



HAL
open science

Construction rapide d'images panoramiques applicables à l'exploration cystoscopique et à l'endoscopie de fluorescence en cancérologie

Yahir Hernandez-Mier

► **To cite this version:**

Yahir Hernandez-Mier. Construction rapide d'images panoramiques applicables à l'exploration cystoscopique et à l'endoscopie de fluorescence en cancérologie. Traitement du signal et de l'image [eess.SP]. Institut National Polytechnique de Lorraine - INPL, 2007. Français. NNT: . tel-00182184

HAL Id: tel-00182184

<https://theses.hal.science/tel-00182184>

Submitted on 25 Oct 2007

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.

Construction rapide d'images panoramiques applicables
à l'exploration cystoscopique et à l'endoscopie de
fluorescence en cancérologie

THÈSE

présentée et soutenue publiquement le 22 octobre 2007

pour l'obtention du

Doctorat de l'Institut National Polytechnique de Lorraine
Spécialité Automatique et Traitement du Signal

par

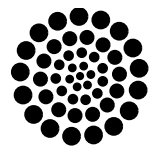
Yahir HERNÁNDEZ MIER

Composition du jury

<i>Rapporteurs :</i>	Jack-Gérard Postaire	LAGIS Lille
	Anne Humeau	ISAIP-ESAIP, LISA Angers
	Michel Paindavoine	Le2i Dijon
<i>Examineurs :</i>	Walter C. P. M. Blondel	CRAN Nancy
	Christian Daul	CRAN Nancy
	François Guillemin	CRAN-CAV
	Didier Wolf	CRAN Nancy



Centre de Recherche en Automatique de Nancy CNRS – UMR 7039
Nancy Université



CONACYT



Mis en page avec la classe thloria.

Remerciements

Cette thèse a été développée au Centre de Recherche en Automatique de Nancy (CRAN), dans le thème Ingénierie pour la Santé (IPS).

Je remercie Monsieur le Professeur Didier Wolf pour m'avoir accueilli au sein du groupe Ingénierie pour la Santé (IPS) du CRAN pour réaliser cette thèse. Je remercie Monsieur Christian Daul pour la correction de ce document.

Je remercie Monsieur le Professeur François Guillemin du Centre Alexis Vautrin (CAV) de lutte contre le cancer qui est à l'origine de ce travail, qui nous a fourni des données patient pour ce travail et qui m'a fait l'honneur de présider mon jury de thèse.

Je tiens à remercier les membres du jury qui m'ont fait l'honneur de participer à l'examen de ce travail.

Monsieur le Professeur Jack-Gérard Postaire, de l'Université Lille 1, pour avoir pris le temps de lire mon rapport et les excellentes critiques constructives qu'il a effectuées.

Monsieur le Professeur Michel Paindavoine, de l'Université de Bourgogne pour l'attention accordée à ce travail et pour sa présence en tant que rapporteur.

Madame le Professeur Anne Humeau, de l'Université d'Angers, pour sa participation à ce jury.

Je remercie le Docteur M.-A. D'Hallewin pour nous fournir des vidéos des cystoscopies utilisées dans cette thèse.

Je remercie tout le personnel du CRAN pour faciliter mon séjour dans ce centre de recherche.

Toute ma reconnaissance à Monsieur W. C. P. M. Blondel pour son guidage, ses conseils et la patience qu'il a démontré pendant la réalisation de ma thèse. Je me sens très identifié avec votre façon de faire de la recherche. Ne faites pas attention aux *outliers* qui pourraient bruyter votre signal. Continuez à développer la robustesse de votre algorithme. C'est grâce à des gens comme vous que je pense que cette expérience a valu la peine. Merci vraiment, j'ai beaucoup appris.

Je tiens à remercier mes amis, pour penser des temps en temps à moi. Je remercie tous les compatriotes mexicains et latino-américains que j'ai connu depuis que je suis arrivé en France. Merci à Manuel, Rosebet, Rubén et Paula, Ramón, Irma, Luisito, Dora, Juanito, Juanote, Noé, pour m'avoir reçu le jour de mon arrivé. Au professeur Martínez pour son support envers les mexicains arrivant à Nancy. Merci, c'est très soulageant de voir de gens familiers lorsqu'on arrive dans un endroit inconnu. À Giovanni et famille, Gilberto, Hugo et Rebeca, Silvia et Pedro, Sinuhé, Joel et Pilar, Diego (bien sûr) et plus récemment, Cédric, Michel, Evaristo, Luis, Ivan, Andrea et Gaby pour les soirées, journées et matchs de foot-ball vécus ensemble. Que de bons moments. À Oscar et famille, Gabriel et famille pour les moments partagés et leur support. Merci à tous pour votre support pendant et après ma soutenance.

Je remercie Mademoiselle Emilie Péry pour m'avoir supporté pendant quatre ans dans une salle de 14 m² pendant au moins 8 heures par jour. Je sais que parfois je peux être difficile. Merci à Sophie pour m'avoir permis d'utiliser ses images acquises sur son fantôme dans cette thèse (même si à la fin elles n'ont pas été utilisées). À Marine pour les réunions SpID et les repas partagés.

Merci à ma famille. À mes chers parents, sœurs et frère, pour m'avoir soutenu si fortement pendant toute la durée de ma thèse. À mes grand-parents, pour leurs conseils et leur soutien. À mes tantes, oncles, cousines et cousins pour penser à moi pendant toute la durée de mon séjour en France.

Je remercie spécialement Yolanda pour partager avec moi des moments inoubliables et pour m'avoir allégé la charge pendant les étapes dures avec son amour et son sourire.

*Je dédie cette thèse à Yolanda
et à ma famille.*

*À la mémoire de
ma grand-mère Hilaria.*

Table des matières

Introduction générale	1
Chapitre 1 Contexte de l'application et problématique	5
1.1 Le cancer et les tissus urothéliaux	5
1.1.1 Données générales sur le cancer	5
1.1.2 Le cancer des tissus épithéliaux	6
1.1.3 Le cancer de la vessie	7
1.1.3.1 Caractéristiques des tissus urothéliaux	7
1.1.3.2 Types et classification des cancers de la vessie	7
1.2 Diagnostic du cancer de la vessie	9
1.2.1 Caractéristiques morphologiques et biochimiques qui peuvent indiquer la présence d'un cancer	9
1.2.2 Évaluation de la qualité d'un examen de dépistage	9
1.2.3 Techniques de dépistage actuelles	10
1.2.4 Techniques de dépistage en cours de développement	11
1.3 Principe du diagnostic optique par fluorescence	12
1.3.1 Le phénomène de fluorescence	12
1.3.2 Diagnostic utilisant l'injection d'une molécule exogène	13
1.3.3 Diagnostic utilisant la fluorescence endogène (autofluorescence)	13
1.3.4 Spectroscopie de fluorescence	14
1.3.5 Imagerie endoscopique de fluorescence	15
1.3.5.1 Systèmes commerciaux d'imagerie endoscopique de fluorescence .	15
1.3.5.2 Le système de cystoscopie de fluorescence développé au CRAN .	18
1.4 Motivation clinique des travaux	20
1.4.1 Supports visuels utilisés dans l'examen cystoscopique actuel	20
1.4.2 Intérêt de la combinaison des informations spatiales et fonctionnelles . . .	21
1.5 Objectifs de cette thèse	22
1.6 Problématique scientifique	22
1.6.1 La qualité des images cystoscopiques	22

1.6.2	Contraintes cliniques imposées sur la méthode de construction des images panoramiques	25
1.6.3	Validation de la méthode proposée	27
1.6.4	Conclusion	28
Chapitre 2 Principales méthodes et applications du mosaïquage d'images		31
2.1	Modèle de formation d'images	31
2.1.1	Projection perspective	31
2.1.2	Distorsion radiale	35
2.2	Le mosaïquage d'images	37
2.2.1	Domaines d'application du mosaïquage d'images	37
2.2.2	Principe du recalage géométrique d'images	40
2.2.3	Définition mathématique du recalage d'images	42
2.2.3.1	Espace de recherche (types de transformation)	42
2.2.3.2	Interpolation des niveaux de gris des images transformées	46
2.2.4	Recalage basé sur les niveaux de gris des images	49
2.2.4.1	Mesures de similarité	49
2.2.4.2	Méthodes d'optimisation	55
2.2.5	Recalage basé sur des primitives images	63
2.2.5.1	Détection des primitives images	63
2.2.5.2	Recalage des primitives images	67
2.2.6	Collage des images recalées	68
2.3	Problèmes d'imagerie spécifiques aux cystoscopes	70
2.3.1	Distorsion radiale	71
2.3.2	Hétérogénéités d'illumination et vignettage	72
2.3.3	Motif de fibres optiques (fibroscope)	73
2.4	Discussion et conclusion	76
Chapitre 3 Solution proposée		79
3.1	Méthode de mosaïquage disponible au laboratoire	79
3.2	Pré-traitements appliqués aux images cystoscopiques	81
3.2.1	Correction des inhomogénéités d'illumination et du motif de fibres optiques	81
3.2.2	Correction de la distorsion radiale	89
3.3	Choix de la méthode de recalage	90
3.3.1	Tests préliminaires effectués sur des images cystoscopiques réelles	90
3.3.1.1	Test d'un détecteur de primitives image (détecteur de Harris)	91
3.3.1.2	Évaluation du degré de recouvrement des images de vidéo-cystoscopie	91

3.3.2	Conclusions tirées des tests	93
3.4	Première méthode de recalage proposée : l'algorithme de Fourier-Mellin	94
3.4.1	Principe de la méthode	94
3.4.2	Application de l'algorithme à des séquences d'images d'un fantôme et cliniques	96
3.4.3	Discussion	98
3.5	Seconde méthode proposée	98
3.5.1	La corrélation avec les transformées de Fourier des images	98
3.5.2	La méthode de recalage par composition inverse	100
3.5.2.1	L'algorithme de recalage de Lucas et Kanade	102
3.5.2.2	L'algorithme de recalage par composition	103
3.5.2.3	Recalage par composition inverse	104
3.5.3	Le collage des images	106
3.6	Synthèse schématique de l'approche proposée	108
Chapitre 4 Validation de la solution proposée		111
4.1	Évaluation de la précision du recalage	111
4.1.1	Méthode	111
4.1.1.1	Test de recalage avec des transformations élémentaires	113
4.1.1.2	Test de recalage avec des transformations élémentaires combinées	114
4.1.1.3	Discussion sur les limites et la précision de recalage	117
4.2	Évaluation de la robustesse du recalage	119
4.2.1	Méthode	119
4.2.2	Discussion et conclusion	122
4.3	Évaluation de la qualité du mosaïquage	122
4.3.1	Instrumentation	122
4.3.2	Tests effectués	123
4.4	Application de la méthode de mosaïquage à des données patient	125
4.5	Sélection d'images pour éviter le surplus de calcul	135
4.6	Combinaison des informations spatiales et fonctionnelles	141
4.7	Discussion et conclusion	141
Conclusion et perspectives		145
Annexe A Développement de l'algorithme de Lucas et Kanade		149
Annexe B Autres images panoramiques		153

Bibliographie

159

Introduction générale

Cette thèse a été effectuée au CRAN (Centre de Recherche en Automatique de Nancy, UMR 7039) dans le groupe thématique IPS (Ingénierie pour la Santé) et plus précisément au sein de l'action SpID (Spectro Imagerie de Diagnostic). Elle s'insère dans l'un des axes de recherche de l'équipe-projet SpID visant à améliorer la sensibilité et la spécificité de l'examen cystoscopique dans la détection du cancer de la vessie. Pour ce faire, des techniques de traitement d'images et de spectroscopie de fluorescence sont développées et utilisées dans cet axe de recherche.

La cystoscopie est actuellement l'examen clinique de référence qui permet de détecter et de diagnostiquer des lésions cancéreuses dans la vessie. Cette détection est effectuée par l'exploration cystoscopique de l'organe en utilisant un éclairage en lumière blanche. Cependant, les stades plus précoces de certaines de ces lésions sont difficilement décelables avec cet examen (lésions *infra-cliniques*). Pourtant, la détection précoce du cancer permet d'augmenter les chances de survie des patients. C'est la raison pour laquelle des techniques de détection de lésions pré-cancéreuses par fluorescence des tissus sont développées par plusieurs équipes de recherche dans le monde.

Le cystoscope est un instrument constitué notamment d'une fine tige (souple ou rigide) qui, insérée par l'urètre, permet l'observation des parois internes de la vessie. Le clinicien effectue l'examen cystoscopique en balayant la surface de l'organe et en prenant comme référence, de son déplacement, quelques repères anatomiques. L'observation de chaque région d'intérêt est réalisée à travers une « fenêtre » (champ de vue) réduite ne permettant de visualiser dans une image qu'une petite surface de l'organe. La taille de cette fenêtre dépend de la distance d'observation et des caractéristiques de l'optique du cystoscope. Pour des observations postopératoires et pour l'archivage des examens, le support photographique et le support vidéo sont communément utilisés. Sur une photographie, seule une petite région d'une zone d'intérêt restera enregistrée. Sur un support vidéo, une région plus grande sera enregistrée, mais avec une grande quantité de données redondantes. De plus, la recherche d'une zone particulière d'intérêt sur la vidéo (enregistrée à un instant donné non connu dans la séquence) peut prendre beaucoup de temps. Par ailleurs, la recherche de la zone d'intérêt avec un magnétoscope peut être fastidieuse pour le clinicien si cette opération est souvent répétée.

Un support visuel contenant des zones larges de la vessie et données sans redondance de données permettrait au clinicien d'observer, en une seule fois, des zones de tissu plus étendues. Voir des zones élargies permettrait de faciliter le repérage de lésions satellites autour d'une lésion principale (une information sur la distribution spatiale des lésions est aussi un indicateur important pour le diagnostic). Ces zones élargies pourraient aussi être utilisées pour réaliser des observations post-opératoires, pour guider des examens ultérieurs et pour stocker les données sur un minimum d'espace mémoire. Ce support visuel peut être obtenu par des méthodes de *mosaïquage d'images*. Le mosaïquage est une technique de traitement d'images qui permet de construire des images panoramiques d'une zone déterminée par le recalage et le « collage » d'une séquence d'images successives dans un repère unique. Le recalage permet d'obtenir le lien géométrique liant deux images. Cette information géométrique peut ensuite être utilisée pour ramener les

images de la séquence dans un système de coordonnées commun (celui de l'image panoramique).

L'objectif général de cette thèse est de développer une méthode de mosaïquage d'images qui permette la construction d'images panoramiques de l'intérieur de la vessie sans affecter de façon importante le protocole d'acquisition. Deux contraintes principales doivent être respectées par notre méthode. D'une part, le temps de construction doit permettre au clinicien d'obtenir des images panoramiques des zones d'intérêt à la fin de l'examen ou au cours de l'examen même. Dans notre laboratoire, un travail précédent a conduit à une méthode de mosaïquage basée sur l'information mutuelle. Cette méthode s'est révélée être précise, mais le temps de calcul nécessaire pour construire une image panoramique (plusieurs heures) limite cette solution à une utilisation « hors ligne ». D'autre part, la méthode de mosaïquage doit être robuste vis-à-vis de la grande variabilité des conditions d'éclairage variant d'un instrument à l'autre, des distorsions inhérentes aux cystoscopes et de la variabilité de « texture » des parois internes de la vessie visibles dans les images cystoscopiques. Les images cystoscopiques sont affectées par des artefacts (par exemple, des inhomogénéités d'illumination) qui doivent être atténuées afin de permettre la mise en œuvre d'une méthode robuste de la méthode de mosaïquage. Nous devons donc étudier et concevoir des méthodes permettant de corriger les images pour être utilisables par une méthode de mosaïquage. Pour proposer une méthode de mosaïquage rapide et robuste nous devons effectuer une étude des techniques récentes les plus adaptées à nos contraintes et aux caractéristiques des images cystoscopiques. Par ailleurs, la méthode proposée doit être automatique pour augmenter la vitesse de calcul et pour éviter l'intervention humaine pour la détection de points homologues. La méthode doit également, être capable d'utiliser uniquement les informations contenues dans les images (pas de modification de l'instrument ou d'ajout d'autres informations).

L'observation des parois internes de la vessie en lumière blanche ne permet pas de déceler toutes les lésions à un stade précoce. Cependant, la détection de ces lésions devient meilleure si des observations en lumière bleue-proche UV (par la fluorescence des tissus) sont effectuées. Nous proposons de combiner les informations spatiales et de détections macroscopiques fournies par des images acquises en lumière blanche avec les informations (de détection de lésions précoces) données par les images de fluorescence. Pour ce faire, nous pouvons utiliser un dispositif cystoscopique spécifique qui est développé au laboratoire et qui permet de faire des acquisitions « quasi-simultanées » des images en lumière blanche et en fluorescence (deux séquences d'une même région). Avec les séquences fournies par ce dispositif nous pouvons construire des images panoramiques « superposables » pour observer en une seule fois des grandes régions où les informations spatiales, macroscopiques et de fluorescence seront combinées.

L'étude du contexte médical, des techniques de base du mosaïquage et le développement de notre méthode de mosaïquage sont présentées dans un document dont l'organisation est la suivante.

Le chapitre 1 présente le contexte applicatif, la motivation clinique, les objectifs de nos travaux et la problématique liée à ces derniers. Dans la dernière partie de ce chapitre nous exposons les contraintes scientifiques liées à l'application.

Le chapitre 2 donne un état de l'art sur les techniques de mosaïquage, les parties qui le composent et ses applications dans divers domaines. Une grande partie de ce chapitre est concentrée sur le recalage d'images qui est au cœur du mosaïquage. Nous présentons les différentes étapes qui composent le recalage d'images et nous abordons les deux approches principales pour recalculer une paire d'images. Ces approches sont les méthodes basées sur l'utilisation directe des niveaux d'intensité des images et les méthodes basées sur la détection de primitives images. Nous étudions dans ce chapitre les mesures de similarité et les méthodes d'optimisation pouvant être utilisés pour le recalage d'images. Nous abordons aussi la technique permettant de ramener les images sur un repère commun (collage des images).

Le chapitre 3 décrit les tests préliminaires qui nous ont permis de choisir une méthode de recalage adaptée, la technique de pré-traitement des images, ainsi que la méthode de recalage et collage d'images que nous proposons dans cette thèse pour la construction d'images panoramiques de la vessie. Dans ce chapitre, nous présentons aussi une technique de recalage que nous avons testée et que nous n'avons pas retenue. À la fin du chapitre nous présentons une synthèse de l'approche retenue.

Le chapitre 4 présente les tests menés pour évaluer et valider notre méthode de mosaïquage. Nous testons dans un premier temps la précision et la robustesse de la méthode de recalage. Puis, nous évaluons la précision des images panoramiques construites par notre méthode de mosaïquage. Une partie importante de ce chapitre est la validation globale de notre méthode par la construction d'images panoramiques à partir d'images cystoscopiques acquises pendant des examens cliniques standard.

Nous terminons ce document par une conclusion qui met en avant les contributions et les perspectives de nos travaux.

1

Contexte de l'application et problématique

La première partie de ce chapitre présente un panorama général sur le cancer de la vessie ainsi que les caractéristiques des lésions cancéreuses qui posent des problèmes de détection précoce et précise par les techniques de diagnostic actuelles. Dans un premier temps, nous rappelons les éléments de diagnostic clinique de la maladie et sa classification. Ensuite, nous présentons les techniques de dépistage/détection actuelles et notamment les méthodes optiques qui font usage de la fluorescence endogène ou exogène. Notre travail s'intégrant dans un projet de recherche visant à développer une approche originale de détection de lésions cancéreuses dans la vessie par endoscopie de fluorescence et mosaïquage d'images, nous présentons un résumé de l'état actuel des systèmes commerciaux de détection de lésions cancéreuses par fluorescence (pour tous les organes). Ensuite, nous présentons une description du système instrumental associé au projet.

La deuxième partie présente les objectifs et la problématique scientifique de cette thèse.

1.1 Le cancer et les tissus urothéliaux

1.1.1 Données générales sur le cancer

Le cancer est un terme générique qui regroupe plus de 100 maladies qui peuvent affecter n'importe quelle partie du corps [WHO, 2006]. Actuellement, le cancer est la maladie la plus mortelle dans le monde. Sur 58 millions de morts par maladie en 2005, 7,6 millions (13 %) sont dues au cancer [WHO, 2006]. En Europe, en 2004, le cancer a tué 1,7 millions de personnes [Boyle and Ferlay, 2005]. En France, en 2005, les cancers ont représenté la deuxième cause de décès (27%), juste derrière les maladies cardio-vasculaires, mais restent la première cause de mort prématurée (37,5%) dans le pays [DREES, 2006].

Le cancer est une maladie qui se caractérise par une prolifération anormale de cellules malignes au sein d'un tissu sain. Les cellules cancéreuses ont acquis des modifications de leur génome induisant une perte de l'apoptose (mort programmée), une perte de la régulation de leur multiplication, un pouvoir d'invasion des structures voisines, une perte de la cohésion entre elles, parmi d'autres changements.

Ces cellules sont des clones d'une cellule dont certaines caractéristiques ont changé. Une tumeur cancéreuse peut être définie comme une population de cellules génétiquement instables qui sont faiblement liées entre elles [Abelev and Kopnin, 2000]. Plus de 85% des cancers prennent naissance dans les couches superficielles (l'épithélium) des organes du corps humain, dont 50%

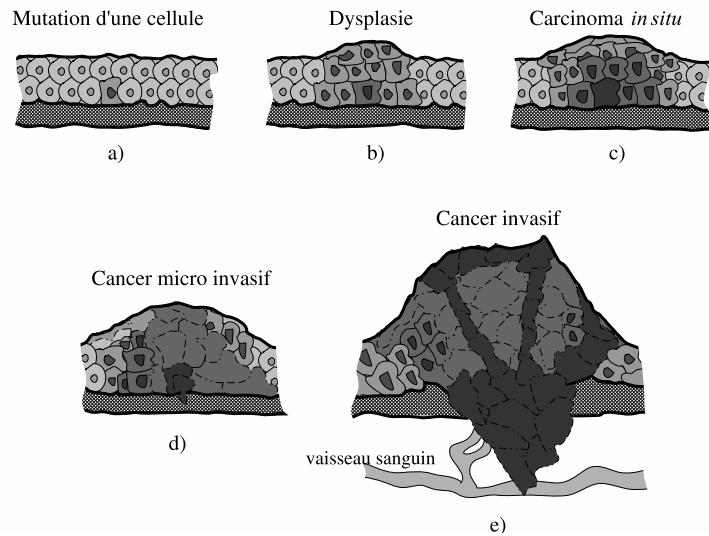


FIG. 1.1 – Représentation simplifiée des différentes étapes d'évolution d'un tissu épithélial cancéreux. a) Mutation d'une cellule. b) Augmentation anormale du nombre de cellules et de la masse du tissu (hyperplasie). c) Développement anormal du tissu (dysplasie). d) Une lésion maligne confinée à l'épithélium apparaît. e) Les cellules malignes se propagent et peuvent envahir d'autres tissus.

au sein des organes creux (vessie, colon, œsophage, bronches, estomac, col de l'utérus).

1.1.2 Le cancer des tissus épithéliaux

L'épithélium est un tissu dont les cellules sont jointives et solidaires les unes des autres grâce à des jonctions intercellulaires. Il n'est pas vascularisé et repose sur une membrane basale au travers de laquelle l'apport des nutriments et l'évacuation des déchets s'effectue grâce au tissu conjonctif sous-jacent. Dans ce type de tissu, les cellules sont souvent polarisées (les deux extrémités opposées sont différentes morphologiquement et biochimiquement). L'épithélium se trouve aussi bien à l'extérieur du corps (la peau) humain, qu'à l'intérieur des organes creux (œsophage, vessie, colon). L'épaisseur des épithéliums varie de quelques dizaines de micromètres à plusieurs centaines suivant les organes (50 pour le colon, environ 100-200 pour l'œsophage).

Le processus initial de formation d'une tumeur de type épithélial est schématisé dans la figure 1.1(b). La *dysplasie* correspond à une étape de mutation où le nombre de cellules et leur développement augmente de façon significativement anormale (figure 1.1(c)). Si l'état des cellules affectées est très anormal, tant au niveau du développement que de l'apparence et que cette prolifération de cellules reste confinée à l'épithélium, elle est appelée cancer *in situ* (CIS)(figure 1.1(d)). La tumeur va ensuite envahir les tissus voisins et les circuits sanguin ou lymphatique. Cette tumeur devient donc *invasive* (figure 1.1(e)). Les cellules peuvent se propager dans les tissus non-homologues (d'autres organes) et former des *métastases*.

Une lésion à un stade pré-cancéreux est une lésion témoignant d'un processus de cancérisation à un stade précoce. Ces lésions sont dites intra épithéliales ou *in situ*. Elle comporte un risque de progression tumorale, mais peut aussi régresser ou persister sans s'aggraver. Une lésion infraclinique est une lésion dont la présence ne provoque pas de symptôme apparent.

1.1.3 Le cancer de la vessie

Parmi les cancers des organes creux, le cancer de la vessie est le 7^{ème} cancer le plus répandu dans le monde, soit 3,2% de l'ensemble des cancers [Tyczynski and Parkin, 2003]. En France, en 2000, selon le CIRC (Centre International de Recherche sur le Cancer), les tumeurs de vessie ont représenté la 5^{ème} cause de cancer et la 3^{ème} de décès par cancer. Leur incidence (nombre de nouveaux cas pendant une période de temps définie) est en augmentation d'environ 1% par an [Chopin and Gattengo, 2001]. L'incidence du cancer de la vessie est trois fois plus élevée chez les hommes que chez les femmes. 80% des tumeurs surviennent entre 50 et 80 ans. 95% des tumeurs de la vessie sont d'origine épithéliale.

1.1.3.1 Caractéristiques des tissus urothéliaux

La vessie est un organe qui fait partie du système urinaire (figure 1.2(a)). Elle a la forme d'une poche qui reçoit l'urine produite par les reins. Elle est constituée d'un tissu musculaire qui en se contractant, permet à l'urine d'être expulsée. Lorsque la vessie est vide, l'épaisseur de sa paroi est d'environ 5 mm. En état de distension moyenne, son épaisseur ne dépasse pas 2 mm. La paroi de la vessie est constituée de plusieurs couches. Une couche de *cellules urothéliales* (aussi appelées *transitionnelles*) recouvre l'intérieur de la vessie. Cette couche, appelé *urothélium* ou *épithélium transitionnel* repose sur une membrane basale. Sous l'urothélium se trouve une fine couche de tissu conjonctif appelée *chorion*. L'épaisseur de cette couche mesure en moyenne 1,4 mm d'épaisseur. La couche suivante, plus profonde, est une zone de tissu musculaire appelée *musculeuse*. Enfin, une couche de tissu adipeux sépare la vessie des autres organes. Sur la figure 1.2(b) sont représentées les couches qui composent la paroi de la vessie.

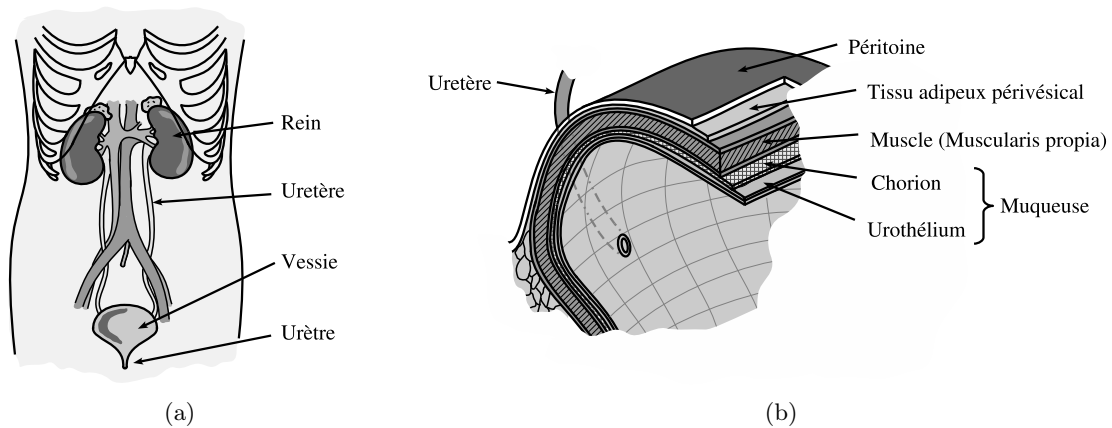


FIG. 1.2 – (a) Schéma montrant les composants principaux du système urinaire. (b) Représentation des couches constituantes de la paroi de la vessie.

1.1.3.2 Types et classification des cancers de la vessie

Du point de vue histologique, deux types de tumeurs peuvent être distingués : les *tumeurs papillaires* et les *lésions planes*.

- **Tumeurs papillaires :**

Non infiltrantes. La plupart des tumeurs de vessie sont superficielles. Selon leur aspect, elles sont appelées polype superficiel, papillome ou tumeur superficielle. Ces tumeurs n'infiltrent pas la *lamina propria*, laissant la membrane basale intacte.

Infiltrantes. Leurs extensions s'enfoncent plus ou moins profondément dans la paroi de la vessie, jusqu'à la couche *muscleuse* (qui contient les fibres musculaires de la vessie) et au delà. L'appréciation du degré d'extension permet de déterminer le stade (figure 1.3).

• **Les lésions tumorales « planes » :**

Carcinome in-situ (CIS). Les lésions urothéliales planes désignent un spectre d'anomalies histologiques intra-épithéliales de l'urothélium plan allant de l'hyperplasie plane au CIS. Le CIS est une tumeur plate, non papillaire, située dans l'épaisseur même de la muqueuse. Le CIS est un cancer épithélial résistant et agressif avec un potentiel évolutif. Le CIS peut être *multifocal* et *diffus*. La présence de CIS proche des tumeurs papillaires est associée à un risque plus important de progression de la tumeur et de récurrence. Il existe un risque élevé de récurrence et de progression parmi des patients avec des lésions de CIS. Dû à ses caractéristiques, le CIS peut passer inaperçu à l'examen cystoscopique habituel.

Le *T* de la classification *TNM*, *Tumor Node Metastasis* en anglais, classe les tumeurs par rapport à leur degré de pénétration dans la paroi de la vessie (figure 1.3).

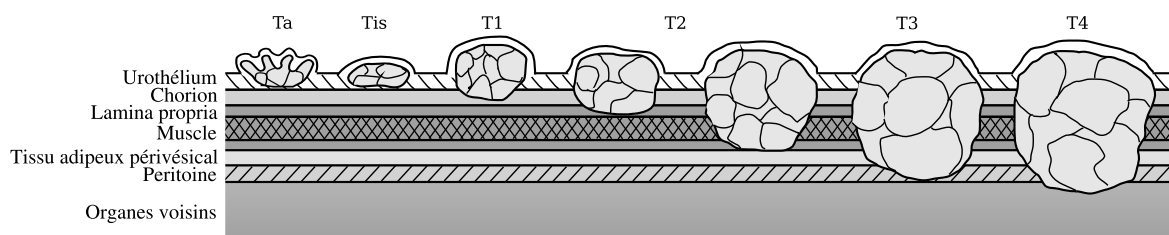


FIG. 1.3 – Différents degrés d'invasion d'une tumeur de vessie.

Ta Tumeur papillaire non-invasive (confinée à la muqueuse).

Tis Carcinoma *in situ* (Tumeur *plane* résistante, confinée à l'urothélium).

T1 La tumeur franchit la *lamina propria*.

T2 La tumeur envahit le muscle de la vessie.

T3 La tumeur envahit le tissu adipeux périvésical.

T4 La tumeur envahit un organe voisin (prostate, utérus, vagin, paroi pelvienne ou paroi abdominale).

Les tumeurs superficielles peuvent poser deux problèmes évolutifs essentiels. D'abord, un risque de récurrence sans progression mais surtout, un risque de récurrence avec progression. La progression peut être définie comme l'aggravation du stade et/ou du grade par rapport à celui de la tumeur initiale. D'une manière générale, plus le stade et le grade sont élevés plus le risque de récurrence et de progression est grand.

1.2 Diagnostic du cancer de la vessie

1.2.1 Caractéristiques morphologiques et biochimiques qui peuvent indiquer la présence d'un cancer

Au diagnostic initial, la plupart des tumeurs se présente sous une forme superficielle dont environ 40% sous forme Ta, 30% sous forme T1 et 2 à 5% de CIS ou de lésions planes (figure 1.3) [Chopin and Gattengo, 2001]. Depuis les premiers stades du développement, les tissus cancéreux vont présenter des différences tant au niveau *macroscopique* qu'au niveau *microscopique*.

Le CIS se développant à l'intérieur de la muqueuse, il est très difficilement décelable par une étude macroscopique de la vessie. Au niveau microscopique, les signes architecturaux et cytologiques qui indiqueraient la présence d'une tumeur sont :

- Une désorganisation de l'architecture de l'urothélium, dont l'importance est variable selon la différenciation tumorale.
- Une infiltration des structures sous-jacentes, dont le degré peut être un indicateur de l'agressivité de la tumeur.
- Une stroma réaction. Le stroma est le réseau conjonctif de support. Le stroma constitue un micro-environnement qui répond au développement de la tumeur.
- Des changements au niveau cytotologique comme des anomalies de la différenciation cellulaire, des irrégularités dans la taille des cellules (*anisocytose*) et dans la taille des noyaux (*anisocaryose*), un rapport nucléo-cytoplasmique (volume noyau/volume cytoplasme) élevé, un *hyperchromatisme* avec des noyaux denses et de la chromatine épaisse ou répartie irrégulièrement, une perte de polarité cellulaire (changements d'orientation), un épaississement de la membrane nucléaire, des nucléoles volumineux, etc.

L'intérêt du dépistage des dysplasies et des CIS réside dans le fait qu'une découverte à ce stade, permet une guérison par un traitement local, soit par une destruction de la lésion, soit par résection locale de la zone affectée.

1.2.2 Évaluation de la qualité d'un examen de dépistage

Un examen clinique à visée diagnostique doit pouvoir indiquer de façon fiable la présence ou l'absence d'une maladie chez un individu. La qualité d'un examen clinique peut être mesurée par l'appréciation de sa *validité intrinsèque*. Cette validité intrinsèque est composée de deux paramètres : la *sensibilité* et la *spécificité*.

La sensibilité et la spécificité peuvent être définies en utilisant les résultats possibles d'un examen clinique : le *vrai positif* (résultat positif d'un test sur un organe ou tissu malade), le *faux positif* (résultat positif d'un test sur un organe ou tissu non malade), le *faux négatif* (résultat négatif d'un test sur un organe ou tissu malade) et le *vrai négatif* (résultat négatif d'un test sur un organe ou tissu non malade). Ces notions sont présentées dans le tableau de contingence 1.1.

TAB. 1.1 – Résultats possibles lors de la mesure de la validité intrinsèque d'un examen clinique.

Résultat du test	Maladie présente	Maladie absente	Nombre de résultats
Test positif	vrai positif : VP	faux positif : FP	positifs : $VP + FP$
Test négatif	faux négatif : FN	vrai négatif : VN	négatifs : $FN + VN$
Nombre de sujets	malades : $VP + FN$	non malades (sains) : $FP + VN$	

La *sensibilité* (Se) est la probabilité d'avoir un test positif lorsque la maladie est présente :

$$Se = \frac{\text{Nombre de vrais positifs}}{\text{Nombre de sujets malades}} = \frac{VP}{VP + FN}. \quad (1.1)$$

La *spécificité* (Sp) est la probabilité d'avoir un test négatif lorsque la maladie est absente :

$$Sp = \frac{\text{Nombre de vrais négatifs}}{\text{Nombre de sujets non malades}} = \frac{VN}{FP + VN}. \quad (1.2)$$

La sensibilité et la spécificité sont indissociables. Prises séparément, elles ne veulent rien dire. Par exemple, un test qui a une sensibilité de 95% mais une spécificité de 6% n'est pas un test valable. Il existe aussi deux autres paramètres issus de l'évaluation de la validité extrinsèque d'un examen clinique. La *valeur prédictive positive* du test est la probabilité d'avoir la maladie quand le test est positif :

$$Vpp = \frac{\text{Nombre de vrais positifs}}{\text{Nombre de résultats positifs}} = \frac{VP}{VP + FP}. \quad (1.3)$$

La *valeur prédictive négative* du test est probabilité de ne pas avoir la maladie quand le test est négatif :

$$Vpn = \frac{\text{Nombre de vrais négatifs}}{\text{Nombre de résultats négatifs}} = \frac{VN}{FN + VN}. \quad (1.4)$$

La sensibilité et la spécificité dépendent uniquement des qualités du test (et éventuellement de l'opérateur). Les valeurs prédictives n'informent pas sur le test lui-même mais sur la situation après lui, c'est-à-dire, la probabilité qu'un sujet après un test positif soit malade (valeur prédictive positive) ou qu'un sujet après un test négatif soit sain (valeur prédictive négative).

1.2.3 Techniques de dépistage actuelles

Le dépistage d'une lésion cancéreuse consiste à repérer à l'aide de tests, les personnes présentant un cancer précoce ou une lésion pré-cancéreuse avant que les signes soient décelables. Il existe plusieurs techniques plus ou moins performantes et contraignantes qui, en détectant des changements aux niveaux biochimiques, morphologiques ou cytologiques, permettent de diagnostiquer une lésion cancéreuse dans la vessie. Les plus courantes sont abordées ci-dessous. Les données sur la sensibilité et la spécificité des traitements sont issues du *Rapport 2001 de l'Association Française d'Urologie (AFU) sur les tumeurs superficielles de la vessie* [Chopin and Gattengo, 2001].

La cytologie urinaire. Dans cet examen, les cellules urothéliales desquamées à partir d'urines mictionnelles ou d'un lavage vésical sont étudiées. C'est l'examen biologique courant le plus performant, sa spécificité étant de 93%. En revanche, son point faible est la sensibilité qui dépend étroitement du grade tumoral : 27% pour les tumeurs de bas grade et 77% pour les tumeurs de haut grade. La cytologie urinaire est donc intéressante pour le diagnostic de tumeurs avancées. Il existe jusqu'à 12% de faux-positifs (inflammation vésicale, lithiase urinaire, etc). Pour éviter ces inconvénients, elle peut être associée à la cytométrie de flux (cf. section 1.2.4), permettant un taux de diagnostic de 95 %. Cependant, son utilisation n'est pas de pratique courante.

L'échographie vésicale sus pubienne. Elle utilise des ultrasons pour chercher des polypes dans la vessie. Cet examen présente une sensibilité faible de 74% mais une bonne spécificité (95%). Compte tenu de la résolution, son exploitabilité est limitée, surtout quand il s'agit de tumeurs planes, situées sur le dôme vésical, de moins de 2 mm de diamètre. Cet examen ne peut donc être utilisé seul pour le diagnostic ou la surveillance. Elle s'effectue par voie sus pubienne et cherche une image caractéristique de végétation endo-luminale, à base d'implantation pédiculée ou sessile.

L'urographie intraveineuse (UIV) de contraste. Elle consiste en une radiographie de l'appareil urinaire après injection d'un produit iodé par voie intraveineuse. Cette technique ne permet de déceler en moyenne que 60% des tumeurs vésicales. Son intérêt principal est le bilan et la surveillance du haut appareil urinaire. En cas de tumeur vésicale suffisamment volumineuse, elle montre typiquement une image de lacune (zone radio-transparente localisée) avec rigidité pariétale partielle. Elle n'est pratiquement plus réalisée depuis l'avènement de l'échographie et du scanner.

La cystoscopie. Elle constitue l'examen de référence. Utilisant un endoscope souple ou rigide, elle permet de voir la tumeur vésicale, d'en déterminer sa localisation, d'en préciser ses caractères morphologiques et d'explorer le reste de la vessie à la recherche de l'existence d'une ou plusieurs autres tumeurs. Nous parlerons de l'examen cystoscopique plus en détails dans la section 1.4.1.

La biopsie. Lorsqu'un tissu anormal est observé, il est nécessaire de prélever un échantillon de tissu pour effectuer une analyse microscopique et faire un diagnostic. Les échantillons utilisés pour la biopsie sont obtenus par ablation chirurgicale par voie endoscopique. Il est nécessaire de prélever des tissus à différents endroits de la vessie (plus ou moins aléatoirement) de façon à tenter de déterminer l'existence de lésions *in situ*. L'analyse histopatologique de l'échantillon prélevé est le seul moyen de donner un diagnostic définitif sur la présence ou non de cancer (« gold standard »).

Ainsi, la cystoscopie reste l'examen de référence pour la surveillance des tumeurs vésicales superficielles. La cytologie urinaire est une excellente méthode de surveillance, mais elle possède une faible valeur prédictive négative, en particulier en présence d'une tumeur de bas grade. Les tests de biologie moléculaire sont encore en évaluation, actuellement ils ne peuvent être utilisés seuls en pratique quotidienne.

1.2.4 Techniques de dépistage en cours de développement

Les travaux de recherche actuels visent à développer des outils de test diagnostique « idéal » qui soient non invasifs, reproductibles, de lecture rapide, réalisables et interprétables par l'urologue. La détection précoce des récidives et progressions, ainsi que l'amélioration de la sensibilité et de la spécificité pour atteindre 100% sont un enjeu majeur.

C'est dans ce cadre, que des techniques dites « alternatives », ont été développées. Comme par exemple, l'utilisation de *marqueurs immunocytologiques* dans la *cytométrie de flux*. Ces tests visent à identifier certaines substances libérées dans l'urine par les cellules cancéreuses. Dans ce test d'*immunofluorescence*, les cellules sont isolées et mises en contact avec des anticorps monoclonaux qui ont préalablement été liés à des marqueurs fluorescents. Ces mêmes anticorps seront couplés aux antigènes présents à la surface des cellules cancéreuses, permettant ainsi leur identification par microscopie de fluorescence. Ce test, en combinaison avec l'utilisation de la cytologie urinaire, offre une sensibilité de 86% et une spécificité de 99%. Cependant ce type

de test n'est toujours pas d'utilisation courante et l'utilisation de la cystoscopie continue d'être nécessaire.

Dans le domaine de l'imagerie endoscopique, avec l'introduction de la vidéo-endoscopie il y a plus de 25 ans et l'amélioration de la résolution des images acquises en cystoscopie (caméras plus performantes), le diagnostic des maladies vésicales s'est beaucoup amélioré. Néanmoins, des *techniques de fluorescence* peuvent encore contribuer à l'obtention d'un diagnostic optimisé *in vivo in situ*. Ces techniques peuvent être utiles non seulement dans le diagnostic précoce de carcinomes, mais aussi dans la détection de stades préliminaires, comme les dysplasies et CIS.

Dans les premières années du développement des techniques de fluorescence, les efforts se sont concentrés sur la *spectroscopie de fluorescence*. Néanmoins, la nécessité d'avoir des images endoscopiques en temps réel permettant de visualiser les différences entre tissus sains et tissus tumoraux, a amené à développer des techniques d'*imagerie de fluorescence*. Il existe deux approches différentes qui utilisent soit des *fluorophores endogènes* (*autofluorescence* ou *fluorescence induite*) soit des *fluorophores exogènes* (cette approche est aussi connue comme *diagnostic photodynamique* ou *PDD*, en anglais). L'objectif des deux approches est le même : la détection et la localisation des lésions invisibles en lumière blanche (dysplasies, cancers *in situ*), qui amènerait à la détermination fiable de l'extension et de la distribution spatiale des lésions.

1.3 Principe du diagnostic optique par fluorescence

Le diagnostic par fluorescence est basée sur l'interaction d'une lumière de longueur d'onde déterminée avec les fluorophores intrinsèques ou extrinsèques présents dans un tissu. Le diagnostic se fait en détectant la présence ou absence d'un fluorophore et/ou son niveau d'intensité. Plusieurs fluorophores peuvent être détectés. Les types de sources de lumière qui peuvent être utilisées sont des sources de lumière incohérente filtrées (lampes à arc) et des sources monochromatiques (laser). La profondeur de pénétration de la lumière dans les tissu dépend de la longueur d'onde utilisé. Celle-ci peut aller de quelques microns dans l'UV à plusieurs centimètres dans le proche IR. Quand le photon excite un fluorophore dans le tissu, il va générer un photon de fluorescence que l'on espère détecter.

1.3.1 Le phénomène de fluorescence

Le phénomène de fluorescence est illustré dans le diagramme des états électroniques ou *diagramme de Jablonski* simplifié montré dans la figure 1.4.

1. Un photon d'énergie $E_p = h \frac{c}{\lambda_{ex}}$ est émis par une source d'excitation externe (lampe ou laser) et absorbé par une molécule, ce qui a pour effet de faire passer cette molécule de son état initial S_0 à un niveau d'énergie électronique supérieur S'_1 .
2. L'état excité dure un temps très court ($\approx 10^{-9}$ s). Le passage de S'_1 à S_1 correspond à une perte d'énergie de l'électron qui se stabilise à une couche métastable (avant de retrouver son orbite originelle). Cette perte d'énergie (Δ_E) entre ces deux niveaux est égale à l'énergie d'excitation E_{ex} moins l'énergie du photon émis (photon de fluorescence).

Pendant cette période, le fluorophore subit des changements conformationnels et il est sujet à de multiples interactions avec son environnement. Ces interactions vont avoir deux conséquences. D'abord, l'énergie de S'_1 est partiellement dissipée (chaleur, transferts entre fluorophores), produisant l'état de relaxation singulet S_1 .

3. Puis le retour de l'état S_1 à S_0 s'effectue par une transition radiative, c'est-à-dire, l'émission d'un photon d'énergie $h \frac{c}{\lambda_{em}} < h \frac{c}{\lambda_{ex}}$ d'où $\lambda_{em} > \lambda_{ex}$. Une quantification de l'énergie

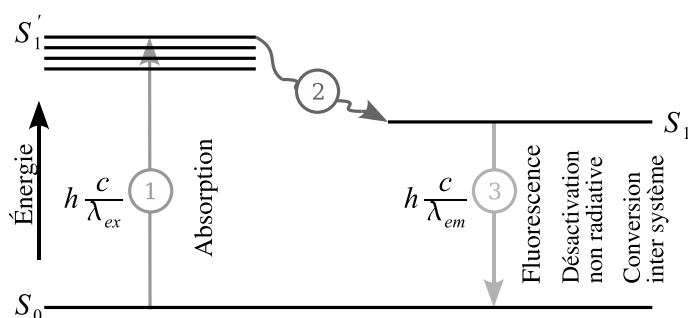


FIG. 1.4 – Diagramme de Jablonski simplifié. h est la constante de Planck, c est la célérité de la lumière, λ_{ex} est la longueur d'onde d'excitation et λ_{em} est la longueur d'onde d'émission de la fluorescence. 1. Absorption d'un photon d'énergie E_p et excitation de la molécule fluorescente à l'état S_1' . 2. Dissipation partielle d'énergie. 3. Dés-excitation radiative et émission d'un photon de fluorescence.

dissipée sous une forme différente de la fluorescence est effectuée en mesurant le *rendement quantique*, qui est le rapport entre les photons émis et les photons absorbés.

En fonction du fluorophore ciblé, la longueur d'onde de la lumière d'excitation devra être choisie. La nature du tissu suspect peut être déterminée en comparant les intensités de fluorescence du tissu dans différentes bandes spectrales.

1.3.2 Diagnostic utilisant l'injection d'une molécule exogène

Parmi les principaux marqueurs fluorescents utilisables en clinique nous trouvons les *photosensibilisants* utilisés pour la PDT (Thérapie photo-dynamique). Ces photosensibilisants sont des substances qui vont s'accumuler de façon préférentielle, au niveau des cellules cancéreuses. L'acide 5-aminolevulinique (5-ALA) est une substance qui suscite de l'intérêt car elle n'est pas un photosensibilisant, mais un précurseur de la biosynthèse de l'hème qui induit une accumulation de porphyrines endogènes, spécialement des protoporphyrines IX (PpIX) dans les tissus d'origine épithéliale. La PpIX est un métabolite fluorescent qui possède un spectre d'excitation centré autour de 400 nm. Dans le cas de la vessie, après application du 5-ALA, il existe une accumulation sélective de PpIX dans les tumeurs urothéliales, produisant un contraste intense entre des lésions cancéreuses qui vont émettre une fluorescence dans le rouge et les tissus normaux qui ne font que réfléchir la lumière d'excitation [Zaak *et al.*, 2005].

1.3.3 Diagnostic utilisant la fluorescence endogène (autofluorescence)

L'autofluorescence des tissus épithéliaux provient de la stimulation des fluorophores endogènes, particulièrement des coenzymes, comme le NADH (Nicotinamide Adénine Dinucléotide Hydrogène), les flavines ou des protéines structurales comme le collagène et l'élastine (figure 1.5). Du fait de leur large spectre d'émission, il existe des superpositions de spectres qui compliquent la distinction entre ces fluorophores. Différents états pathologiques, spécialement tumoraux, mais aussi des inflammations et des ischémies, vont affecter la condition métabolique et oxydative des cellules et par conséquent, leur autofluorescence. De plus, l'épaississement de la muqueuse peut aussi affecter les spectres d'autofluorescence (cf. section 1.2). Après stimulation avec une lumière proche UV-violet, les fluorophores endogènes vont émettre de la lumière dans des longueurs d'onde supérieures (cf. figure 1.5). L'évolution cancéreuse du tissu entraîne des modifications

biochimiques et structurales qui se traduisent par des variations des contributions relatives des différents fluorophores endogènes. Le principal avantage de la technique d'auto fluorescence est qu'il n'est pas nécessaire d'administrer une substance étrangère dans l'organisme du patient. En revanche, il existe des problèmes de détection du fait du faible niveau d'intensité du signal de fluorescence.

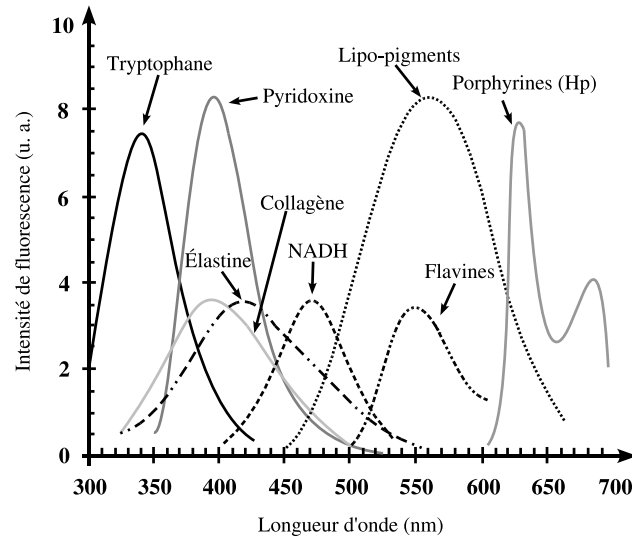


FIG. 1.5 – Spectres normalisés de fluorescence pour quelques fluorophores endogènes (d'après [Wagnières *et al.*, 1998]).

1.3.4 Spectroscopie de fluorescence

La figure 1.6 présente les principaux composants d'un système de spectroscopie de fluorescence.

