. Reisenhofer, B. Tamadazte, N. Andreff, et A. Krupa  
2018. Wavelet and shearlet-based image representations for visual servoing. *The International Journal of Robotics Research*, to be published.
- Dussik, K. T.  
1942. Über die möglichkeit, hochfrequente mechanische schwingungen als diagnostisches hilfsmittel zu verwenden. *Zeitschrift für die gesamte Neurologie und Psychiatrie*, 174(1) :153–168.
- Easley, G., D. Labate, et F. Colonna  
2009. Shearlet-based total variation diffusion for denoising. *IEEE Transactions on Image processing*, 18(2) :260–268.

- Easley, G., D. Labate, et W.-Q. Lim  
2008. Sparse directional image representations using the discrete shearlet transform. *Applied and Computational Harmonic Analysis*, 25(1) :25–46.
- Erickson, B. J., A. Manduca, P. Palisson, K. R. Persons, F. Earnest 4th, V. Savcenko, et N. J. Hangiandreou  
1998. Wavelet compression of medical images. *Radiology*, 206(3) :599–607.
- Espiau, B.  
1994. Effect of camera calibration errors on visual servoing in robotics. In *Experimental robotics III*, Pp. 182–192. Springer.
- Espiau, B., F. Chaumette, et P. Rives  
1992. A new approach to visual servoing in robotics. *IEEE Transactions on Robotics and Automation*, 8(3) :313–326.
- Fang, L., S. Li, D. Cunefare, et S. Farsiu  
2017. Segmentation based sparse reconstruction of optical coherence tomography images. *IEEE transactions on medical imaging*, 36(2) :407–421.
- Fang, L., S. Li, R. McNabb, Q. Nie, A. Kuo, C. Toth, J. Izatt, et S. Farsiu  
2013. Fast acquisition and reconstruction of optical coherence tomography images via sparse representation. *Medical Imaging, IEEE Transactions on*, 32(11) :2034–2049.
- Feddema, J. T. et O. R. Mitchell  
1989. Vision-guided servoing with feature-based trajectory generation (for robots). *IEEE Transactions on Robotics and Automation*, 5(5) :691–700.
- Fercher, A.  
1993. In vivo optical coherence tomography in ophthalmology. In *Medical optical tomography : functional imaging and monitoring*, volume 10311, P. 103110L. International Society for Optics and Photonics.
- Fourier, J.  
1822. *Theorie analytique de la chaleur, par M. Fourier*. Chez Firmin Didot, père et fils.
- Fujimoto, J. G.  
2003. Optical coherence tomography for ultrahigh resolution in vivo imaging. *Nature biotechnology*, 21(11) :1361–1367.
- Gabor, D.  
1946. Theory of communication. part 1 : The analysis of information. *Electrical Engineers - Part III : Radio and Communication Engineering, Journal of the Institution of*, 93(26) :429–441.
- Gangloff, J. A., M. De Mathelin, et G. Abba  
1998. 6 dof high speed dynamic visual servoing using gpc controllers. In *Robotics and*

- Automation, 1998. Proceedings. 1998 IEEE International Conference on*, volume 3, Pp. 2008–2013. IEEE.
- Garcia, C. et G. Tziritas  
1999. Face detection using quantized skin color regions merging and wavelet packet analysis. *IEEE Transactions on multimedia*, 1(3) :264–277.
- Gilbert, A. C., M. J. Strauss, J. A. Tropp, et R. Vershynin  
2007. One sketch for all : fast algorithms for compressed sensing. In *Proceedings of the thirty-ninth annual ACM symposium on Theory of computing*, Pp. 237–246. ACM.
- Grgic, S., M. Grgic, et B. Zovko-Cihlar  
2001. Performance analysis of image compression using wavelets. *IEEE Transactions on industrial electronics*, 48(3) :682–695.
- Haar, A.  
1910. Zur theorie der orthogonalen funktionensysteme. *Mathematische Annalen*, 69(3) :331–371.
- Hadj-Abdelkader, H., Y. Mezouar, N. Andreff, et P. Martinet  
2006a. Decoupled homography-based visual servoing with omnidirectional cameras. In *Intelligent Robots and Systems, 2006 IEEE/RSJ International Conference on*, Pp. 2332–2337. IEEE.
- Hadj-Abdelkader, H., Y. Mezouar, N. Andreff, et P. Martinet  
2006b. Omnidirectional visual servoing from polar lines. In *Robotics and Automation, 2006. ICRA 2006. Proceedings 2006 IEEE International Conference on*, Pp. 2385–2390. IEEE.
- Han, M., D. Chang, C. Kim, B. J. Lee, Y. Zuo, H.-J. Kim, D. Petrisor, B. Trock, A. W. Partin, R. Rodriguez, et al.  
2012. Geometric evaluation of systematic transrectal ultrasound guided prostate biopsy. *The Journal of urology*, 188(6) :2404–2409.
- Häuser, S. et G. Steidl  
2012. Fast finite shearlet transform. *arXiv preprint arXiv :1202.1773*.
- Heikkila, J. et O. Silven  
1997. A four-step camera calibration procedure with implicit image correction. In *Computer Vision and Pattern Recognition, 1997. Proceedings., 1997 IEEE Computer Society Conference on*, Pp. 1106–1112. IEEE.
- Heitzmann, T., C. Doignon, C. Albitar, et P. Graebbling  
2008. Position-based visual servoing using a coded structured light sensor. In *Robotic and Sensors Environments, 2008. ROSE 2008. International Workshop on*, Pp. 126–131. IEEE.