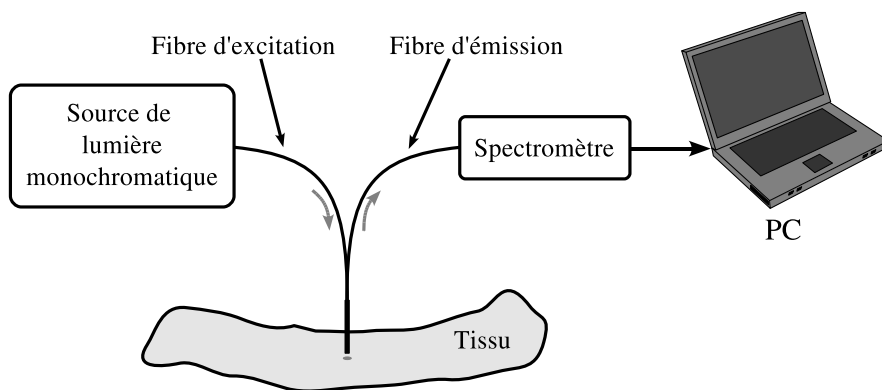


FIG. 1.6 – Principaux composants d'un système de spectroscopie de fluorescence.

Les systèmes de spectroscopie de fluorescence utilisent généralement des sondes à fibres optiques pour exciter le tissu (injection de la lumière UV 300–480 nm) ainsi que pour recueillir (à travers la même fibre ou des fibres distinctes) la lumière émise par le tissu étudié. Un spectromètre permet de faire l'acquisition des intensités en fonction de la longueur d'onde. Des opérations de

soustraction du signal de fond ou des calculs de ratio entre deux ou plusieurs bandes (ou longueurs d'onde particulières) sont effectuées. Des algorithmes spécifiques sont nécessaires pour effectuer une classification des tissus et obtenir une valeur qui pourrait indiquer la présence ou l'absence d'une tumeur dans la région examinée.

Des travaux récents [Heintzelman *et al.*, 2000] [Bourg-Heckly *et al.*, 2000] [Palmer *et al.*, 2003] [Goujon *et al.*, 2003] [Svistun *et al.*, 2004], ont montré des sensibilités pour la spectroscopie de fluorescence entre 75% et 90% et des spécificités entre 45% et 85% lorsque les longueurs d'onde d'excitation sont optimisées. Le principal avantage de cette technique est sa sensibilité due à sa résolution spectrale. De ce fait, seul un petit volume est sondé à chaque mesure. Les fibres utilisées pour mesurer les spectres de fluorescence permettent de sonder un volume d'environ 0,05 mm³ à 0,5 mm³ [Zaak *et al.*, 2005]. Même si plusieurs mesures sont effectuées, la surface totale de tissu examinée en clinique reste très limitée et il existe un risque de manquer des régions de tissu pathologique.

1.3.5 Imagerie endoscopique de fluorescence

De la même façon que pour la biopsie mécanique, il n'est pas envisageable d'effectuer en clinique des biopsies optiques point par point sur toute la surface de l'organe à examiner. Il existe donc un risque de non détection d'une zone malade liée à la faible résolution spatiale de la technique. Ce problème a motivé le développement de systèmes d'imagerie endoscopique de fluorescence. Plusieurs travaux montrent l'intérêt de l'utilisation de la fluorescence en imagerie endoscopique pour améliorer la perception de contraste entre des zones saines et tumorales *in vitro* [Svistun *et al.*, 2004] et *in vivo* [Kara *et al.*, 2004][Zeng *et al.*, 2004]. À l'instar de la spectroscopie de fluorescence, ces images peuvent être utilisées pour calculer un ratio (rapport d'images prises dans des bandes spectrales différentes) qui pourrait révéler des contrastes entre zone saine et suspecte du tissu. Du point de vue clinique, un système endoscopique d'imagerie de fluorescence est très attractif puisque c'est la seule manière de pouvoir explorer de grandes zones de tissu avec une bonne définition (nombre de pixels) et une bonne résolution spatiale (nombre de pixels/cm). En pratique, du fait du faible niveau du signal d'autofluorescence *in vivo*, il est nécessaire d'utiliser une instrumentation spécifique. Un système d'imagerie tenant compte de ces contraintes est constituée des éléments suivants (figure 1.7) :

- Une ou plusieurs sources de lumière large bande (lampes Xénon par exemple) associées à des filtres passe-bande pour sélectionner une bande d'excitation plus ou moins large centrée sur une longueur d'onde spécifique.
- Une ou plusieurs caméras CCD pour acquérir les images en lumière blanche et en fluorescence équipées de filtres passe-bande aux longueurs d'onde d'intérêt.
- Un filtre passe-haut placé sur la voie d'acquisition d'autofluorescence pour filtrer la lumière d'excitation rétro-diffusée.
- Une colonne vidéo-endoscopique conventionnelle (moniteur, enregistreur, fibroscope, cystocope).

1.3.5.1 Systèmes commerciaux d'imagerie endoscopique de fluorescence

Depuis plusieurs années, quelques industriels associés à des équipes de recherche ont développé des solutions technologiques en imagerie endoscopique de fluorescence/autofluorescence qui ont permis d'aboutir à la fabrication et à la commercialisation d'instruments utilisables en clinique.

Le premiers systèmes endoscopiques commercialisés pour la détection du cancer par fluorescence sont le système *D-Ligth* de la société *Karl Storz* (Tuttlingen, Allemagne) et le système

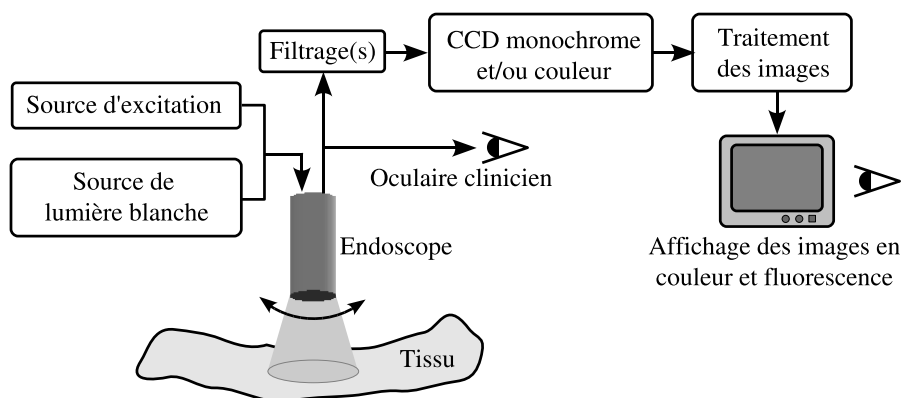


FIG. 1.7 – Schéma qui représente les composants d'un système d'imagerie endoscopique de fluorescence *in vivo*. La lumière blanche ou « bleue » (proche UV-violet) est injectée alternativement par le canal d'illumination de l'endoscope pour exciter le tissu. La lumière réfléchie est récupérée par l'optique de l'endoscope et transmise à un bloc de filtrage qui va sélectionner les bandes de longueurs d'onde d'intérêt. Les images dans ces bandes de longueur d'onde déterminées sont reçues par les caméras puis traitées pour améliorer le contraste des zones de tissu fluorescents avant d'être affichées sur un moniteur.

LIFE de la société *Xillix* (British Columbia, Canada). Ces systèmes sont apparus sur le marché au milieu des années 1990. Ensuite, d'autres ont été proposés, comme le *SAFE-3000* de la société *Pentax* (Japon) et le *DAFE* de la société *Richard Wolf GmbH* (Knittlingen, Allemagne). Les caractéristiques de ces systèmes sont résumés dans le tableau 1.2.

Tous les systèmes mentionnés dans le tableau 1.2 permettent l'observation en temps réel des images endoscopiques en fausses couleurs, formées à partir des rapports entre des images acquises dans des bandes spectrales différentes. De plus, ces systèmes permettent de faire une commutation à la demande entre le mode d'illumination lumière blanche et le mode (auto)fluorescence.

Il est important de noter que parmi ces systèmes, le seul conçu pour une application à la détection de lésions tumorales dans la vessie est le *D-Light* de *Storz*. Les autres systèmes sont plutôt utilisés en bronchoscopie et en gastroentérologie. Néanmoins, la société *Xillix* a annoncé en 1999 le système *LIFE-Bladder*, version modifiée de leur système *LIFE* traditionnel pour des applications en urologie.

Le fonctionnement des systèmes qui ont des applications en urologie est décrit de façon plus détaillée ci-après.

Le système *D-Light*. Pour l'excitation, ce système possède une lampe à arc Xénon couplée à un filtre passe-bas à 400 nm dont la lumière est injectée aux tissus à travers le canal d'illumination d'un fibroscope. La puissance d'excitation avoisine les 80 mW. La fluorescence des tissus est recueillie en utilisant un filtre passe-haut à 450 nm placé sur l'oculaire du fibroscope. Ce filtre permet le passage d'une petite partie de la lumière d'excitation (en plus de la fluorescence) pour améliorer le contraste de la zone observée. L'observation de l'autofluorescence était l'objectif de départ de ce dispositif, mais le fort taux de faux positifs obtenus a fait que la compagnie constructrice recommande l'utilisation d'un photosensibilisant (5-ALA) pour augmenter la fiabilité de détection.

Un avantage de ce dispositif est l'utilisation d'une technologie relativement simple qui diminue le nombre de variables à gérer.

TAB. 1.2 – Systèmes commerciaux d'imagerie de fluorescence/autofluorescence existants. λ_{ex} est la longueur d'onde centrale de la lumière d'excitation utilisée par le système. λ_{em} est la longueur d'onde d'émission centrale à laquelle le système effectue les acquisitions d'images en mode *autofluorescence*.

<i>Compagnie</i>	<i>Système</i>	<i>Longueurs d'onde</i>	<i>Application</i>
Xillix Technologies, British Columbia, Canada	LIFE (Light Induced Fluorescence Endoscopy)	$\lambda_{ex} = 422$ nm $\lambda_{em} = 630$ nm et 480 – 520 nm	Bronches, œsophage, poumons, système GI. Vessie ?
Pentax, Tokio, Japon	SAFE-3000	$\lambda_{ex} = 408$ nm $\lambda_{em} = 520 - 600$ nm	Bronches, larynx
Karl Storz, Tuttlingen, Germany	D-LIGHT	$\lambda_{ex} = 380 - 440$ nm $\lambda_{em} = 475 - 800$ nm	Vessie (mode fluores- cence), utérus, colon, bronches, œsophage, poumons
Richard Wolf GmbH, Knittlingen, Ger- many	DAFE (Diagnostic AutoFluorescence Endoscopy)	$\lambda_{ex} = 390 - 460$ nm $\lambda_{em} = 500 - 590$ nm et 600 – 700 nm	Bronches, œsophage

Le système LIFE. Le système *LIFE* dont l'acronyme veut dire *Light Induced Fluorescence Endoscopy* a été conçu pour la détection de lésions cancéreuses dans les bronches. Pour exciter le tissu, le système utilise un mécanisme de commutation pour sélectionner soit une lumière bleue (un laser à 422 nm) soit une lumière blanche (une lampe xénon). Un capteur CCD intensifié est utilisé pour détecter la lumière dans le rouge et le vert dans le mode fluorescence et une caméra couleur est utilisée pour l'observation d'images dans le mode normal. Le système possède aussi un système informatique d'acquisition de données et un moniteur pour l'affichage. Lorsque le mode fluorescence est sélectionné, la lumière produite par la source laser est projetée sur le tissu au travers du canal d'illumination d'un endoscope. Cette excitation produit une fluorescence des tissus éclairés qui est captée par le capteur CCD intensifié. Les bandes spectrales captées par le système se trouvent l'une dans le vert (480 – 520 nm) et l'autre dans le rouge (> 630 nm). Une fluorescence autour de 500 nm (vert) est observée pour un tissu normal et cette fluorescence diminue de façon appréciable pour un tissu cancéreux, laissant apparaître des taches brunâtres à l'endroit de ces tissus. Cela est dû à une augmentation de l'autofluorescence des porphyrines endogènes présentes en plus grande quantité dans les tissus néoplasiques qui absorbent des longueurs d'onde dans le vert et réfléchissent plus dans le rouge. Même si dans le domaine de la bronchoscopie et de la gastroentérologie il existe plusieurs études qui montrent des augmentations de la sensibilité [Prosst and Gahlen, 2002] il n'existe aucun rapport de l'utilisation réussie de ce dispositif pour la détection de lésions cancéreuses dans la vessie.

Un gros avantage de ce système est l'utilisation de l'autofluorescence des tissus pour détecter les lésions cancéreuses et donc l'application d'un composant exogène n'est pas nécessaire. Cette caractéristique augmente aussi la sensibilité de la détection. Cependant, ce système ne fonctionne que dans le spectre lumineux visible et donc il n'est pas possible d'exciter des fluorophores qui réagissent en dehors de cette bande (comme le *Tryptophane*).

Plusieurs travaux montrent l'intérêt de l'utilisation de l'autofluorescence en imagerie endoscopique pour améliorer la perception de contraste entre zones saines et tumorales des muqueuses sur tissus frais excisés [Svistun *et al.*, 2004] et sur tissus *in vivo in situ* [Kara *et al.*, 2004] [Zeng *et al.*, 2004]. Sur ce point, de nombreuses études recensées par [Kara *et al.*, 2004] ont déjà été menées

sur l'utilisation conjointe de l'imagerie à base d'autofluorescence (EAF, ou endoscopie d'autofluorescence) et de l'endoscopie conventionnelle (ELB, ou endoscopie en lumière blanche) pour la détection de lésions néoplasiques intraépithéliales bronchiques (plusieurs milliers de patients).

Les résultats d'une étude randomisée sur 1173 patients soumis à une bronchoscopie pour la détection de lésions néoplasiques [Häußinger *et al.*, 2005], montrent que la sensibilité de la combinaison de l'ELB et de l'EAF (82,3%) est supérieure à celle obtenue par l'ELB seule (57,9%), mais sa spécificité est moindre : (58,4%) pour l'EAF contre (62,1%) pour l'ELB.

Un recensement concernant le diagnostic de lésions colorectales (dysplasie et carcinome) par spectroscopie et imagerie de fluorescence a été effectuée par [Prosst and Gahlen, 2002]. À l'aide des systèmes de *Xillix* et de *Karl Storz*, plusieurs études montrent que la combinaison de l'EAF et de l'ELB permet d'augmenter significativement la sensibilité (valeurs qui varient entre 83% et 100% suivant l'étude) et la spécificité (entre 70% et 90%).

[Jang *et al.*, 2006] dans une application à la détection de néoplasmes intra-épithéliaux dans les bronches rapportent une sensibilité de 84% et une spécificité de 50% pour l'utilisation combinée de l'EAF et de l'ELB contre 26% et 84% pour l'ELB seule.

Actuellement, des études concernant l'utilisation de l'autofluorescence pour le diagnostic de lésions cancéreuses dans la vessie sont très limitées. Néanmoins, il existe de nombreuses études utilisant des photosensibilisants pour la fluorescence induite qui rapportent des améliorations de la sensibilité (96%) par rapport à l'ELB, mais des spécificités inchangées [D'Hallewin *et al.*, 2002].

En résumé, même si ces études montrent le potentiel de l'EAF pour améliorer la détection de lésions dysplasiques/pré-cancéreuses, il existe encore d'importantes limitations de la technique et plusieurs aspects à améliorer :

- Le grand nombre de faux positifs (et la faible valeur prédictive positive résultante) qui implique l'utilisation de méthodes complémentaires pour améliorer la spécificité.
- La qualité des images de l'EAF est inférieure à celle de l'ELB. Ceci suggère un raffinement des techniques de détection.
- Les techniques d'autofluorescence utilisent principalement des spectres dans le visible ou l'UV, et donc les images résultantes correspondent à des profondeurs limitées de la muqueuse.
- Les méthodes d'autofluorescence sont basées sur la détection des changements biochimiques qui ont seulement un rapport *indirect* avec la dysplasie et la néoplasie. Les études d'anatomopathologie sont basées aussi sur des différences morphologiques. Ceci implique l'utilisation de méthodes complémentaires.

Pour la détection de lésions néoplasiques dans la vessie, le système développé par *Storz* donne une sensibilité accrue par l'utilisation d'un photosensibilisant (ALA) ce qui diminue sensiblement la spécificité de l'examen. Nous ne pouvons pas affirmer que ce soit mauvais car même ces systèmes ne disposent pas aujourd'hui des biopsies. L'amélioration de la sensibilité oriente mieux les biopsies que la simple exploration en lumière blanche.

1.3.5.2 Le système de cystoscopie de fluorescence développé au CRAN

Tous les systèmes d'endoscopie étudiés dans la section précédente permettent d'acquérir des images en lumière blanche et en fluorescence. Cependant, les caractéristiques de ces systèmes permettent d'acquérir des séquences d'images *soit* en mode lumière blanche *soit* en mode lumière bleue. Par conséquent, si l'on veut effectuer en lumière blanche des observations morphologiques ou colorimétriques d'une zone d'intérêt dans un organe et effectuer des observations en lumière bleue de la même zone pour obtenir des informations concernant la fluorescence, il faut acquérir deux séquences vidéo de la même zone (correspondances spatiales des images en lumière blanche

et en fluorescence) mais à des instants différents. Cette modalité n'est pas implémentable telle quelle sur les systèmes commerciaux présentés.

L'équipe SpID du groupe IPS du CRAN, développe actuellement en collaboration avec le BioMoCeTi (Laboratoire de Biophysique Moléculaire, Cellulaire et Tissulaire) à Paris et le LETI (Laboratoire d'Électronique de Technologie de l'Information) du CEA à Grenoble, un système cystoscopique qui doit permettre d'acquérir de façon simultanée des images de type conventionnel (en lumière blanche) et des images en fluorescence/autofluorescence (en lumière bleue) dans trois bandes spectrales différentes. Un premier prototype a déjà été développé par le CRAN et le BioMoCeTi et avait permis de réaliser une première validation du principe [Blondel *et al.*, 2005].

L'originalité de notre système réside dans le fait de pouvoir commuter les sources de lumière blanche et de fluorescence à une cadence idéale de 15–20 images/seconde. Cette commutation permettrait d'acquérir des images en lumière blanche et en fluorescence de façon « quasi-simultanée » (plus exactement multiplexées dans le temps). De cette façon il devient possible de comparer les *informations spatiales* données par les images acquises en lumière blanche et les *informations fonctionnelles* données par les images de fluorescence. Il est important de noter que le développement de l'instrumentation en rapport avec le système d'endoscopie du CRAN ne fait pas partie de cette thèse, néanmoins nous utilisons ce système pour effectuer des tests concernant le traitement d'images. Les paragraphes suivants décrivent notre système.

La partie excitation du système est composée de deux lampes à arc Xénon. Une lampe est couplée à un filtre passe-bande ($410 \text{ nm} \pm 20 \text{ nm}$) pour avoir une lumière dans le bleu. La deuxième lampe est utilisée pour l'illumination en lumière blanche. Le logiciel développé contrôle par une interface numérique un diaphragme mécanique permettant d'injecter alternativement EAF ou EAF+ELB. La lumière est conduite vers le tissu au travers du canal d'illumination d'un cystoscope. Le système de commutation permet d'alterner entre la source lumineuse de lumière blanche et la source de lumière bleue. La lumière réflétee par le tissu est captée par trois capteurs CCD monochromes dont deux sont couplés à des filtres passe-bande amovibles, il est alors possible de placer différents filtres dont les longueurs d'onde varient entre 550 et 590 nm. Les capteurs CCD filtrés imagent la fluorescence émise par les tissus lorsqu'ils sont excités avec la lumière bleue. Un rapport entre les images acquises par les capteurs CCD filtrés est effectué pour améliorer le contraste des zones qui présentent une fluorescence différente.

Une reconstruction colorimétrique numérique est effectuée avec les images captées par les capteurs CCD pendant l'éclairage en lumière blanche pour simuler les bandes RGB d'une image acquise par une caméra couleur. Il est important de noter que l'acquisition des images en fluorescence se fait de façon simultanée dans trois bandes spectrales différentes. Les images captées par les CCD sont envoyées en temps réel vers un ordinateur où s'effectue le post-traitement et l'affichage des images acquises. Actuellement la cadence d'acquisition du système est de 2 images/seconde, mais l'objectif final est d'atteindre 15–20 images/seconde. Les composants du système sont schématisés dans la figure 1.8.

L'intérêt de cet instrument pour notre thèse est que sa vitesse de commutation rend possible d'avoir deux jeux d'images, l'un composé des images acquises sous lumière blanche et un autre composé des images acquises pendant l'excitation en lumière bleue. Ces images sont pratiquement superposables puisqu'elles sont acquises (idéalement) à des intervalles d'environ 80 ms de différence. Ces images vont nous permettre de combiner dans un support panoramique unique, les informations spatiales (images en lumière blanche) et les informations fonctionnelles (images de fluorescence) (cf. chapitre 4).

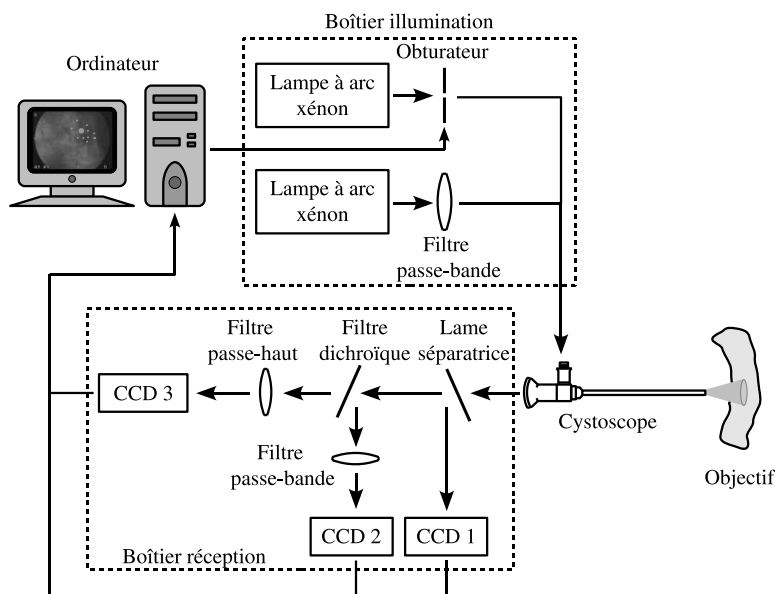


FIG. 1.8 – Schéma qui représente les composants du système d'imagerie endoscopique de fluorescence développé dans notre groupe de travail.

1.4 Motivation clinique des travaux

Cette thèse fait partie d'un projet qui vise la détection précoce de lésions cancéreuses dans la vessie par endoscopie de fluorescence. Comme il a été vu dans la section 1.1.3, les *tumeurs vésicales superficielles* peuvent poser deux problèmes évolutifs importants : un risque de récurrence sans progression, mais surtout, un risque de progression. Une grande partie des dysplasies et des CIS peuvent rapidement progresser et envahir la couche musculaire. Donc, *une détection précoce efficace et localisée est vitale pour pouvoir individualiser et commencer le traitement, ainsi que pour limiter les récurrences après le traitement*. D'autre part, le risque de progression pour les CIS varie fortement suivant le nombre de foyers tumoraux et le risque de récurrence est plus important suivant la taille de la tumeur. Il est donc important de pouvoir *évaluer* de façon effective lors d'un examen clinique, *la taille et la distribution spatiale des lésions cancéreuses* dans la vessie.

La motivation principale de notre thèse est l'amélioration de la détection précoce des tumeurs vésicales et de leur étendue en proposant au clinicien un *nouveau support visuel* sous la forme d'images panoramiques de la vessie construites à partir de séquences d'images cystoscopiques. Les sous-sections suivantes présentent l'état actuel des supports visuels utilisés dans l'examen cystoscopique et expliquent les raisons pour lesquelles une nouvelle représentation est nécessaire et possible.

1.4.1 Supports visuels utilisés dans l'examen cystoscopique actuel

Actuellement, la cystoscopie est l'examen de référence qui permet de faire le diagnostic macroscopique et de réaliser des résections ou des biopsies de la tumeur et des zones suspectes. Elle permet accessoirement (avec l'équipement adéquat) d'enregistrer la séquence vidéo et de prendre des photographies de « petites » zones d'intérêt. Le support photographique permet de conserver une représentation précise de la zone, d'effectuer des comparaisons dans le temps en cas de récurrence et de disposer d'un document objectif médico-légal [Gattengo and Chopin, 2001].

Actuellement, ces photos et les marqueurs placés par le clinicien sur un dessin de la vessie sont les supports utilisés pour garder une trace de l'examen effectué. Le support vidéo est moins utilisé car son coût est élevé, la relecture des séquences vidéo est fastidieuse et prend du temps.

Limites des supports visuels utilisés dans la cystoscopie conventionnelle. Le clinicien note à la main sur un croquis de la vessie les sites où il a observé des anomalies, pris des photos ou réalisé des biopsies. Un enregistrement vidéo peut être utilisé, mais le fait de chercher une image spécifique au sein d'une séquence d'images complète peut s'avérer contraignant (surtout si cette tâche est répétée pour chaque patient). De plus, quand le clinicien réalise la cystoscopie il oriente son appareil et localise les zones qu'il observe. Lorsqu'il regarde une vidéo enregistrée, il n'a pas cette aide et doit reconstruire « de tête » la position et l'orientation du cystoscope. Par ailleurs, le *champ de vision* limité du cystoscope ne permet pas d'estimer de façon précise l'extension d'une lésion cancéreuse ni la distribution spatiale des éventuelles lésions multifocales. La localisation des lésions sur un croquis réalisé par le clinicien est hautement dépendant de son expérience.

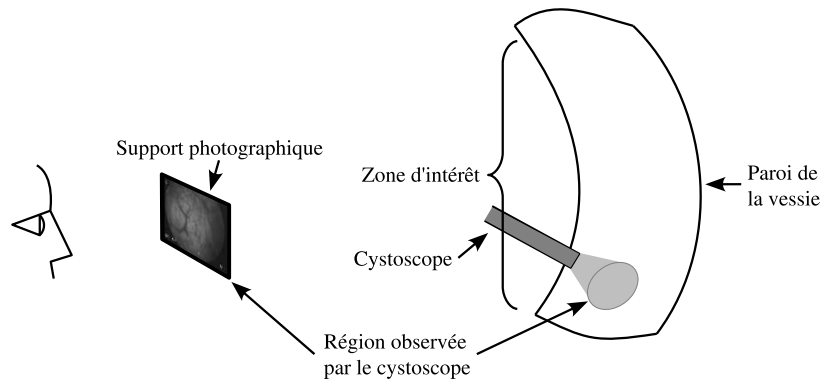


FIG. 1.9 – Illustration de la petite région pouvant être relevée par un support photographique cystoscopique actuel.

1.4.2 Intérêt de la combinaison des informations spatiales et fonctionnelles

La nécessité d'un support plus adapté pour localiser, évaluer et comparer des zones suspectes ainsi que pour préparer et guider des examens postérieurs a été exprimée par les urologues. Du point de vue du diagnostic, il est intéressant pour le clinicien de pouvoir étudier une lésion et la zone qui l'entoure dans son ensemble sans devoir chercher une image particulière ou passer d'une image à la suivante pour passer en revue la totalité de la zone. Le fait de pouvoir observer de façon globale une zone d'intérêt doit permettre de mieux localiser des petites lésions satellites (multifocales) potentiellement à l'origine de récurrences. Une représentation globale qui puisse permettre ce genre d'observation peut être obtenue en utilisant le mosaïquage d'images.

En cystoscopie classique (en lumière blanche), les tumeurs macroscopiques peuvent être détectées et donc traitées, c'est beaucoup moins facile pour les lésions dysplasiques et de type CIS. C'est la raison pour laquelle des biopsies « aléatoires » de la muqueuse vésicale sont effectuées. Les dysplasies « invisibles » sont majoritairement à l'origine de nombreuses récurrences tumorales qui auraient pu être évitées si elles avaient été décelées et détruites.

La cystoscopie de fluorescence en utilisant des photosensibilisants (comme par exemple 5-ALA sous excitation en lumière bleue), permet de détecter des lésions (dysplasies et CIS) qui

ne sont pas visibles en cystoscopie conventionnelle (en lumière blanche). Cette méthode présente une sensibilité proche de 100% mais une spécificité d'environ 60% en moyenne selon les études. Même si la méthode par fluorescence est supérieure à la cystoscopie conventionnelle dans la détection des lésions infra-cliniques (dysplasies et CIS) de la muqueuse vésicale, le taux élevé de faux positifs de la cystoscopie de fluorescence est une limite de l'examen.

La cystoscopie d'autofluorescence présente l'avantage de n'utiliser aucun marqueur exogène en excitant les fluorophores intrinsèques du tissu. L'application de la cystoscopie d'autofluorescence permet d'atteindre des taux de spécificité d'environ 95% [Gattengo and Chopin, 2001], mais des études supplémentaires sont nécessaires pour vérifier ces résultats et disposer d'outils utilisables en clinique.

Donc, un système combinant les *informations spatiales* d'images globales (images panoramiques d'une zone d'intérêt et des zones voisines) et les *informations fonctionnelles* de fluorescence ou d'autofluorescence, peut contribuer à améliorer la détection de lésions néoplasiques de la vessie invisibles ou inaperçues en cystoscopie conventionnelle. Elle devrait aussi améliorer la surveillance après résection.

1.5 Objectifs de cette thèse

Cette thèse s'inscrit dans la poursuite des travaux développés par [Miranda Luna, 2005] lequel a proposé une méthode de correction radiale et de mosaïquage d'images basée sur l'utilisation de l'information mutuelle optimisée par une technique de gradient descendant stochastique. L'un des objectifs de notre thèse est de proposer une méthode de recalage alternative pour améliorer la vitesse de calcul, qui tout en gardant la robustesse, permette de construire des images panoramiques de l'intérieur de la vessie au cours de l'examen clinique.

La construction de ces images panoramiques implique le choix d'une méthode « optimale » pour recalculer et ensuite coller les images de la séquence vidéo-cystoscopique. L'évaluation quantitative de la méthode proposée pour construire les images panoramiques en utilisant une instrumentation adaptée ainsi que son application à des séquences « réelles » acquises en clinique font aussi l'objet des travaux de cette thèse.

Nous proposons également de vérifier la faisabilité de la combinaison des informations spatiales et fonctionnelles dans un support panoramique unique. Cette partie utilise le système développée au sein du projet SpID au CRAN et décrit dans la section 1.3.5.2.

Les objectifs des travaux sont illustrés dans la figure 1.10.

1.6 Problématique scientifique

Dans cette section nous présentons les difficultés que nous devons considérer pour parvenir aux objectifs fixés et satisfaire les contraintes cliniques qui motivent cette thèse.

1.6.1 La qualité des images cystoscopiques

Les caractéristiques des images cystoscopiques dépendent de l'instrument utilisé pour acquérir les images (le cystoscope) et de la qualité de l'examen cystoscopique réalisé par le clinicien.

Effets produits par l'optique du cystoscope. Le cystoscope est un instrument fin qui, inséré au travers de l'urètre, permet d'observer les parois internes de la vessie. Il existe deux types de cystoscopes qui peuvent être utilisés pour un examen clinique : les cystoscopes rigides

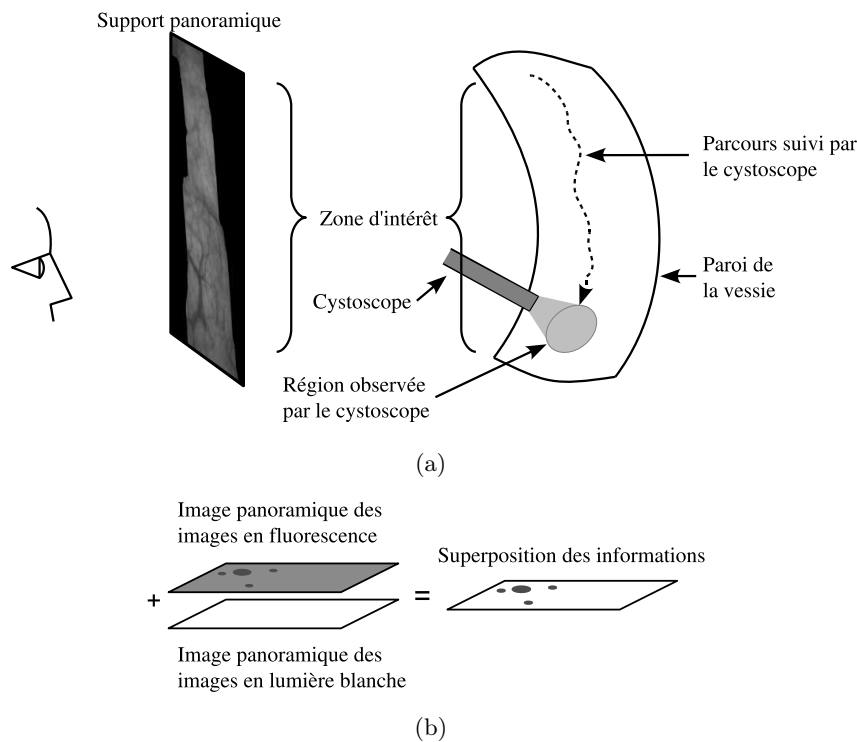


FIG. 1.10 – Représentation schématisée des objectifs fixés. (a) Construction d’images panoramiques de l’intérieur de la vessie. (b) Évaluation de la superposition des informations fonctionnelles données par une image panoramique construite à partir d’images acquises en lumière blanche et des informations spatiales données par une image construite à partir d’images acquises en fluorescence.

et les souples. Le cystoscope rigide est un dispositif composé d’un fin tube métallique à l’intérieur duquel se trouve l’optique qui transmet les images reçues par l’objectif situé dans la partie distale de l’instrument. Le cystoscope souple est un dispositif qui utilise des longs faisceaux de fibres optiques pour transmettre la lumière et les images de la scène observée. La figure 1.11 montre des exemples des ces dispositifs.

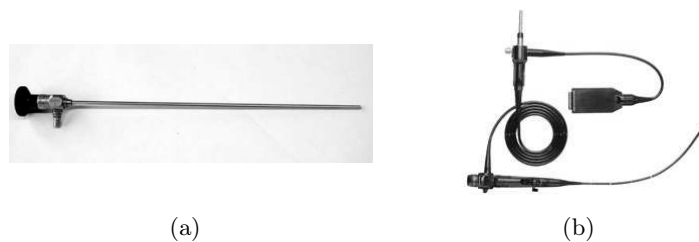


FIG. 1.11 – Les deux types de cystoscope utilisés en clinique. (a) Cystoscope rigide *Karl Storz* 27005BA. (b) Cystoscope souple *Olympus EndoEYE*.

Pour les deux types de cystoscope, l’optique impliquée dans la transmission des images observées, va provoquer des distorsions géométriques et des aberrations photométriques dans les images. Les deux principaux effets introduits par l’optique qui peuvent affecter la méthode de construction d’images panoramiques sont la *distorsion radiale* et le *vignettage*. La distorsion ra-

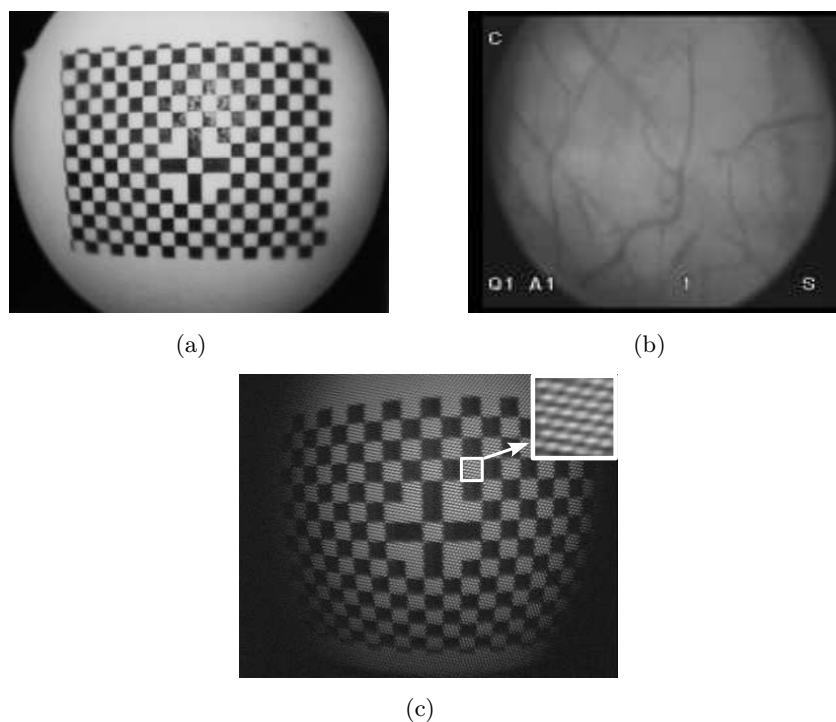


FIG. 1.12 – Défauts d'optique du cystoscope. (a) Distorsion radiale. Les lignes formées par les carreaux devraient être droites. (b) Effet de vignettage. (c) Pattern formée par les fibres optiques lorsqu'un cystoscope souple est utilisé.

diale est un effet optique géométrique qui fait que l'image a un aspect de « barillet » au fur et à mesure de l'augmentation de la distance au centre de l'image. Le vignettage est un effet photométrique qui fait apparaître les bords de l'image plus sombres que le centre. Lorsqu'un cystoscope souple est utilisé, une autre dégradation va affecter les images cystoscopiques. Le faisceau de fibres utilisé par le cystoscope souple fait apparaître sur les images un « pattern » en forme de nid d'abeilles produit par les différents taux de réfraction de lumière entre le cœur et l'enveloppe des fibres optiques. Des exemples de ces deux effets sont montrés dans la figure 1.12. Ces trois effets seront étudiés plus en détails dans la section 2.3

Effets produits pendant l'examen clinique. Pour effectuer une cystoscopie, le clinicien remplit la vessie avec du sérum physiologique. Cela distend les parois de la vessie et permet de les « lisser » pour en effectuer une observation plus aisée. Le clinicien insère lentement l'instrument à travers l'urètre du patient jusqu'à la vessie. Il va prendre comme référence les orifices des uretères et une « bulle d'air ». Cette bulle d'air est une petite quantité d'air introduite avec le cystoscope due à une purge incomplète de la tubulaire servant à remplir la vessie et du cystoscope lui même. La figure 1.13 montre une illustration d'un examen cystoscopique clinique en utilisant un cystoscope rigide.

Du fait de la morphologie de la vessie, du liquide avec lequel elle est rempli, de l'illumination, des réglages et des mouvements effectués par le cystoscope à l'intérieur de la vessie, la qualité des images cystoscopiques peut varier de façon importante dans une même séquence vidéo-cystoscopique et évidemment d'un patient à un autre. Cette variabilité peut être problématique pour la méthode de construction d'images panoramiques. Par exemple, quelques séquences

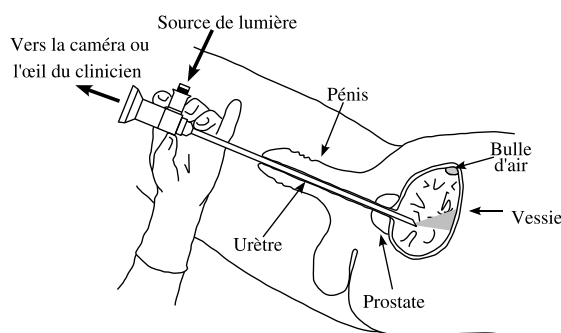


FIG. 1.13 – Illustration de la procédure d’application d’un examen cystoscopique clinique en utilisant un cystoscope rigide.

d’images peuvent avoir des structures très visibles (figure 1.14(a)) tandis que d’autres images en auront très peu (figure 1.14(e)). Il existe des images où le champ de vision est partiellement occulté par le relief des parois latérales (1.14(i)). La vessie étant remplie, on observe communément des éléments en suspension qui circulent dans la zone d’observation du cystoscope (figure 1.14(b)). Le réglage de la focale du cystoscope va affecter les images observées et il est possible de trouver des images floues (figure 1.14(c)). Cependant, des images floues peuvent être aussi produites par des mouvements très rapides du cystoscope (figure 1.14(c)). Parfois, les images seront acquises quand le cystoscope est assez proche des parois et les images acquises auront une très bonne illumination avec une bonne visibilité des grosses structures (figure 1.14(f)), mais il est aussi possible que l’illumination et la distance d’observation du cystoscope produise des images où les détails sont difficilement visibles (figure 1.14(g)). Il est fréquent de trouver sur certaines zones de la vessie, des réflexions importantes qui produisent de petites zones très lumineuses (figure 1.14(h)). Finalement, une caractéristique majeure présente sur toutes les images cystoscopiques est la différence d’illumination entre les différentes régions de l’image. Cette caractéristique est plus évidente sur la figure 1.14(j).

En plus de la variabilité au niveau des caractéristiques des images cystoscopiques, les mouvements « libres » du cystoscope à l’intérieur de la vessie font que, d’une image de la séquence à la suivante, il existe des variations géométriques plus ou moins importantes suivant la vitesse de déplacement du cystoscope, la distance du cystoscope à la paroi de la vessie, le point de vue et l’angle d’observation du cystoscope.

1.6.2 Contraintes cliniques imposées sur la méthode de construction des images panoramiques

Informations disponibles. La position spatiale du cystoscope à l’intérieur de la vessie est inconnue et l’utilisation d’une technique pour permettre de situer le cystoscope à l’intérieur de la vessie n’est pas faisable sans modifier le dispositif traditionnel. En plus, l’utilisation de repères anatomiques pour guider la construction d’images panoramiques n’est pas envisageable parce qu’ils ne sont pas visibles la plupart du temps. La seule information disponible pour construire les images panoramiques sont les images acquises par le cystoscope. Actuellement, lorsque le clinicien veut garder la trace d’une région d’intérêt, il fait l’impression d’une photographie ou enregistre une vidéo. Pour construire les images panoramiques, notre méthode n’utilisera donc que l’information fournie par les images d’une vidéo.

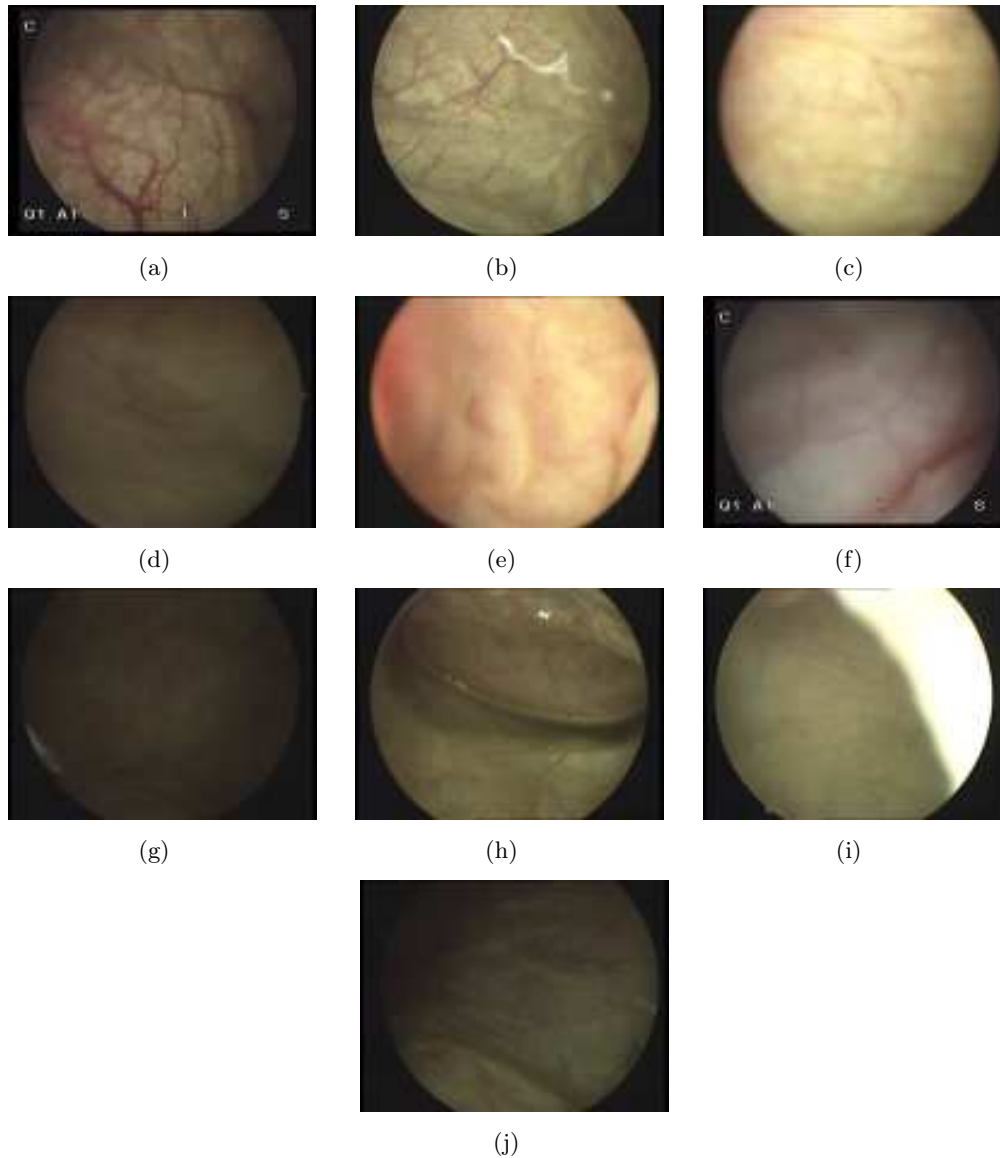


FIG. 1.14 – Quelques exemples d'images pouvant être extraites d'une séquence vidéocystoscopique. (a) Image avec une grande quantité de vaisseaux visibles. (b) Structure qui passe dans le champ de vision du cystoscope. (c) Image floue dû à une mauvaise focalisation du cystoscope. (d) Image floue dû à des déplacements rapides du cystoscope. (e) Image avec très peu de structures visibles. (f) Image prise à une courte distance. (g) Image avec une très faible illumination. (h) Image présentant des reflets sur la bulle d'air. (i) Image avec occlusion. (j) Image où les différences d'illumination sont très évidentes.

Degré de recouvrement entre les images. Si les images panoramiques sont construites à partir du support vidéo enregistré par le clinicien, il faut considérer le *degré de recouvrement* entre les images de la séquence vidéo. La vitesse des mouvements du cystoscope est un facteur important à considérer pour choisir une méthode de construction des images panoramiques parce qu'elle va nous indiquer le degré de recouvrement entre les images. Nous savons que les images cystoscopiques enregistrées sur un support vidéo sont acquises à une vitesse de 25 images par seconde mais nous ne connaissons pas la vitesse maximale de déplacement du cystoscope ni le champ de vue exact. Il est donc nécessaire d'évaluer cette vitesse de déplacement, ainsi que le champ de vue moyen pour pouvoir connaître le recouvrement minimum envisageable entre les images et ainsi aider au choix d'une méthode de construction d'images panoramiques.

Erreur acceptable dans la construction des images panoramiques. Le support visuel proposé doit permettre au clinicien d'effectuer des observations post-opératoires, de le guider dans des examens postérieurs et de lui permettre d'effectuer des comparaisons entre différents examens. Il est donc nécessaire que la qualité de ce support permette d'effectuer ces opérations avec une erreur minimale. Idéalement, l'erreur de recalage entre les images devrait être nulle, mais en pratique cela n'est pas possible. Il sera donc nécessaire de caractériser cette erreur et de justifier une valeur limite acceptable sur la base de données qualitatives objectives.

Temps de calcul disponible. Si le clinicien veut effectuer des observations des zones d'intérêt qu'il a sélectionnées immédiatement après avoir fini d'effectuer l'examen de la vessie, il sera nécessaire que la méthode de construction des images panoramiques construise les images dans un temps qui soit compatible avec la durée de l'examen clinique. Idéalement, la méthode devrait construire les images en temps réel, mais dans un premier temps pour cette thèse, nous allons considérer la durée moyenne d'un examen clinique comme temps limite de construction des images. Les informations trouvées dans la littérature concernant la durée d'un examen cystoscopique sont très variables. Si c'est la première fois que le patient est examiné, il est probable que la durée de l'examen sera plus importante parce qu'il faut balayer de façon « exhaustive » la vessie. Mais, si le clinicien effectue un examen de contrôle sur le patient, cet examen sera plus court.

Présentation des images panoramiques au clinicien La représentation en couleur des images panoramiques au clinicien doit être privilégiée et les artefacts introduits par les différences d'illumination entre les bords des images doivent être minimisés. Les images panoramiques devront avoir la même résolution que les images originales pour conserver le niveau de détail.

La combinaison des informations spatiales et de fluorescence serait mieux observée dans un support panoramique unique. Une zone de tissu plus étendue pourrait être observée avec ce support et aider à localiser des zones de fluorescence différente en une seule observation.

Une extension du support panoramique que nous proposons serait de présenter au clinicien un modèle générique 3D de la vessie sur lequel les images panoramiques seraient projetées.

1.6.3 Validation de la méthode proposée

Il est nécessaire d'effectuer une évaluation quantitative de la méthode de mosaïquage qui sera développée pour connaître ses limites et assurer qu'elle sera compatible avec les séquences réelles. Pour effectuer cette évaluation quantitative, il sera nécessaire de concevoir un protocole et l'instrumentation qui puissent permettre de mesurer les erreurs induites par la méthode de

construction des images panoramiques et les erreurs dans les images résultantes par rapport aux scènes réelles desquelles elles sont une représentation.

Dans une seconde étape, la méthode de recalage doit être appliquée à des séquences réelles et la qualité des images panoramiques doit être évaluée. Cette évaluation sera qualitative puisque les séquences cystoscopiques que nous allons utiliser seront enregistrées pendant un examen clinique conventionnel sans aucune instrumentation spécifique supplémentaire.

D'autre part, l'évaluation de la faisabilité de la combinaison d'informations spatiales et fonctionnelles dans un support unique nécessite un système capable d'effectuer des acquisitions d'images cystoscopiques en lumière blanche et en fluorescence de façon simultanée et contrôlée. Nous utiliserons le système d'endoscopie de fluorescence actuellement disponible au sein du laboratoire (sous-section 1.3.5.2).

1.6.4 Conclusion

La construction d'images panoramiques de l'intérieur de la vessie est un problème très spécifique qui impose le développement d'outils spéciaux et adaptés aux conditions instrumentales et environnementales particulières dans lesquelles sont effectuées les observations. La méthode proposée devra donc tenir compte des contraintes imposées par ces conditions. Comme nous verrons dans la section 2.2.1, les domaines d'application du mosaïquage d'images sont très variés. Cependant, en médecine les applications sont peu nombreuses et pratiquement inexistantes dans le domaine de l'urologie. En tenant compte de la problématique exposée nous pouvons déjà établir une démarche à suivre pour arriver aux objectifs fixés. Il faudra donc :

1. Étudier les caractéristiques des images cystoscopiques pour établir des méthodes de traitement adaptées.
2. Rechercher les méthodes adaptées pour construire les images panoramiques à partir de la seule information dont nous disposons : les images de la séquence cystoscopique.
3. Proposer une présentation des images panoramiques au clinicien.
4. Évaluer les erreurs de construction des images panoramiques issues de l'application de l'algorithme de mosaïquage. Pour cela, il est nécessaire de proposer un protocole pour l'acquisition contrôlée de séquences d'images avec une instrumentation adaptée.
5. Évaluer qualitativement l'application de l'algorithme de mosaïquage aux séquences cystoscopiques « réelles » acquises en clinique.
6. Explorer la combinaison des *informations spatiales* données par une image panoramique construite à partir d'images acquises en cystoscopie conventionnelle et des *informations fonctionnelles* données par une image panoramique construite à partir d'images de fluorescence.

Il existe aussi des caractéristiques que la méthode proposée devra posséder pour avoir une vraie utilité en clinique. La méthode pour construire les images panoramiques devra être :

- Automatique : pour limiter le temps de construction de l'image et pour éviter le besoin de connaissances techniques expertes.
- Précise : pour représenter de façon correcte la géométrie de la région de la vessie qui a été observée.
- Robuste : pour pouvoir gérer la variabilité existante dans les séquences d'images cystoscopiques et les déformations produites par les mouvements du cystoscope à l'intérieur de la vessie.
- Capable de produire des images photo-réalistes : pour représenter correctement les couleurs de la scène observée.

- Efficient : au niveau du coût de calcul pour réduire le temps de construction pour qu'il soit compatible avec la durée d'un examen clinique et pour pouvoir fonctionner sur des matériels informatiques courants.

2

Principales méthodes et applications du mosaïquage d'images

Ce chapitre présente le principe du mosaïquage et du recalage d'images et fait un bilan de l'utilisation de cette technique dans le domaine médical.

Cependant, avant de présenter le principe du recalage d'images, nous allons introduire le modèle décrivant la projection d'une scène 3D acquise par une caméra dans le plan image. La connaissance de ce modèle permet de faciliter la présentation de certains aspects du principe du recalage d'images.

Après avoir expliqué le principe du mosaïquage et du recalage d'images, nous détaillerons les parties du processus du recalage qui sont communes à tous les algorithmes de recalage : choix de la transformation géométrique liant deux images, méthodes d'interpolation des niveaux de gris, optimisation des transformations géométriques et placement des images recalées dans un repère commun à toutes les images. Les autres parties du processus de recalage sont détaillées en classant les algorithmes en deux groupes : ceux basés sur les niveaux de gris et ceux utilisant des primitives images. Des exemples illustratifs permettant de compléter l'état de l'art sont donnés.

La fin du chapitre est consacrée aux défauts caractérisant les images endoscopiques (forte distorsion radiale, hétérogénéités d'illumination). Ces défauts sont à prendre en compte lors du choix d'une méthode de mosaïquage d'images cystoscopiques.

En conclusion de ce chapitre nous faisons un premier tri parmi les méthodes étudiées dans la littérature et nous justifions les pistes à suivre dans le chapitre 3.

2.1 Modèle de formation d'images

2.1.1 Projection perspective

Il existe plusieurs modèles qui décrivent le processus de formation d'images. Le plus simple (et le plus utilisé) est le *modèle sténopé* de formation d'images (les droites dans l'espace sont projetées en droites dans le plan) ou *pinhole camera model* en anglais. Le principe de ce modèle peut être expliqué en considérant une boîte dont l'une des faces possède une petite perforation et dont la face opposée est remplacée par une plaque transparente (plan image). L'image sera formée par les rayons de lumière issus de la scène et passant par le trou. Si le trou est réduit à un point, un photon qui est issu de la scène et qui tombe sur un point particulier de l'image en

passant par le sténopé ne peut suivre qu'une trajectoire rectiligne unique. En réalité, le sténopé a une dimension finie et chaque point du plan image va collecter des rayons de lumière formant un cône. Par ailleurs, les caméras réelles possèdent des lentilles qui ne sont pas prises en compte par le modèle sténopé. Néanmoins, le *modèle sténopé de projection perspective* (aussi connu sous le terme *projection centrale perspective*) permet de modéliser correctement d'un point de vue mathématique une caméra idéale [Forsyth and Ponce, 2003]. Ainsi illustré par la figure 2.1, la projection perspective va créer des images inversées. Mis à part l'inversion, l'image fournie par une caméra est strictement équivalente à celle du plan image. En effet, l'image fournie par une caméra peut-être vue comme une *image virtuelle* qui est la symétrique de celle qui est dans le plan image, ces deux plans étant situés à une même distance de part et d'autre du sténopé. Cette image est formée sur le plan placé en face du sténopé à la même distance que le plan image (figure 2.1).

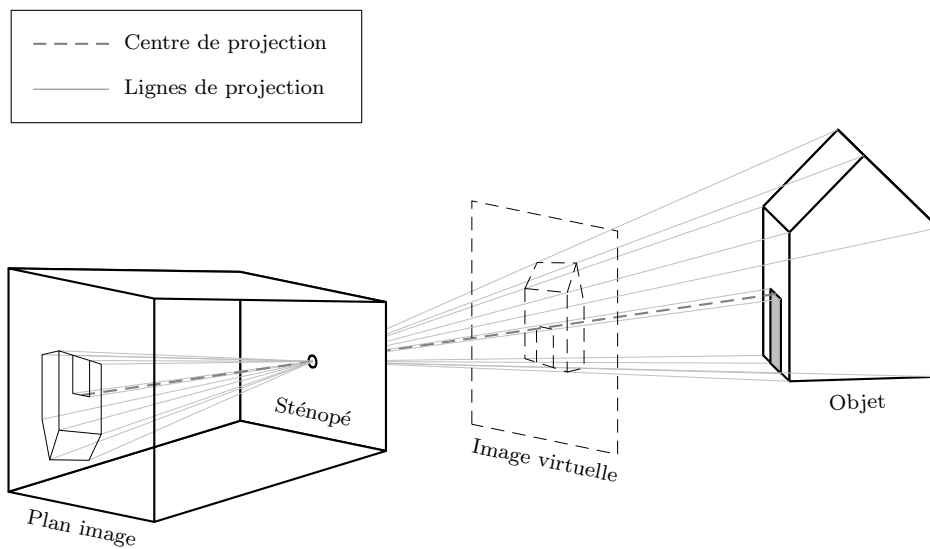


FIG. 2.1 – Représentation du modèle sténopé de formation d'images.

L'image virtuelle de la figure 2.1 définit le plan et le *repère image* $(O, \vec{x}_c, \vec{y}_c)$ de la figure 2.2. Le point O correspond au sténopé (centre optique de la caméra). La ligne perpendiculaire au plan image qui passe par le sténopé est appelée l'*axe optique* et le point C de coordonnées (u_0, v_0) , est le *centre optique projeté dans l'image*. Ce point peut être utilisé comme origine O du repère image $(\vec{x}_{im}, \vec{y}_{im})$, notamment lors de la calibration d'une caméra. À partir de cette représentation, il est possible de calculer la projection d'un point \mathbf{p}_{3D} de coordonnées (x, y, z) dans le repère 3D monde, sur le repère image (O', \vec{u}, \vec{v}) dont l'origine est le coin supérieur gauche de l'image. Les coordonnées (u, v) 2D de la projection sont calculées en passant par le repère (C, x_{im}, y_{im}) et en tenant compte de la position, de l'orientation et des paramètres optiques de la caméra utilisée pour acquérir l'image. Les paramètres qui décrivent la position et l'orientation de la caméra par rapport à la scène sont appelés *paramètres extrinsèques*. Les paramètres liés à l'optique et à la matrice CCD (Charge-Coupled Device) de la caméra (taille des pixels, zoom, distance focale) sont appelés *paramètres intrinsèques*. Ces derniers paramètres sont indépendants de la position de la caméra.

Le processus de projection passe par 3 étapes : (1) le passage d'un point 3D du repère monde dans le repère caméra, (2) la projection du point 3D du repère caméra $(O, \vec{x}_c, \vec{y}_c, \vec{z}_c)$ dans le *repère image* (point 2D), (3) le passage du repère image vers le *repère pixel*. Le lien géométrique

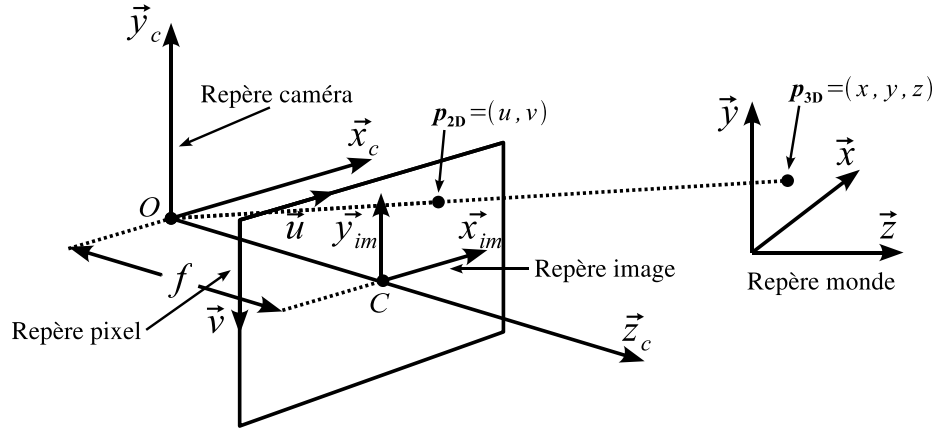


FIG. 2.2 – Projection d'un point dans l'espace 3D sur le plan image 2D.

entre un point dans le repère monde (coordonnées $\mathbf{x} = (x, y, z)$) et le même point dans le repère caméra (coordonnées $\mathbf{x}_c = (x_c, y_c, z_c)$) est défini par une transformation rigide composée d'une rotation et d'une translation 3D et donnée dans l'équation 2.1 en coordonnées homogènes.

$$\mathbf{x}_c = \begin{bmatrix} x_c \\ y_c \\ z_c \\ 1 \end{bmatrix} = \mathbf{K}\mathbf{x} = \begin{bmatrix} & & t_x \\ & \mathbf{R} & t_y \\ & & t_z \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} = \begin{bmatrix} r_{11} & r_{12} & r_{13} & t_x \\ r_{21} & r_{22} & r_{23} & t_y \\ r_{31} & r_{32} & r_{33} & t_z \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix}, \quad (2.1)$$

\mathbf{K} est la matrice des paramètres extrinsèques. $\mathbf{R} = \mathbf{R}_x(\theta)\mathbf{R}_y(\phi)\mathbf{R}_z(\psi)$ est la matrice de rotation composée de rotations successives sur les trois axes et \mathbf{t} est le vecteur de translation. Les rotations qui composent \mathbf{R} peuvent être représentées par les matrices (rotation dans le sens des aiguilles de la montre) :

$$\mathbf{R}_x(\theta) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos \theta & -\sin \theta \\ 0 & \sin \theta & \cos \theta \end{bmatrix}; \quad \mathbf{R}_y(\phi) = \begin{bmatrix} \cos \phi & 0 & \sin \phi \\ 0 & 1 & 0 \\ -\sin \phi & 0 & \cos \phi \end{bmatrix};$$

$$\mathbf{R}_z(\psi) = \begin{bmatrix} \cos \psi & -\sin \psi & 0 \\ \sin \psi & \cos \psi & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad (2.2)$$

avec θ , ϕ et ψ les angles de rotation sur les axes \vec{x} , \vec{y} et \vec{z} respectivement.

Pour obtenir le passage du repère caméra vers le repère image de coordonnées (x_{im}, y_{im}) nous pouvons utiliser le théorème de Thalès (triangles semblables) :

$$\frac{x_{im}}{f} = \frac{x_c}{z_c}$$

$$\frac{y_{im}}{f} = \frac{y_c}{z_c}$$
(2.3)

et donc :

$$x_{im} = f \frac{x_c}{z_c} \quad (2.4)$$

$$y_{im} = f \frac{y_c}{z_c}$$

Nous pouvons reformuler l'équation 2.4 sous la forme de matrice homogène :

$$\alpha \begin{bmatrix} x_{im} \\ y_{im} \\ 1 \end{bmatrix} = \begin{bmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} x_c \\ y_c \\ z_c \\ 1 \end{bmatrix}. \quad (2.5)$$

avec f la distance focale de la caméra et α un réel dont la valeur dépend de la position du point sur la trajectoire rectiligne dans le repère monde.

Ensuite, pour passer du repère image au repère pixel, il faut considérer le restant des paramètres internes de la caméra. Ces derniers sont les dimensions k_x et k_y du CCD de la caméra et les coordonnées (u_0, v_0) du centre optique du repère pixel :

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} k_x & 0 & u_0 \\ 0 & k_y & v_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_{im} \\ y_{im} \\ 1 \end{bmatrix}. \quad (2.6)$$

La matrice \mathbf{G} qui comprend l'ensemble des paramètres intrinsèques de la caméra est construite en combinant l'équation 2.5 et l'équation 2.6 :

$$\alpha \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \mathbf{G} \mathbf{x}_c = \begin{bmatrix} k_x f & 0 & u_0 & 0 \\ 0 & k_y f & v_0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} x_c \\ y_c \\ z_c \\ 1 \end{bmatrix}. \quad (2.7)$$

Une fois que nous avons défini les rapports entre les différents repères du modèle, il est possible de calculer la projection du point 3D sur le repère image 2D (pixel) en combinant l'équation 2.1 (matrice des paramètres extrinsèques K) et l'équation 2.7 (matrice des paramètres intrinsèques G), pour avoir :

$$\alpha \mathbf{u} = \mathbf{GKp} = \mathbf{Mp} = \begin{bmatrix} \beta_x r_{11} + u_0 r_{31} & \beta_x r_{12} + u_0 r_{32} & \beta_x r_{13} + u_0 r_{33} & \beta_x t_x + u_0 t_z \\ \beta_y r_{21} + v_0 r_{31} & \beta_y r_{22} + v_0 r_{32} & \beta_y r_{23} + v_0 r_{33} & \beta_y t_y + v_0 t_z \\ r_{31} & r_{31} & r_{31} & t_z \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix}, \quad (2.8)$$

ou de façon simplifiée :

$$\alpha \mathbf{u} = \begin{bmatrix} \beta_x \mathbf{r}_1 + u_0 \mathbf{r}_3 & s_x t_x + u_0 t_z \\ \beta_y \mathbf{r}_2 + v_0 \mathbf{r}_3 & s_y t_y + v_0 t_z \\ \mathbf{r}_3 & t_z \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix}, \quad (2.9)$$

où la matrice \mathbf{M} est la *matrice homogène de projection de la caméra*, $\beta_x = k_x f$, $\beta_y = k_y f$ sont les distances focales (en pixels) sur \vec{x} et \vec{y} respectivement et \mathbf{r}_1 , \mathbf{r}_2 , \mathbf{r}_3 sont les vecteurs ligne de

la matrice de rotation \mathbf{R} dont les éléments sont

$$\begin{aligned}
 r_{11} &= \cos \phi \cos \psi & r_{31} &= -\sin \phi \\
 r_{12} &= \sin \theta \sin \phi \cos \psi - \cos \theta \sin \psi & r_{32} &= \sin \theta \cos \phi \\
 r_{13} &= \cos \theta \sin \phi \cos \psi + \sin \theta \sin \psi & r_{33} &= \cos \theta \cos \psi \\
 \\
 r_{21} &= \cos \phi \sin \psi \\
 r_{22} &= \sin \theta \sin \phi \sin \psi + \cos \theta \cos \psi \\
 r_{23} &= \cos \theta \sin \phi \sin \psi - \sin \theta \cos \psi
 \end{aligned} \tag{2.10}$$

2.1.2 Distorsion radiale

L'équation 2.8, obtenue à l'aide du modèle sténopé de projection perspective, décrit une caméra idéale sans optique. Elle ne tient pas compte des distorsions dues notamment aux lentilles (la lentille idéale n'existe pas). Une des déformations liées à l'optique d'une caméra est la distorsion radiale qui apparaît pour des distances focales situées dans la plage de 2,8 à 12 mm [Svoboda, 2005]. Les rayons de lumière qui traversent les bords de la lentille (où la distance focale est plus petite), conduisent à une distorsion radiale plus importante que ceux passant au centre de la lentille. Ce comportement est représenté dans la figure 2.3.

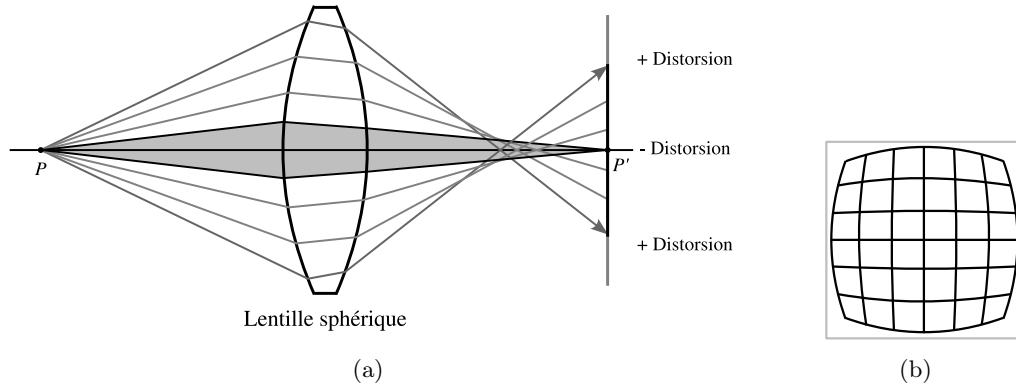


FIG. 2.3 – (a) Représentation de la distorsion radiale produite par une lentille sphérique. Les rayons qui ne passent pas par le centre de la lentille sont déviés des positions idéales dans le plan image. Les zones de la lentille qui sont plus éloignées du centre vont produire des déviations plus importantes que le point de convergence central (schéma modifié de [Forsyth and Ponce, 2003]). (b) Effet que la distorsion produit sur une grille initialement carrée (ligne en gris clair).

Pour modéliser la distorsion radiale affectant un pixel de coordonnées (x'_d, y'_d) nous pouvons utiliser une fonction $f(r)$ de distorsion qui donne la dépendance entre (x'_d, y'_d) et les coordonnées (x', y') du même pixel dans l'image sans distorsion. Dans l'équation 2.11 nous avons $r = \sqrt{x'^2 + y'^2}$:

$$\begin{aligned}
 x'_d &= x' f(r) \\
 y'_d &= y' f(r).
 \end{aligned} \tag{2.11}$$

La fonction $f(r)$ peut être développée en un polynôme supérieur de sorte que l'équation 2.11 devient :

$$\begin{aligned}
 x'_d &= x'(r + k_1 r^3 + k_2 r^5 + k_3 r^7 \dots) \\
 y'_d &= y'(r + k_1 r^3 + k_2 r^5 + k_3 r^7 \dots),
 \end{aligned} \tag{2.12}$$

k_1, k_2, k_3, \dots sont les coefficients qui contrôlent la distorsion radiale. L'équation 2.12 peut être reformulée comme suit :

$$\begin{aligned} x'_d &= x'r(1 + k_1r^2 + k_2r^4 + k_3r^6 \dots) = x' \left[r + \sum_{i=1}^n k_i r^{2i} \right] \\ y'_d &= y'r(1 + k_1r^2 + k_2r^4 + k_3r^6 \dots) = y' \left[r + \sum_{i=1}^n k_i r^{2i} \right], \end{aligned} \quad (2.13)$$

Les équations 2.13 ne tiennent pas compte du centre de distorsion C de coordonnées (u_0, v_0) (voir figure 2.2, le centre de distorsion est aussi la projection du centre optique dans l'image). En récrivant les équations en tenant compte du centre de distorsion nous avons :

$$\begin{aligned} x'_d - u_0 &= (x' - u_0)r(1 + k_1r^2 + k_2r^4 + k_3r^6 \dots) \\ y'_d - v_0 &= (y' - v_0)r(1 + k_1r^2 + k_2r^4 + k_3r^6 \dots). \end{aligned} \quad (2.14)$$

En pratique, les paramètres de distorsion sont difficiles à obtenir de façon analytique. Néanmoins, ils peuvent être obtenus par des méthodes numériques. Il existe plusieurs travaux qui proposent différentes solutions pour trouver les paramètres de distorsion. Pour une première classe de méthodes, le but est effectivement de trouver uniquement les paramètres de distorsion, les autres paramètres de la caméra étant non recherchés. Pour une deuxième classe de méthodes, l'ensemble des paramètres intrinsèques et extrinsèques sont recherchés y compris les paramètres de distorsion k_i (étalonnage de la caméra). En général, ces méthodes utilisent des « patterns » (motifs) de référence ; les plus connues sont celles proposées par [Tsai, 1987], [Zhang, 1998] et [Heikkilä, 2000]. Toutes ces méthodes sont basées sur le modèle sténopé de formation d'images.

La méthode proposée par [Tsai, 1987], comporte deux étapes. Des motifs d'étalonnage (planaires ou non) dont la géométrie est connue *a priori* sont utilisés. Cette méthode considère dans un premier temps que la projection du centre optique est localisée au centre de l'image. Dans la première étape, la distorsion radiale est supposée inexistante et une méthode linéaire (algorithme de moindres carrés) est utilisée pour calculer une partie des paramètres : rotation, translation sur \vec{x} , \vec{y} et les dimensions des CCD. Dans la deuxième étape, une méthode linéaire est utilisée pour obtenir une estimation de la distance focale et de la distance entre le motif et la caméra (translation sur \vec{z}). Ensuite, une méthode de minimisation est utilisée pour affiner les paramètres calculés dans l'étape précédente et pour trouver la distorsion radiale représentée par un seul coefficient de distorsion. Cette méthode peut utiliser une seule ou plusieurs images, mais la correspondance entre les points 3D de l'objet d'étalonnage (positions connues par construction) et leurs homologues dans l'image 2D doit être connue.

Pour la méthode de [Zhang, 1998], il est nécessaire d'utiliser un objet d'étalonnage représentant un damier (ou un échiquier) dont l'image est acquise avec la caméra pour au moins trois positions et orientations. La première étape de l'algorithme est l'extraction des coins des images du motif utilisé pour calculer une transformation projective entre les coins 2D et ceux connus dans le monde 3D. Ensuite, le premier coefficient de distorsion est calculé par une méthode analytique et les coefficients du troisième et cinquième ordre (cf. équation 2.14) sont obtenus par une méthode de moindres carrés. Finalement, une méthode de minimisation (algorithme de Levenberg-Marquardt) est utilisée pour ajuster précisément tous les paramètres.

La méthode de [Heikkilä, 2000] utilise d'abord une méthode linéaire (l'algorithme *Direct Linear Transform*, DLT) pour obtenir une estimation initiale de quelques paramètres (paramètres extrinsèques, distance focale et projection du centre optique) et effectue ensuite une estimation

non linéaire de type moindres carrés (Levenberg-Marquardt) pour ajuster précisément tous les paramètres obtenus auparavant et pour calculer les paramètres de distorsion radiale. La méthode utilise deux coefficients (ceux du premier et troisième ordre) pour obtenir une approximation de la distorsion radiale. Cette méthode peut fonctionner pour une ou plusieurs images, mais la méthode est plus robuste si plusieurs images sont considérées.

Récemment, [Miranda-Luna *et al.*, 2004a] a proposé une méthode simplifiée pour corriger la distorsion radiale des images endoscopiques. Celle-ci, peut aussi être appliquée à d'autres types de dispositifs d'acquisition d'images (caméra à faible distance focale). Cette méthode est basée sur l'acquisition d'une image d'un motif connu, qui est recalée avec une image numérisée du même motif sans distorsion. Dans cette méthode, l'auteur a montré que l'utilisation de quatre coefficients de distorsion suffisaient pour modéliser précisément la distorsion radiale des images endoscopiques. Une transformation projective, quatre coefficients de distorsion et les deux coordonnées du centre de distorsion sont considérés pour minimiser l'erreur entre la pattern de référence et l'image acquise. Pour minimiser l'erreur lors du recalage, la méthode utilise l'information mutuelle [Viola1997] des niveaux de gris des images comme mesure de similarité et le gradient descendant comme stratégie d'optimisation. L'avantage principale de cette méthode est la simplicité d'utilisation, puisque l'acquisition d'une seule image est nécessaire pour calculer les paramètres de distorsion. Elle détermine aussi plus de coefficients que les méthodes précédentes, car avec deux ou trois coefficients [Zhang, 1998], le modèle est moins précis dans le cas d'images endoscopiques (à forte distorsion).

2.2 Le mosaïquage d'images

Après la présentation du modèle de formation d'images, nous sommes en mesure d'aborder la technique qui est au centre de cette thèse : le mosaïquage d'images (*image mosaicing* en anglais). Le mosaïquage d'images est une méthode qui permet de placer l'ensemble des images d'une séquence dans un repère unique pour obtenir une représentation globale 2D d'une scène. Une mosaïque (ou une image panoramique), est donc construite en trouvant les transformations géométriques qui relient les images de la séquence et en les plaçant dans un système de coordonnées commun. Ces liens géométriques sont obtenus par des algorithmes de *recalage d'images*. Les images utilisées dans le mosaïquage doivent avoir un certain degré de superposition, c'est-à-dire, des régions communes. Lorsque le lien entre les images a été établi, il est possible de projeter les images dans un repère commun pour former une vue élargie de la scène par la combinaison des informations contenues dans les images. Dans la suite, nous allons appeler ce processus *collage d'images*. Le mosaïquage d'images est un bon moyen pour augmenter le champ de vision effectif d'une scène et supprimer la redondance des informations (figure 2.4).

Dans la section 2.2.1, nous présentons quelques domaines où des applications du mosaïquage d'images ont été développées. Ensuite, nous allons détailler dans les sections 2.2.2 et 2.2.6, les deux parties principales du mosaïquage : le recalage et le collage d'images.

2.2.1 Domaines d'application du mosaïquage d'images

Application dans le cas général (hors médical). Des applications concernant le mosaïquage d'images dans le domaine de la photographie grand public sont présentés dans [Brown and Lowe, 2003] [Baudisch *et al.*, 2005] [Steedly *et al.*, 2005] où l'objectif est de fabriquer des outils informatiques qui peuvent être utilisés par le grand public pour construire des images panoramiques à partir des images prises par des appareils photo peu onéreux. Le mosaïquage par logiciel est moins coûteux que l'achat d'un dispositif spécifique pour accomplir cette tâche. Les

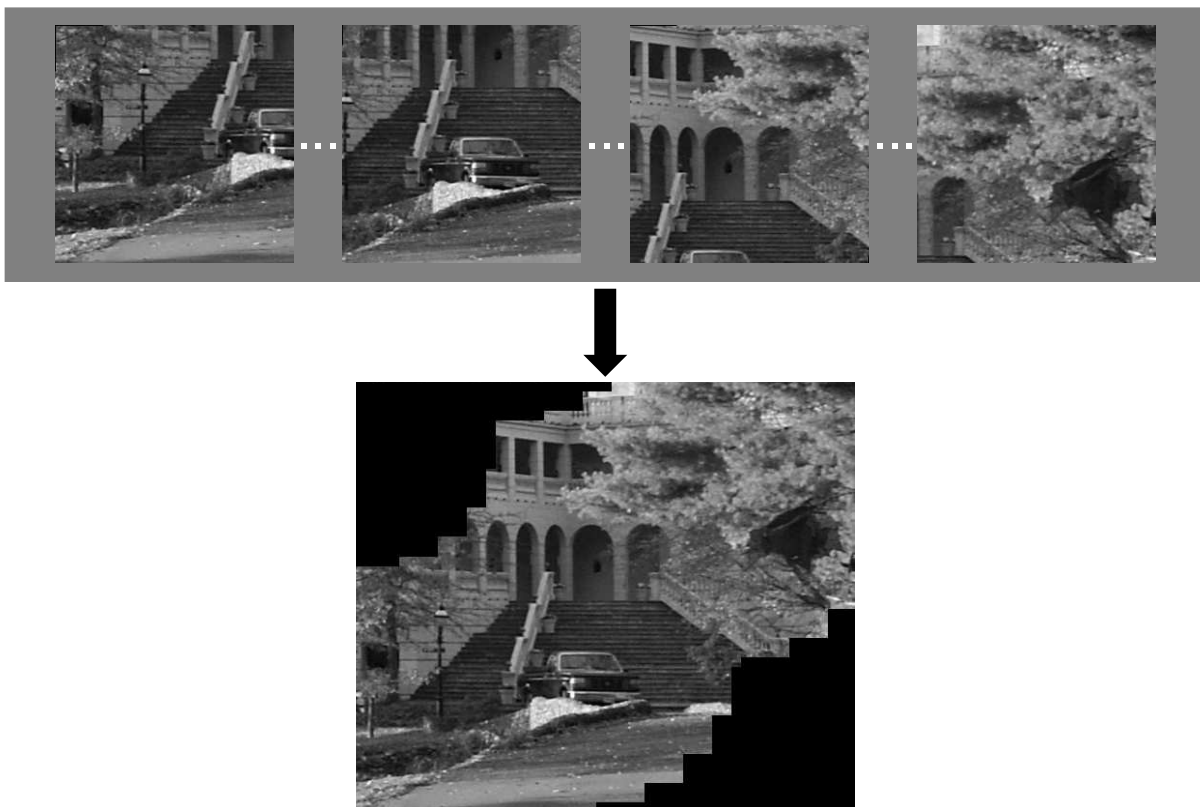


FIG. 2.4 – Principe du mosaïquage d'images. Une image panoramique est formée à partir d'une séquence d'images. Une image panoramique est une représentation concise d'une séquence d'images.

images utilisées comme données d'entrée dans ces applications sont en général de bonne qualité (bien contrastées) et contiennent beaucoup d'informations qui peuvent faciliter la construction d'images panoramiques. En général, des primitives images (régions ou contours bien prononcés) sont utilisés pour aligner les images dans ces approches.

Le mosaïquage d'images a été aussi utilisé pour la stabilisation de séquences vidéo acquises manuellement (compensation du mouvement de l'opérateur) [Litvin *et al.*, 2003] [Matsushita *et al.*, 2005]. Ici, une image est prise comme référence pour aligner un certain nombre d'images (par exemple 4 images avant et 4 après l'image de référence) pour atténuer l'effet de « tremblement » provoqué par l'utilisateur. Comme pour le domaine de fabrication d'images panoramiques grand public, la qualité des images est généralement bonne et elles contiennent beaucoup de primitives images pouvant faciliter le mosaïquage. La plupart des méthodes de stabilisation diminuent la résolution des images pour atteindre le fonctionnement en temps réel. C'est pour cette raison que les séquences stabilisées ont en général une résolution plus faible que les séquences originales [Matsushita *et al.*, 2005].

Un autre domaine d'application est la création d'environnements virtuels, c'est-à-dire des vues panoramiques à 360° destinées à des applications de réalité augmentée [Szeliski, 1996] [Szeliski and Shum, 1997] [Douze, 2004]. Dans ces approches, l'appareil photo est fixé sur un support, de telle façon que seules des rotations selon l'axe perpendiculaire au sol sont possibles. Les degrés de liberté étant réduits, la nature des déformations entre images consécutives est limitée.

La surveillance aérienne et la télédétection sont d'autres domaines d'application du mosaïquage [Zhu *et al.*, 2001] [Fedorov *et al.*, 2002] [Hurskainen and Pellikka, 2005]. Les objectifs sont de réduire la quantité de données à stocker, de pouvoir comparer l'évolution temporelle des zones observées ou, si les images ont été acquises par des capteurs différents, d'avoir des informations complémentaires sur la région observée. Pour ce genre d'applications, les déformations entre les images sont limitées à des translations, des rotations perpendiculaires à l'axe d'acquisition et des changements d'échelle.

L'exploration sous-marine [Rzhanov *et al.*, 2002] [Gracias and Negahdaripour, 2005] est un autre domaine d'application du mosaïquage. Dans ce domaine, les observations effectuées par une sonde sont stockées sous forme d'images panoramiques afin de réduire de façon conséquente l'espace de stockage des données. Pour cette application, la difficulté du mosaïquage est liée au fait que les images sont affectées par l'atténuation et la diffraction de la lumière provoquées par le milieu liquide.

L'exploration spatiale est un domaine classique d'utilisation du mosaïquage [Huntsberger *et al.*, 1999] [Moshir, 2005]. Les images acquises par les sondes spatiales et les télescopes représentant seule une petite région d'une zone d'intérêt plus vaste, des images panoramiques sont construites pour avoir une vision globale d'une partie plus ou moins grande de la zone explorée. Dans ces applications, l'information de position connue pour les dispositifs d'acquisition est utilisée pour guider la construction des images panoramiques.

Le mosaïquage d'images est aussi utilisé pour l'obtention d'images avec super-résolution [Zomet and Peleg, 2000] [Capel, 2001] [Capel and Zisserman, 2003]. Ces approches profitent du fait que dans une séquence vidéo, il existe beaucoup d'information quasi-redondante qui peut être utilisée pour améliorer la résolution d'une région donnée de la séquence en alignant des images, qui auront chacune une information un peu différente qui peut être utilisée pour sur-échantillonner les données.

Application dans le domaine médical. Un domaine d'application du mosaïquage d'images qui est particulièrement intéressant dans le cadre de cette thèse, est naturellement la médecine.

Dans ce domaine, à ce jour, il existe peu d'applications documentées dans la littérature. L'ophtalmologie est le domaine de la médecine dans lequel la plupart des applications du mosaïquage ont été développées. Le mosaïquage d'images est utilisé pour construire des images avec un champ de vue élargi qui facilite le diagnostic et la chirurgie. Ces approches utilisent en général, une méthode qui compare directement les niveaux de gris des images pour trouver des régions superposables entre les images et utilise un modèle à 6 degrés de liberté (modèle affine) pour l'alignement. Toujours dans le domaine de l'ophtalmologie, [Can *et al.*, 2002] segmentent les vaisseaux de la rétine visibles dans les séquences d'images pour localiser des intersections qui représentent l'information utilisable pour construire des images panoramiques. Ces images panoramiques facilitent le diagnostic du clinicien. Les résultats de cette méthode sont hautement dépendants de la capacité de l'algorithme de segmentation à localiser les vaisseaux. Cette approche utilise un modèle à 12 degrés de liberté (modèle non-linéaire) pour modéliser les déformations entre les images et la courbure de la rétine.

Dans le domaine de l'histologie, des applications visant à former des images d'une coupe histologique complète à partir de plusieurs prises de vue au microscope, ont aussi été développées [Hoge *et al.*, 2003] [Rankov *et al.*, 2005]. Dans ces applications, le mouvement du microscope étant limité à des mouvements parallèles à l'objectif du microscope, construire une image panoramique consiste à trouver les translations entre les images.

Des vues élargies augmentant le champ de vue en ultrasonographie sont construites par la méthode décrite dans [Gee *et al.*, 2003]. Cette méthode utilise l'information mutuelle pour aligner des balayages effectués à la main pendant une échographie, et ce pour 8 degrés de liberté (modèle de transformation perspective). Même si les images utilisées dans cette approche sont très bruitées, aucun pré-traitement des images n'est utilisé.

En radiologie, [Yaniv and Joskowicz, 2004] proposent de construire des images panoramiques en utilisant des images obtenues en radiologie de fluorescence pour l'observation des os longs qui ne sont pas visibles en une seule radiographie. Pour aligner les images, ils considèrent qu'il n'existe pas de rotation entre les images et que les transformations entre celles-ci peuvent être représentées par des déplacements parallèles à l'axe de projection des images. Une règle est donc placée à côté de chaque image et l'alignement est fait manuellement en utilisant la graduation de la règle.

Dans [Vercauteren *et al.*, 2005], une séquence d'images de microscopie confocale (fibrée) est utilisée pour construire des images panoramiques devant faciliter la détection du cancer. Le dispositif d'acquisition utilisé dans cette application fournit des séquences d'images dont le lien est modélisé dans une première étape par 3 degrés de liberté (des translations et une rotation perpendiculaire à la sonde d'acquisition) et dans un deuxième temps par un modèle polynomial.

2.2.2 Principe du recalage géométrique d'images

Les algorithmes de recalage d'images (*image registration* en anglais), déterminent la relation géométrique (*transformation spatiale*) liant deux images contenant des informations communes. En d'autres termes, le recalage d'images est l'opération qui situe deux images ou plus contenant des informations homologues, dans un même repère géométrique. Ces images peuvent être représentées en 2D (le type d'image d'intérêt dans cette thèse) ou en 3D. La figure 2.5 montre une représentation simplifiée du lien entre deux images 2D issues d'une même scène 3D.

Si les images à recalcer sont issues de capteurs différents (par exemple des images IRM et des images ultrasonographiques), alors le recalage est dit *multimodal*. En revanche, des images issues d'un même type de capteur font appel à du recalage *monomodal*.

Les applications du recalage d'images sont très vastes. Il est utilisé dans la vision stéréosco-

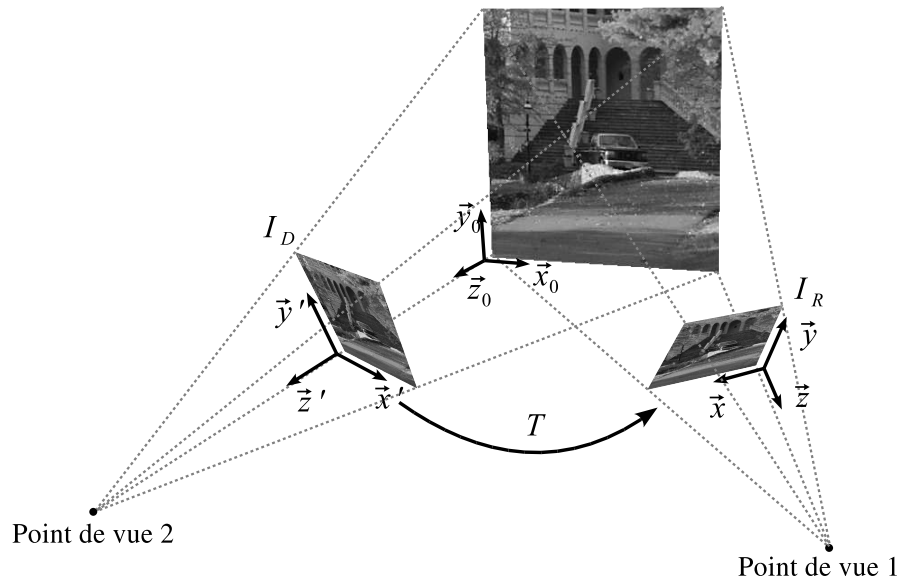


FIG. 2.5 – Représentation simplifiée de l'espace T de transformation 2D–2D pour le recalage entre deux images d'une même scène prises pour deux points de vue différents. T représente un changement de perspective.

rique, la reconnaissance de formes, dans le suivi d'objets, le mosaïquage d'images, l'indexation de bases d'images, *etc.* Étant donné le très vaste champ applicatif du recalage d'images, nous allons aborder seulement quelques exemples significatifs (pour notre travail) du recalage d'images dans le domaine médical. Des aperçus très complets sur les techniques de recalage dans des divers domaines peuvent être trouvées dans [Brown, 1992] et [Zitová and Flusser, 2003], et dans [Maintz and Viergever, 1998] et [Pluim *et al.*, 2003] pour des applications dans le domaine médical.

Le recalage d'images, contrairement au mosaïquage d'images, a trouvé de nombreuses applications dans le domaine de la médecine. La fusion d'images médicales pour le diagnostic, la chirurgie assistée et le positionnement du patient en radiothérapie ne sont que quelques exemples d'application du recalage d'images en médecine.

Le recalage multimodalité d'images ultrasonographiques et de coupes IRM en pondération T2, est utilisé dans [Verhey *et al.*, 2005] pour le diagnostic de dysfonctions pelviennes chez la femme. Après le pré-traitement des images ultrasonographiques par un filtrage adaptatif, ces images sont alignées avec les images IRM de façon manuelle, pour ensuite affiner les résultats de façon automatisée en considérant 3 degrés de liberté (deux translations et une rotation).

Une approche permettant d'observer le fonctionnement de certaines zones du cerveau pour le diagnostic de maladies neuropsychiatriques par le recalage monomodalité de cartes IRM en pondération T2, est présenté par [Wu *et al.*, 2006]. Ces auteurs recalent des images IRM acquises en clinique avec des images issues d'un atlas en utilisant un modèle déformable (non linéaire). Ensuite, ils appliquent un algorithme de segmentation pour séparer les zones d'intérêt.

[Posada *et al.*, 2007] propose une méthode pour améliorer le positionnement des patients en radiothérapie intracrânienne. Ces auteurs effectuent le recalage d'images volumétriques du visage acquises par un capteur 3D (basé sur la lumière structurée) fixé dans la salle de traitement avec des images CT acquises avant le traitement et comprenant la tête du patient. Ils utilisent un modèle rigide 3D (6 paramètres) pour modéliser la transformation entre les images. Cette transformation permet de calculer la position de la tumeur dans la salle de traitement.

Dans [Hoeksma *et al.*, 2005], des images IRM en pondération T1 de patients avec un dysfonctionnement cérébral sont recalées à des images IRM de sujets sains et à des cartes de signaux EEG pour effectuer des comparaisons. Ils utilisent un modèle de transformation non linéaire pour décrire la relation entre les images.

2.2.3 Définition mathématique du recalage d'images

Le processus de recalage de deux images peut être formulé mathématiquement comme suit.

$$\tilde{T} = \arg \operatorname{opt} \sum_T \sum_{\mathbf{x}} \operatorname{sim} [g(I_R(\mathbf{x})), \Omega(g(I_D(T(\mathbf{x}))))]. \quad (2.15)$$

Dans l'équation 2.15, I_D est l'image à recaler avec I_R (respectivement, image à transformer ou source et image de référence ou cible). \tilde{T} est la transformation qui va amener les coordonnées des pixels de l'image I_D dans le système de coordonnées de l'image I_R . Les valeurs des paramètres de cette transformation sont actualisées par une méthode d'optimisation, $\arg \operatorname{opt}$, qui cherche à maximiser ou minimiser une mesure de similarité, sim . Cette dernière est calculée soit à partir de primitives images détectées par une fonction g , soit directement avec les niveaux de gris des images. Dans ce dernier cas, aucune fonction g n'est appliquée. Cet algorithme d'optimisation cherche les valeurs des paramètres de la transformation dans un espace de recherche donné par T . Ω est la méthode d'interpolation associée à la transformation T qui sera appliquée à l'image I_D . \tilde{T} est la transformation qui optimise sim et qui superpose au mieux les parties communes des images.

Pour notre étude, nous avons séparé les méthodes de recalage en deux classes : les méthodes qui n'utilisent pas la fonction g , que nous allons appeler *méthodes de recalage basées sur les niveaux de gris des images* (voir section 2.2.4), et les méthodes qui en font usage, appelées dans notre document *méthodes basées sur la détection de primitives images* (voir section 2.2.5.1).

Même s'il est possible de reconnaître tous les éléments mathématiques dans les deux approches de recalage d'images, certains de ces éléments sont utilisés de façon différente. C'est pour cette raison que nous allons d'abord présenter les éléments du recalage d'images qui sont communs aux deux approches : l'espace de recherche et l'interpolation des images transformées, pour ensuite présenter l'utilisation que chaque approche fait du reste des éléments.

2.2.3.1 Espace de recherche (types de transformation)

L'organe qui nous intéresse (la vessie) est mou et sujet à des déformations élastiques en conditions normales. Le protocole suivi lors d'un examen cystoscopique clinique standard indique que la vessie doit être remplie d'un liquide, de sorte que les parois de l'organe soient sous une tension qui entraîne une réduction de l'amplitude des mouvements de la vessie. Par ailleurs, la vitesse d'acquisition d'un endoscope est de 25–30 images par seconde. Même si l'organe est mou, en réalité la surface de la vessie peut être considérée comme rigide entre deux acquisitions séparées de 40 ms (ou plus). Si l'on suppose que les seules non linéarités de l'instrument peuvent être corrigées (ce qui est effectivement le cas pour la distorsion due à l'optique), alors il est possible d'approximer précisément les transformations non linéaires entre les images par des transformations linéaires. Nous allons donc focaliser notre étude sur les modèles de transformation 2D linéaires.

Pour le texte suivant, les coordonnées homogènes sont représentées par le symbole $\tilde{\cdot}$ (par exemple \tilde{x}). Les coordonnées standards n'ont pas de $\tilde{\cdot}$ (tilde).

Le cas le plus général de transformation linéaire 2D est la *transformation projective* à 8 degrés de liberté (1 axe de rotation dans le plan image, 2 directions de cisaillement (*shearing*

en anglais), 2 rotations en perspective dites hors plan, 2 translations et 1 changement d'échelle isotrope). Néanmoins, pour des mouvements particuliers (plus limités) de la caméra, la mise en correspondance entre les vues de la scène peut être effectuée en utilisant des transformations plus simples. Ces transformations ont moins de degrés de liberté et par conséquent, sont représentées par moins de paramètres. Les transformations géométriques élémentaires (translations, rotations et changements d'échelle) peuvent être combinées pour obtenir des transformations plus ou moins complexes avec différents degrés de liberté.

Translations. Si le mouvement de la caméra consiste seulement en des translations parallèles au plan d'acquisition, le rapport entre les images peut être représenté par une transformation à 2 degrés de liberté. Ce modèle est le plus simple. Ces translations ($\mathbf{x}' = \mathbf{x} + \mathbf{t}$) peuvent être exprimées sous forme matricielle dans un espace projectif :

$$\mathbf{x}' = \begin{bmatrix} \mathbf{I} & \mathbf{t} \end{bmatrix} \tilde{\mathbf{x}}, \quad (2.16)$$

où \mathbf{I} est la matrice unité de taille 2×2 , $\mathbf{t} = (t_x, t_y)^T$ est le vecteur des translations et $\tilde{\mathbf{x}} = (x, y, 1)^T$ est un vecteur 2D en coordonnées *homogènes*. Ce type de transformation est appliqué, par exemple, lors de l'observation au microscope des coupes histologiques [Hoge *et al.*, 2003][Rankov *et al.*, 2005].

Transformation rigide. Un mouvement de la caméra entraînant des translations parallèles au plan de la scène et une rotation autour de l'axe perpendiculaire à ce plan, peut être modélisé par une matrice de *transformation rigide* ou *euclidienne* (ainsi nommée car les distances euclidiennes, le parallélisme et les angles sont préservés) à 3 degrés de liberté. Elle peut être exprimée sous la forme $\mathbf{x}' = \mathbf{R}\mathbf{x} + \mathbf{t}$ ou de façon matricielle :

$$\mathbf{x}' = \begin{bmatrix} \mathbf{R} & \mathbf{t} \end{bmatrix} \tilde{\mathbf{x}}, \quad (2.17)$$

où

$$\mathbf{R} = \begin{bmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{bmatrix}$$

θ est l'angle de rotation dans le plan. \mathbf{R} est une matrice orthonormale de rotation avec $\mathbf{R}\mathbf{R}^T = \mathbf{I}$ et $|\mathbf{R}| = 1$.

Le recalage des coupes en IRM (Imagerie par Résonance Magnétique) est l'un des domaines d'application biomédicale de cette transformation [Calhoun *et al.*, 1998] [Kotsas, 2005].

Similitude. Si des translations le long de l'axe perpendiculaire au plan de la scène (qui se traduisent par des changements d'échelle de l'image) sont ajoutées à la transformation euclidienne, elle devient une *similitude* telle que $\mathbf{x}' = s\mathbf{R}\mathbf{x} + \mathbf{t}$, où s est un facteur d'échelle. Elle peut aussi être écrite comme suit

$$\mathbf{x}' = \begin{bmatrix} s\mathbf{R} & \mathbf{t} \end{bmatrix} \tilde{\mathbf{x}} = \begin{bmatrix} a & -b & t_x \\ b & a & t_y \end{bmatrix} \tilde{\mathbf{x}}, \quad (2.18)$$

où $a = s \cos \theta$ et $b = s \sin \theta$.

Cette transformation à 4 degrés de liberté (2 translations, 1 axe de rotation et 1 changement d'échelle isotrope) préserve les angles et le parallélisme entre les lignes. Des applications dans le *filigranage* d'images [Lin *et al.*, 2001], dans l'imagerie satellitaire [Cheeseman *et al.*, 1994] et dans la photographie aérienne [Xie *et al.*, 2003] ont été développées à l'aide de ce type de transformation.

Transformation affine. Pour cette transformation, le changement d'échelle est asymétrique et donc, un effet de « cisaillement » apparaît. La transformation est définie par la matrice (de

taille 2×3) à 6 degrés de liberté :

$$\mathbf{x}' = \begin{bmatrix} \mathbf{A} & \mathbf{t} \end{bmatrix} \tilde{\mathbf{x}} = \begin{bmatrix} a_{00} & a_{01} & t_x \\ a_{10} & a_{11} & t_y \end{bmatrix} \tilde{\mathbf{x}}. \quad (2.19)$$

La matrice \mathbf{A} peut être représentée par une combinaison de matrices de rotation et une matrice de facteurs d'échelle asymétriques [Hartley and Zisserman, 2004] :

$$\mathbf{A} = \mathbf{R}(\theta)\mathbf{R}(-\varphi)\mathbf{D}\mathbf{R}(\varphi), \quad (2.20)$$

avec θ l'angle de rotation de l'image, φ l'angle de déformation (cf. figure 2.6) et

$$\mathbf{D} = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix}, \quad (2.21)$$

où λ_1 et λ_2 sont les facteurs d'échelle asymétriques.

Cette transformation préserve le parallélisme et les rapports de longueurs entre les lignes ainsi que les rapports de surfaces.

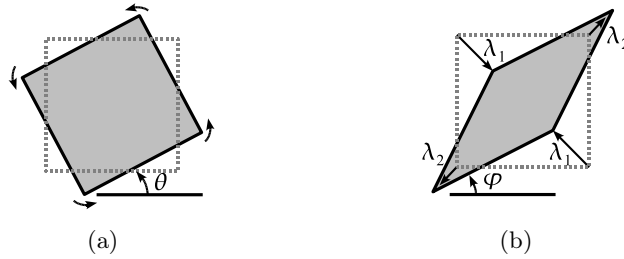


FIG. 2.6 – Déformations issues d'une transformation affine (d'après [Hartley and Zisserman, 2004]). (a) Rotation d'angle θ . (b) Déformation $\mathbf{R}(-\varphi)\mathbf{D}\mathbf{R}(\varphi)$. Il faut noter que les directions des changements d'échelle (λ_1 et λ_2) sont orthogonales.

Des exemples d'application de ce modèle de transformation sont le recalage de contours et d'images échocardiographiques [Dydenko *et al.*, 2003], le recalage d'images CT-T2 [Zhang and Rangarajan, 2004] ou le recalage de documents photographiés à la main avec un appareil photo [Sato *et al.*, 2004].