- Herrmann, F. J., M. P. Friedlander, et O. Yilmaz  
2012. Fighting the curse of dimensionality : Compressive sensing in exploration seismology. *IEEE Signal Processing Magazine*, 29(3) :88–100.
- Herrmann, F. J. et G. Hennenfent  
2008. Non-parametric seismic data recovery with curvelet frames. *Geophysical Journal International*, 173(1) :233–248.
- Hong, J., T. Dohi, M. Hashizume, K. Konishi, et N. Hata  
2004. An ultrasound-driven needle-insertion robot for percutaneous cholecystostomy. *Physics in Medicine & Biology*, 49(3) :441.
- Horn, B. et B. Schunck  
1981. Determining optical flow. *Artificial intelligence*, 17(1-3) :185–203.
- Huang, G., H. Jiang, K. Matthews, et P. Wilford  
2013. Lensless imaging by compressive sensing. In *2013 IEEE International Conference on Image Processing*, Pp. 2101–2105.
- Huang, J.-p., L.-k. Zhu, L.-h. Wang, et W.-l. Song  
2017. Compressed sensing mri using sparsity averaging and fista. *Applied Magnetic Resonance*, 48(8) :749–760.
- Hutchinson, S., G. D. Hager, et P. I. Corke  
1996. A tutorial on visual servo control. *IEEE transactions on robotics and automation*, 12(5) :651–670.
- Iwatsuki, M. et N. Okiyama  
2005. A new formulation of visual servoing based on cylindrical coordinate system. *IEEE Transactions on Robotics*, 21(2) :266–273.
- Kallem, V.  
2009. *Vision-based control on lie groups with application to needle steering*. The Johns Hopkins University.
- Kallem, V., M. Dewan, J. P. Swensen, G. D. Hager, et N. J. Cowan  
2007. Kernel-based visual servoing. In *Intelligent Robots and Systems, 2007. IROS 2007. IEEE/RSJ International Conference on*, Pp. 1975–1980. IEEE.
- Kim, D.-J., R. Lovelett, et A. Behal  
2009. Eye-in-hand stereo visual servoing of an assistive robot arm in unstructured environments. In *Robotics and Automation, 2009. ICRA'09. IEEE International Conference on*, Pp. 2326–2331. IEEE.
- King, E., G. Kutyniok, et W.-Q. Lim  
2013. Image inpainting : theoretical analysis and comparison of algorithms. In *SPIE Optical Engineering+ Applications*, Pp. 885802–885802. International Society for Optics and Photonics.