Transformation projective. Cette transformation est aussi appelée *transformation perspective* ou *homographie*. C'est un modèle de transformation à 8 degrés de liberté. Ceci implique des mouvements de translation, de rotations, de changements d'échelle, ainsi que des changements de perspective de la caméra (rotations hors plan). Elle peut être exprimée comme suit : $\tilde{\mathbf{x}}' \sim \tilde{\mathbf{H}} \tilde{\mathbf{x}}$, où $\tilde{\mathbf{H}}$ est une matrice de taille 3×3 . Celle-ci peut être écrite aussi comme suit :

$$\tilde{\mathbf{x}}' = \begin{bmatrix} h_{00} & h_{01} & h_{02} \\ h_{10} & h_{11} & h_{12} \\ h_{20} & h_{21} & h_{22} \end{bmatrix} \tilde{\mathbf{x}}. \quad (2.22)$$

La matrice de transformation projective \mathbf{H} peut être décomposée en une série de transformations [Hartley and Zisserman, 2004] :

$$\mathbf{H} = \mathbf{H}_S \mathbf{H}_A \mathbf{H}_P = \begin{bmatrix} s\mathbf{R} & \mathbf{t} \\ \mathbf{0}^T & 1 \end{bmatrix} \begin{bmatrix} \mathbf{K} & \mathbf{O} \\ \mathbf{0}^T & 1 \end{bmatrix} \begin{bmatrix} \mathbf{I} & \mathbf{O} \\ \mathbf{v}^T & v \end{bmatrix} = \begin{bmatrix} \mathbf{A} & \mathbf{t} \\ \mathbf{v}^T & v \end{bmatrix} \quad (2.23)$$

avec \mathbf{A} une matrice non-singulière donnée par $\mathbf{A} = s\mathbf{R}\mathbf{K} + \mathbf{t}\mathbf{v}^T$ et \mathbf{K} une matrice triangulaire supérieure normalisée de telle façon que $\det(\mathbf{K}) = 1$. \mathbf{R} est encore une matrice de rotation dans le plan, s est un facteur d'échelle, \mathbf{t} est un vecteur translation et $\mathbf{v} = (v_1, v_2)^T$ est un vecteur contenant les deux premiers éléments de normalisation de la matrice de transformation perspective (h_{20}, h_{21}) . Cette composition est valide pour $v \neq 0$ et elle est unique si s est positif.

Les coordonnées homogènes résultantes doivent être normalisées pour obtenir des coordonnées « inhomogènes » :

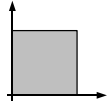
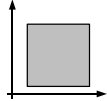
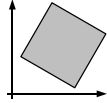
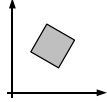
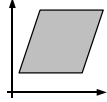
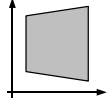
$$x' = \frac{h_{00}x + h_{01}y + h_{02}}{h_{20}x + h_{21}y + h_{22}} \quad \text{et} \quad y' = \frac{h_{10}x + h_{11}y + h_{12}}{h_{20}x + h_{21}y + h_{22}}. \quad (2.24)$$

Les transformations projectives préservent les lignes droites. Ce modèle est utilisé lorsque la caméra effectue des mouvements libres dans un repère tridimensionnel, comme par exemple lors de la prise d'une vidéo [Irani *et al.*, 1996], dans le cas du mosaïquage d'images issues d'exploration sous-marine [Gracias *et al.*, 2002] ou pour des photographies prises avec un appareil photo déplacé manuellement [Zomet and Peleg, 2000].

C'est aussi le type de transformation adapté à notre application (examen cystoscopique clinique), où les images sont acquises par l'extrémité du cystoscope qui se déplace « librement » dans la vessie (cf. chapitre 3).

Le tableau 2.1 donne un résumé des modèles de transformation décrits précédemment.

TAB. 2.1 – Modèles de transformations géométriques linéaires en 2D. Toutes les transformations comprennent les paramètres de la transformation précédente à laquelle nous rajoutons de nouveaux paramètres. Les paramètres t_x , t_y , s , θ sont expliqués dans le texte.

Transformation	Matrice	Effet	Invariants
Identité 0 degré de liberté	$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$		
Translation 2 degrés de liberté	$\begin{bmatrix} 1 & 0 & t_x \\ 0 & 1 & t_y \end{bmatrix}$		Orientation, distances, angles, parallélisme, lignes droites
Rigide 3 degrés de liberté	$\begin{bmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{bmatrix}$		Distances, angles, parallélisme, lignes droites
Similitude 4 degrés de liberté	$\begin{bmatrix} s \cos \theta & -s \sin \theta \\ s \sin \theta & s \cos \theta \end{bmatrix}$		Angles, parallélisme, lignes droites
Affine 6 degrés de liberté	$\begin{bmatrix} a_{00} & a_{01} & a_{02} \\ a_{10} & a_{11} & a_{12} \end{bmatrix}$		Parallélisme, lignes droites
Projective 8 degrés de liberté	$\begin{bmatrix} h_{00} & h_{01} & h_{02} \\ h_{10} & h_{11} & h_{12} \\ h_{20} & h_{21} & h_{22} \end{bmatrix}$		Lignes droites

2.2.3.2 Interpolation des niveaux de gris des images transformées

Un pixel dont les coordonnées sont entières dans l'image à transformer, ne sera pas forcément placé sur des coordonnées entières dans l'image transformée. Par ailleurs, un pixel de l'image à transformer ne correspond pas à un seul pixel de l'image transformée (et inversement). Pour ces raisons, il n'est pas suffisant de recopier les niveaux de gris de l'image à transformer vers l'image transformée.

Lorsqu'une transformation est appliquée à une image, deux méthodes peuvent être utilisées pour déterminer les valeurs d'intensité des pixels de l'image transformée : la méthode *directe* et la méthode *inverse*. Pour les deux méthodes, la première étape consiste à appliquer la transformation aux coordonnées minimum et maximum de l'image (les quatre coins). De cette façon, la surface du polygone à quatre côtes (*bounding box*) sur laquelle l'image transformée va être « reproduite » est connue. Si la méthode directe est utilisée, la transformation est appliquée à toutes les coordonnées des pixels de l'image originale (image source) et le niveau de gris des pixels originaux est copié dans l'image transformée sur les pixels dont les coordonnées sont les entiers les plus proches des coordonnées calculées. Cette opération est illustrée dans la figure 2.7(a). Il existe deux problèmes associés à l'utilisation de cette technique. D'une part, lorsque la transformation est appliquée à l'image originale, certains pixels de l'image transformée peuvent ne correspondre à aucun pixel dans l'image originale. Cela fait apparaître des trous de niveaux de gris dans l'image transformée. D'autre part, un pixel dans l'image transformée peut correspondre à plusieurs pixels dans l'image originale. Donc, les valeurs d'intensité des pixels vont se superposer à une même position dans l'image transformée. La méthode directe ne permet pas d'affecter une et une seule valeur de niveaux d'intensité à la totalité des pixels de l'image transformée.

En raison des inconvénients liés à l'utilisation de la méthode directe, la méthode inverse est généralement préférée. Pour cette méthode, une fois l'extension de l'image transformée connue, la transformation inverse T^{-1} est appliquée à toutes les coordonnées entières des pixels de l'image transformée pour calculer, pour chaque pixel, le niveau de gris correspondant à cette position dans l'image originale. Comme il est probable que cette position ne soit pas entière, il faut alors calculer la valeur de ce pixel par interpolation des valeurs des pixels voisins. Le principe de la méthode inverse est montrée dans la figure 2.7(b). Cette méthode garantit l'absence de trous dans l'image transformée.

Pour calculer la valeur du pixel dans l'image transformée, une méthode d'interpolation qui prend en compte les valeurs d'intensité des pixels dans le voisinage du pixel correspondant dans l'image source est utilisée, les plus connues étant les interpolations *par le plus proche voisin*, *bilinéaire* et *bicubique*.

Interpolation par le plus proche voisin. Cette méthode d'interpolation est la plus simple. Ici, la valeur du pixel cherchée est obtenue, comme son nom l'indique, en prenant la valeur du pixel qui sera le plus proche de la position calculée sur l'image originale.

Interpolation bilinéaire. Cette méthode considère un voisinage de taille 2×2 pixels autour de la coordonnée d'intérêt calculée (voir figure 2.8). Une moyenne pondérée est donc calculée avec les valeurs d'intensité de ces 4 pixels. C'est une extension de l'interpolation linéaire au cas bi-dimensionnel. La méthode consiste à appliquer une interpolation linéaire d'abord dans une direction et ensuite dans l'autre. Si les quatre valeurs des pixels les plus proches autour de la position d'intérêt sont connues, la valeur du pixel de coordonnées (p_x, p_y) dans l'image transformée peut être calculée comme suit. Soit (x, y) les coordonnées entières du pixel situé immédiatement en haut et à gauche du pixel de coordonnées non entières ramené de l'image

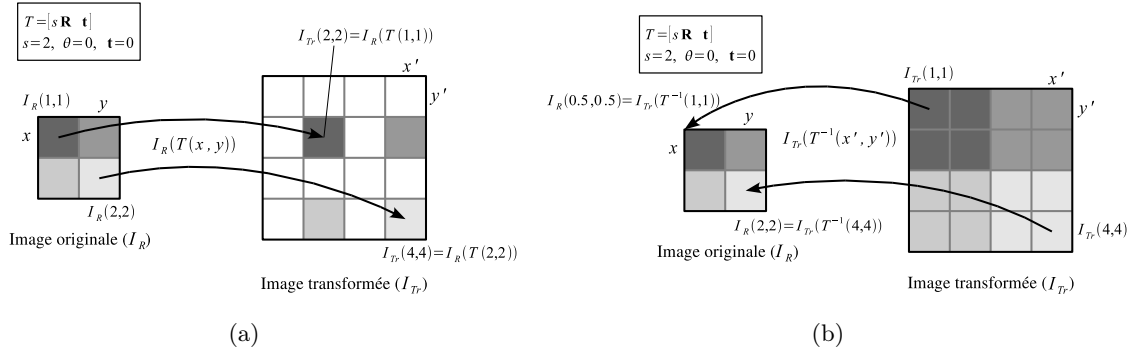


FIG. 2.7 – Illustrations des deux approches possibles pour transformer une image : directe (a) et inverse (b). I_R est l'image originale et I_{T_r} est l'image transformée. T est la transformation appliquée aux coordonnées des images. La transformation prise comme exemple ici est une similitude qui va multiplier par deux la taille de l'image ($s = 2$). (a) Transformation de l'image en appliquant la méthode directe. (b) Transformation par la méthode inverse avec une interpolation par le plus proche voisin.

transformée dans l'image à transformer avec T^{-1} (figure 2.8). L'interpolation selon l'axe \vec{x} est effectué avec la relation 2.25.

$$I_R(p_x, y) = (1 - a)I_R(x, y) + aI_R(x + 1, y) \quad (2.25)$$

$$I_R(p_x, y + 1) = (1 - a)I_R(x, y + 1) + aI_R(x + 1, y + 1)$$

où a est la distance, dans la direction \vec{x} , du point dont l'intensité doit être calculée à la position du pixel pris comme référence (x, y) . Avec ces deux valeurs, il est maintenant possible de calculer le niveau de gris $I_R(p_x, p_y)$ du point de coordonnées entières (p_x, p_y) dans l'image transformée en interpolant les niveaux de gris selon \vec{y} :

$$I_R(p_x, p_y) = (1 - b)I_R(p_x, y) + bI_R(p_x, y + 1). \quad (2.26)$$

où b est la distance, dans la direction \vec{y} .

Le résultat de cette méthode est visuellement beaucoup plus lisse (moins pixelisé) que celui obtenu par la méthode du plus proche voisin.

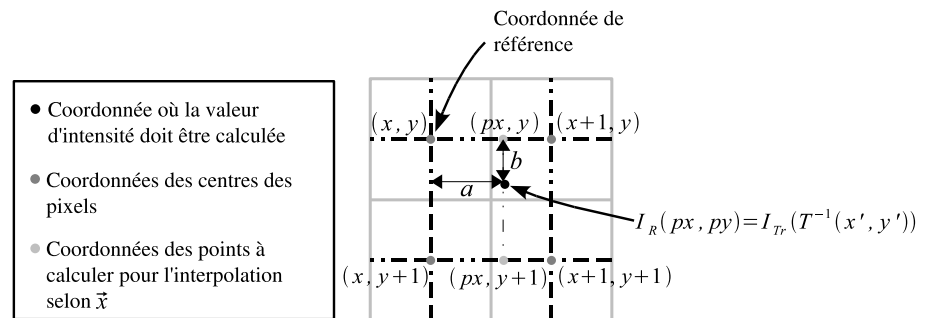


FIG. 2.8 – Interpolation bilinéaire.

Interpolation bicubique. Cette méthode prend en compte un voisinage de taille 4×4 . Les seize pixels les plus proches du pixel dont l'intensité doit être calculée définissent ce voisinage. Comme montré dans la figure 2.9, une moyenne pondérée des seize niveaux de gris est calculée en utilisant un polynôme du 3^{ème} degré. Comme pour l'interpolation bilinéaire, une première interpolation cubique est appliquée d'abord selon une première direction et ensuite selon l'autre en considérant cette fois-ci quatre pixels par ligne. Les coordonnées entières (x, y) sont encore celles du pixel situé immédiatement en haut et à gauche du pixel de coordonnées (p_x, p_y) dont le niveau de gris $I_R(p_x, p_y)$ doit être calculé. a est la distance selon \vec{x} entre les points de coordonnées (x, y) et (p_x, p_y) . L'interpolation selon \vec{x} est calculée comme suit :

$$\begin{aligned}
 I_R(p_x, y) &= -a(1-a)I_R(x-1, y) + (1-2a^2+a^3)I_R(x, y) + \\
 &\quad a(1+a-a^2)I_R(x+1, y) - a^2(1-a)I_R(x+2, y) \\
 I_R(p_x, y-1) &= -a(1-a)I_R(x-1, y-1) + (1-2a^2+a^3)I_R(x, y-1) + \\
 &\quad a(1+a-a^2)I_R(x+1, y-1) - a^2(1-a)I_R(x+2, y-1) \\
 I_R(p_x, y+1) &= -a(1-a)I_R(x-1, y+1) + (1-2a^2+a^3)I_R(x, y+1) + \\
 &\quad a(1+a-a^2)I_R(x+1, y+1) - a^2(1-a)I_R(x+2, y+1) \\
 I_R(p_x, y+2) &= -a(1-a)I_R(x-1, y+2) + (1-2a^2+a^3)I_R(x, y+2) + \\
 &\quad a(1+a-a^2)I_R(x+1, y+2) - a^2(1-a)I_R(x+2, y+2)
 \end{aligned}
 \tag{2.27}$$

Avec les quatre valeurs obtenues, la valeur de $I_R(p_x, p_y)$ peut être calculée comme suit :

$$\begin{aligned}
 I_R(p_x, p_y) &= -b(1-b)I_R(p_x, y-1) + (1-2b^2+b^3)I_R(p_x, y) + \\
 &\quad b(1+b-b^2)I_R(p_x, y+1) - b^2(1-b)I_R(p_x, y+2).
 \end{aligned}
 \tag{2.28}$$

La distance b est définie sur la figure 2.9.

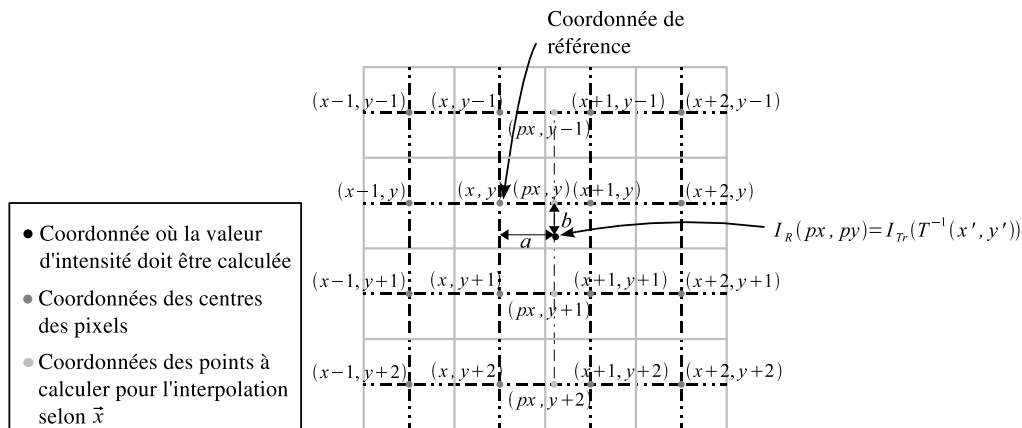


FIG. 2.9 – Principe de l'interpolation bicubique.

La figure 2.10 montre un exemple du résultat de l'application des méthodes d'interpolation décrites ci-dessus. L'interpolation bicubique fournit un rendu visuel légèrement amélioré par rapport à l'interpolation bilinéaire. Cependant, elle requiert beaucoup plus de temps de calcul

que cette dernière. Pour la plupart des applications de recalage d'images, l'interpolation bilinéaire offre le meilleur compromis entre rendu visuel et temps de calcul. L'interpolation par le plus proche voisin peut être préférée si les valeurs d'intensité originales doivent absolument être préservées.

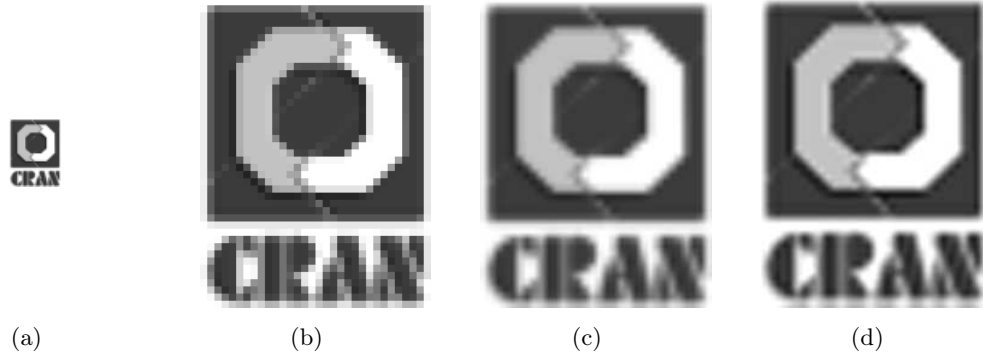


FIG. 2.10 – Résultat des différentes méthodes d'interpolation pour une transformation constituée d'un changement d'échelle d'un facteur 6. (a) Image originale. (b) Interpolation de (a) par le plus proche voisin. (c) Interpolation bilinéaire de (a). (d) Interpolation bicubique de (a).

2.2.4 Recalage basé sur les niveaux de gris des images

Pour définir une méthode de recalage basée sur les niveaux de gris des images, il faut d'abord choisir une mesure de similarité permettant effectivement de comparer les images. Cette mesure sera sélectionnée en tenant compte des caractéristiques des images (mono ou multimodalités, bruits présents sur les images, déformations géométriques, *etc.*). Une fois la mesure de similarité choisie, une technique de *recherche* de la similarité optimale doit être mise en place. Pour mesurer le degré de similarité de deux images, des méthodes basées sur la comparaison directe des niveaux d'intensité, la corrélation des images, la transformée de Fourier ou des fonctions statistiques ont été proposées. Ces techniques sont décrites ci-dessous.

2.2.4.1 Mesures de similarité

Différences de niveau de gris. La façon la plus simple de recalcr deux images est de comparer directement leurs niveaux de gris pixel par pixel. Soit $I_R(\mathbf{x})$ l'image de référence et $I_D(\mathbf{x})$ l'image à recalcr sur I_R , avec $\mathbf{x} = (x, y)^T$. Pour effectuer le recalcr, nous pouvons calculer le minimum de la *Somme des Différences au Carré*, (*SDC*), entre les images,

$$SDC = \sum_{\mathbf{x}} [I_D(T(\mathbf{x})) - I_R(\mathbf{x})]^2, \quad (2.29)$$

où T est la transformation appliquée aux coordonnées de l'image source $I_D(\mathbf{x})$. Ce genre de mesure est bien adaptée à des applications où les images sont issues d'un même capteur (monomodalité) avec des conditions d'éclairage très proches entre les images à recalcr.

Corrélation croisée. Une alternative à la différence d'intensité entre les images est l'utilisation d'une *corrélation croisée* qui permettra de maximiser pixel à pixel le produit de deux images,

$$CC = \sum_{\mathbf{x}} I_R(\mathbf{x})I_D(T(\mathbf{x})), \quad (2.30)$$

Un problème lié à la corrélation est que si une région dans l'image I_D a des valeurs d'intensité très supérieures au reste de l'image, il est très probable que le maximum de la fonction de corrélation sera faussé par cette partie de l'image. C'est pour cela, qu'en général la *corrélation croisée normalisée* est préférée :

$$CCN = \frac{\sum_{\mathbf{x}} [I_R(\mathbf{x}) - \hat{I}_R] [I_D(T(\mathbf{x})) - \hat{I}_D]}{\sqrt{\sum_{\mathbf{x}} [I_R(\mathbf{x}) - \hat{I}_R]^2 [I_D(T(\mathbf{x})) - \hat{I}_D]^2}}, \quad (2.31)$$

où

$$\begin{aligned} \hat{I}_R &= \frac{1}{N} \sum_{\mathbf{x}} I_R(\mathbf{x}), \\ \hat{I}_D &= \frac{1}{N} \sum_{\mathbf{x}} I_D(T(\mathbf{x})) \end{aligned}$$

sont les moyennes des images et N est le nombre de pixels dans les images. Le résultat de la corrélation croisée normalisée est toujours compris entre -1 et 1. Cependant, cette mesure est non définie si l'une des images a une variance égale à zéro (rare en pratique). Cette mesure de similarité est souvent utilisée pour aligner des images dont le lien est décrit par des translations 2D, mais peut également être utilisée s'il existe des variations légères de rotation et de changement d'échelle. Cette mesure de similarité est utilisée lors du recalage d'images monomodale. La *CCN* est moins sensible aux changements d'illumination que la *SDC*.

Recalage basé sur la transformée de Fourier. Lorsque entre deux images, les composantes dominantes de la transformation sont des translations, une approche basée sur la transformée de Fourier peut être suffisante. Le principe de cette technique repose sur la propriété de translation de la transformée de Fourier. En effet, la transformée de Fourier \mathcal{F} d'une image $I_D(\mathbf{x} + \mathbf{t})$ ayant subi une translation $\mathbf{t} = (t_x, t_y)$, aura la même image des modules que celle de l'image originale $I_D(\mathbf{x})$, mais l'image des phases variera proportionnellement avec le déplacement,

$$\mathcal{F}\{I_D(\mathbf{x} + \mathbf{t})\} = \mathcal{F}\{I_D(\mathbf{x})\} e^{-2\pi j\mathbf{t}\mathbf{f}} = \mathcal{I}_D(\mathbf{f}) e^{-2\pi j\mathbf{t}\mathbf{f}} \quad (2.32)$$

où $\mathbf{f} = (u, v)^T$ est le vecteur des fréquences spatiales de l'image et $\mathcal{I}_D(\mathbf{f})$ dénote la transformée de Fourier de l'image $I_D(\mathbf{x})$.

Une autre propriété utile de la transformée de Fourier est qu'une convolution dans le domaine spatial correspond à une multiplication dans le domaine fréquentiel (théorème de Plancherel). Donc, la transformée de Fourier de la corrélation croisée des images peut être exprimée comme suit :

$$\mathcal{F}\left\{\sum_{\mathbf{x}} I_R(\mathbf{x}) I_D(\mathbf{x} + \mathbf{t})\right\} = \mathcal{F}\{I_R(\mathbf{t}) \otimes I_D(\mathbf{t})\} = \mathcal{I}_R(\mathbf{f}) \mathcal{I}_D^*(\mathbf{f}), \quad (2.33)$$

où \otimes dénote l'opération de convolution et $\mathcal{I}_D^*(\mathbf{f})$ est le conjugué complexe de $\mathcal{I}_D(\mathbf{f})$. En résumé, pour évaluer de façon efficace la corrélation croisée entre les images sur toute la gamme de valeurs possibles de \mathbf{t} , la transformée de Fourier des images $I_R(\mathbf{x})$ et $I_D(\mathbf{x} + \mathbf{t})$ est calculée, pour ensuite, effectuer le produit de la transformée de la première image par le conjugué de la deuxième image. Finalement, la transformée de Fourier inverse est calculée. L'algorithme de la transformée rapide de Fourier (*FFT, Fast Fourier Transform*) permet de calculer la transformée d'une image de taille $N \times M$ en $O(NM \log NM)$ opérations (N et M étant puissances de 2). Ceci est sensiblement plus rapide que les $O(N^2 M^2)$ opérations requises par une recherche exhaustive dans les parties communes des images à recalculer.

Corrélation des phases. Une variante de la corrélation croisée dans le domaine de Fourier, aussi utilisée pour l'estimation du déplacement entre images, est la *corrélation des phases* (CC_{CP}) [Kuglin and Hines, 1975]. Dans cette méthode, le spectre des deux images est « nettoyé » en divisant le produit des transformées (équation 2.33) par les modules des transformées avant d'appliquer la transformée inverse de Fourier,

$$CC_{CP}(\mathbf{t}) = \mathcal{F}^{-1} \left\{ \frac{\mathcal{I}_R(\mathbf{f})\mathcal{I}_D^*(\mathbf{f})}{\|\mathcal{I}_R(\mathbf{f})\| \|\mathcal{I}_D(\mathbf{f})\|} \right\}. \quad (2.34)$$

Dans l'idéal, le résultat est une impulsion unique localisée à la valeur correcte de \mathbf{t} . Cette caractéristique simplifierait l'estimation du déplacement.

Transformée de Fourier-Mellin. Même si le recalage basé sur la transformée de Fourier est utilisé pour estimer des déplacements entre les images, il peut être utilisé aussi, sous certaines conditions, pour calculer des rotations et des changements d'échelle isotropes sans aucun déplacement. D'abord, il faut estimer uniquement la rotation entre les images. Si le lien géométrique entre deux images est une rotation $\mathbf{R}(\hat{\theta})$ où $\hat{\theta}$ est le déplacement angulaire, tel que

$$I_D(\mathbf{R}\mathbf{x}) = I_R(\mathbf{x}), \quad (2.35)$$

alors ces images peuvent être ré-échantillonnées en coordonnées *polaires* en prenant comme origine le centre des images (figure 2.11),

$$\bar{I}_R(\rho, \theta) = I_R(\rho \cos \theta, \rho \sin \theta) \quad \text{et} \quad \bar{I}_D(\rho, \theta) = I_D(\rho \cos \theta, \rho \sin \theta), \quad (2.36)$$

où

$$\rho = \sqrt{(x - x_c)^2 + (y - y_c)^2} \quad \text{et} \quad \theta = \tan^{-1} \left(\frac{y - y_c}{x - x_c} \right), \quad (2.37)$$

avec x_c et y_c les coordonnées du centre des images. De ces équations il peut être déduit que :

$$\bar{I}_D(\rho, \theta + \hat{\theta}) = \bar{I}_R(\rho, \theta), \quad (2.38)$$

où le symbole $\bar{}$ représente les images ré-échantillonnées.

Il est alors possible de calculer ce décalage angulaire en utilisant la corrélation basée sur la transformée de Fourier (équation 2.33) ou sur la corrélation des phases (équation 2.34).

Si entre les deux images, il existe aussi un facteur d'échelle \hat{s} , ce changement peut être représenté en utilisant des coordonnées logarithmiques sur l'axe ρ des coordonnées polaires :

$$I_D(e^{\hat{s}} \mathbf{R}\mathbf{x}) = I_R(\mathbf{x}), \quad (2.39)$$

avec $\hat{s} = \log \hat{\rho}$ le déplacement sur l'axe $\log \rho$ qui représente le changement d'échelle. En faisant le ré-échantillonnage des images en coordonnées *logarithmiques-polaires* (figure 2.12), nous avons

$$\bar{I}_R(s, \theta) = I_R(e^s \cos \theta, e^s \sin \theta) \quad \text{et} \quad \bar{I}_D(s, \theta) = I_D(e^s \cos \theta, e^s \sin \theta), \quad (2.40)$$

avec $s = \log \rho$, $\bar{I}_R(s, \theta)$ et $\bar{I}_D(s, \theta)$ les images ré-échantillonnées en coordonnées logarithmiques-polaires, telles que

$$\bar{I}_D(s + \hat{s}, \theta + \hat{\theta}) = \bar{I}_R(s, \theta). \quad (2.41)$$

Il est maintenant possible d'appliquer la corrélation basée sur la transformée de Fourier pour trouver les déplacements angulaire et en échelle entre les images.

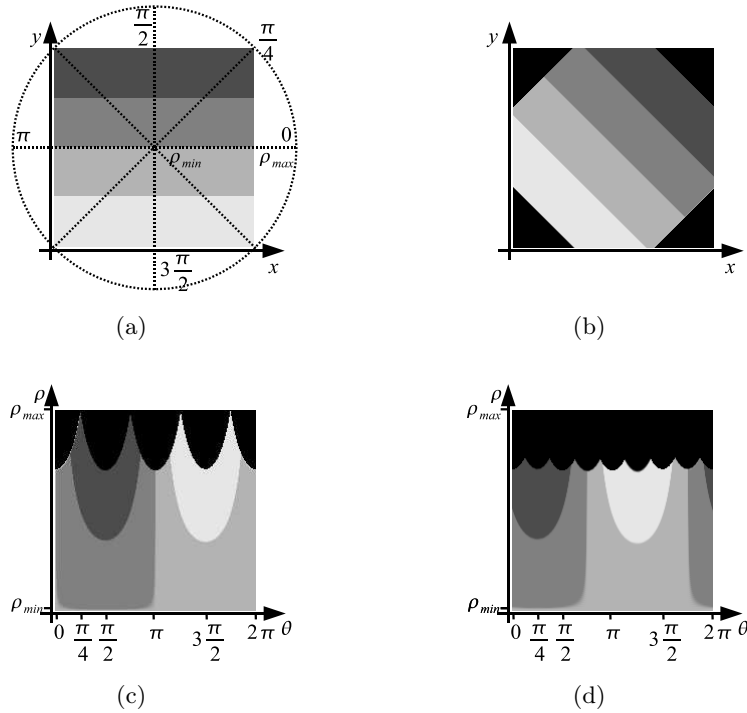


FIG. 2.11 – Ré-échantillonnage d'images en coordonnées *polaires*. (a) Image originale. (b) Image après une rotation de -45° . (c) Ré-échantillonnage de l'image originale en coordonnées polaires. (d) Image avec rotation en coordonnées polaires. Le décalage sur l'axe θ entre les images (c) et (d) est visible. Ce déplacement peut être obtenu par corrélation croisée en utilisant la FFT.

Pour le cas où une translation existe aussi entre les images (pour former un modèle de transformation de similitude),

$$I_D(e^{\hat{s}} \mathbf{R}\mathbf{x} + \mathbf{t}) = I_R(\mathbf{x}), \quad (2.42)$$

il est aussi possible de récupérer cette translation mais il faut appliquer une procédure spécialement adapté à ce cas. Sur la base d'un travail de [De Castro and Morandi, 1987], [Reddy and Chatterji, 1996] ont développé un algorithme pour estimer les paramètres d'une similitude. D'abord, il faut obtenir la transformée de Fourier des images, pour ne retenir que les amplitudes des spectres. Cette amplitude est censée être invariante aux déplacements dans le domaine spatial (si les spectres sont cycliques). Lorsque l'échelle et la rotation sont connues, une image peut être obtenue avec l'inverse de ces paramètres appliquée à l'image à transformer pour ensuite estimer le déplacement par la corrélation basée sur la transformée de Fourier.

Pour que cet algorithme fonctionne, il est nécessaire que le déplacement entre les images soit petit (quelques pixels seulement) et que le contenu informatif dans les images soit important.

Critère de Woods. Le recalage d'images multimodales (images issues de capteurs basés sur des principes physiques différents) est très souvent nécessaire, spécialement dans le domaine médical, mais compliqué du fait des différences de nature, de taille ou de résolution de ces images. Par exemple, la comparaison entre des images anatomiques et fonctionnelles d'un patient permet d'améliorer et/ou de préciser le diagnostic. Dans ce cas, une approche basée sur une mesure statistique de la distribution des pixels dans l'image est préférée. Le critère de Woods [Woods *et al.*, 1993] est une mesure de similarité initialement créée pour le recalage IRM-TEP. Ce critère

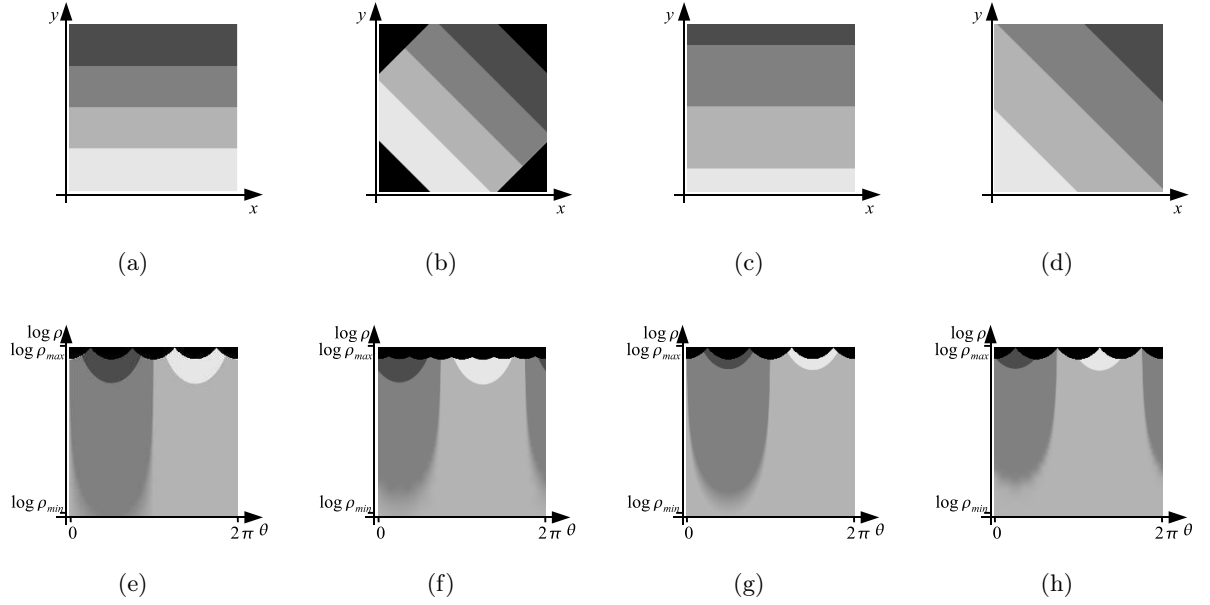


FIG. 2.12 – Ré-échantillonnage d'images en coordonnées *log-polaires*. Les zones en noir représentent les couples (ρ, θ) non contenus dans les images (a) et (b). (a) Image originale. (b) Image après une rotation de -45° . (c) Un grossissement de l'image (a) avec un facteur d'échelle de 1,5. (d) Image (a) après rotation de -45° et changement d'échelle de facteur 1,5. (e) Image originale ré-échantillonnée en coordonnées log-polaires. (f) Image en coordonnées log-polaires de l'image (b). Il existe un déplacement sur l'axe θ . (g) Un déplacement sur l'axe $\log \rho$ (variation de l'amplitude) est visible lors du ré-échantillonnage de l'image (c) en coordonnées log-polaires. (h) Déplacements sur l'axe θ et $\log \rho$ pour l'image avec rotation et changement d'échelle. Ces déplacements peuvent être obtenus par corrélation croisée en utilisant la FFT.

suppose que les pixels ayant une intensité D dans l'image I_D correspondent tous à des pixels d'intensité R dans l'image I_R si les deux images sont recalées. De façon simplifiée, le critère de Woods indique que l'histogramme conjoint des images sera moins étalé lorsque les images seront recalées. Ce comportement est illustré dans la figure 2.13. Ce critère peut être calculé de la façon suivante

$$CW(I_R|I_D) = \sum_R \frac{\sigma_{I_D|R}}{m_{I_D|R}} p_{RD} \quad (2.43)$$

avec

$$m_{I_D|R} = \frac{1}{p_R} \sum_D D p_{RD} \quad \text{et} \quad (2.44)$$

$$\sigma_{I_D|R} = \sqrt{\frac{1}{p_R} \sum_D (D - m_{I_D|R})^2 p_{RD}}$$

où R et D sont respectivement, les niveaux de gris des images I_R et I_D , $m_{I_D|R}$ et $\sigma_{I_D|R}$ sont respectivement, une moyenne et une variance conditionnelles, c'est-à-dire, la moyenne et la variance des niveaux de gris D de l'image I_D correspondants tous à un même niveau de gris R dans l'image I_R . p_R est la probabilité d'un niveau d'intensité R dans l'image I_R . p_{RD} est la probabilité jointe des niveaux d'intensité R et D . $CW(I_R|I_D)$ est minimal lorsque les images sont recalées.

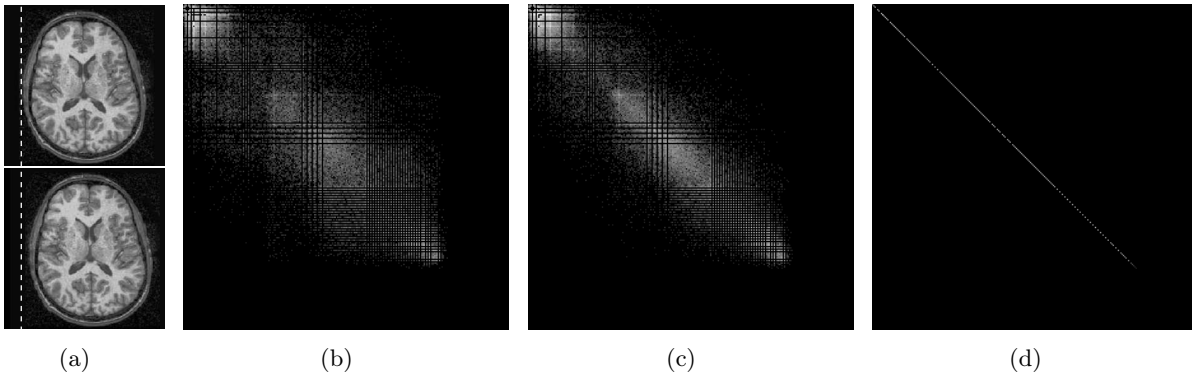


FIG. 2.13 – Histogramme joint de deux images IRM. (a) Deux images IRM décalées l'une par rapport à l'autre (translation sur l'axe horizontal). (b) Histogramme joint pour un déplacement de 4 pixels. (c) Histogramme joint pour un déplacement de 2 pixels. (d) Histogramme lorsque les images sont recalées.

Information mutuelle. L'information mutuelle prend son origine dans la théorie de l'information. C'est une mesure de la dépendance statistique entre deux groupes de données. L'information mutuelle entre deux variables aléatoires I_R et I_D (les images à recalcer) est donnée par

$$IM(I_R, I_D) = H(I_R) - H(I_R|I_D) = H(I_R) + H(I_D) - H(I_R, I_D), \quad (2.45)$$

avec

$$\begin{aligned} H(I_R) &= - \sum_{\mathbf{R}} p_{\mathbf{R}} \ln(p_{\mathbf{R}}) \\ H(I_D) &= - \sum_{\mathbf{D}} p_{\mathbf{D}} \ln(p_{\mathbf{D}}) \\ H(I_R, I_D) &= - \sum_{\mathbf{R}} \sum_{\mathbf{D}} p_{\mathbf{RD}} \ln(p_{\mathbf{RD}}) \end{aligned} \quad (2.46)$$

où $H(I_R)$ et $H(I_D)$ sont les *entropies* des niveaux de gris des images I_R et I_D respectivement et $H(I_R|I_D)$ est l'entropie conditionnelle. Les entropies $H(I_R)$ et $H(I_D)$ prises de façon séparée, mesurent la complexité des images, tandis que l'entropie jointe $H(I_R, I_D)$ mesure la quantité d'information que les images I_R et I_D apportent en même temps. $p_{\mathbf{R}}$ et $p_{\mathbf{D}}$ sont des fonctions de densité de probabilité des niveaux de gris \mathbf{R} et \mathbf{D} . $p_{\mathbf{RD}}$ est une fonction de probabilité jointe. Si les images sont proches, une image fournit la même information que l'autre et l'entropie jointe est minimale. Pour des applications pour lesquelles l'information mutuelle est effectivement adaptée, l'information mutuelle est maximale si les images sont recalées. Dans ce cas, la maximisation de l'information mutuelle conduit à la superposition des images. L'un des premiers travaux concernant l'utilisation de l'information mutuelle est celui développé par [Viola and Wells III, 1997]. Ils décrivent l'utilisation de cette technique pour recalcer des images IRM et pour recalcer des objets 3D à une scène réelle. Les auteurs utilisent la fenêtre de Parzen pour calculer les fonctions de densité de probabilité. En utilisant une méthode similaire, [Miranda-Luna *et al.*, 2004b] construisent des images panoramiques à partir d'images endoscopiques. Cette dernière méthode est expliquée avec plus de détails dans la section 3.1.

Actuellement, le recalage basé sur l'information mutuelle est une technique très utilisée pour le recalage d'images multimodales. L'inconvénient de cette mesure est le coût en terme de calcul,

car la détermination de l'information mutuelle est longue par rapport aux mesures de similarité présentées précédemment.

2.2.4.2 Méthodes d'optimisation

Pour choisir une méthode d'optimisation, il existe plusieurs critères à prendre en compte : il faut considérer la forme de l'espace des paramètres (convexe ou non, avec un seul extremum global avec ou sans plusieurs extrema locaux), il faut tenir compte du fait que la fonction à optimiser est analytique ou si sa dérivée doit être calculée numériquement, il faut savoir si les conditions initiales sont proches de la solution et il faut respecter le temps de calcul disponible pour l'application.

Dans la suite, nous allons appeler $f(\mathbf{p})$ la fonction à optimiser, où \mathbf{p} est un vecteur à N dimensions qui contient les paramètres à optimiser. Nous allons considérer aussi que la minimisation de la fonction $f(\mathbf{p})$ est équivalente à la maximisation de la fonction $-f(\mathbf{p})$.

Un algorithme d'optimisation est un processus itératif par lequel nous cherchons de nouvelles valeurs de \mathbf{p}_k , à chaque itération k , l'inégalité suivante devant être vérifiée : $f(\mathbf{p}_{k+1}) < f(\mathbf{p}_k)$. À partir de conditions initiales (paramètres \mathbf{p}_0), nous allons, à chaque itération, ajouter des variations $\Delta\mathbf{p}$ aux paramètres \mathbf{p}_k tel que $\mathbf{p}_{k+1} = \mathbf{p}_k + \Delta\mathbf{p}$. Pour définir un critère d'arrêt du processus de recherche, nous pouvons établir un seuil minimal de variation ϵ de la fonction $f(\mathbf{p})$ en dessous duquel l'extremum est supposé atteint. Cette idée peut être formulée mathématiquement par : $\|f(\mathbf{p}_{k+1}) - f(\mathbf{p})\| < \epsilon$ ou $\nabla f(\tilde{\mathbf{p}}) \rightarrow 0$, avec $\tilde{\mathbf{p}}$ les paramètres optimisés. Beaucoup de méthodes d'optimisation utilisent le gradient de la fonction à optimiser. Pour les méthodes faisant usage des dérivées, nous définissons le gradient de la fonction $\nabla f(\mathbf{p})$ comme suit :

$$\nabla f(\mathbf{p}) = \begin{bmatrix} \frac{\partial f}{\partial p_1} \\ \frac{\partial f}{\partial p_2} \\ \vdots \\ \frac{\partial f}{\partial p_n} \end{bmatrix}, \quad (2.47)$$

avec n le nombre total de paramètres. Ce vecteur est aussi connu sous le nom de *Jacobien* de la fonction. Le *Hessien* H de la fonction est une matrice composée des dérivés secondes (ou des approximations de ces dérivées) de la fonction par rapport aux paramètres \mathbf{p} :

$$H = \nabla^2 f(\mathbf{p}) = \begin{bmatrix} \frac{\partial^2 f}{\partial p_1^2} & \frac{\partial^2 f}{\partial p_1 \partial p_2} & \cdots & \frac{\partial^2 f}{\partial p_1 \partial p_n} \\ \frac{\partial^2 f}{\partial p_2 \partial p_1} & \frac{\partial^2 f}{\partial p_2^2} & \cdots & \frac{\partial^2 f}{\partial p_2 \partial p_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial p_n \partial p_1} & \frac{\partial^2 f}{\partial p_n \partial p_2} & \cdots & \frac{\partial^2 f}{\partial p_n^2} \end{bmatrix}. \quad (2.48)$$

Cependant, selon le problème, le calcul du Jacobien dans le cadre d'une optimisation peut nécessiter plus de temps que le calcul de la valeur de la fonction en elle-même. Le calcul du Hessien peut parfois, prendre beaucoup plus de place mémoire et de temps de calcul que le Jacobien, surtout lorsque le vecteur \mathbf{p} est grand. Les techniques d'optimisation qui n'utilisent pas le Hessien, requièrent souvent plus d'itérations que celles qui l'utilisent et sont plus lentes et moins fiables [Moreno Velo *et al.*, 2001].

Pour estimer les incréments à ajouter à l'estimation actuelle des paramètres \mathbf{p} , nous allons utiliser la relation suivante [Press *et al.*, 1992] :

$$\mathbf{p}_{k+1} = \mathbf{p}_k + \lambda_k \mathbf{d}_k, \quad (2.49)$$

avec \mathbf{d}_k un vecteur fixant la direction de recherche de l'extremum de la fonction et λ l'amplitude du « pas » à faire dans cette direction.

Pour présenter les méthodes d'optimisation nous allons considérer une *fonction quadratique*, dont la forme généralisée pour N -dimensions est donnée par l'équation suivante :

$$f(\mathbf{p}) = \mathbf{a} + \mathbf{b}^T \mathbf{p} + \frac{1}{2} \mathbf{p}^T \mathbf{A} \mathbf{p}. \quad (2.50)$$

où \mathbf{a} et \mathbf{b} sont des vecteurs et \mathbf{A} une matrice contenant des nombres réels différents de zéro.

Ci-dessous, nous présentons une vue d'ensemble pour quelques méthodes d'optimisation utilisées dans la littérature et concernant le recalage d'images.

Méthodes de premier ordre : gradient descendant. La méthode itérative la plus simple pour trouver l'extremum d'un espace de paramètres construit pour une mesure de similarité est la méthode du gradient descendant ou *steepest descent*, en anglais. Pour cette méthode le gradient de la fonction à optimiser doit pouvoir être estimé. La direction de recherche dans l'espace de paramètres est donnée par

$$\mathbf{d}_{k+1} = -\nabla f(\mathbf{p}_k), \quad (2.51)$$

qui est l'approximation ou une estimation exacte de la dérivée au point \mathbf{p} de l'espace de recherche et qui va nous indiquer la direction avec la variation la plus importante dans l'espace des paramètres.

À chaque itération, la nouvelle direction de recherche sera orthogonale à la direction du pas précédent, c'est-à-dire $\nabla f(\mathbf{p}_k) \cdot \nabla f(\mathbf{p}_{k-1}) = 0$ [Capel, 2001]. Si nous considérons une fonction 2D, le chemin suivi par cette méthode pour arriver à l'extremum aura une forme de zigzag. Ce comportement peut être observé dans la figure 2.14, qui est un grossissement de la figure 2.16(b). En général, cette méthode d'optimisation est extrêmement lente et inefficace si l'amplitude du pas λ est constante. C'est pour cela qu'en pratique, des paramètres qui modifient l'amplitude du pas λ à chaque itération sont ajoutés [Ericsson and Åström, 2003] [Alterovitz *et al.*, 2004].

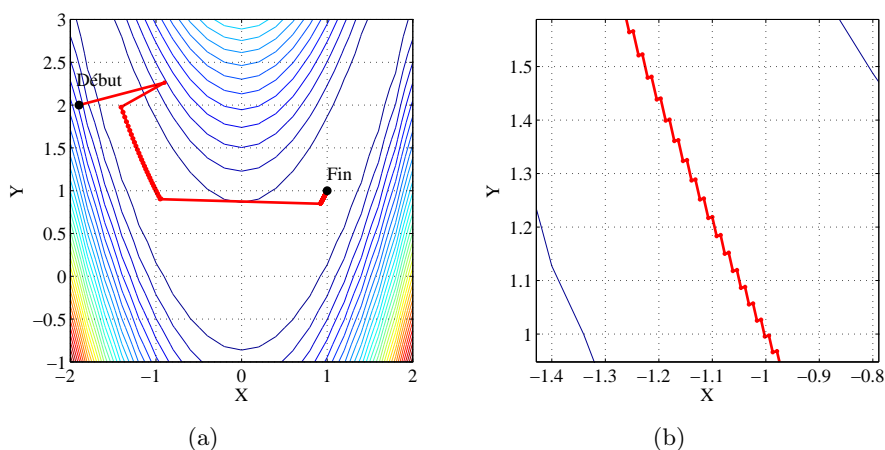


FIG. 2.14 – (a) Minimisation de la fonction de Rosenbrock (cf. figure 2.16) par la méthode du gradient descendant. (b) Vue agrandie d'une zone de la figure (a) qui montre que les déplacements effectués par la méthode du gradient descendant sont effectués dans des directions orthogonales.

Méthode de premier ordre améliorée : gradient conjugué. La raison pour laquelle la méthode du gradient descendant converge de façon lente est qu'elle effectue des déplacements orthogonaux entre eux. Par conséquent, deux recherches consécutives dans la même direction dans l'espace de paramètres sont effectuées. L'idée de la méthode du gradient conjugué est de faire en sorte que les directions de recherche \mathbf{d}_k soient dépendantes des directions de recherche précédentes. Le terme « conjugué » est utilisé pour indiquer que deux vecteurs \mathbf{d}_k et \mathbf{d}_m , sont orthogonaux par rapport à une matrice définie positive (toutes ses valeurs propres sont positives) ; par exemple \mathbf{A} dans l'équation 2.50, de sorte que $\mathbf{d}_k^T \cdot \mathbf{A} \cdot \mathbf{d}_m = 0$. Si nous considérons une fonction quadratique N -dimensionnelle, des directions construites par cette méthode peuvent amener vers l'extremum en N itérations. Les directions conjuguées sont données par

$$\mathbf{d}_{k+1} = -\nabla f(\mathbf{p}_k) + \beta_k \mathbf{d}_k, \quad (2.52)$$

où β_k peut être calculé par la formule de Fletcher-Reeves [Press *et al.*, 1992] :

$$\beta = \frac{\mathbf{g}_{k+1}^T \cdot \mathbf{g}_{k+1}}{\mathbf{g}_k^T \cdot \mathbf{g}_k}. \quad (2.53)$$

avec $\mathbf{g}_k = \nabla f(\mathbf{p}_k)$. Pour plus de détails sur la méthode, se référer à [Press *et al.*, 1992].

En terme de convergence (nombre d'itérations), la méthode d'optimisation par gradient conjugué, est de loin supérieure à la méthode du gradient descendant. Cette méthode est particulièrement adaptée à l'optimisation de fonctions quadratiques avec un nombre de paramètres N grand. [Cuzol *et al.*, 2005] ont utilisé cette méthode d'optimisation pour recalibrer des images IRM pondérées en T2 pour l'étude de la sclérose en plaques.