- Kobayashi, Y., M. Suzuki, A. Kato, M. Hatano, K. Konishi, M. Hashizume, et M. G. Fujie  
2012. Enhanced targeting in breast tissue using a robotic tissue preloading-based needle insertion system. *IEEE Transactions on Robotics*, 28(3) :710–722.
- Kragic, D.  
2001. Visual servoing for manipulation : Robustness and integration issues. *IEEE transactions on robotics and automation*, P. 230.
- Krupa, A., G. Fichtinger, et G. D. Hager  
2007. Full motion tracking in ultrasound using image speckle information and visual servoing. In *Proceedings 2007 IEEE International Conference on Robotics and Automation*, Pp. 2458–2464.
- Krupa, A., D. Folio, C. Novales, P. Vieyres, et T. Li  
2016. Robotized tele-echography : an assisting visibility tool to support expert diagnostic. *IEEE Systems Journal*, 10(3) :974–983.
- Krupa, A., J. Gangloff, C. Doignon, M. de Mathelin, G. Morel, J. Leroy, L. Soler, et J. Marescaux  
2003. Autonomous 3-d positioning of surgical instruments in robotized laparoscopic surgery using visual servoing. *IEEE Trans. on Robotics and Automation*, 19(5) :842–853.
- Kundur, D. et D. Hatzinakos  
1998. Digital watermarking using multiresolution wavelet decomposition. In *Acoustics, Speech and Signal Processing, 1998. Proceedings of the 1998 IEEE International Conference on*, volume 5, Pp. 2969–2972. IEEE.
- Kutyniok, G. et D. Labate  
2012a. Introduction to shearlets. In *Shearlets*, Pp. 1–38. Springer.
- Kutyniok, G. et D. Labate  
2012b. *Shearlets : Multiscale analysis for multivariate data*. Springer Science & Business Media.
- Kutyniok, G. et W.-Q. Lim  
2010. Image separation using wavelets and shearlets. In *International Conference on Curves and Surfaces*, Pp. 416–430. Springer.
- Kutyniok, G., W.-Q. Lim, et R. Reisenhofer  
2016. Shearlab 3d : Faithful digital shearlet transforms based on compactly supported shearlets. *ACM Transactions on Mathematical Software (TOMS)*, 42(1) :5.
- Kutyniok, G. et P. Petersen  
2015. Classification of edges using compactly supported shearlets. *Applied and Computational Harmonic Analysis*.

- Kutyniok, G., M. Shahram, et X. Zhuang  
2012. Shearlab : A rational design of a digital parabolic scaling algorithm. *SIAM Journal on Imaging Sciences*, 5(4) :1291–1332.
- Labate, D., W.-Q. Lim, G. Kutyniok, et G. Weiss  
2005. Sparse multidimensional representation using shearlets. *Optics & Photonics*, 2005 :59140U–59140U.
- Le Gall, D.  
1991. Mpeg : A video compression standard for multimedia applications. *Communications of the ACM*, 34(4) :46–58.
- Lebed, E., P. J. Mackenzie, M. V. Sarunic, et M. F. Beg  
2010. Rapid volumetric oct image acquisition using compressive sampling. *Optics express*, 18(20) :21003–21012.
- Lee, D., N. Koizumi, K. Ota, S. Yoshizawa, A. Ito, Y. Kaneko, Y. Matsumoto, et M. Mitsuishi  
2007. Ultrasound-based visual servoing system for lithotripsy. In *Intelligent Robots and Systems, 2007. IROS 2007. IEEE/RSJ International Conference on*, Pp. 877–882. IEEE.
- Leitgeb, R., C. K. Hitzenberger, et A. F. Fercher  
2003. Performance of fourier domain vs. time domain optical coherence tomography. *Opt. Express*, 11(8) :889–894.
- Li, H., B. Manjunath, et S. K. Mitra  
1995. Multisensor image fusion using the wavelet transform. *Graphical models and image processing*, 57(3) :235–245.
- Li, T., A. Krupa, et C. Collewet  
2011. A robust parametric active contour based on fourier descriptors. In *2011 18th IEEE International Conference on Image Processing*, Pp. 1037–1040.
- Lieb Gott, H., A. Basarab, D. Kouame, O. Bernard, et D. Friboulet  
2012. Compressive sensing in medical ultrasound. In *Ultrasonics Symposium (IUS), 2012 IEEE International*, Pp. 1–6.
- Lippiello, V., B. Siciliano, et L. Villani  
2007. Position-based visual servoing in industrial multirobot cells using a hybrid camera configuration. *IEEE Transactions on Robotics*, 23(1) :73–86.
- Liu, X. et J. U. Kang  
2010. Compressive sd-oct : the application of compressed sensing in spectral domain optical coherence tomography. *Optics express*, 18(21) :22010–22019.

- Ludwig, G. D. et F. W. Struthers  
1949. Considerations underlying the use of ultrasound to detect gallstones and foreign bodies in tissue. Technical report, NAVAL MEDICAL RESEARCH INST BETHESDA MD.
- Lustig, M., D. Donoho, et J. M. Pauly  
2007. Sparse mri : The application of compressed sensing for rapid mr imaging. *Magnetic resonance in medicine*, 58(6) :1182–1195.
- Lyapunov, A. M.  
1992. The general problem of the stability of motion. *International journal of control*, 55(3) :531–534.
- Malis, E. et F. Chaumette  
2000. 2d1/2 visual servoing with respect to unknown objects through a new estimation scheme of camera displacement. *International Journal of Computer Vision*, 37(1) :79–97.
- Malis, E., Y. Mezouar, et P. Rives  
2010. Robustness of image-based visual servoing with a calibrated camera in the presence of uncertainties in the three-dimensional structure. *IEEE Transactions on Robotics*, 26(1) :112–120.
- Mallapragada, V., N. Sarkar, et T. K. Podder  
2011. Toward a robot-assisted breast intervention system. *IEEE/ASME Transactions on Mechatronics*, 16(6) :1011–1020.
- Mallat, S.  
1989. A theory for multiresolution signal decomposition : the wavelet representation. *IEEE transactions on pattern analysis and machine intelligence*, 11(7) :674–693.
- Mao, J. et W. Chao  
2010. Pedestrian detection via wavelet fractal signature and support vector machine. In *2010 International Conference on Computer Application and System Modeling (ICCASM 2010)*, volume 5, Pp. V5–111–V5–115.
- Marchand, E. et C. Collewet  
2010. Using image gradient as a visual feature for visual servoing. In *2010 IEEE/RSJ International Conference on Intelligent Robots and Systems*, Pp. 5687–5692.
- Marchand, É., F. Spindler, et F. Chaumette  
2005. Visp for visual servoing : a generic software platform with a wide class of robot control skills. *IEEE Robotics & Automation Magazine*, 12(4) :40–52.
- Mardanpour, M. et M. A. Z. Chahooki  
2016. Robust transparent image watermarking with shearlet transform and bidia-gonal singular value decomposition. *AEU-International Journal of Electronics and Communications*, 70(6) :790–798.

- Marquardt, D. W.  
1963. An algorithm for least-squares estimation of nonlinear parameters. *Journal of the society for Industrial and Applied Mathematics*, 11(2) :431–441.
- Martin, V., M. Chabert, et B. Lacaze  
2008. An interpolation-based watermarking scheme. *Signal Processing*, 88(3) :539–557.
- Marturi, N., B. Tamadazte, S. Dembélé, et N. Piat  
2016. Image-guided nan positioning scheme for sem. *IEEE Transactions on Automation Science and Engineering*.
- Máthé, K., L. Buşoniu, L. Barabás, C. I. Iuga, L. Miclea, et J. Braband  
2016. Vision-based control of a quadrotor for an object inspection scenario. In *2016 International Conference on Unmanned Aircraft Systems (ICUAS)*, Pp. 849–857.
- Mayya, M. et C. Doignon  
2011. Visual tracking of small animals based on real-time level set method with fast infra-red thermographic imaging. In *Robotic and Sensors Environments (ROSE), 2011 IEEE International Symposium on*, Pp. 60–64. IEEE.
- Mebarki, R., A. Krupa, et F. Chaumette  
2010. 2-d ultrasound probe complete guidance by visual servoing using image moments. *IEEE Transactions on Robotics*, 26(2) :296–306.
- Meyer, Y.  
1986. Ondelettes, fonctions splines et analyses graduées. *Lectures given at the University of Torino, Italy*.
- Mezouar, Y., H. H. Abdelkader, P. Martinet, et F. Chaumette  
2004. Central catadioptric visual servoing from 3d straight lines. In *Intelligent Robots and Systems, 2004.(IROS 2004). Proceedings. 2004 IEEE/RSJ International Conference on*, volume 1, Pp. 343–348. IEEE.
- Michel, H. et P. Rives  
1993. *Singularities in the determination of the situation of a robot effector from the perspective view of 3 points*. PhD thesis, INRIA.
- Michelson, A. A. et E. W. Morley  
1887. On the relative motion of the earth and of the luminiferous ether. *Sidereal Messenger*, vol. 6, pp. 306-310, 6 :306–310.
- Nadeau, C. et A. Krupa  
2011. Intensity-based direct visual servoing of an ultrasound probe. In *2011 IEEE International Conference on Robotics and Automation*, Pp. 5677–5682.
- Nadeau, C. et A. Krupa  
2013. Intensity-based ultrasound visual servoing : Modeling and validation with 2-d and 3-d probes. *IEEE Transactions on Robotics*, 29(4) :1003–1015.