Méthodes du deuxième ordre (ou méthodes de Newton). Une alternative à la méthode du gradient conjugué et qui a aussi une vitesse de convergence plus élevée que la méthode du gradient descendant repose sur l'utilisation des dérivées secondes (ou le *Hessien*) de la fonction à optimiser. Ces méthodes considèrent une fonction $f(\mathbf{p} + \Delta\mathbf{p})$ pour laquelle nous cherchons les incréments $\Delta\mathbf{p}$ à ajouter aux paramètres \mathbf{p} pour obtenir $\nabla f(\mathbf{p} + \Delta\mathbf{p}) = 0$ (un extremum). Nous pouvons faire un développement limité de la fonction en série de Taylor (au second ordre) pour avoir :

$$f(\mathbf{p} + \Delta\mathbf{p}) \approx f(\mathbf{p}) + \nabla f(\mathbf{p})^T \Delta\mathbf{p} + \frac{1}{2} (\Delta\mathbf{p})^T \nabla^2 f(\mathbf{p}) \Delta\mathbf{p}. \quad (2.54)$$

En dérivant la fonction $f(\mathbf{p})$ par rapport à $\Delta\mathbf{p}$ nous obtenons :

$$\nabla f(\mathbf{p} + \Delta\mathbf{p}) \approx \nabla f(\mathbf{p}) + \nabla^2 f(\mathbf{p}) \Delta\mathbf{p} = \nabla f(\mathbf{p}) + H \Delta\mathbf{p}, \quad (2.55)$$

avec H le Hessien de la fonction $f(\mathbf{p})$. Lorsqu'un extremum est atteint, $\nabla f(\mathbf{p} + \Delta\mathbf{p}) = 0$ et donc $\nabla f(\mathbf{p}) + H \Delta\mathbf{p} = 0$, d'où

$$\Delta\mathbf{p} = -H^{-1} \nabla f(\mathbf{p}). \quad (2.56)$$

La valeur $\Delta\mathbf{p}$ est désignée par le terme « pas de Newton » [Triggs *et al.*, 2000]. Nous pouvons maintenant calculer la valeur de \mathbf{d}_{k+1} :

$$\mathbf{d}_{k+1} = -H_k^{-1} \cdot \nabla f(\mathbf{p}_k). \quad (2.57)$$

Puisque le calcul des dérivées secondes requiert des temps importants, le Hessien est calculé généralement en utilisant les valeurs du gradient. Cette idée est utilisée par des méthodes d'optimisation comme les algorithmes de *Broyden-Fletcher-Goldfarb-Shanno* et *Davidon-Fletcher-Powell*. Ces méthodes sont connues aussi sous le terme « méthodes *quasi-Newton* ». L'idée de

base de ces méthodes est de construire à chaque itération k des bonnes *approximations* Q_k^{-1} de l'inverse du Hessien H . Donc, de façon similaire à la méthode Newton, nous avons :

$$\mathbf{d}_{k+1} = -Q_k^{-1} \cdot \nabla f(\mathbf{p}_k). \quad (2.58)$$

Lors du choix des paramètres initiaux, nous pouvons choisir classiquement $\mathbf{d}_0 = f(\mathbf{p}_0)$ et $Q_0^{-1} = \mathbf{I} \cdot \mathbf{d}_0$. Dans la première approximation du Hessien, \mathbf{I} est la matrice identité. Une étude sur ces méthodes et leur application au recalage d'images médicales est présentée par [Maes *et al.*, 1999].

Dans les cas où le problème d'optimisation peut être exprimé au sens des moindres carrés, la méthode d'optimisation de Gauss-Newton peut être appliquée. Cette méthode calcule l'approximation Q du Hessien en considérant que la fonction à minimiser est représentée par $f^2(\mathbf{p})$, la dérivée première de cette dernière étant $2f(\mathbf{p})\nabla f(\mathbf{p})$ et la dérivée seconde $2\nabla f(\mathbf{p})^T \nabla f(\mathbf{p}) + f(\mathbf{p})\nabla^2 f(\mathbf{p})$. La fonction $f(\mathbf{p})$ tend vers 0 lorsque les paramètres se trouvent proches de la valeur optimale et dans ce cas la dérivée seconde se réduit à $G \approx 2\nabla f(\mathbf{p})^T \nabla f(\mathbf{p})$. Cette approximation est dite « l'approximation de Gauss-Newton du Hessien ». En faisant la substitution dans l'équation de la méthode de Newton (équation 2.57) nous obtenons l'équation du pas de la méthode de Gauss-Newton :

$$\mathbf{d}_{k+1} = -G_k^{-1} \cdot \nabla f(\mathbf{p}_k). \quad (2.59)$$

L'algorithme de Gauss-Newton peut diverger si le Hessien de la fonction à optimiser en un point $\mathbf{x} = \mathbf{p}$ donné n'est pas défini positif (par exemple si le point \mathbf{x} est un « point selle »). Pour palier ce problème, l'algorithme de *Levenberg-Marquardt* introduit un terme adaptatif. Cette méthode propose d'ajouter une matrice $\lambda \mathbf{I}$ (\mathbf{I} est une matrice identité et λ un paramètre de régularisation) à l'approximation du Hessien par la méthode de Gauss-Newton. Le pas pour cette méthode est ainsi défini par

$$\mathbf{d}_{k+1} = - [G_k^{-1} + \lambda \mathbf{I}] \cdot \nabla f(\mathbf{p}_k). \quad (2.60)$$

Loin de l'extremum de la fonction, le paramètre λ est petit et l'équation 2.60 peut être approximée par : $\mathbf{d}_{k+1} = -G_k^{-1} \cdot \nabla f(\mathbf{p}_k)$. Cette forme nous amènera rapidement dans la région de l'extremum. À proximité de la solution, en fixant λ à une grande valeur, le terme $\lambda \mathbf{I}$ devient dominant et l'équation 2.60 devient $\mathbf{d}_{k+1} = -\nabla f(\mathbf{p}_k)$ qui correspond au pas de la méthode du gradient descendant, qui trouve de façon précise l'extremum de la fonction.

Les efficacités des méthodes de Gauss-Newton et Levenberg-Marquardt sont comparés dans [Baker and Matthews, 2002] dans une application au suivi d'objets.

Méthodes de degré zéro (ou méthodes sans dérivées). Lorsque le gradient de la fonction à optimiser ne peut pas être estimé parce que l'opération est trop coûteuse en temps ou parce que son évaluation est impossible, des méthodes d'optimisation n'utilisant pas de dérivée doivent être utilisées. Un exemple de ces méthodes est l'algorithme d'optimisation du *simplexe* de Nelder-Mead, qui utilise un ensemble d'évaluations de la fonction à optimiser pour changer à chaque itération un des paramètres. Dans cette méthode, un polyèdre à $N + 1$ sommets (le simplexe) est calculé, N étant la taille du vecteur paramètre \mathbf{p} à optimiser et les valeurs de la fonction $f(\mathbf{p}_i)$ définissant les sommets ($i \in [0, N]$). Par exemple, si la fonction à optimiser a deux paramètres, alors le simplexe est un triangle. Le point de départ de l'algorithme est un simplexe dont les sommets définissent les valeurs initiales des paramètres à optimiser. D'abord, la fonction est évaluée à chaque sommet pour identifier le sommet avec la valeur de $f(\mathbf{p})$ la plus grande. En général, les premiers pas de la méthode vont déplacer le point dont la valeur $f(\mathbf{p})$ est la plus grande vers le côté opposé du simplexe. Ces changements de sommets sont appelés

des « réflexions » (figure 2.15). Mais, si la forme de l'espace de recherche le permet, la méthode va effectuer des « expansions » du simplexe pour faire des déplacements plus grands. Lorsque le simplexe atteint une zone de forme convexe (pour une minimisation), la méthode va effectuer des « contractions » du simplexe dans la direction transversale pour essayer de continuer la descente. Un critère d'arrêt communément utilisé est la taille du simplexe : lorsqu'elle est plus petite qu'un seuil établi, la méthode s'arrête. Le choix de la symétrie (réflexions, expansions, contractions) est effectué selon les valeurs de $f(\mathbf{p})$.

Des applications de la méthode du simplexe au recalage multimodalité d'images sont décrites dans [Soman *et al.*, 2003] et dans [Posada *et al.*, 2004] [Posada *et al.*, 2007]. Les méthodes sans dérivées sont en général plus lentes que les méthodes décrites précédemment, mais leur avantage principal est justement qu'il n'est pas nécessaire de calculer la dérivée de la fonction à optimiser. Le simplexe est très précis pour des conditions initiales proches de la solution.

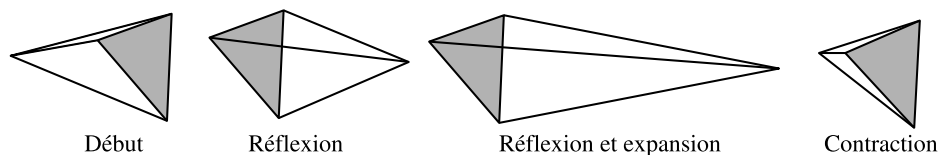


FIG. 2.15 – Règles de base pour les déplacements d'un simplexe à 4 sommets (nombre N de paramètres égal à 3). Les règles sont la réflexion, l'expansion et la contraction, mais il est possible d'effectuer des combinaisons de ces mouvements. Les trois sommets du triangle en gris sont les valeurs utilisées pour calculer le quatrième et nouveau sommet.

Méthodes statistiques. Toutes les méthodes décrites ci-dessus sont sensibles à des extrema locaux prononcés non recherchés qui peuvent perturber la recherche de l'extremum global. Les méthodes d'optimisation statistiques peuvent permettre de converger vers l'extremum global parce qu'elles génèrent des configurations très différentes dans l'espace des paramètres, c'est-à-dire réparties sur l'ensemble de l'espace de recherche. Une façon d'élargir l'occupation de cet espace est de générer des configurations aléatoires pour choisir ensuite les meilleures d'entre elles. Ceci est réalisé par les algorithmes de recherche aveugle (*blind search* en anglais) mais dont la vitesse de convergence est extrêmement lente.

Une solution plus efficace est l'utilisation des algorithmes de type *recuit simulé* [Kirkpatrick *et al.*, 1983]. Leur fonctionnement se base sur l'analogie du processus de recuit d'un matériau en métallurgie. Le recuit implique le réchauffage et le refroidissement lent d'un matériau pour augmenter la taille de ses cristaux et réduire ses défauts. La chaleur fera sortir les atomes de ses états d'énergie initiaux pour les faire osciller de façon aléatoire sur des états d'énergie supérieurs. Le refroidissement lent permet aux atomes de retrouver des configurations avec un état d'énergie plus bas que l'état initial. Ainsi, la partie essentielle du processus est le refroidissement lent du matériau pour permettre aux atomes de se redistribuer pendant qu'ils perdent leur mobilité. Par analogie avec ce processus, nous définissons la fonction à minimiser comme étant l'énergie E du système et un paramètre de contrôle T qui est la *température* du système. À partir d'une énergie initiale E_0 (définissant les paramètres initiaux) prise aléatoirement dans l'espace de recherche et d'une température initiale T_0 « élevée », la méthode va modifier la solution précédente au cours des itérations de sorte que la température décroisse lentement. Ces modifications provoquent des variations ΔE de l'énergie du système. Si cette variation améliore la solution, les paramètres de la fonction sont retenus. Dans le cas contraire, elle est acceptée avec une probabilité $p(E)$ donnée

par la distribution de probabilité de Boltzmann :

$$p(E) \approx e^{-\frac{\Delta E}{kT}}, \quad (2.61)$$

avec k la constante de Boltzmann. L'acceptation d'une solution éloignée de la solution optimale permet à la méthode d'explorer un espace de recherche plus étendu et d'éviter d'être bloquée par un extremum local. Cette méthode est utilisée dans [Black and Anandan, 1991] pour l'estimation du mouvement dans une séquence d'images, tandis que [Johansson, 2006] applique le recuit simulé au mosaïquage d'images.

Il existe d'autres méthodes statistiques d'optimisation telles que les « *algorithmes génétiques* ». Ces algorithmes s'inspirent des règles de l'évolution biologique comme l'hérédité, la mutation, la sélection et le croisement ou recombinaison de gènes. Un algorithme génétique est un processus itératif qui maintient une population d'*individus*. Ces individus sont des solutions possibles à un problème d'optimisation. Chaque itération de l'algorithme est appelée *génération*. À chaque itération, les individus seront classés par rapport à leur proximité de la solution. Basée sur cette classification, une nouvelle population de solutions possibles est formée en utilisant des opérateurs génétiques spécifiques. Chaque individu est représenté par une chaîne ou *chromosome*. Chaque chaîne est composée de caractères (*gènes*) qui ont des valeurs spécifiques (*allèles*). Le placement des caractères dans la chaîne est important. Ces positions spécifiques sont appelées *loci* (pluriel de *locus* en latin, placement). Chaque individu possède une valeur d'aptitude (*fitness value* en anglais), qui est une mesure de la qualité de la solution représentée par l'individu.

Étape de *sélection*. À partir d'une population initiale de P individus (possibles solutions), chaque individu est considéré comme un parent d'un autre individu dont la probabilité de sélection r est définie par :

$$r = 4^{\frac{1}{P-1}}. \quad (2.62)$$

Cette méthode simule un processus de sélection naturelle, où les individus les plus « adaptés » ont plus de probabilité de se reproduire, tandis que les moins « adaptés » peuvent mourir avant la reproduction. De cette façon, les meilleures solutions vont être retenues.

Étape de *croisement*. Après la sélection, le croisement est effectué entre les parents pour créer deux descendants. Les chromosomes (chaînes de caractères) des parents sont divisées en deux parties. Le premier descendant est créé en combinant la première moitié du premier parent et la deuxième moitié du deuxième. Le second descendant est créé avec la première moitié du premier parent et la première moitié du deuxième.

Durant la *mutation*, les descendants peuvent être, de façon aléatoire, modifiés par une mutation qui va altérer leurs chromosomes. La mutation permet à l'algorithme de chercher dans un espace de paramètres plus étendu et d'éviter de rester dans un extremum local de l'espace de recherche.

Étape de *substitution de la population*. Finalement, la création de deux descendances augmente la population de deux individus (passage de P à $P + 2$). Étant donné que nous voulons maintenir la population à une taille constante P , il est nécessaire d'éliminer deux individus. Alors, les deux individus avec l'aptitude la plus basse sont éliminés de la population.

Le critère d'arrêt de l'algorithme correspond à la situation où il n'y a plus d'améliorations de la solution après un nombre de générations W déterminé.

Des exemples d'application de cette méthode peuvent être trouvés dans [Lomonosov *et al.*, 2006] pour le recalage de surfaces 3D et [Huang *et al.*, 2005] pour le recalage d'images CT et des surfaces 3D.

Combinaison des méthodes. Parmi les méthodes d'optimisation précédentes, seule la méthode de Levenberg-Marquardt est basée sur une combinaison de méthodes différentes pour améliorer la vitesse et la précision de convergence. Mais, il est tout à fait possible de combiner autrement différentes méthodes pour tirer partie des avantages de plusieurs algorithmes.

Par exemple, [Wang and Tian, 2005] réalisent une telle combinaison dans le cadre d'une application de recalage d'images IRM pondérées en T1 et d'images CT. Pour ce faire, les auteurs proposent de maximiser l'information mutuelle, calculée à partir du gradient des images, par une méthode d'optimisation qui utilise successivement le recuit simulé (pour se rapprocher du maximum global) et une technique de gradient descendant (pour trouver précisément la solution). Les auteurs montrent que leur approche est moins sensible au maxima locaux que d'autres méthodes qui utilisent une seule méthode d'optimisation.

[Telenczuk *et al.*, 2006] proposent un deuxième exemple d'association de techniques d'optimisation pour recalibrer des images moléculaires (imagerie à rayons X) avec des images de micrographie électronique. Pour ce faire, ils proposent d'utiliser l'information mutuelle optimisée par un algorithme génétique couplé à une méthode de gradient descendant. Ils obtiennent de meilleurs résultats en combinant ces méthodes qu'en les appliquant séparément.

Un autre exemple, sont les travaux de [Posada *et al.*, 2004] [Posada *et al.*, 2007], qui combinent une méthode de type gradient descendant pour se rapprocher rapidement de la solution et la méthode du simplexe pour arriver de façon précise à l'extremum. Cette méthode est utilisée dans le cadre du recalage d'images volumétriques du visage et d'images CT pour positionner le patient dans le repère d'une machine de radiothérapie. Cette combinaison de méthodes d'optimisation permet de converger plus rapidement et plus précisément vers la solution que les méthodes prises de façon isolée.

Exemple du comportement des méthodes. Un exemple illustrant le comportement des méthodes décrites prises séparément (sauf les méthodes statistiques) est donné dans la figure 2.16. La fonction à optimiser est la « célèbre » fonction de *Rosenbrock* en 2D : $f(x, y) = 100(y - x^2)^2 + (1 - x)^2$; $x \in [-2; 2]$, $y \in [-1; 3]$ (voir figure 2.16(a)). Cette fonction est caractérisée par des pentes qui varient peu autour du minimum. Elle possède un seul minimum (nul) en $(x, y) = (1, 1)$. Le point de départ de la minimisation est situé en $(x, y) = (-1.9, 2)$. Le tableau 2.2 donne les résultats obtenus par les méthodes d'optimisation testées, ceci en termes de nombre d'itérations et de coordonnées finales obtenues à la fin du processus de minimisation.

TAB. 2.2 – Résultats obtenus pour la minimisation de la fonction de Rosenbrock pour plusieurs méthodes d'optimisation : GD – Gradient descendant, DFP – Newton, approximation du Hessien par la méthode de Davidon-Fletcher-Powell, GN – Gauss-Newton, LM – Levenberg-Marquardt et NM – Simplexe (Nelder-Mead).

Méthode	Coord. théoriques (x, y)	Coord. calculées (x, y)	Itérations
GD	(1, 1)	(0.9982, 0.9964)	1380
DFP	(1, 1)	(0.9997, 0.9994)	45
GN	(1, 1)	(1, 1)	11
LM	(1, 1)	(1, 1)	18
NM	(1, 1)	(1, 1)	114

Dans ce cas précis, les méthodes de Gauss-Newton et de Levenberg-Marquardt sont celles qui sont les plus précises et qui nécessitent le moins d'itérations. Cependant, il est important de noter que cet exemple ne doit pas être pris comme une généralisation du comportement des algorithmes

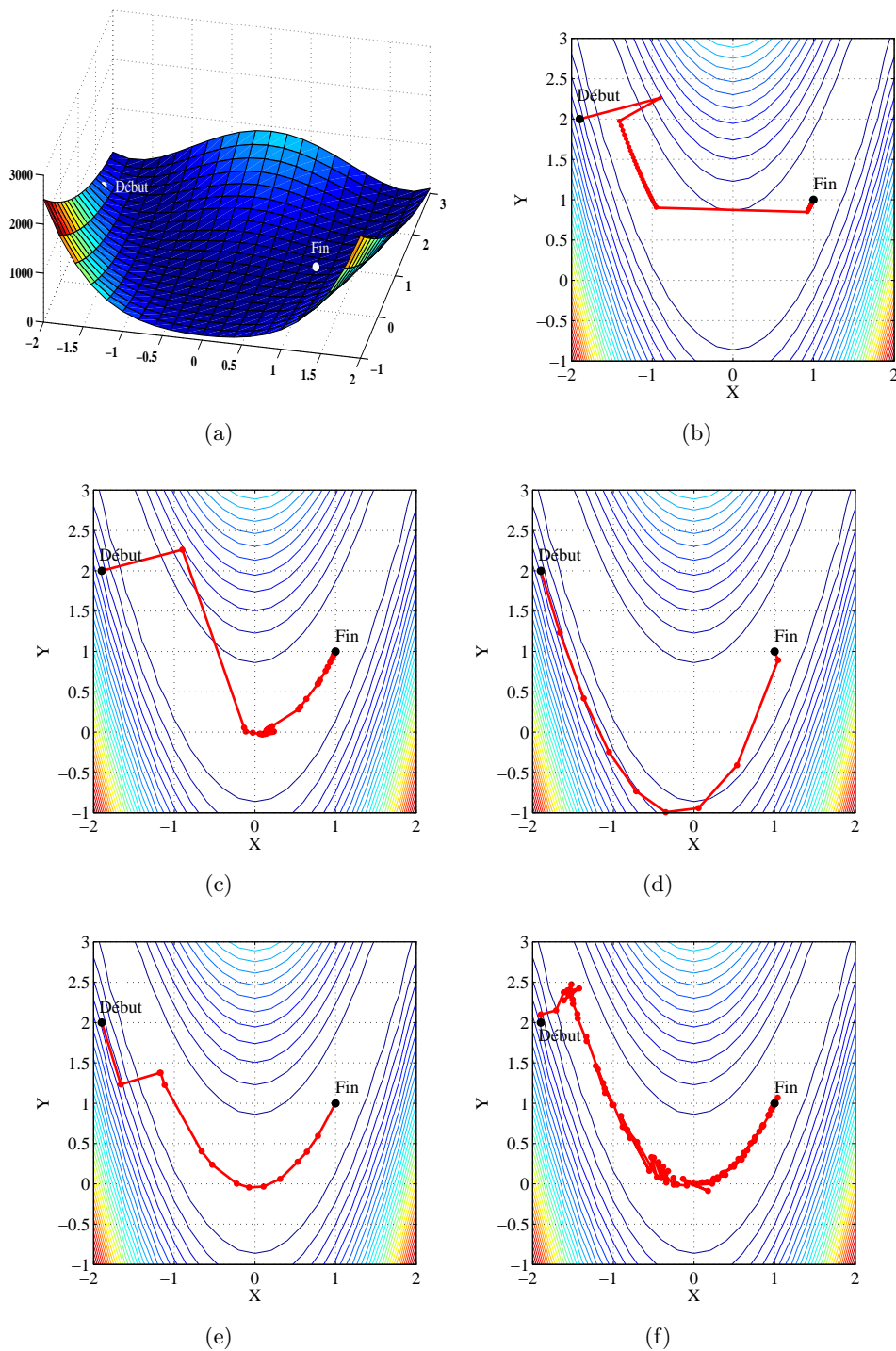


FIG. 2.16 – (a) Fonction de Rosenbrock utilisée pour tester les méthodes d'optimisation. (b) Minimisation par la méthode du gradient descendant. (c) Méthode de Newton avec approximation du Hessien par la méthode de Davidon-Fletcher-Powell. (d) Méthode de Gauss-Newton. (e) Méthode de Levenberg-Marquardt. (f) Méthode du Simplexe (Nelder-Mead).

d'optimisation. Le choix d'une méthode dépendra de l'application et des caractéristiques de la fonction à optimiser.

2.2.5 Recalage basé sur des primitives images

La deuxième approche pour recalibrer les images est basée sur la détection de primitives images. Ces méthodes utilisent des techniques spécifiques de segmentation adaptées à l'application. Les algorithmes de segmentation sont représentés par g dans la définition mathématique du recalage d'images donnée par l'équation 2.15. Ces primitives images à extraire peuvent être des coins, des contours, des régions, *etc.* Il est possible d'extraire des primitives images des deux images à recalibrer, mais il est aussi possible d'en extraire uniquement dans une seule image pour les « comparer » directement aux niveaux d'intensité de l'autre image. Dans le cas plus classique où deux images sont segmentées, une distance entre les primitives est souvent calculée pour mesurer leur degré de superposition (distance euclidienne, distance de Hausdorff, carte de distances, *etc.*). Les paramètres de l'espace de recherche sont à nouveau donnés par la matrice de la transformation géométrique liant les images. Cette méthode de recalage a été utilisée depuis longtemps dans la vision stéréo [Matthies and Okutomi, 1989] et est de plus en plus utilisée dans le mosaïquage d'images [Zoghalmi *et al.*, 1997] [Brown and Lowe, 2003] [Heikkilä and Pietikäinen, 2005].

La première étape dans cette approche est l'extraction des primitives images.

2.2.5.1 Détection des primitives images

Le choix d'un type de primitives images à détecter dépend de plusieurs facteurs. Ainsi, les caractéristiques du capteur sont à prendre en compte (choix d'une primitive la moins sensible possible au bruit caractéristique du capteur). Le choix de la mesure de similarité est à faire en fonction du type et de la quantité de structures présentes dans l'image (lignes, cercles, *etc.*). Souvent, les méthodes de détection de primitives image cherchent sur toute l'image des « candidats » pour ensuite décider si l'information détectée est une primitive image du type souhaité ou pas. Les primitives résultantes sont des régions de l'image qui forment des points isolés, des courbes continues (point connectés) ou des régions connectées. La caractéristique la plus importante d'un détecteur de primitives image est la *répétabilité*, c'est-à-dire, sa capacité pour détecter la même primitive dans plusieurs images de la même scène. Un grand nombre d'algorithmes de détection a été développé. Ces détecteurs varient selon le type de primitive à détecter, la complexité de calcul et la répétabilité. Au final, la qualité du recalage dépend des performances du détecteur de primitives.

Nous présentons, ci-dessous quelques primitives images communément utilisées dans la littérature pour recalibrer des images.

Contours. Les contours sont des régions d'une image correspondant à une forte variation d'intensité. En 1D, le profil d'un tel changement d'intensité aurait de façon idéale la forme d'un échelon, ou pour être plus proche de la réalité, d'une rampe plus ou moins étalée (figure 2.17). Une façon évidente de détecter les variations de niveaux de gris est de dériver le profil 1D. Dans les images, la dérivée doit être remplacée par un gradient qui peut être calculé à partir de deux dérivées (composantes orthogonales, typiquement selon les lignes et colonnes des images). Deux informations importantes qui peuvent être calculées à partir des composantes $\frac{\partial I(\mathbf{x})}{\partial x}$ et $\frac{\partial I(\mathbf{x})}{\partial y}$ du gradient sont l'amplitude et l'orientation du gradient. Étant donnée une image $I(\mathbf{x})$, le module

de son gradient s'exprime comme suit :

$$|\nabla I(\mathbf{x})| = \sqrt{\left(\frac{\partial I(\mathbf{x})}{\partial x}\right)^2 + \left(\frac{\partial I(\mathbf{x})}{\partial y}\right)^2}, \quad (2.63)$$

et l'orientation est donnée par :

$$\angle \nabla I(\mathbf{x}) = \arctan \frac{\frac{\partial I(\mathbf{x})}{\partial x}}{\frac{\partial I(\mathbf{x})}{\partial y}}. \quad (2.64)$$

Un maximum local de l'amplitude du gradient va représenter un point du contour dans l'image. Quand la dérivée première représente un maximum, la dérivée seconde doit avoir une valeur nulle. Ceci, est illustré dans la figure 2.17. C'est pour cette raison qu'une approche alternative dans la détection de contours est de trouver les zéros dans les images des dérivées secondes. L'opérateur différentiel utilisé pour effectuer cette opération est le *Laplacien* :

$$\nabla^2 I(\mathbf{x}) = \frac{\partial^2 I(\mathbf{x})}{\partial x^2} + \frac{\partial^2 I(\mathbf{x})}{\partial y^2}. \quad (2.65)$$

En pratique, des approximations des dérivées des images sont obtenues par la convolution des images avec des filtres. Les filtres approximants des gradients sont aussi appelés opérateurs de gradient. Parmi les plus utilisés, nous pouvons citer l'opérateur de *Roberts*, de *Prewitt*, de *Sobel*, de *Kirch* et de *Robinson*.

Une amélioration de la détection des contours par la méthode du gradient a été proposé par *Canny* [Canny, 1986]. Cette méthode vise à développer un filtre optimal pour détecter un échelon affecté par du bruit blanc. Dans l'article original, Canny propose une série de critères qu'un détecteur de contours optimal doit remplir. Le premier critère, est d'avoir après la détection, un rapport signal (gradient au niveau des contours) à bruit le plus élevé possible (distorsion contour/bruit). Le deuxième critère est que la distance entre les points du contour détecté et les points correspondants dans l'image (contour réel) doit être minimale. Le troisième critère est d'avoir une seule réponse pour chaque contour. Canny a démontré que la dérivée première est le filtre simple qui approxime au mieux le filtre optimal.

Le travail de [Mäkelä *et al.*, 2001] est un exemple de l'application du filtre de Canny à l'extraction des contours d'images IRM qui sont recalés avec des images TEP via une transformation non linéaire. La superposition de ces données doit faciliter le diagnostic de maladies coronaires. [Gay-Bellile *et al.*, 2006] utilisent le détecteur de Canny pour extraire les contours d'une image 2D qui sera recalée avec un modèle 3D déformable en minimisant la différence d'intensité entre les contours et le modèle 3D.

Représenter une image par ses contours, a l'avantage de réduire considérablement la quantité de données tandis que la plupart des informations importantes est retenue. De cette façon il est possible de diminuer le temps requis pour recalcr une paire d'images, mais il est aussi possible que la qualité du recalage en soit affectée.

Coins. Un coin peut être vu comme des variations de niveaux de gris pour lesquelles il y a deux directions dominantes différentes dans un voisinage local centré sur le point d'intérêt. Le terme *coin* est un terme générique. D'ailleurs, un coin peut aussi bien faire partie d'une intersection de droites que de la rencontre de deux segments de droite ou bien encore correspondre à l'intensité maximum dans une courbe.

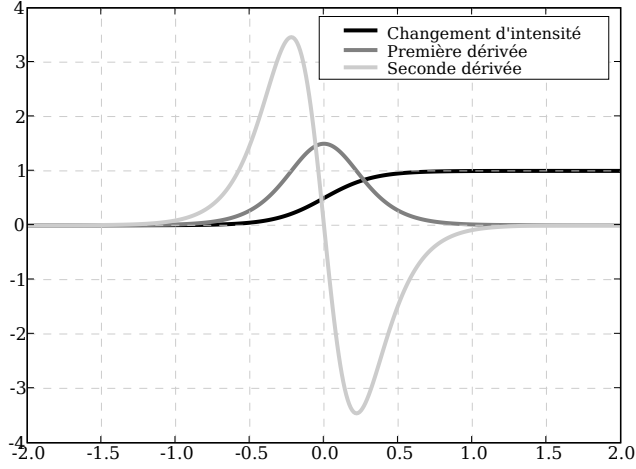


FIG. 2.17 – Représentation du profil d'un changement d'intensité (en noir). En gris foncé, la première dérivée. En gris clair, la dérivée seconde .

Une des méthodes les plus utilisées pour la détection des coins est le détecteur de Harris [Harris and Stephens, 1988]. C'est une méthode basée sur le travail de Moravec [Moravec, 1979]. Le détecteur de Harris utilise l'auto-corrélation locale d'une image. Cette auto-corrélation locale, mesure les changements locaux des niveaux de gris au travers d'une petite fenêtre qui est déplacée pour parcourir toute l'image. Soit $I(\mathbf{x})$ le pixel superposé par le centre de la fenêtre. La fonction d'auto-corrélation est calculée avec les niveaux de gris de l'ensemble des pixels de la fenêtre W . Le vecteur $\mathbf{u} = (\Delta x, \Delta y)^T$ permet de couvrir l'ensemble de la fenêtre W et représente des déplacements relatifs à $I(\mathbf{x})$ avec $\mathbf{x} = (x, y)^T$. Une valeur de corrélation est calculée pour tout $I(\mathbf{x})$ comme suit :

$$C(\mathbf{x}) = \sum_W [I(\mathbf{x} - \mathbf{u}) - I(\mathbf{x})]^2. \quad (2.66)$$

Les niveaux de gris de W sont en général pondérés par une Gaussienne centrée sur $I(\mathbf{x})$. En faisant un développement limité du premier ordre en séries de Taylor de l'expression sur \mathbf{u} nous obtenons : $\mathbf{u}^T \mathbf{M}_H \mathbf{u}$, avec \mathbf{M}_H une matrice des dérivées secondes de la région analysée.

$$\mathbf{M}_H = \begin{bmatrix} \sum_W \frac{\partial^2 I(\mathbf{x})}{\partial x^2} & \sum_W \frac{\partial I(\mathbf{x})}{\partial x} \frac{\partial I(\mathbf{x})}{\partial y} \\ \sum_W \frac{\partial I(\mathbf{x})}{\partial x} \frac{\partial I(\mathbf{x})}{\partial y} & \sum_W \frac{\partial^2 I(\mathbf{x})}{\partial y^2} \end{bmatrix}. \quad (2.67)$$

Ce détecteur utilise les valeurs propres λ_0 et λ_1 de la matrice \mathbf{M}_H pour vérifier si W recouvre un coin. Si $\lambda_0 \approx 0$ et $\lambda_1 \approx 0$, il n'y a pas d'intersection sur la coordonnée \mathbf{x} . Si $\lambda_0 \approx 0$ et $\lambda_1 > 0$, cela veut dire qu'un contour a été trouvé. Si $\lambda_0 > 0$ et $\lambda_1 > 0$ et les deux valeurs sont différentes, une intersection a été trouvée. Une approche plus simple, qui évite de calculer les valeurs propres, pour indiquer quels points seraient potentiellement des coins a été proposée aussi par [Harris and Stephens, 1988]. Les auteurs proposent d'utiliser M_H comme suit :

$$R = \det(\mathbf{M}_H) - \alpha \text{trace}(\mathbf{M}_H)^2 = \lambda_0 \lambda_1 - \alpha (\lambda_0 + \lambda_1)^2 \quad (2.68)$$

où α est une valeur déterminée de façon empirique. Si $R \gg 0$, il existe un coin ; si $R \ll 0$, il existe un bord ; si $|R| < 0$ c'est une région où il n'y a pas de changements d'intensité. [Harris and Stephens, 1988] proposent $\alpha = 0.06$ pour une application du suivi d'objets. [Shi and Tomasi,

1994] proposent $\min(\lambda_0, \lambda_1)$ pour améliorer la détection des coins dans le cadre du suivi du mouvement à partir d'une séquence d'images.

Les points détectés par l'algorithme de Harris ne sont pas invariants aux changements d'échelle. [Linderberg1998] modifie le détecteur de Harris pour le rendre invariant aux changements d'échelle. D'autres algorithmes développés pour détecter des coins sont : SUSAN (*Smallest Univalve Segment Assimilating Nucleus*) [Smith and Brady, 1997] développé dans le cadre d'une application au suivi d'objets, l'algorithme FAST (*Features from Accelerated Segment Test*) [Rosten and Drummond, 2005] développé pour la modélisation 3D en temps réel et le détecteur de [Wang and Brady, 1995] utilisé pour l'extraction de mouvement à partir d'une séquence d'images. Un résumé, ainsi qu'une comparaison expérimentale des détecteurs d'intersections ont été effectués par [Schmid *et al.*, 2000].

Régions. Les régions sont des primitives images très utilisées dans le recalage d'images médicales. Elles sont obtenues par des algorithmes de segmentation dont le but est de détecter des structures anatomiques particulières. Les *snakes* (contours actifs) sont un exemple de modèle déformable très utilisé. Un snake est un contour qui est attiré par la primitive à détecter. Par exemple, s'il faut délimiter le pourtour d'une région, le gradient de l'image peut être vu comme une énergie qui attire le contour (énergie externe au contour). Une énergie interne au contour (liée à l'allongement et à la déformation du contour) est définie selon l'application pour introduire des connaissances *a priori* de la forme sur la région. La minimisation de la somme des énergies (internes et externes) fait converger le snake vers les bords de la région. Par exemple, [Davatzikos, 1996] utilise un contour déformable pour segmenter des images IRM du cerveau qui seront ensuite recalées entre elles pour comparer les différences entre un individu sain et un malade. Une autre application des modèles déformables est celle de [Thompson and Toga, 1996], qui utilisent ces modèles pour segmenter des surfaces dans des images IRM qui sont recalées entre elles. Ce recalage est utilisé dans le cadre de l'étude de la maladie d'Alzheimer.

Une autre approche de segmentation en régions totalement différente de la précédente repose sur les *descripteurs invariants*. Les points segmentés par les détecteurs de contours et les détecteurs de coins ne représentent qu'une partie très petite de l'information par rapport à la taille de l'image. Cette information peut être complétée en prenant en compte la région autour des points détectés par ces détecteurs. Les approches basées sur les descripteurs invariants utilisent en général un détecteur de coins pour trouver des points de départ pour ensuite effectuer des statistiques sur les niveaux de gris des régions englobant les coins détectés. Ces statistiques conduisent à des descripteurs de régions qui sont invariants aux changements d'échelle et aux transformations affines. Cette invariance peut s'avérer très utile lors du recalage d'images caractérisées par des différences en termes d'échelle et même de perspective. Une façon simple d'obtenir une invariance aux changements d'échelle est de chercher des maxima des niveaux de gris (comparer) à des échelles différentes. Ces différences d'échelles peuvent être représentées dans une pyramide de Laplace où chaque niveau est une image correspondant à la différence de deux images filtrées par des gaussiennes (*DoG*) [Lowe, 2004]. Un exemple de descripteur invariant aux transformations affines (et même à quelques transformations en perspective) est le détecteur SIFT (*Scale Invariant Feature Transform*) proposé par [Lowe, 2004]. D'autres algorithmes développés récemment et qui conduisent à une invariance aux transformations affines sont GLOH (*Gradient Location and Orientation Histogram*) [Mikolajczyk and Schmid, 2004] et SURF (*Speeded Up Robust Features*) [Bay *et al.*, 2006].

2.2.5.2 Recalage des primitives images

Comme pour les méthodes basées sur les niveaux d'intensité des images, il faut définir une mesure de similarité qui représente un degré de ressemblance de deux primitives images. Mesurer une distance entre primitives images permet par exemple d'estimer le degré de superposition (donc de ressemblance) de ces primitives. La norme euclidienne est classiquement utilisée pour mesurer la distance entre deux points. Très peu gourmande en ressources, elle est très utilisée pour des applications à la vision par ordinateur [Brown, 1992].

Pour le recalage de contours et des surfaces, il existe d'autres mesures de distance. Une approche classique repose sur l'utilisation de cartes de distances. Dans une carte de distances, les niveaux de gris des pixels superposés par le contour valent zéro. Les niveaux de gris des autres pixels sont directement liés à la plus courte distance du pixel au point de contour le plus proche. Ces distances peuvent être calculées rapidement via du filtrage récursif. Ces cartes permettent de calculer de façon très efficace la distance entre deux contours, la carte étant calculée avec un contour alors que le deuxième contour est superposé à la carte. La moyenne des niveaux de gris superposés par ce contour donne une distance moyenne entre les deux contours. [Fitzgibbon, 2001] propose l'utilisation d'une carte de distances codant des distances euclidiennes pour recalculer des nuages de points 2D et 3D.

Une autre mesure très utilisée pour mesurer la distance entre contours ou surfaces est la distance de Hausdorff. Pour deux ensembles de points $A = \{a_1, \dots, a_n\}$ et $B = \{b_1, \dots, b_n\}$ représentant 2 surfaces ou contours, la distance de Hausdorff est définie par

$$D_H(A, B) = \max(d_H(A, B), d_H(B, A)), \quad (2.69)$$

avec

$$d_H(A, B) = \max_{a \in A} \min_{b \in B} \|a - b\|, \quad (2.70)$$

où $d_H(A, B)$ est une distance de Hausdorff *dirigée*. [Chmielewski *et al.*, 2002] utilisent cette distance pour recalculer des images portales et de simulation en imagerie à rayons X. Récemment [Posada *et al.*, 2007] proposent une méthode basée sur la minimisation de la distance de Hausdorff pour le recalage d'images 3D (point à la surface du visage) et de données CT, dans le cadre du positionnement du patient lors de sessions de radiothérapie intracrânienne (point sur la tête du patient).

Dans le cas de séquences vidéo, le mouvement local prédominant des primitives images est souvent la translation (déplacement entre deux images consécutives) [Shi and Tomasi, 1994]. Si les primitives choisies sont des coins, il est possible d'utiliser une mesure de similarité comme celles présentées dans la section 2.2.4.1. En effet, la *différence d'intensité* ou la *corrélation croisée normalisée* peuvent par exemple être utilisées de façon directe pour comparer les intensités des pixels situés dans des petites régions au voisinage des primitives images. Si les points à recalculer se trouvent dans des images non consécutives d'une séquence, leur apparence peut varier fortement. Dans ce cas, il pourra s'avérer nécessaire d'utiliser un modèle de transformation plus complexe, comme un modèle de transformation affine pour lequel les descripteurs invariants deviennent indispensables.

Comme pour le recalage utilisant tous les niveaux d'intensité des images, il faut, après avoir établi une mesure de similarité reliant les primitives images, choisir la transformation géométrique adaptée à l'application. Ces transformations seront optimisées pour réduire la distance entre les primitives images. À la fin du processus de recalage, la méthode d'optimisation va délivrer la transformation globale existant entre les images.

Si beaucoup de primitives sont détectées, la manière la plus simple de trouver tous les points correspondants entre deux images est de comparer toutes les primitives trouvées dans une image avec toutes les primitives de l'autre image. Cette méthode requiert des temps de calcul trop importants. Le temps de calcul peut être réduit en utilisant des approches *indexées* comme les *kd trees*, utilisés dans [Brown and Lowe, 2003]. Une fois l'ensemble des correspondances potentielles calculées, il faut trier celles-ci pour garder uniquement celles qui vont produire le recalage le plus précis. Dans cette optique, l'approche la plus utilisée (surtout dans le recalage d'images monomodale 2D-2D [Zitová and Flusser, 2003]) est appelée *RANSAC* (*RANdom SAmple Consensus*) [Fischler and Bolles, 1981].

Le problème le plus important dans l'implémentation de ces algorithmes est la détection précise des primitives images et la mise en correspondance des points communs dans les images à recalier.

2.2.6 Collage des images recalées

Lorsque les transformations entre les images consécutives d'une séquence ont été obtenues, il faut ramener (coller) les images dans un repère commun en utilisant ces transformations. Soit I_i une séquence constituée de N images, avec $i = 1 \dots N$, N le nombre total d'images. Pendant le recalage de la séquence, à chaque itération i , une transformation $T_{i,i-1}^r$, appelée ici *relative* est obtenue entre deux images successives I_i et I_{i-1} . Pour ramener les images dans un repère commun (l'image panoramique), il faut calculer les transformations qui vont indiquer la position de chaque image sur ce repère. Pour ce faire, à chaque itération les transformations relatives sont *cumulées* à partir d'une image choisie comme référence pour calculer une transformation *globale*. Si l'image de référence est la première image de la séquence (ici I_1) nous avons :

$$T_i^g = \prod_{i=1}^N T_{i,i-1}^r = T_{i,i-1}^r \times T_{i-1}^g, \quad (2.71)$$

où T_i^g est la matrice de transformation globale qui ramène l'image I_i dans le repère de I_1 . Pour la première image I_1 , la matrice de transformation globale précédente $T_{i-1}^g = T_0^g$ sera égale à la matrice unité. La taille de l'image sur laquelle les images recalées seront collées est calculée en transformant les coins d'une des images de la séquence (en supposant que toutes les images de la séquence ont la même taille) avec la dernière transformation globale. Ensuite, les images de la séquence pourront être placées sur cette image en utilisant les transformations globales calculées pour chaque image.

Lorsque les images ont été placées dans le repère global, il est probable que l'application d'une fonction de lissage soit nécessaire pour éliminer ou atténuer les bords visibles entre les images ramenées dans l'image panoramique. Il existe plusieurs méthodes pour lisser les bords. La façon la plus simple est d'utiliser la moyenne calculée entre un pixel de l'image à coller $I_m(\mathbf{x})$ et le pixel correspondant de l'image panoramique $I_p(\mathbf{x})$ pour toute valeur d'intensité différente de zéro. Si la moyenne arithmétique est utilisée, nous avons :

$$I_p(\mathbf{x}) = \frac{I_p(\mathbf{x}) + I_m(\mathbf{x})}{2}; \quad \forall I_p(\mathbf{x}) \neq 0. \quad (2.72)$$

Une méthode qui permet une meilleure atténuation des bords, est basée sur un moyennage pondéré des pixels. La pondération peut être réalisée par plusieurs fonctions, comme par exemple une fonction de forme pyramidale, où le poids des pixels diminue avec la distance à partir du

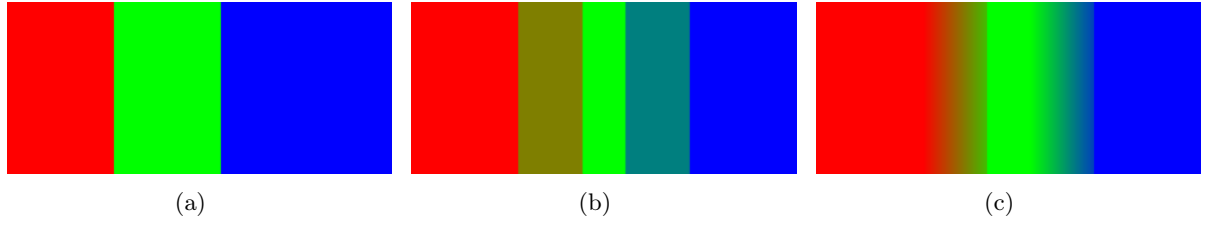


FIG. 2.18 – Fabrication d'une image panoramique en utilisant trois images représentées par trois rectangles de différentes couleurs. (a) Collage d'images en faisant une simple superposition des images de la séquence. (b) Moyennage non pondéré des pixels des images. (c) En appliquant un moyennage pondéré aux pixels des images, le résultat visuel est amélioré. Cette dernière méthode est à retenir pour la plupart d'applications.

centre de l'image :

$$w_p = \left[\frac{c_x - |(c_x - x)|}{c_x} \right] \left[\frac{c_y - |(c_y - y)|}{c_y} \right] \quad (2.73)$$

avec w_p le poids du pixel de coordonnées (x, y) dans l'image et (c_x, c_y) les coordonnées du centre de l'image.

La figure 2.18 montre les résultats obtenus pour trois techniques qui sont couramment utilisées pour former des images panoramiques.

Correction de l'erreur cumulée pendant le collage. Il est évident qu'en utilisant la technique de collage décrite ci-dessus, l'erreur de recalage entre images consécutives est cumulée à chaque nouvelle itération de l'accumulation (produit) des matrices relatives. Cette erreur cumulée sera plus évidente sur les dernières images collées, car leur position calculée sera plus éloignée de la position idéale que pour les premières images. Pour corriger ce problème, [Guillén González *et al.*, 1998], proposent une méthode dont l'idée principale réside dans une répartition plus ou moins homogène des erreurs sur toute la séquence par une méthode d'optimisation. Pour que cette méthode puisse être utilisée, il est absolument nécessaire que la séquence d'images fasse une boucle. [Miranda Luna, 2005] propose une méthode similaire qui considère les transformations de la première et de la dernière image d'une boucle formée dans la séquence recalée et collée. Il utilise la première image de la boucle (I_i) et la dernière (I_{i+m} , avec m , le nombre d'images à l'intérieur de la boucle). L'auteur part du principe que la transformation globale T_{i+m}^g pour la dernière image de la boucle et la transformation relative $T_{i,i+m}^r$ entre la première et la dernière image de la boucle, doivent idéalement être égales. [Miranda Luna, 2005] base son algorithme sur le fait que les erreurs réalisées sur les transformations relatives $T_{k,k+1}$ sont toutes très faibles. Pour corriger l'accumulation des erreurs et aussi améliorer le rendu visuel de l'image panoramique, l'auteur fait tendre les valeurs des paramètres de la matrice T_{i+m}^g vers celles de la matrice relative $T_{i,i+m}^r$ en modifiant légèrement les valeurs de l'ensemble des matrices $T_{k,k+1}^r$ avec $k \in [i, m-1]$. Pour implémenter cette idée, il faut minimiser l'expression suivante :

$$E_p = [T_{i,i+m}^r - T_{i+m}^g]^2 + \sum_{k=i}^{i+m-1} [T_{k,k+1}^{r,a} - T_{k,k+1}^r]^2 \quad (2.74)$$

avec $T_{k,k+1}^{r,a}$ les matrices de transformation relatives après l'ajustement des valeurs des paramètres (initialement $T_{k,k+1}^{r,a} = T_{k,k+1}^r \forall k$). $T_{i,i+m}^r$ est fixe alors que la valeur de T_{i+m}^g varie avec les $T_{k,k+1}^{r,a}$. E_p est ensuite minimisée par une technique de gradient descendant. Cette réduction de l'erreur

de la position des images, affecte légèrement l'erreur de recalage entre les images, mais cela est compensé par l'amélioration visuelle de l'image panoramique.

En utilisant cette technique, [Miranda Luna, 2005] réduit fortement (quelques dizaines de pixels pour quelques cas) l'erreur de position entre la première et la dernière images de la boucle.

2.3 Problèmes d'imagerie spécifiques aux cystoscopes

Nous présentons dans cette section, les défauts affectant spécifiquement les images endoscopiques. Ces défauts, qui ont un impact sur la qualité du recalage d'images, doivent être pris en compte.

Un cystoscope est un endoscope spécialement conçu pour l'étude des voies urinaires. C'est un instrument constitué d'une tige fine, qui en passant par l'urètre permet l'observation des parois internes de la vessie. Il existe deux types de cystoscope : les cystoscopes *rigides* et les cystoscopes *souples*.

Pour les cystoscopes rigides, l'optique nécessaire à la transmission des images de l'objectif vers l'oculaire est placée à l'intérieur même du dispositif (figure 2.19). La lumière illuminant la scène est conduite jusqu'à la sortie de l'endoscope par un faisceau de fibres optiques. La figure 2.19 montre un cystoscope rigide et les éléments optiques principaux qui le composent.

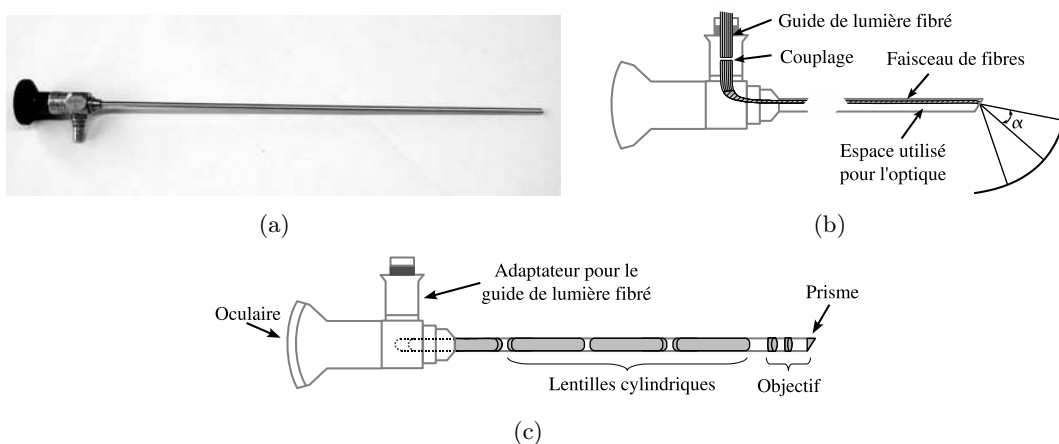


FIG. 2.19 – Représentation schématique des composants internes d'un cystoscope rigide. (a) Cystoscope *Karl Storz 27005BA*. (b) Canal illumination. α représente un demi de l'angle d'ouverture du faisceau d'illumination. (c) Optique incluse dans le cystoscope.