- Nadeau, C., A. Krupa, J. Petr, et C. Barillot  
2016. Moments-based ultrasound visual servoing : From a mono- to multiplane approach. *IEEE Transactions on Robotics*, 32(6) :1558–1564.
- Nakhaeinia, D., P. Payeur, et R. Laganière  
2014. Adaptive robotic contour following from low accuracy rgb-d surface profiling and visual servoing. In *Computer and Robot Vision (CRV), 2014 Canadian Conference on*, Pp. 48–55. IEEE.
- Nayar, S. K., S. A. Nene, et H. Murase  
1996. Subspace methods for robot vision. *IEEE Transactions on Robotics and Automation*, 12(5) :750–758.
- Nesterov, Y. et al.  
2007. Gradient methods for minimizing composite objective function.
- Neubach, Z. et M. Shoham  
2010. Ultrasound-guided robot for flexible needle steering. *IEEE Transactions on Biomedical Engineering*, 57(4) :799–805.
- Nunez, J., X. Otazu, O. Fors, A. Prades, V. Pala, et R. Arbiol  
1999. Multiresolution-based image fusion with additive wavelet decomposition. *IEEE Transactions on Geoscience and Remote sensing*, 37(3) :1204–1211.
- Oren, M., C. Papageorgiou, P. Sinha, E. Osuna, et T. Poggio  
1997. Pedestrian detection using wavelet templates. In *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Pp. 193–199.
- Ourak, M., A. D. Simone, B. Tamadazte, G. J. Laurent, A. Menciassi, et N. Andreff  
2016a. Automated in-plane oct-probe positioning towards repetitive optical biopsies. In *2016 IEEE International Conference on Robotics and Automation (ICRA)*, Pp. 4186–4191.
- Ourak, M., B. Tamadazte, et N. Andreff  
2016b. Partitioned camera-oct based 6 dof visual servoing for automatic repetitive optical biopsies. In *2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Pp. 2337–2342.
- Ourak, M., B. Tamadazte, O. Lehmann, et N. Andreff  
2016c. Wavelets-based 6 dof visual servoing. In *2016 IEEE International Conference on Robotics and Automation (ICRA)*, Pp. 3414–3419.
- Pajares, G. et J. M. De La Cruz  
2004. A wavelet-based image fusion tutorial. *Pattern recognition*, 37(9) :1855–1872.
- Pejoski, S., V. Kafedziski, et D. Gleich  
2015. Compressed sensing mri using discrete nonseparable shearlet transform and fista. *IEEE Signal Processing Letters*, 22(10) :1566–1570.

- Pierce, M. C., J. Strasswimmer, B. H. Park, B. Cense, et J. F. de Boer  
2004. Advances in optical coherence tomography imaging for dermatology. *Journal of investigative dermatology*, 123(3) :458–463.
- Plumbley, M. D., T. Blumensath, L. Daudet, R. Gribonval, et M. E. Davies  
2010. Sparse representations in audio and music : from coding to source separation. *Proceedings of the IEEE*, 98(6) :995–1005.
- Puliafito, C. A., M. R. Hee, C. P. Lin, E. Reichel, J. S. Schuman, J. S. Duker, J. A. Izatt, E. A. Swanson, et J. G. Fujimoto  
1995. Imaging of macular diseases with optical coherence tomography. *Ophthalmology*, 102(2) :217–229.
- Qaisar, S., R. Bilal, W. Iqbal, M. Naureen, et S. Lee  
2013. Compressive sensing : From theory to applications, a survey. *Communications and Networks, Journal of*, 15(5) :443–456.
- Qu, X., X. Cao, D. Guo, C. Hu, et Z. Chen  
2010a. Compressed sensing mri with combined sparsifying transforms and smoothed  $l_0$  norm minimization. In *Acoustics Speech and Signal Processing (ICASSP), 2010 IEEE International Conference on*, Pp. 626–629. IEEE.
- Qu, X., W. Zhang, D. Guo, C. Cai, S. Cai, et Z. Chen  
2010b. Iterative thresholding compressed sensing mri based on contourlet transform. *Inverse Problems in Science and Engineering*, 18(6) :737–758.
- Quinsac, C., A. Basarab, et D. Kouamé  
2012. Frequency domain compressive sampling for ultrasound imaging. *Advances in Acoustics and Vibration*, 2012.
- Reddy, B. S. et B. N. Chatterji  
1996. An fft-based technique for translation, rotation, and scale-invariant image registration. *IEEE transactions on image processing*, 5(8) :1266–1271.
- Reisenhofer, R., J. Kiefer, et E. King  
2016. Shearlet-based detection of flame fronts. *Experiments in Fluids*, 57(3) :1–14.
- Richa, R., R. Sznitman, R. Taylor, et G. Hager  
2011. Visual tracking using the sum of conditional variance. In *Intelligent Robots and Systems (IROS), 2011 IEEE/RSJ International Conference On*, Pp. 2953–2958. IEEE.
- Rieder, P. et J. A. Nossek  
1997. Implementation of orthogonal wavelet transforms and their applications. In *Application-Specific Systems, Architectures and Processors, 1997. Proceedings., IEEE International Conference on*, Pp. 489–498. IEEE.

- Rioul, O. et M. Vetterli  
1991. Wavelets and signal processing. *IEEE signal processing magazine*, 8(4) :14–38.
- Romberg, J.  
2008. Imaging via compressive sampling. *Signal Processing Magazine, IEEE*, 25(2) :14–20.
- Sardy, S., P. Tseng, et A. Bruce  
2001. Robust wavelet denoising. *IEEE Transactions on Signal Processing*, 49(6) :1146–1152.
- Satat, G., M. Tancik, et R. Raskar  
2017. Lensless imaging with compressive ultrafast sensing. *IEEE Transactions on Computational Imaging*, 3(3) :398–407.
- Schuman, J. S., M. R. Hee, A. V. Arya, T. Pedut-Kloizman, C. A. Puliafito, J. G. Fujimoto, et E. A. Swanson  
1995. Optical coherence tomography : a new tool for glaucoma diagnosis. *Current opinion in ophthalmology*, 6(2) :89–95.
- Shannon, C. E.  
1948. A mathematical theory of communication. *Bell system technical journal*, 27(3) :379–423.
- Shih, M.-Y. et D.-C. Tseng  
2005. A wavelet-based multiresolution edge detection and tracking. *Image and Vision Computing*, 23(4) :441–451.
- Shirai, Y. et H. Inoue  
1973. Guiding a robot by visual feedback in assembling tasks. *Pattern recognition*, 5(2) :99–108.
- Simoncelli, E. et W. Freeman  
1995. The steerable pyramid : a flexible architecture for multi-scale derivative computation. In *ICIP (3)*, Pp. 444–447.
- Smith, D. S., J. C. Gore, T. E. Yankeelov, et E. B. Welch  
2012. Real-time compressive sensing mri reconstruction using gpu computing and split bregman methods. *International journal of biomedical imaging*, 2012.
- Strang, G. et T. Nguyen  
1996. *Wavelets and filter banks*. SIAM.
- Sturm, P. F. et S. J. Maybank  
1999. On plane-based camera calibration : A general algorithm, singularities, applications. In *Computer Vision and Pattern Recognition, 1999. IEEE Computer Society Conference on.*, volume 1, Pp. 432–437. IEEE.