Dans les cystoscopes souples, la transmission de la lumière illuminant la scène et les images revenant à l'oculaire se fait à travers un faisceau de fibres optiques. L'optique qui rend possible l'observation des images est située à l'extérieur de l'instrument. La figure 2.20 montre les principales parties qui forment un cystoscope souple.

Pendant un examen clinique, une caméra est généralement placée à l'extrémité proximale (l'oculaire) de l'instrument. Cette caméra va envoyer les images vers les dispositifs d'affichage et d'enregistrement. Les cystoscopes rigides, du fait de leur construction fournissent des images de meilleure qualité en terme de netteté que les cystoscopes souples. Les cystoscopes souples sont plus confortables pour le patient et permettent au clinicien une observation plus aisée certains régions de la vessie (il es possible de faire de la retro vision). Le choix d'un type de cystoscope dépend principalement de l'expérience du clinicien qui fait l'examen.

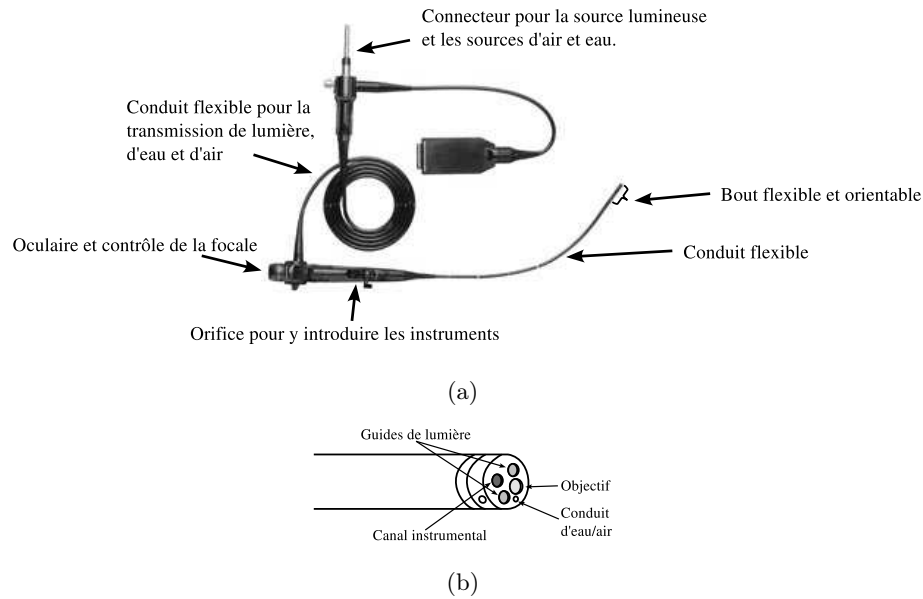


FIG. 2.20 – Cystoscope souple (fibroscope) et sa partie distale. (a) Cystoscope souple *Olympus* et les éléments principaux qui le composent. (b) Schéma qui montre avec détail le bout distal du cystoscope et les orifices qui permettent le passage de la lumière, d'instruments, d'eau ou d'air et la transmission des images vers l'oculaire.

2.3.1 Distorsion radiale

Le modèle sténopé de projection perspective présenté dans la section 2.1.1 ne modélise pas complètement la fonction de transfert d'un cystoscope. Même si la distorsion due à un endoscope est prise en compte par ce modèle, d'autres défauts dus à l'optique ne sont pas intégrés dans ce modèle de formation d'images (voir section 2.3.2 et 2.3.3).

L'optique utilisée par les cystoscopes, rigides ou souples, produit des distorsions non linéaires des images acquises. Nous avons vu dans la section 2.1.2, un modèle de distorsion radiale qui complète le modèle sténopé de formation d'images pour des dispositifs utilisant des lentilles. Ce modèle est directement transposable aux cystoscopes. Nous avons aussi présenté quelques approches couramment utilisées dans la littérature pour corriger cette distorsion dans des images acquises par des dispositifs d'usage général (appareils photo, vidéo-caméras). Néanmoins, il existe des travaux qui proposent des méthodes spécifiquement développées pour corriger la distorsion radiale des images acquises par un endoscope. Ces méthodes sont décrites ci-après.

[Palm *et al.*, 1998] proposent une méthode pour corriger la distorsion des images acquises par un laryngoscope (un endoscope spécialement conçu pour examiner le larynx). Sa méthode utilise un motif formé de lignes horizontales et verticales dont une image est acquise par le laryngoscope. Après extraction des points de croisement des lignes dans l'image avec distorsion, ils identifient tous les points qui appartiennent à une même ligne et minimisent la distance entre ces points et leur position idéale (ces points sont situés sur une ligne dans une image sans distorsion). Ils utilisent un modèle radial à 3 coefficients et le centre de distorsion pour modéliser la distorsion (voir équation 2.14). Un gradient descendant est utilisé pour ajuster les paramètres de distorsion qui minimisent la distance entre les points homologues des lignes distordues et non déformées. Les auteurs ajustent uniquement les paramètres de distorsion radiale sans tenir compte de la position de l'instrument durant l'acquisition du motif. C'est pour cette raison que les résultats

obtenus sont faussés en partie par des paramètres liés à la perspective.

[Asari *et al.*, 1999] proposent d'acquérir une image d'un motif composé de points équidistants avec un endoscope. Sans appliquer de pré-traitement, ils localisent les centres des points et minimisent la distance euclidienne entre ces centres et des lignes qui représentent leur position idéale. Ils considèrent un modèle polynomial d'ordre 4 et le centre de distorsion pour déformer les points. Cette méthode ne tient pas compte d'autres distorsions (affine ou perspective) et les résultats sont également affectés par des déformations perspective dues à la position de l'endoscope.

[Zhang *et al.*, 2000] acquièrent avec un endoscope une image d'un motif composé de points qui sont segmentés pour obtenir les coordonnées de leurs centres. Ces coordonnées sont comparées aux coordonnées extraites d'un pattern sans distorsion. La différence au carré de ces coordonnées est minimisée en ajustant un polynôme d'ordre 5 et le centre de distorsion. Les auteurs ajoutent une contrainte qui fait que les lignes formées par les points restent parallèles. De cette façon, ils obtiennent des images corrigées dans lesquelles les lignes orthogonales dans la scène le sont aussi dans les images corrigées.

Une autre méthode proposée par [Miranda-Luna *et al.*, 2004b] et [Miranda Luna, 2005] utilise un motif connu qui est acquis par un endoscope. Ce motif acquis est recalé avec un motif sans distorsion en utilisant l'information mutuelle. Cette méthode utilise un modèle polynomial d'ordre 4 plus le centre de distorsion pour modéliser la distorsion radiale. Mais, la méthode ajuste aussi un modèle de transformation perspective (8 paramètres) pendant le recalage des images. De cette façon, la correction effectuée produit des images où les lignes du motif sont orthogonales.

2.3.2 Hétérogénéités d'illumination et vignettage

Un autre défaut des images cystoscopiques est lié au fait que l'optique est composée de lentilles. Ce défaut connu sous le nom de *vignettage* obscurcit les bords des images alors que les centres sont bien éclairés. Les causes de ces hétérogénéités d'illumination peuvent être de nature *mécanique* ou de nature *optique*. Cependant, quelque soit la cause du vignettage, l'effet produit au niveau de l'illumination est le même.

Le vignettage est une diminution de la lumière (sur les bords de l'image) particulièrement présente avec les optiques constituées d'objectifs à grand angle d'ouverture et à celles utilisées dans des dispositifs compacts. À cause de leur faible distance focale, ces optiques ont une tendance plus importante au vignettage optique que les objectifs dont la distance focale est plus grande. (voir figure 2.21).

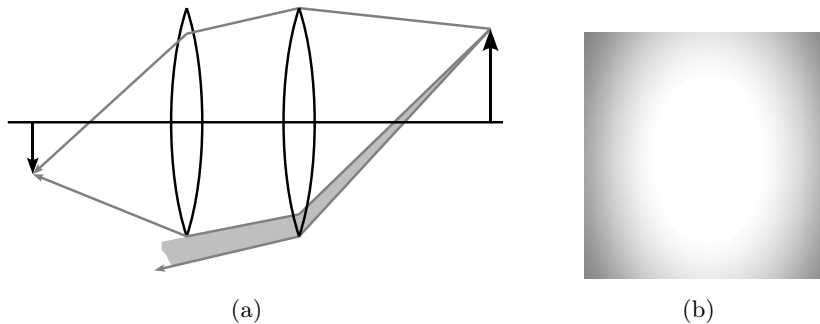


FIG. 2.21 – (a) Représentation de l'effet de vignettage dans un système optique à deux lentilles. La partie ombrée du rayon lumineux n'atteint jamais la deuxième lentille. (b) Exemple de l'effet de vignettage sur une image monochrome.

De plus, dans les images endoscopiques, d'autres hétérogénéités d'illumination peuvent aussi être produites pour des raisons non liées à l'optique. Par exemple, des ombres produites par les parois de la vessie sont également à l'origine d'inhomogénéités dans les images. Dans beaucoup d'applications concernant le traitement d'images et spécialement dans le recalage d'images, ces hétérogénéités d'illumination peuvent faire échouer les algorithmes et doivent être compensées.

Les hétérogénéités d'illumination peuvent être modélisées comme des changements d'intensité comprenant beaucoup de basses fréquences en comparaison avec la composition spectrale du signal utile. La représentation la plus simple des hétérogénéités d'illumination est donnée par un modèle comprenant un terme multiplicatif et un autre terme qui est additif :

$$I_{di}(\mathbf{x}) = I_o(\mathbf{x}) I_m(\mathbf{x}) + I_a(\mathbf{x}), \quad (2.75)$$

où $I_o(\mathbf{x})$ est l'image sans hétérogénéités d'illumination, $I_m(\mathbf{x})$ est la composante d'illumination multiplicative, $I_a(\mathbf{x})$ la composante additive et $I_{di}(\mathbf{x})$ est l'image affectée par les hétérogénéités d'illumination. En pratique, il est difficile de connaître l'apport de chaque composante et en général les solutions proposées cherchent à construire des approximations de la composante d'illumination dans son ensemble (par exemple en considérant que le « fond » de l'image comprend essentiellement les composantes d'illumination).

Une des solutions qui ont été proposées pour corriger ces hétérogénéités est la modélisation des hétérogénéités d'illumination par des fonctions quadratiques ou polynomiales [Vovk *et al.*, 2007]. Le modèle calculé pour une image de référence représente le « fond » de l'image qui sera soustrait à chacune des images. Cette approche est satisfaisante quand les différences d'illumination ne varient pas au sein d'une séquence d'images, ce qui n'est pas le cas des séquences endoscopiques.

Une autre solution, qui compense les différences d'illumination de chaque image, est de filtrer les très basses fréquences par un filtre gaussien dont la fréquence de coupure est suffisamment basse pour laisser passer les moyennes et hautes fréquences. Pour obtenir une image ne comprenant que les plus basses fréquences, il est nécessaire d'utiliser une valeur d'écart type élevée. Par conséquent, la taille du masque de filtrage sera grande (figure 2.22).

L'image obtenue va représenter le fond de l'image qui sera soustrait de l'image originale. La figure 2.22 donne un exemple d'image filtrée par un filtre gaussien avec différentes valeurs de σ (écart type). Le profil d'une image corrigée en illumination est donné par la figure 2.23. L'estimation des hétérogénéités d'illumination par filtrage passe-bas est utilisée dans [Leong *et al.*, 2003] [Miranda-Luna *et al.*, 2004b] [Pirrone and Gambino, 2006]. C'est aussi la méthode que nous allons utiliser dans cette thèse, car cette méthode est un compromis acceptable d'un point de vue efficacité et temps de calcul (si des techniques de filtrage adéquates sont utilisées).

2.3.3 Motif de fibres optiques (fibroscope)

Un problème propre aux cystoscopes souples (fibrosopes) est la présence dans les images d'un motif dû au faisceau de fibres optiques qui transmet les données à la matrice CCD. Ce motif correspond à un pattern périodique issu des bords des fibres optiques et qui se superpose au signal utile. Ce motif peut être vu comme un bruit qui peut perturber le recalage des images. La figure 2.24 montre des images prises avec les deux types de cystoscope. Un pattern de fibres optiques est visible sur l'image acquise avec le cystoscope souple.

Dans la littérature, il existe peu de travaux traitant de l'atténuation du pattern de fibres. Récemment, [Winter *et al.*, 2006] ont proposé une méthode qui réalise une analyse automatique du spectre de Fourier d'une image acquise par un fibroscope. Cette image est utilisée pour calculer la fréquence de coupure d'un filtre passe-bas. En effet, l'essentiel des composantes des signaux liés au pattern ont une fréquence plus élevée que celle des composantes des signaux utiles.

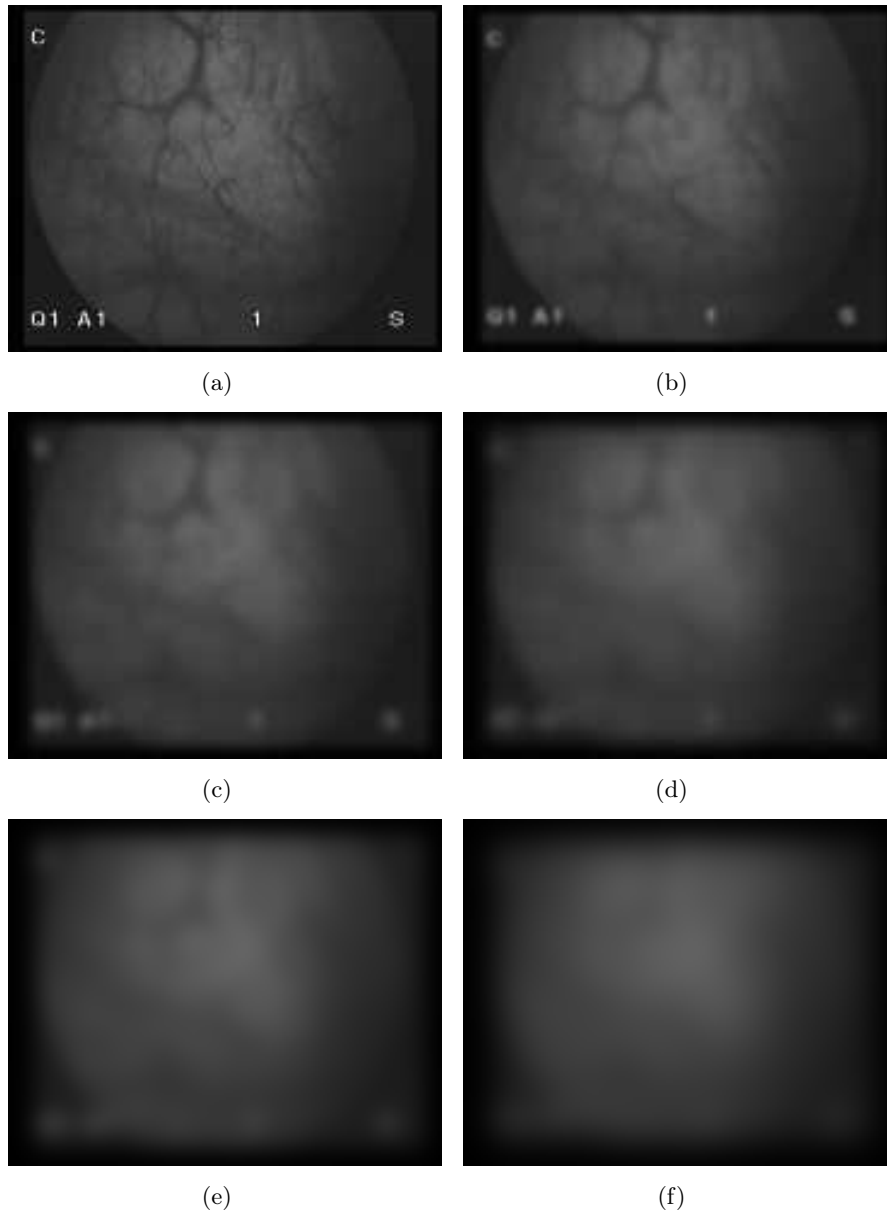


FIG. 2.22 – Estimation de l'hétérogénéité d'illumination d'une image calculée avec un filtre gaussien. (a) Image originale. (b) à (f). Images filtrées avec des filtres gaussiens dont les écarts-types valent respectivement $\sigma = 5, 10, 15, 20$ et 30 . Il est possible d'observer comment les détails de l'image disparaissent pour ne garder au final qu'une estimation de l'hétérogénéité de l'illumination.

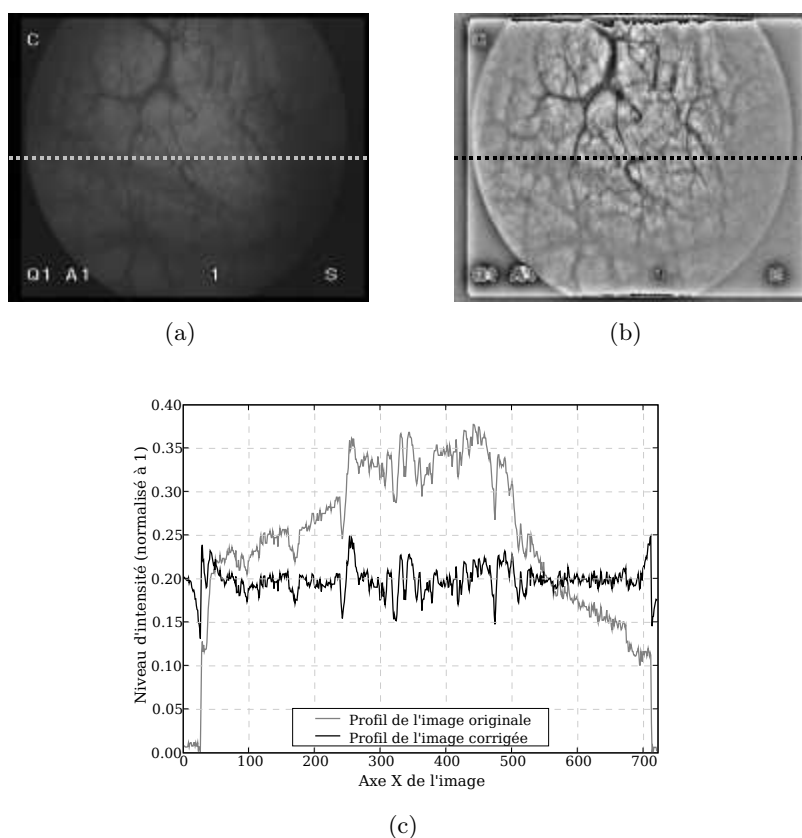


FIG. 2.23 – (a) Image originale extraite d’une séquence cystoscopique pour laquelle les hétérogénéités d’illumination sont évidentes. (b) Image où les hétérogénéités d’illumination ont été corrigées par la soustraction de la composante d’illumination obtenue par un filtrage passe-bas gaussien ($\sigma = 20$). Les niveaux d’intensité de cette image ont été normalisés ($\in [0; 1]$). (c) Profil des niveaux de gris des images (a) (en trait gris) et (b) (en trait noir).

Pour ce faire, les auteurs détectent dans l’image de l’amplitude du spectre de Fourier les pics correspondant au motif de fibres optiques. Les positions des pics sont utilisées pour construire des masques circulaires et en forme d’étoile pour filtrer le pattern de fibres. Le principal atout de leur méthode est sa capacité à construire des masques adaptatifs de façon automatique.

Une méthode similaire pour atténuer le motif de fibres optiques a été proposée par [Miranda-Luna *et al.*, 2004b]. Le processus d’analyse du bruit commence par l’acquisition d’une image par un fibroscope. Le motif de fibres optiques visible sur l’image est isolé et son spectre de Fourier est analysé pour y localiser les composantes fréquentielles appartenant au motif. Le masque de filtrage utilisé par cette méthode est un masque gaussien dont l’écart type σ est choisi de telle façon que les composantes des fréquences du motif soient à l’extérieur du masque dans le domaine de Fourier. L’application du filtrage se fait dans le domaine spatial. Le résultat est une image sans le motif de fibres mais qui sera un peu floue. Cette méthode ne détecte pas de façon automatique les composantes de fréquence, néanmoins le résultat de la détection est similaire en terme de qualité visuelle à celui obtenu par [Winter *et al.*, 2006]. Il est à noter que la grille de fibres optiques est constante pour un fibroscope donné. Un ajustement manuel de la fréquence de coupure du filtre passe-bas peut être réalisé une fois pour toutes pour un instrument. Une méthode non automatisée n’est donc pas pénalisante.

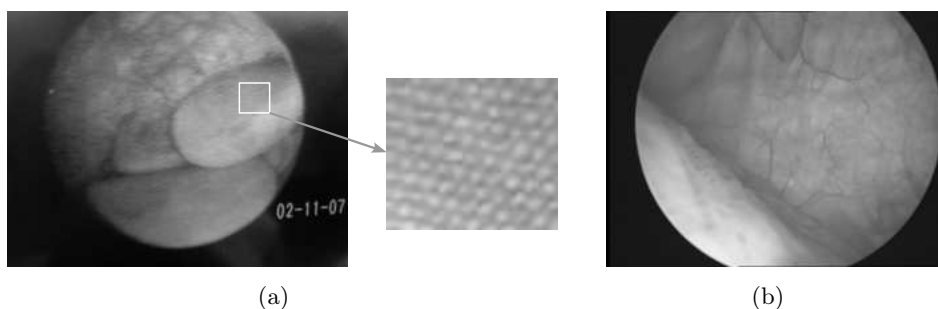


FIG. 2.24 – Images de l'intérieur de la vessie acquises avec les deux types de cystoscopes. (a) Image acquise avec un cystoscope souple. Un pattern formé par le faisceau de fibres optiques est visible surtout sur la partie droite de l'image. (b) Image acquise avec un cystoscope rigide (optique non fibrée).

2.4 Discussion et conclusion

Une conclusion importante de ce chapitre est le constat que le mosaïquage d'images est peu exploité dans le cas d'organes creux, ou plus particulièrement dans la vessie. Même si le recalage d'images est très répandu dans de nombreux domaines médicaux, le mosaïquage d'images endoscopiques est une application originale.

Il est clair que les défauts affectant les images cystoscopiques et qui peuvent être problématiques pour une méthode de mosaïquage d'images doivent être réduits au maximum. Pour traiter le problème des hétérogénéités d'illumination, nous devons obtenir une approximation du fond de l'image qui sera soustraite de celle-ci. Dans la littérature nous n'avons pas trouvé d'alternatives à cette méthode de soustraction du fond de l'image. Nous allons donc nous inspirer de cette approche. Pour l'atténuation du pattern de fibres, l'approche basée sur l'analyse du motif pour ensuite filtrer celui-ci semble la plus adéquate. Les résultats de la méthode de [Winter *et al.*, 2006] et de la méthode présentée dans [Miranda-Luna *et al.*, 2004b] montrent que le motif de fibres est atténué considérablement avec cette méthode. Pour la correction radiale des images cystoscopiques, nous allons appliquer la méthode développée dans [Miranda-Luna *et al.*, 2004a], car les résultats obtenus avec cette méthode montrent que cette approche est bien adaptée aux images endoscopiques et l'étalonnage de l'endoscope ne nécessite pas de mire particulière ni un dispositif précis de positionnement de l'instrument.

Pour le choix de la méthode de recalage il existe des contraintes liées à l'application et que nous devons prendre en compte.

- Les séquences sont acquises avec une vitesse de 25 images/seconde.
- Les images à recaler sont toutes acquises avec un même capteur (images monomodales).
- Il existe une variabilité non négligeable au niveau des images pour un même patient lorsque les données sont acquises avec des matériels différents.
- Il existe une grande variabilité au niveau des données d'un patient à l'autre.
- La qualité, les conditions, l'illumination pendant l'acquisition ne sont pas constante.
- Les images peuvent être affectées par un pattern de fibres optiques superposé si elles sont acquises avec un fibroscope.

À partir de ces observations nous devons définir la méthode de recalage que nous allons utiliser. L'expérience de notre laboratoire nous a appris que les méthodes qui utilisent directement les niveaux d'intensité des images sont bien adaptées aux séquences endoscopiques [Miranda Luna, 2005]. Par ailleurs, la vitesse d'acquisition de 25 images/s associée au fait que le mouvement

de l'endoscope est de l'ordre de quelques mm/s conduit à un recouvrement important de deux images consécutives. Il faudra effectuer des tests sur des séquences cystoscopiques réelles pour caractériser leur taux de recouvrement.

Nous allons toutefois aussi tester une méthode basée sur des primitives image, puisque les images ne sont pas dénuées entièrement de structures (vaisseaux sanguins). Cela pourrait permettre d'augmenter la vitesse ou la robustesse du recalage.

La méthode de recalage doit fonctionner quelque soit les mouvements de l'endoscope durant un examen clinique standard. La surface de la paroi interne de la vessie pouvant être considérée comme rigide entre deux images consécutives de la séquence, un modèle de transformation perspective (8 degrés de liberté) doit être utilisé pour le recalage. Si une méthode basée sur les niveaux d'intensité est choisi, il faut décider d'une mesure de similarité. Étant donné la nature monomodale des images et le modèle de transformation à considérer, une mesure simple, comme par exemple la somme des différences des niveaux de gris des pixels au carré, peut être suffisante.

Le temps de calcul est une contrainte prioritaire fixée par le clinicien. Précédemment, [Miranda Luna, 2005] a proposé une méthode de recalage basée sur l'information mutuelle. Cette méthode a donné de bons résultats en termes de robustesse et précision de mosaïquage dans le cas de l'endoscopie. Cependant, le temps de calcul du recalage de cette méthode (~ 1 minute par paire d'images) est important pour construire une image panoramique à partir d'une vidéo (25 images/seconde). Une mesure simple de la distance (une soustraction) réduirait considérablement le temps de calcul. Il faut cependant évaluer une telle mesure en termes de précision et de robustesse.

Pour tester la méthode de recalage basée sur des primitives image, nous allons utiliser le détecteur préféré dans le mosaïquage d'images 2D [Schmid *et al.*, 2000] : le détecteur de Harris. Les images des séquences cystoscopiques présentent des vaisseaux dont les changements d'intensité des bords par rapport au fond des images pourraient être détectés par le détecteur de Harris comme étant des coins. Une fois les points détectés, nous pourrions ensuite les apparier en utilisant une mesure de similarité comme la corrélation ou la somme des différences au carré et si nécessaire, nous pourrions utiliser une méthode d'appariement robuste (comme RANSAC).

Du côté de la méthode d'optimisation à utiliser, une grande partie des algorithmes de recalage en 2D, utilisent des méthodes de deuxième ordre [Zitová and Flusser, 2003]. La méthode de pré-traitement appliquée aux images pourra aider à obtenir un espace des paramètres convexe qui faciliterait la tâche de l'algorithme d'optimisation. Même s'ils ne doivent pas être pris comme une généralisation, les résultats de l'exemple montré dans la sous-section 2.2.4.2, montrent que les algorithmes de deuxième ordre sont rapides et précis. Pour cette raison, ces méthodes seront utilisées dans cette thèse.

Finalement, pour pouvoir tester une technique de correction de l'erreur de collage des images, il faudra que le parcours effectué par le cystoscope fasse une ou plusieurs boucles. De cette façon nous pourrions détecter les images aux croisements, à partir desquelles l'accumulation de l'erreur de recalage pourra être répartie sur toutes les images.

3

Solution proposée

Ce chapitre propose des solutions aux problèmes médicaux et scientifiques posés dans la section 1.6. Dans un premier temps, nous décrivons la méthode de construction d'images panoramiques des parois internes de la vessie développée au laboratoire. L'étude bibliographique menée tout au long du chapitre 2 montre que cette méthode est actuellement la seule permettant de construire des images panoramiques de sections de la vessie. Dans un deuxième temps, pour augmenter la robustesse du recalage, nous présentons une méthode de pré-traitement appliquée aux images cystoscopiques. Puis, nous présentons des tests effectués sur des images extraites d'une séquence vidéo de cystoscopie. Un premier test évalue l'adéquation d'un détecteur de primitives aux images cystoscopiques d'une séquence réelle. Un deuxième test permet d'évaluer le degré de recouvrement entre images consécutives de la séquence en utilisant l'algorithme de recalage disponible au laboratoire. Les résultats de ces tests nous aideront à concevoir la méthode de recalage la plus adaptée à notre application. Ensuite, nous présentons les différentes approches testées pour recaler deux images de façon « rapide » pour déterminer celle qui est la mieux adaptée à notre problème. Finalement, un résumé de l'approche proposée est présenté.

3.1 Méthode de mosaïquage disponible au laboratoire

Nos travaux de thèse s'inscrivent dans la continuité d'un projet portant sur le mosaïquage d'images endoscopiques initialement engagé avec la thèse de [Miranda Luna, 2005]. Ce dernier a proposé une méthode de correction radiale et de mosaïquage d'images endoscopiques utilisant un algorithme de recalage basé sur l'information mutuelle (mesure de similarité statistique qui utilise les niveaux de gris des images) et le gradient descendant stochastique pour maximiser cette information. Dans son travail, l'auteur utilise un modèle de transformation 2D/2D perspective (cf. chapitre 2) pour recaler les images.

Cette méthode est basée sur l'approche appelée EMMA (*EMphirical entropy Manipulation and Analysis*) proposée dans [Viola and Wells III, 1997]. L'information mutuelle $IM(I'_R(\mathbf{x}), I'_D(\mathbf{x}))$ est définie dans l'équation 3.1 en fonction des entropies empiriques $H(I'_R(\mathbf{x}))$ et $H(I'_D(\mathbf{x}))$ et de l'entropie empirique jointe $H(I'_R(\mathbf{x}), I'_D(\mathbf{x}))$ où $I'_R(\mathbf{x})$ et $I'_D(\mathbf{x})$ sont respectivement les parties communes de l'image de référence $I_R(\mathbf{x})$ et de l'image à transformer $I_D(\mathbf{x})$. L'image à transformer I_D est transformée par la transformation T pour la superposer à l'image de référence I_R , c'est-à-dire $I_D(T(\mathbf{x}))$. Les entropies des images sont calculées à partir des fonctions de densité

de probabilité des niveaux de gris des pixels (voir équations 3.2).

$$IM(I'_R(\mathbf{x}), I'_D(\mathbf{x})) = H(I'_R(\mathbf{x})) + H(I'_D(\mathbf{x})) - H(I'_R(\mathbf{x}), I'_D(\mathbf{x})) \quad (3.1)$$

avec

$$\begin{aligned} H(I'_R(\mathbf{x})) &= - \sum_{z=z_{min}}^{z_{max}} p_R(z) \ln(p_R(z)) \\ H(I'_D(\mathbf{x})) &= - \sum_{z'=z_{min}}^{z_{max}} p_D(z') \ln(p_D(z')) \\ H(I'_R(\mathbf{x}), I'_D(\mathbf{x})) &= - \sum_{z=z_{min}}^{z_{max}} \sum_{z'=z_{min}}^{z_{max}} p_{RD}(z, z') \ln(p_{RD}(z, z')) \end{aligned} \quad (3.2)$$

où z et z' désignent les niveaux de gris des pixels des images I'_R et I'_D . $p_R(z)$, $p_D(z')$ et $p_{RD}(z, z')$ sont respectivement les fonctions de densité de probabilité des niveaux de gris de I'_R , de I'_D et de la fonction de densité de probabilité jointe. z_{min} et z_{max} sont la plus petite et la plus grande valeur des niveaux de gris. Pour une image codée sur un octet non signé $z_{min} = 0$ et $z_{max} = 255$.

L'information mutuelle est maximale lorsque les images sont recalées. Un algorithme d'optimisation basé sur un gradient descendant stochastique est utilisé pour trouver la transformation \tilde{T} qui maximise l'information mutuelle. Pour optimiser l'information mutuelle, les paramètres de la transformation projective T sont actualisés à chaque itération :

$$T_{k+1} = T_k + \lambda \frac{\partial IM(I'_R, I'_D)}{\partial T_k}, \quad (3.3)$$

avec k l'indice d'itération et λ un pas de convergence qui diminue au fur et à mesure que l'algorithme s'approche de la solution. La valeur initiale de T est la matrice identité.

Les fonctions de densité de probabilité des images sont calculées à partir d'échantillons des niveaux de gris en utilisant une fenêtre de Parzen. Cette méthode permet d'approximer précisément les fonctions de densité de probabilité réelles par des fonctions analytiques. Cette densité de probabilité est utilisée pour calculer l'entropie empirique des images ainsi que leur dérivée par rapport aux paramètres de transformation. En effet, la fenêtre de Parzen est une méthode d'estimation non paramétrique qui utilise une somme de gaussiennes pour modéliser des fonctions exponentielles non uni-modales. Les équations analytiques et différentiables ainsi obtenues conduisent à une optimisation plus précise et robuste que celle effectuée avec des dérivées partielles purement numériques. Pour calculer la fenêtre, des échantillons A et B sont sélectionnés aléatoirement dans la région commune des images z_A et z_B (A^{I_R} et B^{I_R} pour l'image I_R et A^{I_D} et B^{I_D} pour l'image I_D). Les valeurs de z_R sont utilisées pour définir les valeurs moyennes des gaussiennes et les valeurs de z_D sont utilisées pour optimiser leur écart type au cours du processus de recalage. [Miranda Luna, 2005] a montré expérimentalement que 260 gaussiennes suffisaient pour trouver de façon robuste la transformation qui maximise l'information mutuelle des images cystoscopiques qu'il a utilisées. Cette méthode est détaillée dans [Miranda-Luna *et al.*, 2007].

Les résultats obtenus par cette implémentation de l'information mutuelle ont montré que cette approche produit des images panoramiques visuellement cohérentes affectées par des erreurs faibles (distance moyenne de 2,4 pixels entre les positions idéales et calculées des pixels). Cependant, en utilisant cette approche, le temps de calcul nécessaire pour construire une image panoramique (environ 1 minute pour le recalage de deux images sur un *Pentium IV*, 2,4 GHz, algorithme programmé en langage C) restreint l'algorithme de mosaïquage à une utilisation « hors

ligne ». C'est la raison principale pour laquelle nous avons décidé de rechercher une autre approche de mosaïquage d'images qui permette de construire des images panoramiques des parois internes de la vessie de façon beaucoup plus rapide tout en restant suffisamment robuste.

Nos recherches bibliographiques montrent qu'il n'existe actuellement aucun algorithme de mosaïquage applicable à la vessie ou plus généralement pour les organes creux du corps humain. La seule approche testée, validée, connue et dont nous disposons pour recaler et coller des images de l'intérieur de la vessie est la méthode développée par [Miranda Luna, 2005]. C'est pour cette raison que nous avons utilisé l'approche basée sur l'information mutuelle pour effectuer nos tests préliminaires sur les images cystoscopiques (présentés dans la sous-section 3.3.1). Cette méthode est pour nous l'algorithme de référence servant à évaluer les résultats obtenus par l'approche proposée dans notre thèse.

La méthode de distorsion radiale utilisée pour corriger les images endoscopiques est celle décrite dans [Miranda-Luna *et al.*, 2004a] et dans ce chapitre (voir section 3.2.2).

3.2 Pré-traitements appliqués aux images cystoscopiques

Le travail proposé dans [Miranda Luna, 2005], a montré que les images des séquences cystoscopiques doivent être pré-traitées pour pouvoir être utilisées efficacement par un algorithme de recalage. D'ailleurs, lors du développement des méthodes, nous avons travaillé avec des images issues d'un fibroscope dans lesquelles apparaît un motif dû aux fibres optiques. Ce motif doit également être atténué avant le recalage des images.

Les solutions retenues pour traiter les problèmes liés spécifiquement aux endoscopes sont décrites dans les sections suivantes. La partie de l'analyse et construction de la solution de pré-traitement proposés dans le cadre de cette thèse a fait l'objet de deux communications [Miranda-Luna *et al.*, 2004b] [Miranda-Luna *et al.*, 2005] et fait partie de la méthode générale de mosaïquage développée dans [Blondel *et al.*, 2005].

3.2.1 Correction des inhomogénéités d'illumination et du motif de fibres optiques

Les inhomogénéités d'éclairage inhérentes aux endoscopes (cf. 2.3.2) et les variations au niveau de l'anatomie et des propriétés de réflexion de la lumière des parois de la vessie perturbent fortement les conditions d'illumination de la scène au cours de la séquence d'images. Cependant, comme l'anatomie et la réflexion sur les parois ne varient que très peu localement (d'une image à l'autre), seules les inhomogénéités d'éclairage doivent être corrigées. Comme justifié dans le paragraphe suivant, une représentation des inhomogénéités d'illumination peut être obtenue en filtrant les images avec un filtre passe-bas. Cette approximation des variations de l'illumination sera soustraite à l'image originale pour atténuer au maximum les effets de l'éclairage inhomogène. Nous avons choisi un filtre gaussien pour ses propriétés (séparabilité et récursivité) qui permettent de construire des filtres passe-bas optimisés. Un filtre passe-bas gaussien est calculé avec :

$$G_{\sigma}(\mathbf{x}) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}}, \quad (3.4)$$

avec σ l'écart type ou l'ouverture du filtre.

Des régions de 256×256 pixels centrés sur les images de taille 720×576 pixels sont extraites des séquences endoscopiques. En analysant les spectres de Fourier des images originales (figure 3.1(b)) et « filtrées passe-bas » nous avons pu vérifier que l'essentiel des composantes fréquentielles liées aux inhomogénéités d'éclairage sont situés dans l'intervalle $[0, 0,012N]$ où $N = 256$ est la

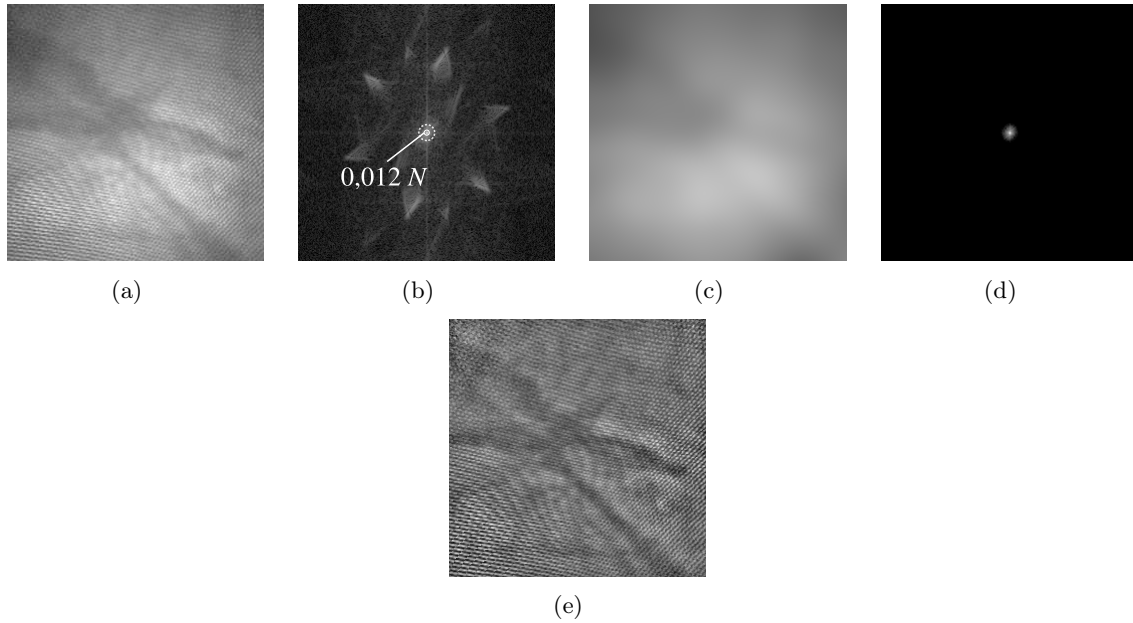


FIG. 3.1 – (a) Image originale. (b) Spectre de Fourier de l'image (a). Le petit cercle en trait continu au centre correspond à la valeur du σ_{ill} ($0,012N$) utilisé pour obtenir une estimation des inhomogénéités d'éclairage et le cercle en trait pointillé montre l'extension du masque gaussien ($3\sigma_{ill}$). (c) Image après filtrage : approximation des variations d'illumination. (d) Spectre de Fourier de l'image filtrée. (e) Soustraction pixel à pixel de l'image (a) à l'image (c). Cette image a été normalisée à 1. Nous pouvons observer que les variations d'illumination sont atténuées.

largeur de l'image, alors que les composantes du signal utile (texture de la vessie) et du motif de fibres sont nettement en dehors de cet intervalle. Soit σ_{ill} l'écart type d'une gaussienne dans le domaine de Fourier. Nous avons vérifié expérimentalement que si $\sigma_{ill} = 0,012N$, tous les signaux sont fortement atténués, sauf ceux liés aux inhomogénéités d'éclairage (figure 3.1(c)). Une image « sans » inhomogénéités est obtenue en soustrayant l'image « filtrées passe-bas » de l'image originale (figure 3.1(e)).

La taille du masque dans le domaine spatial est donc $\frac{1}{0,012} \times \frac{1}{0,012}$, soit environ 83×83 pixels. Cette taille de masque est importante et le filtrage dans le domaine spatial avec un noyau de convolution non séparable serait très lent. Nous avons proposé une méthode pour augmenter la vitesse du filtrage. Nous y reviendrons plus loin dans cette section.

Le pattern de hautes fréquences formé par les fibres optiques produit des motifs, généralement un quadrillage ou des séries de lignes ou des courbes (voir figure 3.2(a)). Pour éliminer le pattern de fibres sur les images, nous avons localisé les composantes fréquentielles correspondant à ces motifs dans le spectre de Fourier. Ainsi indiquées par les flèches dans la figure 3.2(b), l'essentiel des composantes spectrales des motifs sont des pics situés à un diamètre supérieur à $0,052N$. Pour atténuer le pattern de fibres, nous avons utilisé un filtre gaussien dont l'écart type σ_{patt} dans le domaine de Fourier est fixé à $\sigma_{patt} = 0,052N$. Le résultat du filtrage est montré dans les figures 3.2(c) et 3.2(d). Nous constatons que la plus grande partie du pattern de fibres a disparu.

Dans le domaine spatial, la taille du masque gaussien équivalent est $\frac{1}{0,052} \times \frac{1}{0,052}$ soit environ 19×19 pixels.

Finalement, pour atténuer à la fois les patterns de fibres et les inhomogénéités d'illumination,

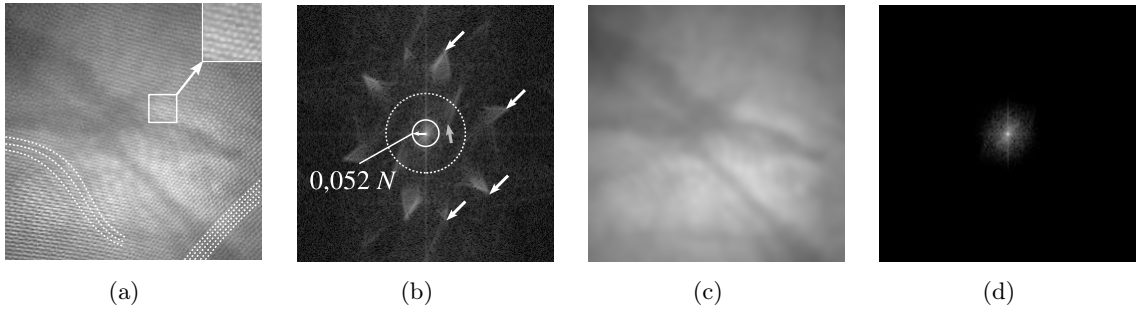


FIG. 3.2 – (a) Image originale qui montre une vue agrandie du pattern de fibres et les motifs en forme de lignes. (b) Spectre de Fourier de l’image (logarithme de l’amplitude). Les flèches blanches montrent les composantes de fréquences correspondant aux fibres. La flèche grise montre la région du spectre correspondant aux motifs. Le cercle en trait continu montre la valeur de σ_{patt} ($0,052N$) utilisé et le cercle en trait pointillé montre l’extension du masque gaussien ($3\sigma_{patt}$). (c) Résultat du filtrage appliqué à l’image. Il peut être observé que le pattern de fibres et les motifs formés ne sont plus présents. (d) Spectre de Fourier de l’image après filtrage.

la méthode à suivre consistera à soustraire de l’image sans le pattern de fibres (filtrée avec une gaussienne d’écart type σ_{patt}) l’image des inhomogénéités d’illumination (filtrée avec une gaussienne d’écart type σ_{ill}).

Du point de vue informatique, il est plus simple d’effectuer le filtrage dans le domaine spatial que dans le domaine fréquentiel, car dans le domaine fréquentiel nous devons fabriquer un masque pour apodiser les images et éviter « les rebondissements » produits par la non périodicité des images (phénomène de Gibbs). Cela ajoutera d’autres paramètres à régler pendant le filtrage des images. Il est évident que, compte tenu de la taille des masques, une convolution 2D dans le domaine spatial serait très coûteuse en ressources informatiques. Afin de réduire le temps de calcul, nous proposons d’utiliser une approche pyramidale [Burt and Adelson, 1983]. Pour construire une pyramide de Gauss, nous considérons une image I et un masque gaussien bi-dimensionnel g_σ appliquée de manière séparable sur les lignes et les colonnes dont la forme est donnée par l’équation 3.4 et 3.6. Le problème des bords est résolu par la technique de symétrisation. L’image du niveau k de la pyramide de Gauss G_k est construite en convoluant l’image du niveau $k - 1$ avec le masque gaussien (lignes puis colonnes). Le même masque est utilisé pour tous les niveaux, le niveau 0 étant celui de l’image originale :

$$G_0 = I \tag{3.5}$$

$$G_k = (G_{k-1} \otimes g_\sigma) \downarrow 2 = \text{REDUCE}[G_{k-1}]$$

avec \otimes l’opération de convolution et \downarrow la décimation (sous-échantillonnage) de l’image. Le facteur de décimation est en général égal à 2. Nous effectuons les convolutions en utilisant pour chaque niveau de la pyramide le même masque séparable de 5 coefficients qui est proposé dans [Burt and Adelson, 1983]. Les coefficients représentent une approximation binomiale précise de la fonction gaussienne (le triangle de Pascal) tout en étant discrets et finis [Burt and Adelson, 1983] :

$$g_\sigma = ([1 \ 4 \ 6 \ 4 \ 1] / 16)^T \times [1 \ 4 \ 6 \ 4 \ 1] / 16 \tag{3.6}$$

L’opération décrite dans l’équation 3.5 permet de « remonter » les niveaux dans la pyramide gaussienne. Cette opération (filtrage passe-bas plus un sous-échantillonnage de facteur 2) correspond à l’opérateur REDUCE dans [Burt and Adelson, 1983]. Pour « descendre » dans la pyramide

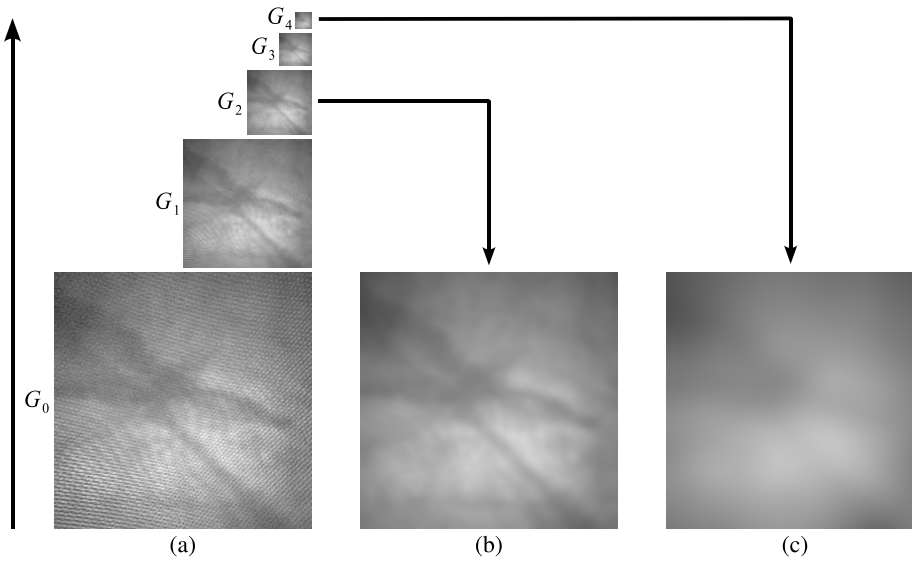


FIG. 3.3 – Filtrage passe-bas des images en utilisant une approche pyramidale. (a) Pyramide de Gauss à 4 niveaux construite à partir de l’image originale G_0 . (b) L’élimination du pattern de fibres est effectuée en remontant et en descendant 2 niveaux dans la pyramide. (c) L’approximation des inhomogénéités d’illumination est obtenue en remontant et en descendant 4 niveaux dans la pyramide.

de Gauss, nous appliquons l’opération inverse (EXPAND dans [Burt and Adelson, 1983]), mais cette fois en sur-échantillonnant les images (le sur-échantillonnage représenté par \uparrow) :

$$G_k = 4 (G_{k+1} \uparrow 2 \otimes g_\sigma) = \text{EXPAND}[G_{k+1}]. \quad (3.7)$$

Le même masque g_σ est aussi utilisé pour interpoler les niveaux de gris des pixels des images lors du changement d’échelle.

Le filtrage des hautes fréquences liées au pattern de fibres et l’approximation des inhomogénéités d’éclairage est effectué en « montant » dans la pyramide jusqu’à un niveau déterminé et ensuite en « redescendant » dans la pyramide jusqu’à la taille de l’image d’origine. Dans notre cas, pour construire l’approximation de la composante d’illumination des images, il suffit de remonter et descendre 4 niveaux dans la pyramide pour avoir un effet similaire à celui obtenu par un filtrage par convolution dans le domaine spatial avec un masque de 83×83 pixels. Pour obtenir un effet similaire au filtrage avec le masque de taille 19×19 pixels utilisé pour atténuer les hautes fréquences introduites par les fibres optiques, il faut remonter et descendre 2 niveaux dans la pyramide gaussienne. Le processus de filtrage des images est illustré dans la figure 3.3. Comme il est observé dans l’équation 3.5, à chaque niveau, un nouveau filtrage avec le même masque est appliqué aux images qui ont été déjà filtrées. Sous-échantillonner les images d’un facteur 4 et conserver le même filtre pour chaque niveau revient à doubler l’étalement du masque gaussien sur des images qui auraient la même taille à chaque niveau k . Convolver des images de plus en plus petites avec un filtre gaussien dont l’écart type est « petit » et constant est moins coûteux en temps de calcul que la convolution de l’image originale avec une gaussienne ayant un grand écart type.