- Sublett, J. W., B. J. Dempsey, et A. C. Weaver  
1995. Design and implementation of a digital teleultrasound system for real-time remote diagnosis. In *Computer-Based Medical Systems, 1995., Proceedings of the Eighth IEEE Symposium on*, Pp. 292–298. IEEE.
- Swanson, E., J. Izatt, M. Bee, D. Huang, C. Lin, C. Puliafito, et J. Fujimoto  
1992. Optical coherence tomography. In *LEOS '92, Conference Proceedings. IEEE Lasers and Electro-Optics Society, 1992 Annual Meeting*, Pp. 656–657.
- Swanson, E. A., J. A. Izatt, M. R. Hee, D. Huang, C. P. Lin, J. S. Schuman, C. A. Puliafito, et J. G. Fujimoto  
1993. In vivo retinal imaging by optical coherence tomography. *Opt. Lett.*, 18(21) :1864–1866.
- Tahri, O. et F. Chaumette  
2003. *Determination of moment invariants and their application to visual servoing*. PhD thesis, INRIA.
- Tahri, O., A. Y. Tamtsia, Y. Mezouar, et C. Demonceaux  
2015. Visual servoing based on shifted moments. *IEEE Transactions on Robotics*, 31(3) :798–804.
- Tamadazte, B. et N. Andreff  
2014. Weakly calibrated stereoscopic visual servoing for laser steering : Application to phonomicrosurgery. In *Intelligent Robots and Systems (IROS 2014), 2014 IEEE/RSJ International Conference on*, Pp. 743–748. IEEE.
- Tamtsia, A. Y., O. Tahri, Y. Mezouar, H. Djalo, et E. Tonye  
2013. New results in images moments-based visual servoing. In *Robotics and Automation (ICRA), 2013 IEEE International Conference on*, Pp. 5271–5276. IEEE.
- Tearney, G. J., M. E. Brezinski, B. E. Bouma, S. A. Boppart, C. Pitris, J. F. Southern, et J. G. Fujimoto  
1997. In vivo endoscopic optical biopsy with optical coherence tomography. *Science*, 276(5321) :2037–2039.
- Teulière, C. et E. Marchand  
2014. A dense and direct approach to visual servoing using depth maps. *IEEE Transactions on Robotics*, 30(5) :1242–1249.
- Usevitch, B. E.  
2001. A tutorial on modern lossy wavelet image compression : foundations of jpeg  
2000. *IEEE signal processing magazine*, 18(5) :22–35.
- Usher, K., P. Ridley, et P. Corke  
2003. Visual servoing of a car-like vehicle-an application of omnidirectional vision. In *Robotics and Automation, 2003. Proceedings. ICRA'03. IEEE International Conference on*, volume 3, Pp. 4288–4293. IEEE.

- Vahrenkamp, N., S. Wieland, P. Azad, D. Gonzalez, T. Asfour, et R. Dillmann  
2008. Visual servoing for humanoid grasping and manipulation tasks. In *Humanoid Robots, 2008. Humanoids 2008. 8th IEEE-RAS International Conference on*, Pp. 406–412. IEEE.
- Vasanawala, S., M. Murphy, M. Alley, P. Lai, K. Keutzer, J. Pauly, et M. Lustig  
2011. Practical parallel imaging compressed sensing mri : Summary of two years of experience in accelerating body mri of pediatric patients. In *Biomedical Imaging : From Nano to Macro, 2011 IEEE International Symposium on*, Pp. 1039–1043.
- Vitrani, M.-A., M. Baumann, D. Reversat, G. Morel, A. Moreau-Gaudry, et P. Mozer  
2016. Prostate biopsies assisted by comanipulated probe-holder : first in man. *International journal of computer assisted radiology and surgery*, 11(6) :1153–1161.
- Vitrani, M.-A., H. Mitterhofer, G. Morel, et N. Bonnet  
2007. Robust ultrasound-based visual servoing for beating heart intracardiac surgery. In *Robotics and Automation, 2007 IEEE International Conference on*, Pp. 3021–3027. IEEE.
- Vitrani, M.-A., G. Morel, et T. Ortmaier  
2005. Automatic guidance of a surgical instrument with ultrasound based visual servoing. In *Robotics and Automation, 2005. ICRA 2005. Proceedings of the 2005 IEEE International Conference on*, Pp. 508–513. IEEE.
- Wang, Y., J. Cao, et C. Yang  
2011. Recovery of seismic wavefields based on compressive sensing by an l 1-norm constrained trust region method and the piecewise random subsampling. *Geophysical Journal International*, 187(1) :199–213.
- Weiss, L., A. Sanderson, et C. Neuman  
1987. Dynamic sensor-based control of robots with visual feedback. *IEEE Journal on Robotics and Automation*, 3(5) :404–417.
- Wild, J. J. et J. M. Reid  
1952. Application of echo-ranging techniques to the determination of structure of biological tissues. *Science*, 115(2983) :226–230.
- Wilson, W. J., C. W. Hulls, et G. S. Bell  
1996. Relative end-effector control using cartesian position based visual servoing. *IEEE Transactions on Robotics and Automation*, 12(5) :684–696.
- Xie, H., L. E. Pierce, et F. T. Ulaby  
2002. Sar speckle reduction using wavelet denoising and markov random field modeling. *IEEE Transactions on geoscience and remote sensing*, 40(10) :2196–2212.
- Xu, T. et W. Wang  
2009. A compressed sensing approach for underdetermined blind audio source se-

- paration with sparse representation. In *Statistical Signal Processing, 2009. SSP'09. IEEE/SP 15th Workshop on*, Pp. 493–496. IEEE.
- Yang, J., X. Liao, X. Yuan, P. Llull, D. J. Brady, G. Sapiro, et L. Carin  
2015. Compressive sensing by learning a gaussian mixture model from measurements. *IEEE Transactions on Image Processing*, 24(1) :106–119.
- Yi, S., D. Labate, G. Easley, et H. Krim  
2008. Edge detection and processing using shearlets. In *2008 15th IEEE International Conference on Image Processing*, Pp. 1148–1151. IEEE.
- Yi, S., D. Labate, G. R. Easley, et H. Krim  
2009. A shearlet approach to edge analysis and detection. *IEEE Transactions on Image Processing*, 18(5) :929–941.
- Yin, W., S. Osher, D. Goldfarb, et J. Darbon  
2008. Bregman iterative algorithms for  $\ell_1$ -minimization with applications to compressed sensing. *SIAM Journal on Imaging sciences*, 1(1) :143–168.
- Z. Zhang, Z.  
2000. A flexible new technique for camera calibration. *IEEE Transactions on pattern analysis and machine intelligence*, 22(11) :1330–1334.
- Zhou, M. et H. Wei  
2006. Face verification using gaborwavelets and adaboost. In *Pattern Recognition, 2006. ICPR 2006. 18th International Conference on*, volume 1, Pp. 404–407. IEEE.





## Résumé

L'asservissement visuel est un procédé consistant à utiliser l'information visuelle obtenue par un capteur afin de commander un système robotique. Ces informations, appelées primitives visuelles peuvent être d'ordre 2D ou 3D. Le travail présenté ici porte sur une nouvelle approche 2D utilisant des primitives directes : les décompositions de l'image en ondelettes ou en shearlets. Ces représentations présentent en effet l'avantage de décrire l'image sous différentes formes, mettant l'accent soit sur les basses fréquences de l'image, soit sur les hautes fréquences selon plusieurs directions. Les travaux de cette thèse montrent la précision et la robustesse de l'approche utilisant la décomposition en shearlets dans le cadre de l'imagerie échographique. Néanmoins, sa contribution majeure est l'élaboration d'une commande permettant d'utiliser au choix les ondelettes ou les shearlets ainsi que la validation de cette méthode sur caméra monoculaire et sur capteur de type tomographie par cohérence optique dans différentes conditions d'utilisation. Cette méthode présente des performances significatives en termes de précision et de robustesse et ouvre la porte vers une utilisation couplée de l'asservissement visuel et de l'acquisition comprimée.

**Mots clefs :** Ondelette, Shearlet, Représentation multi-échelle, Asservissement visuel direct, Acquisition comprimée

## Abstract

A visual servoing scheme consists of a closed-loop control approach which uses visual information feedback to control the movement of a robotic system. This data, called visual features, can be 2D or 3D. This thesis deals with the development of a new generation of 2D direct visual servoing methods in which the signal control inputs are the coefficients of a multiscale image representation. Specially, we consider the use of multiscale image representations that are based on discrete wavelet and shearlet transformations. This kind of representations allows us to obtain several descriptions of the image based on either low or high frequencies levels. Indeed, high coefficients in the wavelet or in the shearlet transformation of the image correspond to image singularities. This thesis has begun with the development of a shearlet-based visual servoing for ultrasound imaging that has performed well in precision and robustness for this medical application. Nevertheless, the main contribution is a framework allowing us to use several multi-scale representations of the image. It was then tested with conventional white light camera and with an optical coherence tomography imaging system with nominal and unfavorable conditions. Then, the wavelet and the shearlet based methods showed their accuracy and their robustness in several conditions and led to the use of both visual servoing and compressed sensing as the main perspective of this work.

**Keywords :** Wavelet, Shearlet, Multiscale representation, Direct visual servoing, Compressed sensing