Le résultat du filtrage avec une pyramide de Gauss est très similaire au résultat obtenu avec un filtrage passe-bas de l’image originale avec des masques de convolution gaussiens séparables

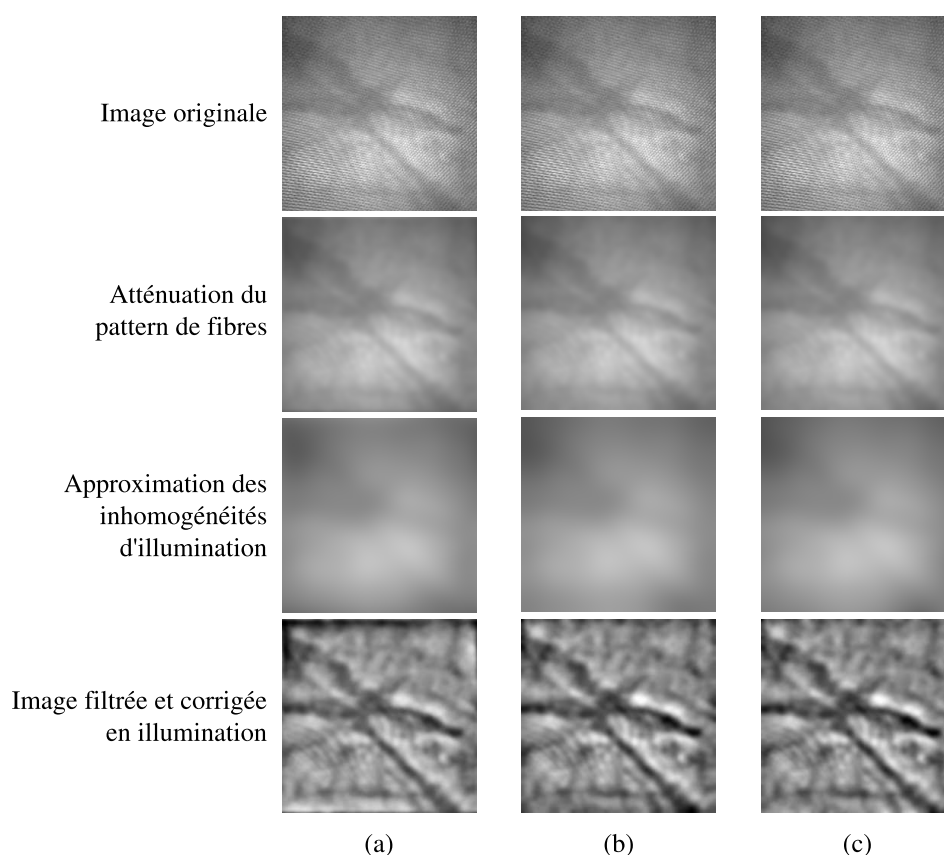


FIG. 3.4 – Filtrage du motif de fibres et des inhomogénéités d’illumination pour trois approches différentes. Il est possible d’observer que les résultats des trois approches sont visuellement similaires. (a) Filtrage dans le domaine de fréquentiel (DF). (b) Filtrage dans le domaine spatial (DS). (c) Filtrage avec une pyramide de Gauss (domaine spatial, DS_p).

ou au résultat du filtrage de cette image dans le domaine fréquentiel. La figure 3.4 montre le résultat du filtrage en appliquant les trois techniques de filtrage : filtrage par convolution dans le domaine spatial d’une image avec une gaussienne (DS), d’un filtrage dans le domaine fréquentiel (DF, transformée de Fourier pour l’obtention du spectre de l’image, produit de ce spectre par un noyau gaussien, puis transformée de Fourier inverse) et un filtrage dans le domaine spatial par une pyramide de Gauss (DS_p). La raison principale pour laquelle nous avons proposé cette méthode de filtrage alternative est la diminution du temps de calcul nécessaire pour le filtrage classique dans le domaine spatial. Les temps de calcul pour le filtrage d’une image dans le domaine de la fréquence, pour le filtrage de l’image dans le domaine spatial avec un masque séparable et pour le filtrage avec l’approche pyramidale ont été comparés. Ces algorithmes ont été programmés en *Scilab V 4.0* sur un ordinateur équipé d’un *Pentium 4*, 3,2 GHz et de 2 GB de mémoire RAM. Nous pouvons observer dans le tableau 3.1 que le temps de calcul pour le filtrage en utilisant l’approche pyramidale est beaucoup plus faible que celui d’une convolution de l’image avec une gaussienne. Nous pouvons aussi observer que le filtrage dans le domaine fréquentiel est un peu plus rapide tant que les dimensions de l’image sont des multiples d’une puissance de 2 (efficacité maximale de l’algorithme FFT). Néanmoins, le filtrage utilisant l’approche pyramidale est, au niveau de la programmation de l’algorithme, plus simple à mettre en œuvre.

Un filtrage dans le domaine fréquentiel avec une gaussienne permet de connaître facilement

TAB. 3.1 – Temps de calcul en secondes pour l’atténuation du pattern de fibres et la correction des inhomogénéités d’illumination sur une image cystoscopique en utilisant trois méthodes différentes : filtrage dans le domaine de la fréquence (DF), filtrage dans le domaine spatial (DS) et filtrage dans le domaine spatial en utilisant une approche pyramidale (DSp).

	DF	DS	DSp
Temps de calcul (s)	0,28	3,31	0,31

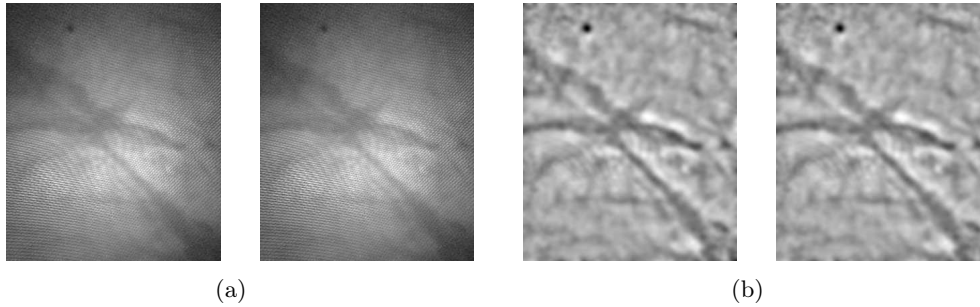


FIG. 3.5 – Deux images successives d’une séquence acquise avec un cystoscope fibré. (a) Images originales sans pré-traitement. (b) Images filtrées (atténuation du pattern de fibres et des inhomogénéités d’illumination).

l’atténuation (en relatif) de chaque composante fréquentielle. Connaître et maîtriser cette atténuation est nettement plus compliqué dans le cas de la pyramide de Gauss, puisque le niveau de filtrage est lié au nombre de niveaux utilisés dans la pyramide gaussienne. Néanmoins, nous avons observé, que pour notre application, il suffit de proposer 4 niveaux de filtrage pour traiter effectivement toutes les images cystoscopiques acquises avec un cystoscope souple ou rigide. D’ailleurs, l’approximation des variations d’illumination des images est toujours construite en remontant et descendant 4 niveaux dans la pyramide gaussienne. Seul le niveau de filtrage du pattern de fibres est ajustable. Les niveaux de filtrage que nous proposons pour le pattern sont : *aucun*, pas de filtrage ; *faible*, construit en remontant et en descendant 1 niveau dans la pyramide gaussienne ; *moyen*, considère 2 niveaux de la pyramide ; *fort*, construit avec le niveau 3 de la pyramide.

Nous avons observé que le pré-traitement des images cystoscopiques est essentiel dans notre application pour le bon fonctionnement des algorithmes de recalage. Pour montrer l’effet que le pattern de fibres et les inhomogénéités d’illumination produisent sur le recalage, nous avons calculé deux mesures de similarité (l’information mutuelle et la valeur absolue de la différence d’intensité des pixels homologues) pour des régions extraites de deux images acquises avec un fibroscope (figure 3.5). Nous connaissons *a priori* la translation entre les deux images : 8 pixels selon \vec{x} (t_x) et -47 pixels selon \vec{y} (t_y) (aucune autre transformation géométrique ne distingue les deux images). Si nous essayons de trouver le maximum (pour l’information mutuelle) ou le minimum (pour la valeur absolue des différences) sans pré-traitement, les translations trouvées sont fausses : $(t_x, t_y)=(-1, -50)$ pour l’information mutuelle et $(t_x, t_y)=(0, -46)$ pour la différence d’intensité. Mais si nous appliquons le pré-traitement sur les images avant de calculer les mesures de similarité nous obtenons le résultat correct : $(t_x, t_y)=(8, -47)$ pour les deux mesures de similarité. Les figures 3.6 et 3.7 montrent l’effet que ces artefacts provoquent respectivement, sur l’information mutuelle et sur la valeur absolue de la différence d’intensité entre les images.

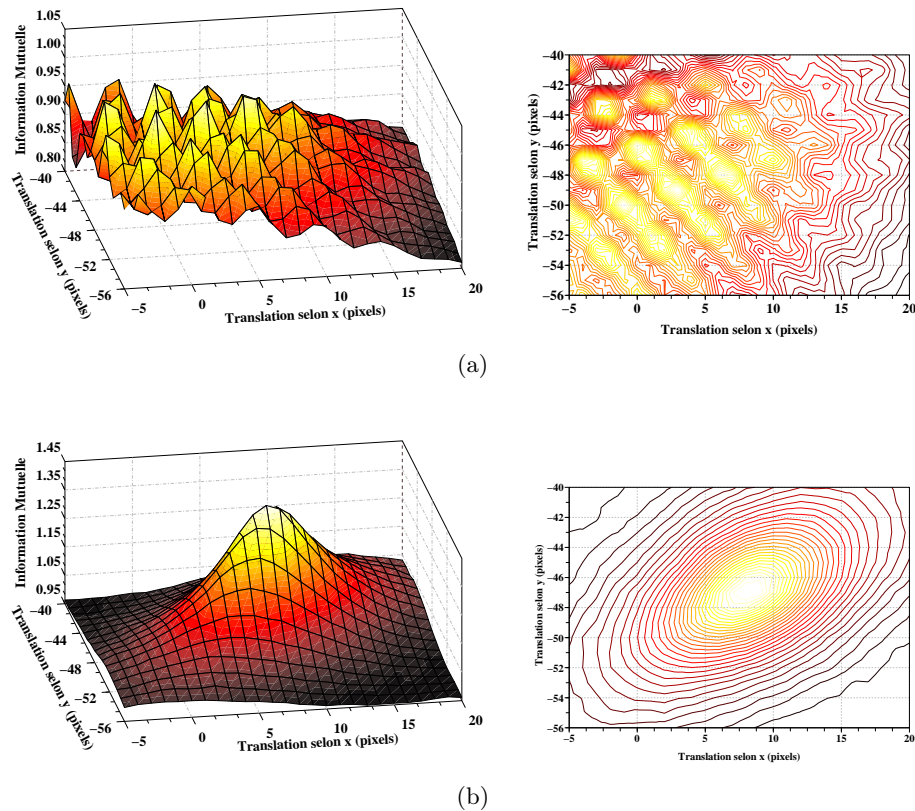


FIG. 3.6 – Effet produit par le pattern de fibres et par les inhomogénéités d’illumination sur l’information mutuelle. (a) Espace de paramètres calculé avec les images bruitées (sans pré-traitement). La surface de similarité est affectée par les artefacts : il existe de nombreux pics d’amplitude ou maxima locaux et le maximum trouvé ne correspond pas à la vraie valeur du déplacement entre les images. (b) Fonction calculée avec les images pré-traitées. Nous pouvons observer qu’il existe un seul maximum global qui correspond à la valeur correcte de translation (8, -47) et que la surface de similarité est convexe.

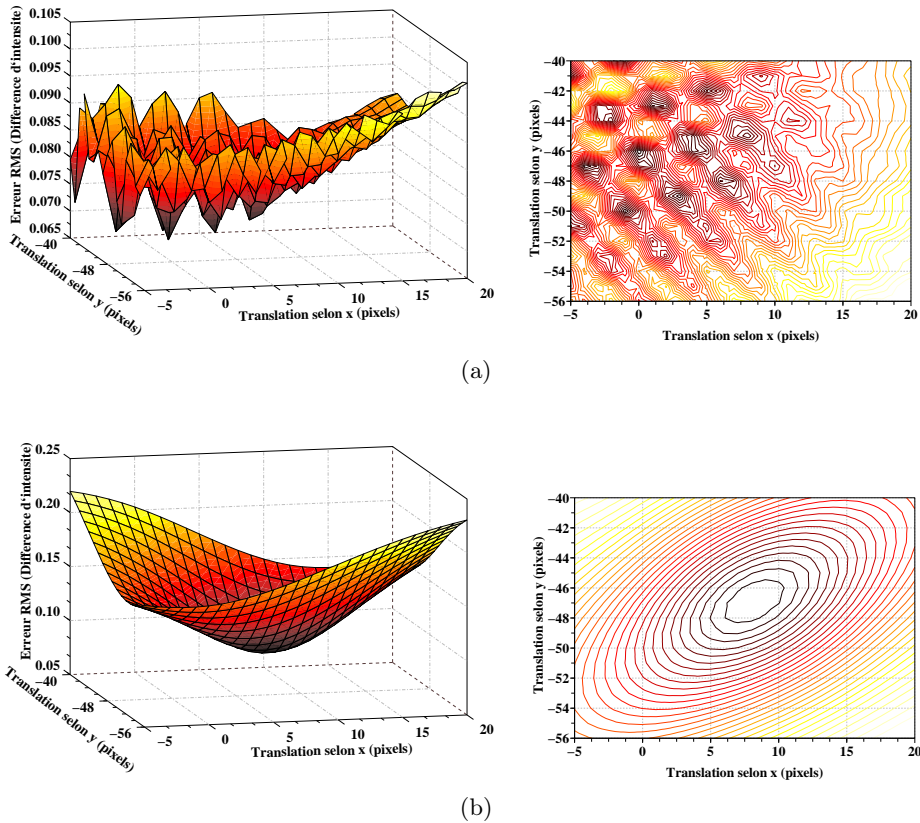


FIG. 3.7 – Effet produit par le pattern de fibres et par les inhomogénéités d’illumination sur la valeur absolue de la différence d’intensité entre les images. (a) Lorsque les images sans pré-traitement sont utilisées, les valeurs de déplacement données par le minimum de la fonction ne correspond pas aux valeurs correctes. Il existe de nombreux pics/vallées qui correspondent à des minima locaux. (b) Les valeurs correctes de translation entre les images (8, -47) sont trouvées lorsque les images sont pré-traitées avec notre méthode puisqu’il existe un seul minimum dans la surface qui est convexe.

3.2.2 Correction de la distorsion radiale

Nous utilisons la méthode de calibrage développée dans [Miranda-Luna *et al.*, 2004b] pour corriger la distorsion radiale introduite par l'optique du cystoscope. Une image sans distorsion I_R représentant une mire (par exemple un damier) est générée par un logiciel. Cette image est ensuite imprimée sur une feuille de papier. Une acquisition de cette mire est effectuée avec un cystoscope qui fournit une image I_D comprenant les distorsions dues à l'optique. L'image I_D est ensuite recalée avec l'image I_R . Les paramètres obtenus par ce recalage permettent de corriger les distorsions de l'optique (voir figure 3.8).

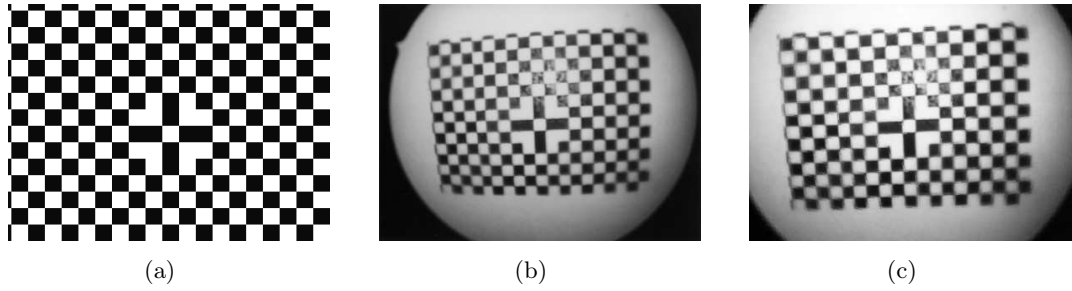


FIG. 3.8 – Correction de la distorsion radiale pour un cystoscope rigide (*Storz 57002*) à l'aide d'une mire. (a) Mire originale à laquelle est comparée l'image acquise. (b) Mire acquise par le cystoscope. (c) Image de la mire après application de la correction radiale avec les paramètres calibrés (coefficients polynomiaux $k_1 = 8,68 \times 10^{-4}$, $k_2 = 2 \times 10^{-8}$, $k_3 = 4 \times 10^{-11}$, $k_4 = 3 \times 10^{-14}$ et centre de distorsion (52.58, 79.10)).

L'algorithme de recalage optimise de façon itérative la ressemblance des images I_R et I_D en utilisant l'information mutuelle comme mesure de similarité. Cette mesure de similarité sera maximisée en ajustant un modèle à 9 paramètres de la transformation perspective (liés aux différences des points de vue des images I_R et I_D) et 6 paramètres de distorsion radiale (4 coefficients polynomiaux $[k_1, k_2, k_3, k_4]$ et les coordonnées du centre de distorsion (x_0, y_0) , (cf. sous-section 2.1.2). Au fur et à mesure du recalage, les coordonnées des pixels de l'image avec distorsion seront corrigées via la matrice de transformation radiale T_r (correction des distorsions) et la matrice de transformation perspective T_p (correction de la différence de point de vue).

$$\begin{bmatrix} x \\ y \end{bmatrix} = T_p \left[T_r \begin{bmatrix} x_d \\ y_d \end{bmatrix} \right], \quad (3.8)$$

Dans l'équation 3.8, $(x_d, y_d)^T$ représente les coordonnées de l'image acquise par le cystoscope et (x, y) sont les coordonnées après correction de la distorsion. L'image corrigée en distorsion est ensuite comparée à l'image sans distorsion $I_R(\mathbf{x})$ en calculant l'information mutuelle.

La maximisation de l'information mutuelle se fait par la méthode décrite dans la section 3.1 avec I'_D l'image corrigée :

$$I'_D(\mathbf{x}) = I_D(T_p(T_r(\mathbf{x}))). \quad (3.9)$$

Une fois l'instrument calibré, les valeurs obtenues pour les paramètres de distorsion radiale seront appliquées à toutes les images pour corriger la distorsion radiale. Il est à noter que seule la matrice T_r est utile pour cette correction, la matrice T_p donnant simplement la position de l'endoscope au moment de la calibration. Les paramètres de distorsion doivent être calculés une fois pour un endoscope donné.

Un cystoscope de la société *Storz* (*Storz* 57002) a été calibré en utilisant la méthode décrite ci-dessus. Le résultat de la correction appliquée au pattern acquis avec le cystoscope rigide (voir figure 3.8(b) pour l'étalonnage) est montré dans la figure 3.8(c).

3.3 Choix de la méthode de recalage

La première étape dans la construction de notre méthode de mosaïquage a été le choix de l'algorithme de recalage le plus adapté aux séquences d'images cystoscopiques.

Pendant un examen clinique, la position spatiale du cystoscope n'est pas connue et ce quelle que soit l'image de la séquence. Comme il a été vu dans la section 2.1.1, le mouvement « libre » d'un endoscope (dans ce cas un cystoscope) à l'intérieur de la vessie va produire des séquences dont le rapport géométrique entre images consécutives peut être décrit par une transformation perspective à laquelle une distorsion radiale est ajoutée. Il est donc naturel de choisir une transformation 2D/2D perspective pour recaler les images.

Dans la section 2.2.2 nous avons vu que les deux principales approches permettant de recaler une paire d'images, sont les méthodes qui utilisent directement les niveaux de gris des images et les méthodes qui utilisent les primitives images. Les deux approches possèdent chacune des avantages et des inconvénients qui les rendent plus ou moins adaptées à notre application.

Les méthodes basées sur l'utilisation de primitives images permettent de recaler des images présentant un faible taux de recouvrement. Certains travaux ont permis le recalage de données avec des taux de recouvrement avoisinant 25% du total de la surface des images [Capel, 2001]. Ces méthodes sont souvent utilisées pour des séquences d'images prises par un appareil photo et peuvent même être utilisées pour la reconnaissance de zones communes dans des images faisant partie d'une séquence désordonnée [Brown and Lowe, 2003]. Les primitives images doivent cependant être suffisamment contrastées pour pouvoir être segmentées de façon robuste.

Lorsque des primitives images ne peuvent pas être segmentées, il faut utiliser les niveaux de gris pour recaler les images. Dans les méthodes de recalage basées sur les niveaux de gris, en général, chaque pixel de l'image contribue au recalage. Ces méthodes peuvent travailler avec des images plus ou moins floues ou avec des gradients de niveaux de gris qui varient lentement (images sans contours marqués). Il existe une grande variété de mesures statistiques de similarité qui permettent d'estimer la ressemblance d'images de diverses modalités et ce dans le cadre d'un recalage monomodal où multimodal.

Le principal désavantage des méthodes basées sur les niveaux de gris est lié au grand taux de recouvrement des images nécessaire pour un recalage robuste. En effet, les distances entre points images homologues doivent généralement ne pas dépasser quelques pixels ou dizaines de pixels. Néanmoins, ces méthodes peuvent être utilisées pour la construction d'images panoramiques à partir de séquences de vidéo-cystoscopie pour lesquelles le taux de recouvrement est naturellement important (cf. section 3.3.1.2).

Nous devons donc considérer que le taux de recouvrement entre les images, l'adéquation d'un détecteur de primitives à un type d'image donné et la nature de la mesure de similarité sont des facteurs importants que nous devons prendre en compte pour choisir la méthode de recalage la plus adaptée à la construction d'images panoramiques de l'intérieur de la vessie.

3.3.1 Tests préliminaires effectués sur des images cystoscopiques réelles

Les tests présentés dans les sections suivantes ont été effectués pour nous aider à choisir la méthode de recalage la plus appropriée aux images cystoscopiques. Nous présentons d'abord le

test effectué avec un détecteur de primitives images. Ensuite, nous présentons l'évaluation du recouvrement entre les images, pour finalement tirer les conclusion de ces tests.

3.3.1.1 Test d'un détecteur de primitives image (détecteur de Harris)

Nous avons appliqué le détecteur de coins de Harris (voir section 2.2.5.1) à des paires d'images choisies au hasard parmi 10 séquences d'images cystoscopiques cliniques (une paire d'images de chaque séquence). Les images ont été pré-traitées par la méthode décrite dans la section 3.2.1. L'algorithme de Harris utilisé pour ce test a été implémenté au moyen des fonctions de la bibliothèque de traitement d'images *OpenCV*. La valeur du paramètre de contrôle α a été fixée à 0,06 comme dans l'article original [Harris and Stephens, 1988]. Des exemples de résultats obtenus sont montrés dans la figure 3.9. Nous observons que des coins sont détectés sur toutes les images. Cependant, les points détectés se concentrent parfois sur une région de l'image laissant d'autres (grandes) régions sans aucune primitive détectée. Sur des images floues, la détection des primitives produit aussi un grand nombre de coins qui après avoir été détectés sur la première image, n'apparaissent plus sur la deuxième.

3.3.1.2 Évaluation du degré de recouvrement des images de vidéo-cystoscopie

Dans notre application, le degré de recouvrement est lié à la cadence d'acquisition des images cystoscopiques (25 images/seconde) et à la vitesse de déplacement du cystoscope à l'intérieur de la vessie. Pour évaluer le taux de recouvrement, nous avons testé 10 séries d'images extraites de 6 séquences vidéo enregistrées pendant des examens cystoscopiques cliniques (6 patients) à des dates distinctes et avec des matériels différents. Le nombre d'images composant chaque séquence sélectionnée varie entre 150 et 511 images. Nous avons utilisé l'algorithme de recalage basé sur l'information mutuelle développé au sein du laboratoire (décrit dans la section 3.1) pour trouver les transformations entre images consécutives des séquences. Des sous-images de 256 x 256 pixels, centrées sur les images originales sont extraites pour former des séquences qui seront traitées par l'algorithme de recalage. Aucun autre pré-traitement n'est appliqué à ces sous-images. Pour chaque paire d'images, l'algorithme de recalage enregistre dans un fichier ASCII les valeurs des transformations relatives (transformation reliant deux images consécutives). Sachant que les translations sont les paramètres de transformation qui varient de façon prédominante dans les séquences de vidéo-cystoscopie, nous avons retenu uniquement les *valeurs absolues* de ces translations (sans considérer les autres paramètres de perspective) pour évaluer le recouvrement entre les images. Le nombre total de paires d'images est de 3860. La distribution des valeurs absolues de translation sur les axes \vec{x} et \vec{y} est montrée dans la figure 3.10

La distribution des valeurs de translations trouvées dans les séquences testées (figure 3.10) montre que ces valeurs sont toujours inférieures à 50 pixels. D'ailleurs, les translations inférieures ou égales à 10 pixels concentrent à elles seules 92 et 89% des valeurs de translations selon \vec{x} et \vec{y} respectivement. Les valeurs moyennes des translations sur \vec{x} et \vec{y} sont 3,20 et 4,04 pixels respectivement avec un écart type de 3,34 et 4,37 pixels. Pour les séquences testées, la valeur maximale de translation trouvée sur l'axe \vec{y} est de 43,56 pixels. Cette valeur apparaît une seule fois et elle est trouvée dans une séquence dont la particularité est de présenter les valeurs de translations les plus importantes parmi les séquences testées. En considérant la taille des images utilisées (256x256 pixels) et la valeur de translation maximale obtenue, nous pouvons dire que pour les séquences utilisées dans ce test, le recouvrement entre deux images est toujours supérieur à 80% (82,98%). Dans sa thèse, [Miranda Luna, 2005] a montré que la méthode de recalage basée sur les niveaux de gris qu'il utilise travaille de façon optimale (au niveau de la précision) lorsque

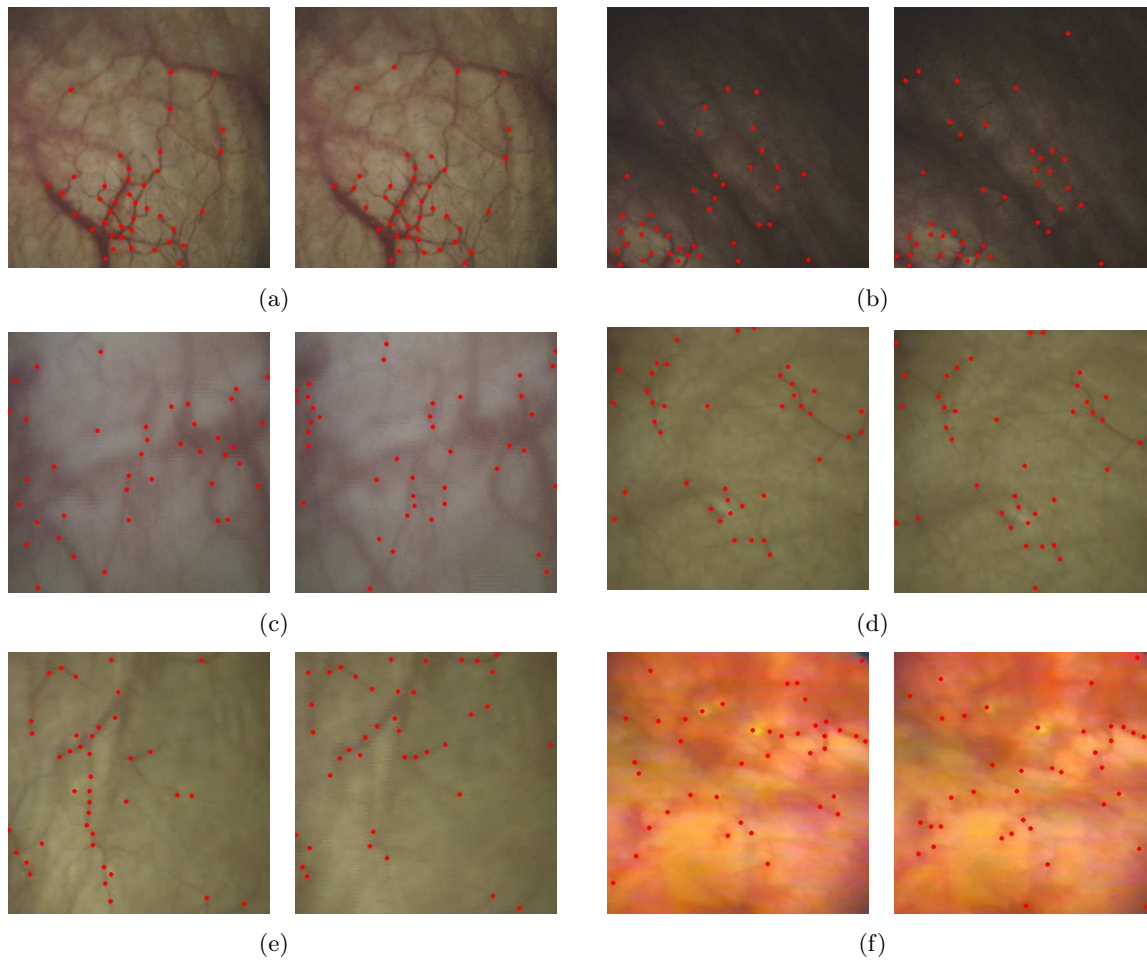


FIG. 3.9 – Exemples des points détectés par l’algorithme de Harris sur des paires d’images consécutives pré-traitées et issues de 6 séquences cystoscopiques différentes (a ... f). Les coins détectés sont marqués par des points rouges. (a) Images bien illuminées avec des structures bien définies. Les points détectés se concentrent sur une petite région de l’image. (b) Images avec des régions de fortes différences d’éclairage. Il n’y a pas de points détectés dans les régions obscures et les points détectés sur la première image ont des points homologues sur la deuxième image seulement dans la partie plus claire. (c) Images bien illuminées et avec des structures visibles mais légèrement floues. Les points détectés sur la première image ont des points homologues sur la deuxième image seulement sur la partie gauche des images. (d) Images avec une bonne illumination mais avec peu de primitives. Les mêmes points sont détectés sur les deux images, mais ils se concentrent seulement sur les structures fortement contrastées. (e) La deuxième image est plus floue que la première. La localisation des points détectés varie beaucoup entre les deux images. (f) Images où la qualité de l’acquisition est médiocre. Beaucoup de points localisés sur la première image ne le sont plus sur la deuxième.

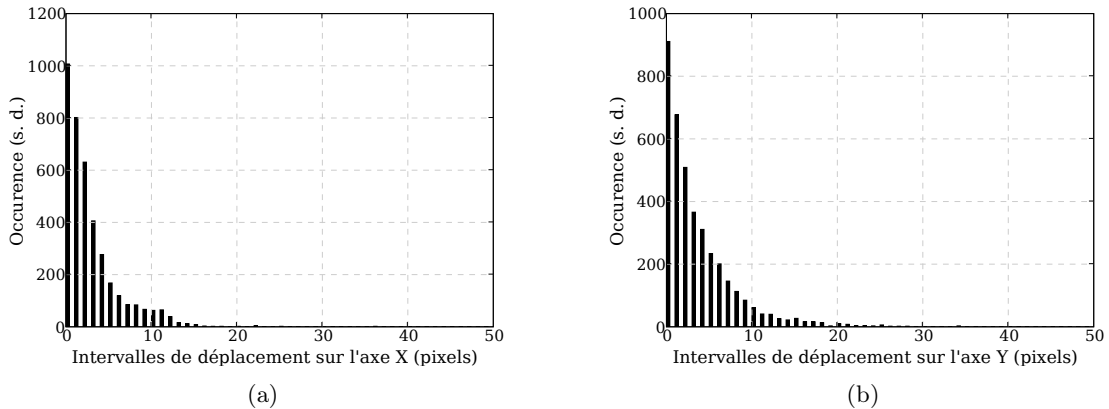


FIG. 3.10 – Distribution des valeurs absolues des translations pour dix séquences d’images (3868 valeurs de translation). Les valeurs de translation ont été arrondies à l’entier inférieur. (a) Distribution des translations selon l’axe \vec{x} (déplacement). (b) Distribution des translations selon l’axe \vec{y} .

le recouvrement entre les images est supérieur ou égal à 90% de la taille de l’image et qu’il arrive à recalcr des images dont le recouvrement est supérieur ou égal à 80%.

3.3.2 Conclusions tirées des tests

Le résultat du test effectué en utilisant le détecteur de Harris a montré que l’utilisation d’une méthode de recalage se basant sur des primitives images de type « coins » n’est pas adaptée à une application aux séquences de vidéo-cystoscopie. En effet, le test a montré que les caractéristiques détectées sur une première image peuvent ne plus apparaître sur l’image adjacente. Effectuer un recalage en utilisant seulement les points détectés ne conduirait pas à un recalage robuste. La même remarque peut être formulée pour les contours, car ces primitives ne sont pas mieux détectées que les coins.

Les résultats du test de recouvrement entre les images d’une séquence cystoscopique montrent que la plupart du temps, le recouvrement entre les images est supérieur à 4/5 de l’image. Cela montre qu’une méthode de recalage pouvant gérer des petits taux de recouvrement entre les images (méthode basée sur l’utilisation de primitives image) n’est pas nécessaire et que l’utilisation d’un algorithme basé sur l’utilisation des niveaux de gris des images est envisageable dans le cas de séquences cystoscopiques.

En tenant compte des caractéristiques des images cystoscopiques, des tests effectués et des travaux menés au laboratoire (basés sur l’information mutuelle, donc des niveaux de gris) [Miranda Luna, 2005], nous avons considéré qu’une méthode de recalage basée sur l’utilisation des niveaux d’intensité des images est bien adaptée à une application à des séquences de vidéo-cystoscopie.

3.4 Première méthode de recalage proposée : l'algorithme de Fourier-Mellin

3.4.1 Principe de la méthode

Nous avons présenté dans la section précédente les raisons pour lesquelles nous avons choisi d'utiliser une approche de recalage basée sur les niveaux de gris des images. Nous devons aussi choisir la mesure de similarité et la méthode d'optimisation qui nous permettront de recalculer avec une bonne précision les séquences d'images cystoscopiques et ce avec une vitesse de calcul compatible avec la durée d'un examen clinique.

D'après la littérature, nous savons que les méthodes de recalage basées sur l'utilisation des niveaux de gris et capables de travailler avec un modèle de transformation perspective sont toutes des méthodes itératives. Si l'extremum (le minimum ou le maximum) de la mesure de similarité se trouve éloigné du point de départ de l'optimisation, il sera nécessaire d'effectuer beaucoup d'itérations avant de pouvoir trouver la solution. Par ailleurs, dans cette situation, il existe un risque de non convergence vers l'extremum global si des extrema locaux existent. Cette configuration de l'espace de paramètres pourrait au mieux ralentir sensiblement l'algorithme ou, au pire, le faire échouer. C'est pour cette raison que nous avons décidé d'utiliser une méthode non itérative capable de nous donner une première approximation de la transformation existant entre les images. Cette dernière pourra être prise comme point de départ pour un algorithme itératif affinant le recalage avec un modèle de transformation perspective. Le travail présenté dans cette section a été décrit dans une communication [Hernandez-Mier *et al.*, 2005].

La première méthode de recalage basé sur les niveaux de gris que nous avons testée est basée sur une mesure de similarité qui utilise la transformée de Fourier discrète : l'algorithme de Fourier-Mellin (cf. section 2.2.4.1). La transformée de Fourier discrète peut être calculée rapidement (en $O(NM \log NM)$ opérations pour une image de taille $N \times M$ avec N et M des puissances de 2) en utilisant la transformée de Fourier rapide (FFT). Cet algorithme permet de trouver les paramètres de transformation d'une *similitude* (transformation définie par 4 paramètres : angle de rotation dans le plan, facteur d'échelle et translations sur \vec{x} et \vec{y}).

Dans cette approche, nous considérons une image de référence $I_R(\mathbf{x})$ et une image à transformer $I_D(\mathbf{x})$ qui sera transformée pour la superposer à I_R à l'aide d'une similitude T_S .

La base de cet algorithme est la propriété de translation de la transformée de Fourier. Le *spectre* des images (module de la transformée de Fourier) calculé avec l'algorithme de Fourier-Mellin est invariant aux translations. De plus, ces translations peuvent être découplées des rotations et des changements d'échelle dans le cas d'une similitude. Dans notre approche, nous appliquons une apodisation (un masque) aux images en les multipliant avec un masque de Butterworth avant de calculer leurs transformées de Fourier. Multiplier un signal par une fonction porte (en 2D un filtre passe-bas circulaire centré sur l'origine) dans le domaine spatial permet d'atténuer les « rebonds » dans le domaine fréquentiel (phénomène de Gibbs) dus dans notre cas aux discontinuités de l'image. La fonction de transfert du masque de Butterworth est donnée par l'équation suivante :

$$W_{ap}(x, y) = \frac{1}{1 + \left(\frac{D_s(x, y)}{\sigma_{ap}}\right)^{2n}} \quad (3.10)$$

avec $D_s(x, y) = \sqrt{x^2 + y^2}$ la distance entre le centre de l'image et le point de coordonnées (x, y) , σ_{ap} l'écart type du masque et n l'ordre de la fonction de Butterworth. Les transformées de Fourier qui ont été multipliées par le masque sont données par :

$$\mathcal{I}_{R_B}(\mathbf{f}) = \mathcal{F}\{I_R(\mathbf{x})W_{ap}(\mathbf{x})\} \quad \text{et} \quad \mathcal{I}_{D_B}(\mathbf{f}) = \mathcal{F}\{I_D(\mathbf{x})W_{ap}(\mathbf{x})\}, \quad (3.11)$$

avec $\mathbf{f} = (u, v)^T$ les coordonnées dans le domaine fréquentiel.

L'étape suivante est de passer du repère cartésien à un repère en coordonnées logarithmiques-polaires (voir section 2.2.4.1). Avant de réaliser ce changement de repère il est nécessaire d'éliminer les plus basses fréquences du spectre de Fourier des images. Cette opération est réalisée par un filtrage passe-haut dans le domaine fréquentiel. ce filtrage est effectué en multipliant les transformées de Fourier des images par la fonction de transfert d'un filtre de Butterworth :

$$\mathcal{I}_{R_{BH}}(\mathbf{f}) = \mathcal{I}_{R_B}(\mathbf{f})W_{PH}(\mathbf{f}) \quad \text{et} \quad \mathcal{I}_{D_{BH}}(\mathbf{f}) = \mathcal{I}_{D_B}(\mathbf{f})W_{PH}(\mathbf{f}), \quad (3.12)$$

avec

$$W_{PH}(u, v) = \frac{1}{1 + \left(\frac{\sigma_{PH}}{D_f(u, v)}\right)^{2n}} \quad (3.13)$$

et $D_f(u, v) = \sqrt{u^2 + v^2}$.

Les modules des spectres (ici indiqués par $\|\cdot\|$) sont ensuite passés dans un repère en coordonnées logarithmiques-polaires :

$$\bar{\mathcal{I}}_{R_{lp}}(s, \theta) = \|\mathcal{I}_{R_{BH}}(e^s \cos \theta, e^s \sin \theta)\| \quad \text{et} \quad \bar{\mathcal{I}}_{D_{lp}}(s, \theta) = \|\mathcal{I}_{D_{BH}}(e^s \cos \theta, e^s \sin \theta)\|, \quad (3.14)$$

Dans cette représentation, le déplacement angulaire $\hat{\theta}$ (rotation) et le déplacement en échelle \hat{s} existant entre les images seront représentés comme par des translations selon les axes θ et $s = \log \rho$:

$$\bar{\mathcal{I}}_{D_{lp}}(s + \hat{s}, \theta + \hat{\theta}) = \bar{\mathcal{I}}_{R_{lp}}(s, \theta). \quad (3.15)$$

Ces translations peuvent aussi être calculées par un algorithme de corrélation. Dans notre approche, nous avons décidé d'utiliser la corrélation des phases pour estimer les déplacements angulaire et en échelle entre les images car le pic correspondant est relativement bien prononcé. Pour trouver ces déplacements, il est donc nécessaire d'obtenir les transformées de Fourier des spectres en coordonnées log-polaires :

$$\mathcal{I}_{R_{lp}}(\mathbf{f}) = \mathcal{F} \{ \bar{\mathcal{I}}_{R_{lp}}(s, \theta) \} \quad \text{et} \quad \mathcal{I}_{D_{lp}}(\mathbf{f}) = \mathcal{F} \{ \bar{\mathcal{I}}_{D_{lp}}(s, \theta) \}, \quad (3.16)$$

afin de calculer ensuite la corrélation des phases

$$CC_{CP}(s, \theta) = \mathcal{F}^{-1} \left\{ \frac{\mathcal{I}_{R_{lp}}(\mathbf{f})\mathcal{I}_{D_{lp}}^*(\mathbf{f})}{\|\mathcal{I}_{R_{lp}}(\mathbf{f})\| \|\mathcal{I}_{D_{lp}}(\mathbf{f})\|} \right\} \quad (3.17)$$

où * dénote un complexe conjugué. Les coordonnées (s_{pic}, θ_{pic}) du maximum global dans CC_{CP} vont nous indiquer le déplacement selon les deux axes des images. Pour obtenir l'équivalent en degrés α du déplacement angulaire, il est nécessaire d'établir la relation entre le nombre d'éléments N_θ sur l'axe θ de l'image en coordonnées log-polaires et θ_{pic} : $\alpha = \theta_{pic}(2\pi/N_\theta)$. Pour obtenir le facteur d'échelle équivalent au déplacement sur l'axe s , il suffit d'associer la valeur de s correspondant à la position donnée par s_{pic} .

Avec l'inverse du paramètres de rotation α et du facteur d'échelle s calculés, nous pouvons construire une matrice de transformation (similitude T_{S_1}) sans translation qui va transformer les coordonnées de l'image $I_D(\mathbf{x})$ comme suit :

$$I_{D_s}(\mathbf{x}) = I_D(T_{S_1}(\mathbf{x})), \quad (3.18)$$

avec

$$T_{S_1} = \begin{bmatrix} s \cos \alpha & -s \sin \alpha & 0 \\ s \sin \alpha & s \cos \alpha & 0 \end{bmatrix}.$$

TAB. 3.2 – Résultats pour une séquence de 5 images acquises pour le fantôme. *Théo* indique la valeur théorique (et exacte) des paramètres de transformation appliqués aux images et *Calc* la valeur des paramètres de transformation calculés par notre version de l’algorithme de Fourier-Mellin. Le temps de calcul par paire d’images est de 0,82 s pour cette séquence.

Paire	Rotation (°)		Échelle		t_x		t_y	
	<i>Théo</i>	<i>Calc</i>	<i>Théo</i>	<i>Calc</i>	<i>Théo</i>	<i>Calc</i>	<i>Théo</i>	<i>Calc</i>
1	5,0	5,62	1,0	1,0	28	29	87	87
2	10,0	9,84	1,0	1,0	68	68	-18	-19
3	30,0	29,53	1,0	1,0	-46	-48	81	78
4	0,0	0,0	1,0	1,0	97	97	35	35

La translation entre l’image I_{D_S} et l’image I_R est finalement calculée à partir d’une corrélation dans le domaine de Fourier. À partir des transformées de Fourier des ces images, nous pouvons ainsi calculer :

$$CC(\mathbf{x}) = \mathcal{F}^{-1} \{ \mathcal{I}_R(\mathbf{f}) \mathcal{I}_D^*(\mathbf{f}) \}. \quad (3.19)$$

Le pic maximum dans cette fonction donne la position (x_{pic}, y_{pic}) qui indique le déplacement (t_x, t_y) existant entre les images. Avec ces informations nous complétons la matrice de transformation de similitude qui donne le lien géométrique entre les images $I_R(\mathbf{x})$ et $I_D(\mathbf{x})$.

3.4.2 Application de l’algorithme à des séquences d’images d’un fantôme et cliniques

La méthode de recalage basée sur l’algorithme de Fourier-Mellin a été intégrée dans algorithme de mosaïquage d’images qui a été implémenté avec *Scilab V4.0* et la bibliothèque de fonctions de traitement d’images *SIP V0.4* sur une plate-forme *GNU/Linux, Kernel 2.6.15*. Les paramètres des masques de Butterworth ont été ajustés de façon empirique, par des approximations successives : $\sigma = 0,625N$, $n = 8$ (voir équation 3.10) et $\sigma = 0,15N$, $n = 6$ (voir équation 3.13) pour des images de dimension $N \times N$.

Les premiers tests que nous avons réalisés ont été effectués avec des sous-images extraites d’une image. Différentes similitudes relient les sous-images extraites. Les liens géométriques entre ces dernières était parfaitement connus par construction. Le fantôme utilisé dans ce test est la photographie d’une vessie de porc qui a été incisée, ouverte et « mise à plat ». Une petite séquence de 5 images extraites de cette photographie de vessie de porc a été utilisée pour construire une image panoramique. Les résultats pour cette séquence sont donnés dans le tableau 3.2 et la figure 3.11.

Nous avons effectué d’autres tests sur des séquences « fantôme » où nous connaissions les paramètres de transformation appliqués qu’il fallait récupérer et où aucun bruit n’est impliqué. Sur tous ces tests, notre implémentation de l’algorithme de Fourier-Mellin arrivait à trouver les paramètres de transformation avec des erreurs moyennes faibles : $\pm 0,8^\circ$ pour la rotation, $\pm 0,05$ pour le facteur d’échelle, ± 2 pixels pour la translation sur \vec{x} et ± 3 pixels pour la translation sur \vec{y} .

Cependant, lorsque nous avons essayé d’utiliser l’algorithme avec des séquences d’images réelles extraites de cystoscopies cliniques, l’algorithme échouait totalement. Les valeurs de rotation et d’échelle obtenues étaient impossibles à trouver du fait de l’apparition de nombreux pics dans la fonction de corrélation. Nous avons observé que l’étape de l’algorithme qui faisait échouer le recalage était la corrélation des phases des modules des spectres de Fourier dans le repère en

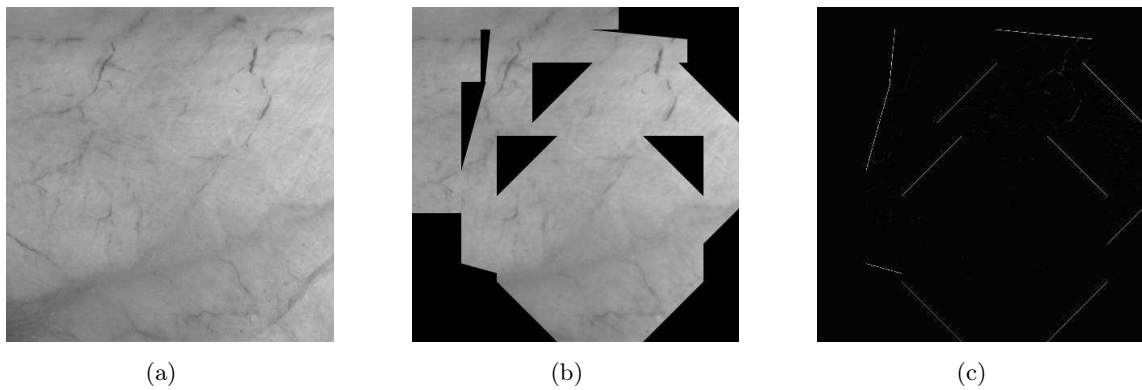


FIG. 3.11 – Résultat du mosaïquage d'images pour la séquence d'images extraites d'une photographie de vessie de porc (voir tableau 3.2). (a) Photographie de la vessie de porc de laquelle les images ont été extraites. (b) Image panoramique formée avec les 5 images extraites. (c) Différence de position entre l'image originale et l'image panoramique calculée avec les niveaux de gris des pixels.

coordonnées logarithmiques-polaires. Dans la figure 3.12 nous pouvons observer un exemple de fonction de corrélation obtenue pour des images cliniques qui illustre l'effet que le bruit présent dans les images cliniques produit sur la corrélation. Alors que pour le fantôme un maximum correspondant effectivement au couple $(\rho_{pic}, \theta_{pic})$ à trouver peut être détecté, pour les données cliniques il existe une grande quantité de maxima qui « cachent » le maximum correspondant à la rotation et au facteur d'échelle corrects.

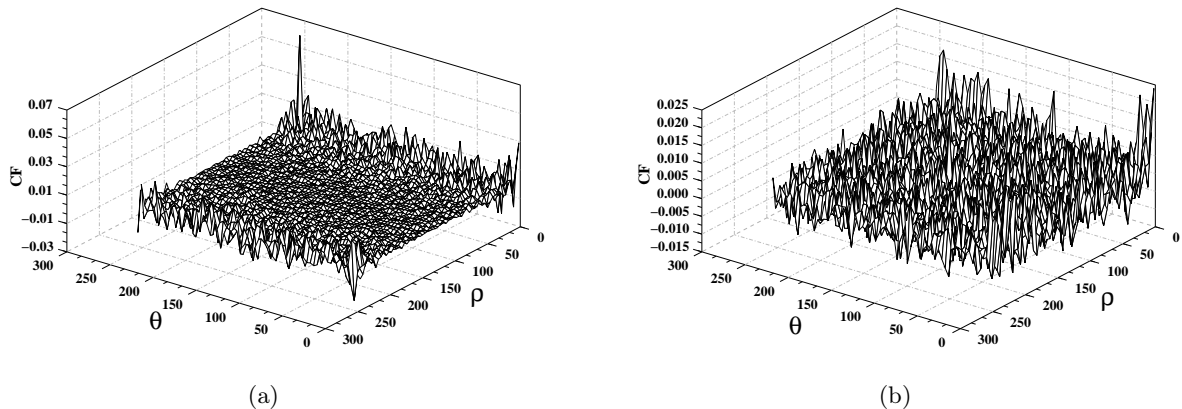


FIG. 3.12 – (a) Fonctions de corrélation des phases pour une paire d'images extraites d'une photo de vessie de porc. (b) Fonction de corrélation des phases pour une paire d'images cliniques.

Nous avons aussi observé que les variations des niveaux de gris d'images adjacentes d'une séquence suffisaient à altérer l'allure de l'amplitude des spectres de Fourier et par conséquent, la fonction de corrélation des phases. Nous pensons que les déformations des images qui ne sont pas considérées par le modèle de transformation de similitude (comme des changements du point de vue) affectent fortement le résultat de l'algorithme.

3.4.3 Discussion

Les premiers résultats obtenus avec l'algorithme de Fourier-Mellin avec les images extraites d'un fantôme de vessie étaient intéressants en termes de précision et de vitesse de calcul. Cependant, les résultats de son utilisation pour des images cliniques ont montré qu'il n'est pas adapté au recalage d'images cystoscopiques réelles. Nous avons identifié plusieurs facteurs affectant la performance de l'algorithme. D'abord, nous avons constaté que les paramètres des fenêtres construites avec la fonction de Butterworth (approximation de fonctions portes) affectent de façon importante le comportement général du recalage. Même si en variant ces paramètres, il est possible d'améliorer les résultats, il serait nécessaire de les adapter à chaque paire d'images, ce qui n'est pas envisageable en pratique. Ensuite, nous avons observé que même des petites variations des niveaux d'intensité des images affectent les résultats de l'algorithme. La correction des inhomogénéités d'éclairage n'est pas parfaite et ne conduit pas aux niveaux de gris optimaux. Un autre problème de cette approche est que les images cystoscopiques sont déformées en réalité par des transformations plus complexes (changements de perspective) qu'une similitude. Des similitudes ont été testées car les changements de perspective sont en moyenne très faibles entre deux images consécutives et ce même pour un examen réel. Cependant lors de changements de trajectoires du cystoscope, les changements de perspective deviennent trop importants et le recalage échoue avec une similitude.

Après avoir constaté que cette approche n'était pas adaptée à notre application, nous avons décidé d'orienter nos travaux vers une autre solution.

Cependant, ces travaux nous ont permis de montrer que la corrélation dans le domaine de Fourier peut initialiser le recalage à partir de conditions initiales accélérant le calcul. En effet, même si la rotation et le facteur d'échelle de la similitude ne peuvent pas être calculés de façon robuste, la corrélation dans le domaine de Fourier fournit des translations 2D proches de la solution. Cela permet de réduire considérablement le nombre des itérations d'un algorithme d'optimisation qui trouve précisément la solution. Nous avons observé que la corrélation dans le domaine de Fourier est suffisamment robuste vis-à-vis des déformations existant entre les images d'une séquence cystoscopique (voir résultats de la section 3.5).

3.5 Seconde méthode proposée

Dans cette section nous présentons la méthode de recalage retenue. En plus d'une étape de pré-traitement, nous proposons une corrélation effectuée avec les transformées de Fourier des images (voir section 3.5.1) délivrant des translations servant de valeurs initiales à un algorithme de recalage itératif qui utilise un modèle de transformation perspective (voir section 3.5.2). La dernière section de ce chapitre, présente de façon globale et synthétique les éléments qui composent notre approche.

3.5.1 La corrélation avec les transformées de Fourier des images

Pendant les tests effectués avec l'algorithme de Fourier-Mellin, nous nous sommes aperçus que l'obtention des paramètres de translation entre deux images cystoscopiques pré-traitées en effectuant une corrélation avec leurs transformées de Fourier est possible, même en présence d'autres transformations, telles que des rotations, facteur d'échelle et perspective. Étant donné que les translations sont les transformations qui prédominent dans une séquence cystoscopique, nous pensons que cette technique peut fournir un bon point de départ pour un algorithme itératif qui va calculer les paramètres du modèle de transformation perspective qui, avec les translations

initiales, donne le lien géométrique entre images consécutives. Pour tester cette idée, nous avons pris deux images successives extraites d'une séquence de cystoscopie clinique. Nous avons pris deux images parce que de cette façon nous avons des données plus réalistes qui comprennent les défauts liés à l'acquisition et aux différences d'éclairage. Après leur pré-traitement, nous avons transformé géométriquement une des deux images de sorte que la matrice perspective entre les deux images soit connue exactement par construction. Nous connaissons la translation initiale des images selon \vec{x} et \vec{y} entre les images grâce à l'estimation réalisée par l'algorithme de recalage basé sur l'information mutuelle. Les valeurs des translations sont 1,42 pixels selon \vec{x} et 0,78 pixels selon \vec{y} . Pour obtenir les valeurs de translations « théoriques » nous ajoutons des valeurs de translation aux translations initiales. Les autres paramètres indépendants appliqués sont la rotation dans le plan image, le changement d'échelle et les translations (une similitude).

Les paramètres de translation trouvés avec la corrélation basée sur la transformée de Fourier sont comparés aux paramètres de translation appliqués en même temps que des rotations et des changements d'échelle. Dans nos tests, nous avons fixé les valeurs théoriques de translation à $(x, y) = (9, -11)$ pixels. Les valeurs de rotation dans le plan varient entre -3° et 3° . Le facteur d'échelle varie entre 0,9 et 1,1. Les résultats de ces tests sont donnés dans le tableau 3.3.

TAB. 3.3 – Résultats de la corrélation dans le domaine de Fourier pour des similitudes connues *a priori*. Les valeurs théoriques des translations sont $(x, y) = (9, -11)$ (cellules ombrées). t_x calc, t_y calc sont les valeurs de translation (en pixels) obtenues par l'algorithme de corrélation.

Rotation ($^\circ$)	-3	-3	-3	-3	-3	-2	-2	-2	-2	-2	-1	-1	-1	-1	0	0	0	
Facteur d'échelle	0,9	0,95	1	1,05	1,1	0,9	0,95	1	1,05	1,1	0,9	0,95	1	1,05	1,1	0,9	0,95	1
t_x calc	8	4	-2	-8	-15	12	7	1	-5	-11	16	12	6	0	-6	20	15	9
t_y calc	-5	-11	-19	-26	-33	-2	-9	-16	-23	-30	-1	-7	-14	-21	-28	2	-4	-11

Rotation ($^\circ$)	0	0	0	1	1	1	1	1	2	2	2	2	2	3	3	3	3	3
Facteur d'échelle	1	1,05	1,1	0,9	0,95	1	1,05	1,1	0,9	0,95	1	1,05	1,1	0,9	0,95	1	1,05	1,1
t_x calc	9	3	-2	18	13	6	0	-5	16	9	2	-3	-8	15	8	0	-6	-10
t_y calc	-11	-18	-25	-1	-7	-15	-22	-28	-2	-10	-17	-24	-30	-5	-12	-20	-27	-33

Dans ce tableau, nous pouvons observer que plus les facteurs d'échelle et les rotations de la similitude sont importants, plus les valeurs de la translation calculée sont éloignées des valeurs théoriques. L'erreur la plus importante (24 pixels selon \vec{x} et 22 pixels selon \vec{y}) intervient lorsque nous appliquons en plus des translations, une rotation de -3° et un facteur d'échelle 1,1. Nous observons aussi, que lorsque le facteur d'échelle est supérieur à 1, la précision de la corrélation est plus affectée que lorsqu'il est inférieur à 1. Ce test nous apprend que l'utilisation de la corrélation dans le domaine fréquentiel pour donner un point de départ à l'algorithme itératif est possible si les transformations entre les images consécutives d'une séquence sont petites. Nous pouvons dire que pour des rotations entre -2° et 2° et des facteurs d'échelle entre 0,9 et 1,05, la corrélation peut donner une approximation exploitable des valeurs de translation entre les images.

Pour confirmer que la corrélation peut être utilisée dans le recalage de séquences cystoscopiques réelles, nous avons effectué un deuxième test. Nous avons appliqué l'algorithme de corrélation dans le domaine fréquentiel sur les dix séquences cystoscopiques cliniques utilisées pour évaluer le degré de recouvrement des images (cf. section 3.3.1.2). Pour évaluer la précision des translation obtenues, nous avons pris comme référence les valeurs de translation obtenues par l'algorithme de recalage basé sur l'information mutuelle. Les courbes présentées dans la figure 3.13 montrent les résultats obtenus pour trois séquences sur dix. Les résultats présentés pour les

trois séquences sont aussi représentatifs des résultats obtenus pour les 7 autres séquences.

L'algorithme de recalage basé sur l'information mutuelle est très précis. Les valeurs des translations délivrées par cet algorithme peuvent être considérées comme exactes et peuvent donc être prises comme des références auxquelles les résultats de l'algorithme proposé seront comparés. L'allure des courbes est très similaire à celle des courbes obtenues par l'algorithme de recalage basé sur l'information mutuelle. Cependant, nous avons observé que pour des valeurs de translation importantes, comme par exemple, autour de la paire 50–51 dans la figure 3.13(c), les valeurs obtenues sont très différentes. Cette différence est essentiellement due aux limites en termes de translations entre deux images (besoin d'un recouvrement $> 90\%$ de l'image) de l'algorithme basé sur l'information mutuelle. La valeur moyenne de la différence entre les translations obtenues par les deux algorithmes pour les dix séquences (3868 valeurs de translations) est de 0,04 pixels selon \vec{x} et de 0,02 pixels selon \vec{y} , avec un écart type de 2,33 et 2,54 pixels respectivement. La différence maximale est de 34,22 pixels sur \vec{x} et 34,44 pixels sur \vec{y} .

En considérant les résultats des tests, particulièrement ceux du second, nous pouvons conclure que la corrélation dans le domaine fréquentiel fournit des conditions initiales en termes de translation qui sont proches de la solution à trouver.

De plus, le second test nous a permis de prouver ce que nous avons déduit par observation : la transformation prédominante dans les séquences cystoscopiques cliniques est la translation. Il est évident que d'autres transformations sont présentes, mais leurs valeurs « peu dominantes » nous permettent d'utiliser efficacement la corrélation pour fournir une première estimation des translations entre les images.

3.5.2 La méthode de recalage par composition inverse

L'approche la plus directe pour recaler deux images consiste en la comparaison des niveaux de gris des pixels homologues des images à recaler. Nous proposons d'utiliser le carré de la différence des niveaux de gris pour mesurer le degré de similarité des images. Cette mesure de similarité est calculée rapidement puisqu'il suffit d'effectuer la soustraction pixel à pixel des images. Étant donné la nature monomodale des images cystoscopiques (les images à recaler sont issues d'un même dispositif) et l'amplitude (plutôt faible) des transformations entre des images consécutives, cette mesure de similarité semble adaptée à notre application. Grâce au pré-traitement que nous appliquons aux images, les niveaux de gris de l'image à recaler sont proportionnels aux niveaux de gris de l'image de référence et avoir un minimum global bien prononcé est possible. Le point de départ fourni par la corrélation dans le domaine fréquentiel permet à l'algorithme de recalage de démarrer plus près de la solution et donc, potentiellement de converger plus rapidement.

Cette section explique la méthode utilisée pour recaler deux images dont le lien géométrique est régi par une transformation perspective. Cette méthode nous la présentons aussi dans [Hernández-Mier *et al.*, 2005] et [Hernández-Mier *et al.*, 2006]. La méthode de recalage que nous proposons est basée sur le travail développé par [Baker and Matthews, 2002]. Ils utilisent le carré des différences d'intensité comme mesure de similarité. Cette dernière est minimisée par un algorithme d'optimisation de deuxième degré (Gauss-Newton). Cette méthode est utilisée dans le cadre du suivi de visages appliqué à la détection de la fatigue chez le conducteur. Leur méthode est appelée *recalage d'images par « composition inverse »* (*Inverse Compositional Algorithm* en anglais). La méthode de recalage de [Baker and Matthews, 2002] s'inspire et améliore, d'un point de vue de la robustesse et de la vitesse, la méthode de [Lucas and Kanade, 1981]. Pour ce faire, cette dernière est modifiée de sorte que certaines étapes, coûteuses en temps de calcul et ré-calculées à chaque itération, ne soient calculées qu'une seule fois et ré-utilisées à chaque itération. Pour expliquer l'algorithme de recalage par composition inverse nous allons rappeler

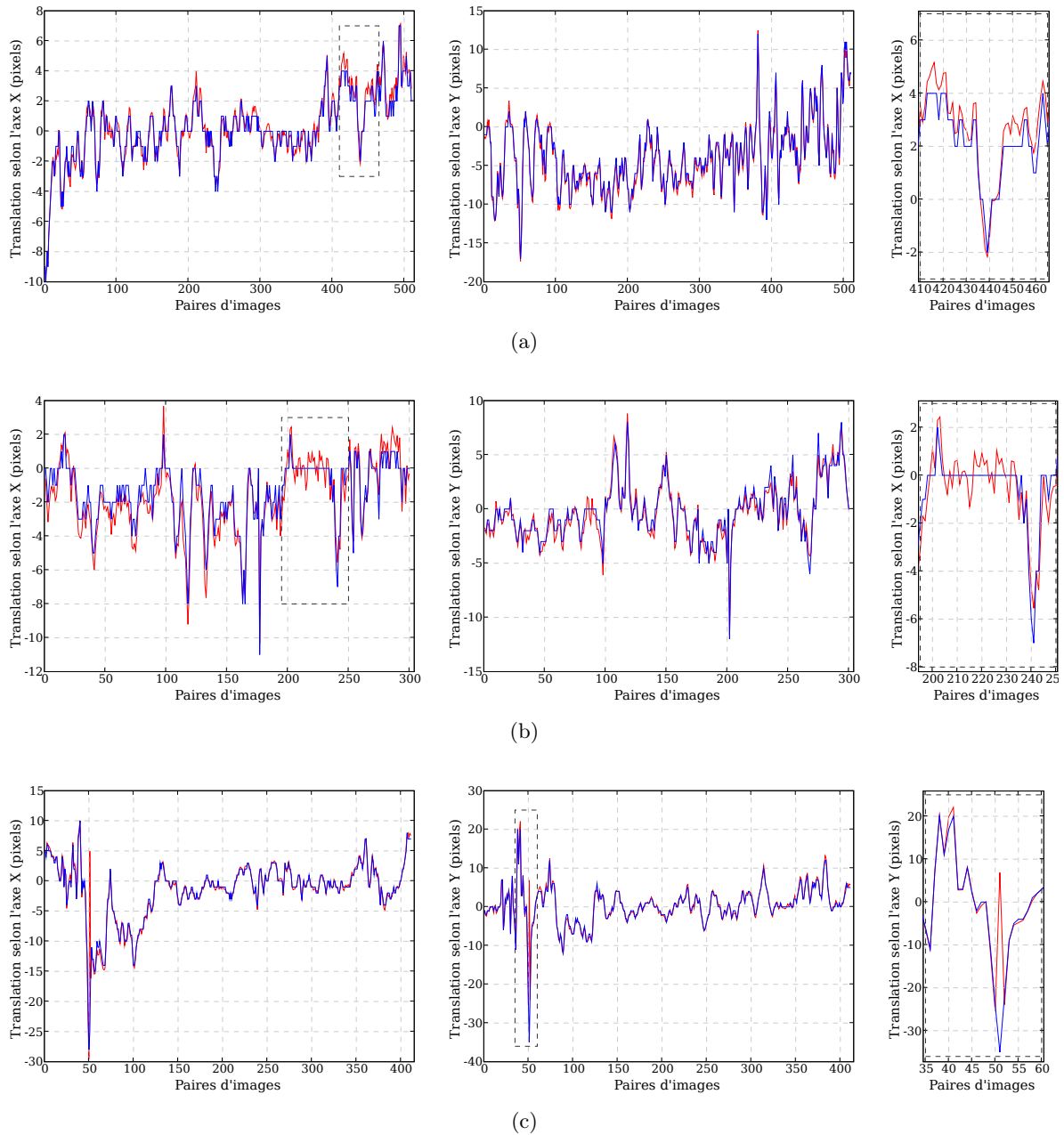


FIG. 3.13 – Comparaison des valeurs de translations obtenues par l’algorithme de corrélation dans le domaine fréquentiel (ligne bleue) et celles obtenues par l’algorithme de recalage basé sur l’information mutuelle (ligne rouge) pour des images consécutives de trois séquences cystoscopiques cliniques. Les petites fenêtres à droite sont des grossissements des zones indiquées par les rectangles en trait pointillé. (a) Valeurs de translation pour une séquence de 511 images. (b) Translations pour une séquence composée de 301 images. (c) Valeurs de translation pour une séquence de 412 images. Dans cette dernière séquence, il peut être observé qu’autour de la paire d’images 50–51 les valeurs obtenues par les deux algorithmes sont très différentes. La différence est due aux valeurs de translation qui sont importantes pour ces paires d’images (taux de recouvrement inférieure à 90% de la taille de l’image).

les algorithmes sur lesquels il est basé. Nous présentons ces méthodes sous forme d'algorithme pour faciliter l'appréciation des améliorations proposées par [Baker and Matthews, 2002]. Les détails des étapes des algorithmes sont développés dans l'annexe A.

3.5.2.1 L'algorithme de recalage de Lucas et Kanade

Cet algorithme a été développé pour déterminer le flot optique entre images d'une séquence dans le cadre d'une application à la vision stéréo. Dans l'article original, l'algorithme de [Lucas and Kanade, 1981] ne prend en compte que les translations. Cet algorithme peut cependant être étendu à d'autres paramètres.

L'algorithme recalcule une image $I_D(\mathbf{x})$ sur une image de référence $I_R(\mathbf{x})$, avec $\mathbf{x} = (x, y)^T$ un vecteur colonne contenant les coordonnées des pixels. La transformation des coordonnées est notée $T(\mathbf{x}; \mathbf{p})$. Cette notation signifie que les coordonnées \mathbf{x} vont être transformées par une matrice T constituée d'une liste de paramètres \mathbf{p} , où $\mathbf{p} = (p_1, \dots, p_n)^T$ est le vecteur contenant les paramètres de transformation. $n = 8$ pour une transformation perspective. Cette transformation sera appliquée aux pixels de coordonnées \mathbf{x} de l'image I_D , c'est-à-dire $I_D(T(\mathbf{x}; \mathbf{p}))$, en « bougeant » les pixels de façon itérative jusqu'à obtenir une image la plus similaire possible à l'image de référence. Pour résumer, nous allons minimiser la différence des niveaux de gris des images. Pour ce faire, nous utilisons le carré de la somme de l'erreur des pixels homologues des deux images :

$$\sum_{\mathbf{x}} [I_D(T(\mathbf{x}; \mathbf{p})) - I_R(\mathbf{x})]^2. \quad (3.20)$$

L'algorithme de [Lucas and Kanade, 1981] considère que les paramètres de transformation \mathbf{p} initiaux sont connus et cherche ensuite les incréments qui, ajoutés ou soustraits à ces paramètres, feront que l'image I_D devienne de plus en plus similaire à l'image I_R :

$$\sum_{\mathbf{x}} [I_D(T(\mathbf{x}; \mathbf{p} + \Delta\mathbf{p})) - I_R(\mathbf{x})]^2. \quad (3.21)$$

Cette fonction coût est minimisée par rapport à $\Delta\mathbf{p}$ et les paramètres de transformation \mathbf{p} sont actualisés à chaque itération avec les valeurs de $\Delta\mathbf{p}$ ainsi calculées (voir équation 3.22) jusqu'à la convergence de \mathbf{p} vers la solution :

$$\mathbf{p} \leftarrow \mathbf{p} + \Delta\mathbf{p}, \quad (3.22)$$

Généralement, il est admis que l'algorithme a convergé lorsque le module du vecteur $\Delta\mathbf{p}$ est en-dessous d'un seuil ϵ pré-défini, c'est-à-dire $\|\Delta\mathbf{p}\| \leq \epsilon$.

Les incréments de $\Delta\mathbf{p}$ sont obtenus par une optimisation non-linéaire du deuxième ordre (Gauss-Newton). D'abord, la mesure de similarité est linéarisée par un développement en série de Taylor du premier degré de $I_D(T(\mathbf{x}; \mathbf{p} + \Delta\mathbf{p}))$ dans $\Delta\mathbf{p}$:

$$\sum_{\mathbf{x}} \left[I_D(T(\mathbf{x}; \mathbf{p})) + \nabla I_D \frac{\partial T}{\partial \mathbf{p}} \Delta\mathbf{p} - I_R(\mathbf{x}) \right]^2, \quad (3.23)$$

avec

$$\nabla I_D = \left(\frac{\partial I_D(\mathbf{x})}{\partial x}, \frac{\partial I_D(\mathbf{x})}{\partial y} \right) \quad (3.24)$$

$$\frac{\partial T}{\partial \mathbf{p}} = \begin{pmatrix} \frac{\partial T_x}{\partial p_1} & \dots & \frac{\partial T_x}{\partial p_n} \\ \frac{\partial T_y}{\partial p_1} & \dots & \frac{\partial T_y}{\partial p_n} \end{pmatrix} \quad (3.25)$$

étant respectivement, le gradient de l'image défini entièrement par la dérivée de l'image selon \vec{x} et \vec{y} et la dérivée première des coordonnées transformées par rapport aux paramètres de transformation (le *Jacobien*). Le gradient de l'image peut facilement être estimé par des opérateurs de convolution où par des différences centrales selon chaque axe.

Puisque nous travaillons avec un modèle de transformation en perspective (8 paramètres), les coordonnées transformées sont exprimées comme suit (x' et y' sont les coordonnées homogènes et w le facteur de normalisation en coordonnées non homogènes) :

$$\begin{bmatrix} \frac{x'}{w} \\ \frac{y'}{w} \end{bmatrix} = \begin{bmatrix} \frac{p_1x + p_2y + p_3}{p_7x + p_8y + 1} \\ \frac{p_4x + p_5y + p_6}{p_7x + p_8y + 1} \end{bmatrix}. \quad (3.26)$$

Les dérivées partielles par rapport aux paramètres de transformation (voir l'équation 3.25) sont :

$$\frac{\partial T}{\partial \mathbf{p}} = \frac{1}{p_7x + p_8y + 1} \begin{bmatrix} x & y & 1 & 0 & 0 & 0 & \frac{-x(p_1x + p_2y + p_3)}{p_7x + p_8y + 1} & \frac{-y(p_1x + p_2y + p_3)}{p_7x + p_8y + 1} \\ 0 & 0 & 0 & x & y & 1 & \frac{-x(p_4x + p_5y + p_6)}{p_7x + p_8y + 1} & \frac{-y(p_4x + p_5y + p_6)}{p_7x + p_8y + 1} \end{bmatrix} \quad (3.27)$$

Avec ces équations nous avons les éléments pour minimiser l'équation 3.23. Cette minimisation est un problème de moindres carrés dont la solution peut être obtenue en calculant la dérivée première de la fonction coût (équation 3.23) selon $\Delta \mathbf{p}$:

$$2 \sum_{\mathbf{x}} \left[\nabla_{I_D} \frac{\partial T}{\partial \mathbf{p}} \right]^T \left[I_D(T(\mathbf{x}; \mathbf{p})) + \nabla_{I_D} \frac{\partial T}{\partial \mathbf{p}} \Delta \mathbf{p} - I_R(\mathbf{x}) \right]. \quad (3.28)$$

En annulant le deuxième terme de cette équation et en isolant $\Delta \mathbf{p}$ nous avons

$$\Delta \mathbf{p} = H^{-1} \sum_{\mathbf{x}} \left[\nabla_{I_D} \frac{\partial T}{\partial \mathbf{p}} \right]^T [I_R(\mathbf{x}) - I_D(T(\mathbf{x}; \mathbf{p}))] \quad (3.29)$$

avec

$$H = \sum_{\mathbf{x}} \left[\nabla_{I_D} \frac{\partial T}{\partial \mathbf{p}} \right]^T \left[\nabla_{I_D} \frac{\partial T}{\partial \mathbf{p}} \right]. \quad (3.30)$$

H est l'approximation que l'algorithme de Gauss-Newton fait du Hessien (matrice des dérivées secondes) et qui va contrôler les pas effectués par l'algorithme à chaque itération.

En résumé, l'algorithme de Lucas-Kanade applique de façon itérative les équations 3.29 et 3.22 jusqu'à ce que l'estimation des paramètres permette de converger vers la solution. Les étapes suivies par la méthode de Lucas-Kanade sont présentées dans l'algorithme 1.

Étant donné que les paramètres \mathbf{p} varient à chaque itération, le gradient de l'image ∇_{I_D} et la transformation $T(\mathbf{x}; \mathbf{p})$ doivent être ré-calculées à chaque itération.

3.5.2.2 L'algorithme de recalage par composition

Un désavantage de l'algorithme de Lucas et Kanade est lié à son efficacité en terme de temps de calcul. [Shum and Szeliski, 1997] ont proposé une alternative qui est moins coûteuse en calculs. En effet, ils proposent de minimiser par rapport à $\Delta \mathbf{p}$ la fonction coût de l'équation 3.31 plutôt que de minimiser la fonction coût de l'équation 3.20.

$$\sum_{\mathbf{x}} [I_D(T(T(\mathbf{x}; \Delta \mathbf{p}); \mathbf{p})) - I_R(\mathbf{x})]^2, \quad (3.31)$$

Algorithme 1 L'algorithme de Lucas-Kanade

Faire :

- 1 : Transformation de I_D avec $T(\mathbf{x}; \mathbf{p})$ pour calculer $I_D(T(\mathbf{x}; \mathbf{p}))$
- 2 : Calcul de l'image erreur $[I_R(\mathbf{x}) - I_D(T(\mathbf{x}; \mathbf{p}))]$
- 3 : Transformation du gradient ∇I_D avec $T(\mathbf{x}; \mathbf{p})$
- 4 : Évaluation du Jacobien $\frac{\partial T}{\partial \mathbf{p}}$ avec les paramètres $(\mathbf{x}; \mathbf{p})$
- 5 : Calcul du produit $\nabla I_D \frac{\partial T}{\partial \mathbf{p}}$
- 6 : Calcul de la matrice Hessienne en utilisant l'équation 3.30
- 7 : Calcul de $\sum_{\mathbf{x}} \left[\nabla I_D \frac{\partial T}{\partial \mathbf{p}} \right]^T [I_R(\mathbf{x}) - I_D(T(\mathbf{x}; \mathbf{p}))]$
- 8 : Calcul de $\Delta \mathbf{p}$ en utilisant l'équation 3.29
- 9 : Mise à jour des paramètres $\mathbf{p} \leftarrow \mathbf{p} + \Delta \mathbf{p}$

Tant que : $\|\Delta \mathbf{p}\| \leq \epsilon$

Au lieu d'ajouter des incréments aux paramètres de la transformation (voir équation 3.22), la transformation T est mise-à-jour à chaque itération comme suit :

$$T(\mathbf{x}; \mathbf{p}) \leftarrow T(\mathbf{x}; \mathbf{p}) \circ T(\mathbf{x}; \Delta \mathbf{p}), \quad (3.32)$$

Le mot *compositional* ou « par composition » est utilisé par [Baker and Matthews, 2002] pour illustrer le fait que l'expression dans l'équation 3.32 est composée de deux transformations. Le symbole \circ représente un produit matriciel :

$$T(\mathbf{x}; \mathbf{p}) \circ T(\mathbf{x}; \Delta \mathbf{p}) \equiv T(T(\mathbf{x}; \Delta \mathbf{p}); \mathbf{p}). \quad (3.33)$$

Cela veut dire que nous allons appliquer une transformation aux coordonnées \mathbf{x} avec les paramètres $\Delta \mathbf{p}$ pour ensuite appliquer une nouvelle transformation de paramètres \mathbf{p} aux coordonnées déjà transformées. Le résultat sera une *composition* de 2 transformations.

Pour minimiser l'équation 3.31, il faut la linéariser avec un développement en série de Taylor au premier ordre, pour aboutir à l'équation 3.34.

$$\sum_{\mathbf{x}} \left[I_D(T(T(\mathbf{x}; \mathbf{0}); \mathbf{p})) + \nabla I_D(T(\mathbf{x}; \mathbf{p})) \frac{\partial T}{\partial \mathbf{p}} \Delta \mathbf{p} - I_R(\mathbf{x}) \right]^2. \quad (3.34)$$

$T(\mathbf{x}; \mathbf{0})$ est la transformation initiale appliquée à I_D . Il existe deux différences entre l'équation 3.23 et son équivalente *par composition* (équation 3.34). La première est que le gradient de $I_D(\mathbf{x})$ (qui est transformé après avec $T(\mathbf{x}; \mathbf{p})$) est remplacé par le gradient de $I_D(T(\mathbf{x}; \mathbf{p}))$ (le gradient est calculé sur l'image déjà transformée). La deuxième est que le Jacobien $\frac{\partial T}{\partial \mathbf{p}}$ est évalué pour les paramètres $(\mathbf{x}; \mathbf{p})$ (qui changent à chaque itération) dans l'équation 3.23, mais il est évalué dans $(\mathbf{x}; \mathbf{0})$ dans l'équation 3.34. Ces différences font que le Jacobien devient constant et qu'il peut être calculé au départ et ré-utilisé à chaque itération. En considérant ces différences, il est possible de construire un algorithme de recalage plus rapide. Un résumé de l'algorithme de recalage par composition est donné dans l'algorithme 2.

3.5.2.3 Recalage par composition inverse

[Baker and Matthews, 2002] ont observé que l'étape la plus coûteuse en temps de calcul est la détermination à chaque itération de la matrice des dérivées secondes (le Hessien). Ils proposent de faire plusieurs modifications de l'algorithme de recalage par composition (tout en restant

Algorithme 2 L'algorithme de recalage par composition (Shum-Szeliski)

Calcul avant :4 : Évaluation du Jacobien $\frac{\partial T}{\partial \mathbf{p}}$ avec les paramètres $(\mathbf{x}; \mathbf{0})$ **Faire :**1 : Transformation de I_D avec $T(\mathbf{x}; \mathbf{p})$ pour calculer $I_D(T(\mathbf{x}; \mathbf{p}))$ 2 : Calcul de l'image erreur $[I_R(\mathbf{x}) - I_D(T(\mathbf{x}; \mathbf{p}))]$ 3 : Calcul du gradient $\nabla I_D(T)$ de l'image $I_D(T(\mathbf{x}; \mathbf{p}))$ 5 : Calcul du produit $\nabla I_D(T) \frac{\partial T}{\partial \mathbf{p}}$

6 : Calcul de la matrice Hessienne en utilisant l'équation 3.30

7 : Calcul de $\sum_{\mathbf{x}} \left[\nabla I_D(T) \frac{\partial T}{\partial \mathbf{p}} \right]^T [I_R(\mathbf{x}) - I_D(T(\mathbf{x}; \mathbf{p}))]$ 8 : Calcul de $\Delta \mathbf{p}$ en utilisant l'équation 3.299 : Mise à jour des paramètres par $T(\mathbf{x}; \mathbf{p}) \leftarrow T(\mathbf{x}; \mathbf{p}) \circ T(\mathbf{x}; \Delta \mathbf{p})$ **Tant que :** $\|\Delta \mathbf{p}\| \leq \epsilon$

équivalent à l'algorithme original en termes de robustesse et précision) pour avoir le Hessian constant ré-utilisable à chaque itération.

Les auteurs ont montré qu'en inversant le rôle des images de référence I_R et à transformer I_D , il est possible de diminuer les temps de calcul. Pour ce faire, ils proposent de minimiser l'expression

$$\sum_{\mathbf{x}} [I_R(T(\mathbf{x}; \Delta \mathbf{p})) - I_D(T(\mathbf{x}; \mathbf{p}))]^2, \quad (3.35)$$

par rapport à $\Delta \mathbf{p}$, pour ensuite mettre à jour *la transformation* comme suit :

$$T(\mathbf{x}; \mathbf{p}) \leftarrow T(\mathbf{x}; \mathbf{p}) \circ T(\mathbf{x}; \Delta \mathbf{p})^{-1}. \quad (3.36)$$

La seule différence avec l'algorithme par composition est que pour la mise à jour de la transformation, il faut effectuer l'inversion de la matrice de transformation $T(\mathbf{x}; \Delta \mathbf{p})$ avant d'effectuer la multiplication combinaison avec l'estimation actuelle de la transformation $T(\mathbf{x}; \mathbf{p})$.

Pour minimiser l'équation 3.35, il faut effectuer un développement en séries de Taylor du premier ordre :

$$\sum_{\mathbf{x}} \left[I_R(T(\mathbf{x}; \mathbf{0})) + \nabla I_R \frac{\partial T}{\partial \mathbf{p}} \Delta \mathbf{p} - I_D(T(\mathbf{x}; \mathbf{p})) \right]^2, \quad (3.37)$$

En considérant à nouveau que $T(\mathbf{x}; \mathbf{0})$ est une matrice identité, la solution à ce problème de moindres carrés est :

$$\Delta \mathbf{p} = H^{-1} \sum_{\mathbf{x}} \left[\nabla I_R \frac{\partial T}{\partial \mathbf{p}} \right]^T [I_D(T(\mathbf{x}; \mathbf{p})) - I_R(\mathbf{x})], \quad (3.38)$$

où H est la matrice Hessienne dans laquelle I_D est remplacée par I_R :

$$H = \sum_{\mathbf{x}} \left[\nabla I_R \frac{\partial T}{\partial \mathbf{p}} \right]^T \left[\nabla I_R \frac{\partial T}{\partial \mathbf{p}} \right]. \quad (3.39)$$

Les démonstrations de l'équivalence entre les trois algorithmes sont présentées dans [Baker and Matthews, 2002]. Nous avons vérifié de façon expérimentale que les trois algorithmes donnent des résultats très similaires en termes de précision. Nous avons pu vérifier que la méthode proposée par [Baker and Matthews, 2002] est effectivement plus rapide tout en étant robuste et précise. C'est pour cette raison que la méthode de recalage par composition inverse est la méthode que nous utilisons pour notre algorithme de mosaïquage d'images cystoscopiques.

Algorithme 3 L'algorithme de recalage par composition inverse (Baker-Matthews)

Calcul avant :

- 3 : Évaluation du gradient ∇I_R de l'image de référence I_R
- 4 : Évaluation du Jacobien $\frac{\partial T}{\partial \mathbf{p}}$ avec les paramètres $(\mathbf{x}; \mathbf{0})$
- 5 : Calcul du produit $\nabla I_R(T) \frac{\partial T}{\partial \mathbf{p}}$
- 6 : Calcul de la matrice Hessienne en utilisant l'équation 3.39

Faire :

- 1 : Transformation de I_D avec $T(\mathbf{x}; \mathbf{p})$ pour calculer $I_D(T(\mathbf{x}; \mathbf{p}))$
- 2 : Calcul de l'image erreur $[I_D(T(\mathbf{x}; \mathbf{p})) - I_R]$
- 7 : Calcul de $\sum_{\mathbf{x}} \left[\nabla I_D(T) \frac{\partial T}{\partial \mathbf{p}} \right]^T [I_D(T(\mathbf{x}; \mathbf{p})) - I_R(\mathbf{x})]$
- 8 : Calcul de $\Delta \mathbf{p}$ en utilisant l'équation 3.38
- 9 : Mise à jour des paramètres par $T(\mathbf{x}; \mathbf{p}) \leftarrow T(\mathbf{x}; \mathbf{p}) \circ T(\mathbf{x}; \Delta \mathbf{p})^{-1}$

Tant que : $\|\Delta \mathbf{p}\| \leq \epsilon$

3.5.3 Le collage des images

Le recalage des images fournit les matrices des transformations relatives T_i^r , qui donnent le lien géométrique entre la $(i - 1)^{\text{ème}}$ et la $i^{\text{ème}}$ image. Les matrices T_i^r permettent de calculer les transformations globales T_i^g qui sont utilisées pour placer les images dans le repère global de l'image panoramique. Si le repère de l'image panoramique est celui de la première image de la séquence, alors la transformation appliquée à la première image « à projeter » dans le repère global est la matrice identité. Pour actualiser la matrice T_i^g à chaque recalage d'une paire $(i - 1, i)$ d'images nous devons cumuler les matrices relatives T_i^r :

$$T_i^g = T_i^r \times T_{i-1}^g = \prod_{n=0}^{i-1} T_{i-n}^r, \quad (3.40)$$

avec $i = 1 \dots N - 1$, où N est le nombre total d'images et $T_0^g = \mathbf{I}$ (matrice identité 3 x 3). Cette opération considère que la première image de la séquence recalée est prise comme référence.

Une fois les matrices de transformations cumulées, nous pouvons donc ramener toutes les images dans le repère de l'image panoramique 2D (voir figure 3.14). Ramener les images dans ce repère revient à « projeter » les images acquises pour différents points de vue dans un plan 2D commun. Puisque, d'une part, le cystoscope ne permet pas de calculer la forme 3D des surfaces acquises et d'autre part, la position de l'endoscope n'est pas connue au cours de l'acquisition, l'image panoramique représentant les parois internes de la vessie ne peut être qu'une image 2D contenant la projection de l'organe (voir figure 3.14).

La projection des images sur une surface plane n'est pas la seule façon de présenter l'image panoramique au clinicien. L'idéal serait de présenter un modèle 3D de la vessie à l'intérieur duquel, l'image panoramique serait plaquée sur les parois en prenant comme référence des point anatomiques facilement identifiables par le clinicien. Le travail de thèse présenté ici se limite à la construction des images panoramiques projetées sur une surface plane. Cependant, des travaux pour une représentation 3D de la vessie sont effectués dans le cadre d'une autre thèse en cours au laboratoire. Cela fait aussi partie des perspectives de nos travaux (voir conclusion générale).

La projection des images dans l'image panoramique 2D introduit des effets de bord et ce même si les paires d'images sont exactement recalées. Ces bords, qui correspondent aux bords des images, sont dus aux différences d'éclairage d'une même région de la scène qui est placée à deux endroits différents dans les images consécutives. Ces effets de bord doivent être atténués au

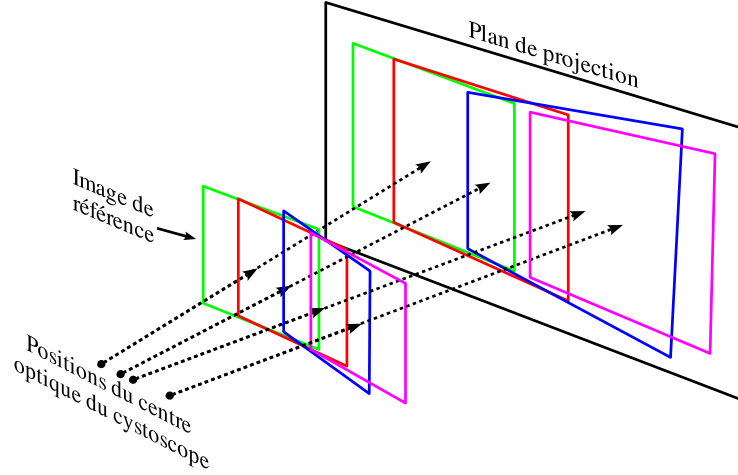


FIG. 3.14 – Illustration de la projection des images recalées sur une surface plane.

maximum pour assurer au clinicien une qualité visuelle suffisante. Pour atténuer ces effets, les niveaux de gris des pixels situés sur les bords des images sont recalculés par une pondération des niveaux de gris des pixels des régions communes de l'image à projeter et de l'image panoramique en cours de construction. Le principe de la correction des niveaux de gris des images projetées repose sur le fait qu'au centre des images, la scène est illuminée sans atténuation, alors que cette atténuation augmente continuellement en se rapprochant des bords. Plus un pixel est situé au bord de l'image à projeter, plus son niveau de gris est compensé en le moyennant fortement avec les niveaux de gris situés dans le voisinage dans l'image panoramique 2D. Ainsi, une fonction de pondération en forme de gaussienne donne plus de poids aux pixels qui se trouvent proches du centre de l'image projetée et diminue de façon exponentielle leur poids au fur et à mesure que sa position s'éloigne du centre. Les valeurs des pixels qui se trouvent déjà projetés sur le plan de projection (l'image panoramique) $I_{pano}(\mathbf{x})$, auront un poids w_{pano} complémentaire au poids w_{proj} des pixels de l'image projetée $I_{proj}(\mathbf{x})$ ($w_{pano} + w_{proj} = 1$). La valeur d'un pixel dans l'image panoramique $I_{pano}(\mathbf{x})$ est calculée comme suit :

$$I_{pano}(\mathbf{x}) = I_{pano}(\mathbf{x})w_{pano} + I_{proj}w_{proj}, \quad (3.41)$$

avec

$$w_{proj} = 0.9e^{-\frac{r}{2\sigma_{proj}^2}} + 0.1 \quad (3.42)$$

$$w_{pano} = 1 - w_{proj},$$

où w_{proj} et w_{pano} sont respectivement, le poids des valeurs des pixels de l'image I_{proj} et de l'image I_{pano} , σ_{proj} l'écart type de la fonction gaussienne et $r = \sqrt{(x - x_0)^2 + (y - y_0)^2}$ la distance radiale du pixel $\mathbf{x} = (x, y)^T$ à partir du centre de l'image $\mathbf{x}_0 = (x_0, y_0)^T$. La valeur typique de σ_{proj} utilisée dans notre approche est $\sigma_{proj} = \sqrt{(x_{max} - x_0)^2 + (y_{max} - y_0)^2}$, avec (x_{max}, y_{max}) , la position du pixel le plus éloigné du centre de l'image. La constante 0.1 de l'équation 3.42 assure un minimum de contribution des niveaux de gris de l'image projetée aux niveaux de gris de l'image panoramique.

3.6 Synthèse schématique de l'approche proposée

Les étapes (pré-traitement, recalage et collage) composant l'algorithme proposé pour obtenir des images panoramiques à partir de séquences vidéo-cystoscopiques de vessie sont résumées dans le schéma-bloc de la figure 3.15.

Du point de vue algorithmique, notre approche est divisée en deux grands blocs. Le premier bloc comprend le pré-traitement, la corrélation et le recalage des images cystoscopiques. Le second regroupe les opérations nécessaires au collage des images sur l'image panoramique.

La première étape dans la partie du pré-traitement des images est le changement du codage des couleurs des images (passage d'un espace RGB vers un espace YUV), pour ne retenir que la composante de luminance Y . Le recalage est effectué avec les niveaux de gris de l'image de luminance. L'image au final est présentée en couleur. La deuxième étape est la correction de la distorsion radiale des images avec les données de correction calculées hors ligne (en dehors de l'examen) par la méthode décrite dans la sous-section 3.2.2. La fin de cette partie est l'étape de filtrage des images corrigées en distorsion en utilisant l'approche pyramidale (cf. sous-section 3.2) pour atténuer les différences d'éclairage et le pattern de fibres optiques.

Pour obtenir une première estimation des valeurs des translations entre les images, une région de 256×256 pixels est extraite des images corrigées en distorsion et filtrées. Les transformées de Fourier de ces images sont utilisées pour en effectuer la corrélation. La position du pic maximum de la fonction de corrélation va indiquer de façon approximative la translation existant entre les images. L'opération de corrélation est décrite dans la sous-section 3.5.1.

Les valeurs de translation (x_{CC}, y_{CC}) obtenues par la corrélation sont utilisées ensuite pour initialiser l'algorithme de recalage itératif par composition inverse. Cette étape va minimiser la différence d'intensité entre des sous-régions (400×400 pixels) extraites des images filtrées en optimisant une transformation perspective (section 3.5.2). En pratique, dans l'algorithme programmé, nous avons établi deux critères d'arrêt. Le premier considère que l'algorithme a convergé lorsque les incréments des translations selon X et Y sont inférieurs à 1×10^{-4} pixels (0,0001 pixels, l'image "ne bouge plus"). Le deuxième considère qu'il n'existe plus d'amélioration des paramètres ou que l'algorithme n'a pas convergé si le nombre d'itérations dépasse 60 itérations. Le résultat de cet optimisation sera une matrice de transformation T_i^r (matrice relative) pour chaque paire d'images. Ces matrices sont cumulées pour construire les matrices de transformation globales T_i^g correspondantes. Le bloc de pré-traitement et recalage des images termine cette étape.

La première étape du bloc de collage des images découpe la région centrale des images à la taille utilisée par l'algorithme de recalage itératif (400×400 pixels). Cette région est ensuite ramenée dans le repère de l'image panoramique en utilisant la matrice de transformation globale calculée pour chaque image. Une fonction de pondération gaussienne est utilisée pour coller les images sur l'image panoramique existante (cf. sous-section 3.5.3).

Notre algorithme a été implémenté en langage C en utilisant la bibliothèque de traitement d'images $OpenCV$. Le programme résultant est entièrement automatique du point de vue de la construction d'images panoramiques. Il suffit que l'utilisateur introduise 7 paramètres : le chemin du dossier où se trouve la séquence d'images, le nom générique des images, le nombre de chiffres dans le nom séquentiel des images, le format des images (png, tif, jpg ou tif), la taille de la région à considérer dans les images, le niveau de filtrage à appliquer (1-4) pour atténuer le motif de fibres (voir section 3.2.1) et finalement, le nombre d'images à traiter.

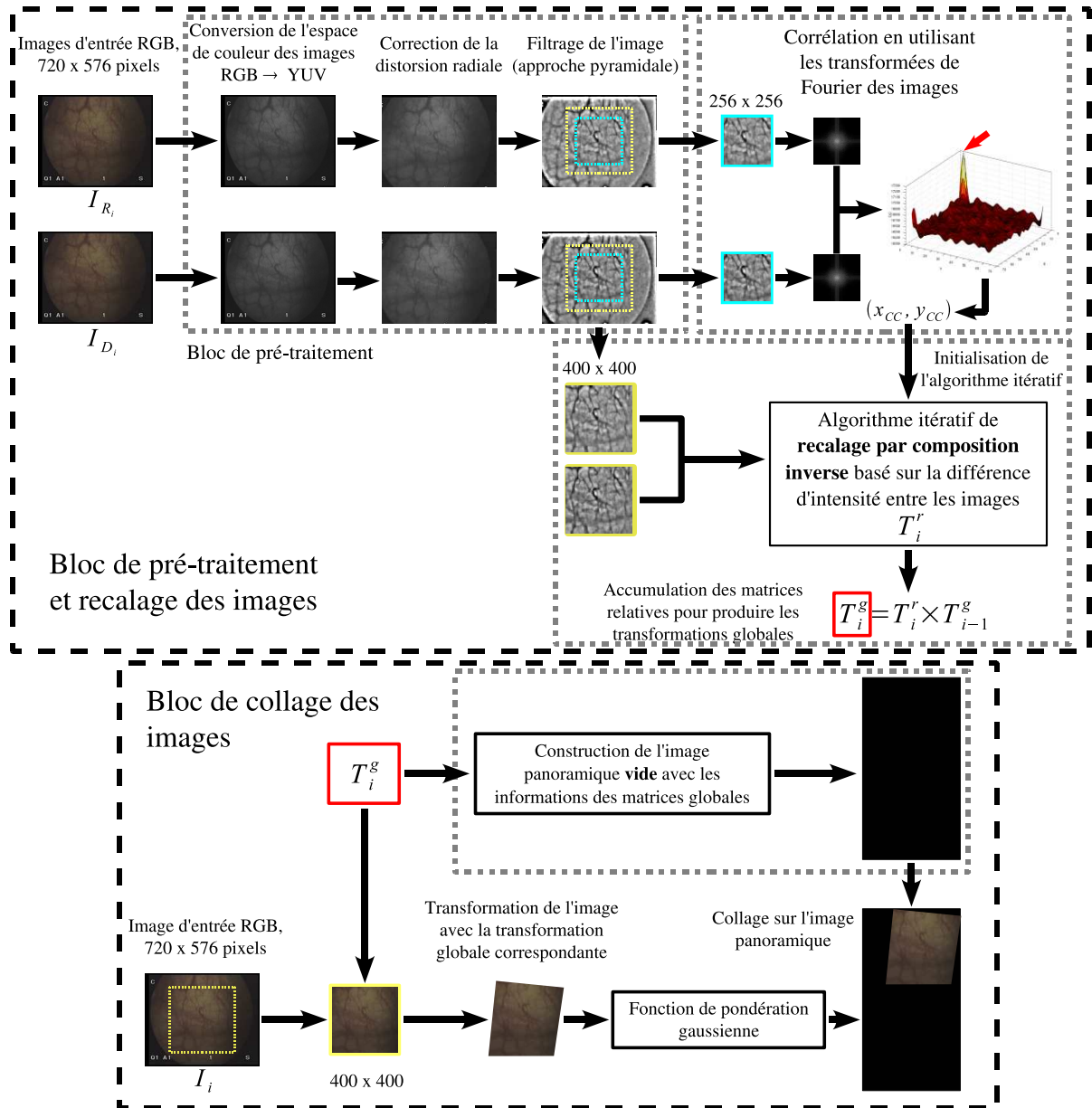


FIG. 3.15 – Schéma général de l'approche algorithmique proposée. Notre approche est composée de deux blocs principaux. Le premier bloc comprend le pré-traitement et le recalage des images. Le deuxième bloc regroupe le collage des images en utilisant les matrices de transformation globales produites par le premier bloc.

4

Validation de la solution proposée

Ce chapitre présente les matériels et la méthodologie utilisés pour valider notre solution algorithmique de construction d'images panoramiques des parois de la vessie. La première section présente la méthode suivie pour évaluer les limites et la précision de l'algorithme de recalage qui est au cœur de notre méthode de mosaïquage. Nous testons aussi dans cette section la robustesse de l'algorithme vis-à-vis du bruit (optique, électronique) simulé. La section suivante traite de l'instrumentation et des méthodes utilisées pour l'évaluation de la précision de la construction des images panoramiques. Ensuite, nous présentons les images panoramiques obtenues pour des données patient, ainsi qu'une comparaison des paramètres de translation obtenus avec notre méthode et ceux obtenus par la méthode de recalage précédemment développée au laboratoire. Finalement, nous évaluons la faisabilité de la combinaison des informations spatiales fournies par une image panoramique construite à partir d'images acquises en lumière blanche et des informations spectrales fournies par une image panoramique construite à partir d'images de fluorescence.

4.1 Évaluation de la précision du recalage

Le recalage est la partie la plus importante de la méthode de mosaïquage. S'il ne fonctionne pas correctement, l'image panoramique résultante ne sera pas cohérente par rapport à la scène réelle. Il est donc important d'évaluer sa précision et de connaître les valeurs maximales des paramètres de transformation permettant un recalage correct des images.

4.1.1 Méthode

Pour évaluer la précision du recalage nous avons utilisé trois images extraites de trois séquences cystoscopiques cliniques. Les images choisies présentent des niveaux d'illumination et des textures très différents. Pour tous les tests, nous avons extrait une région d'intérêt de taille 256×256 pixels de la partie centrale des images dont la taille d'origine est de 720×576 pixels. Les images utilisées pour nos tests ont été référencées par les chiffres romains **I**, **II** et **III** selon la figure 4.1.

Pour évaluer les valeurs maximales des paramètres de transformation permettant un recalage robuste, nous avons appliqué des transformations connues sur les images. Dans un premier temps, nous appliquons des transformations élémentaires indépendantes, à savoir des translations, des

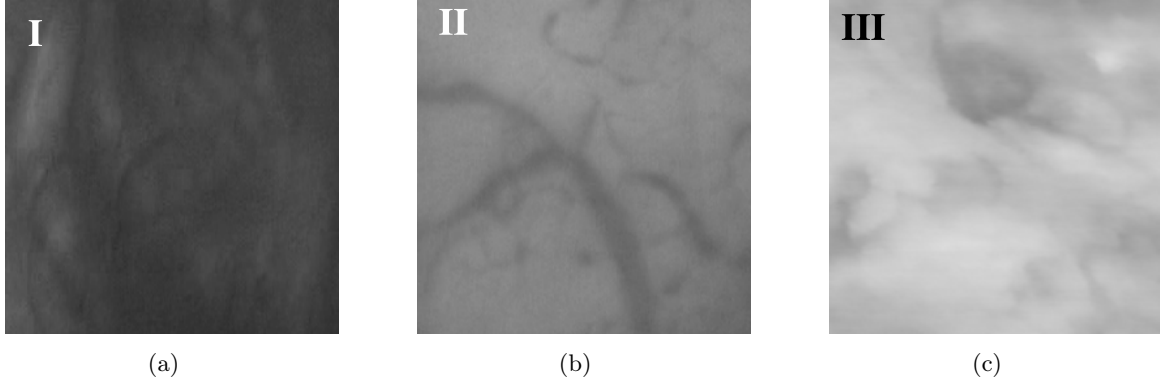


FIG. 4.1 – Images utilisées pour évaluer la précision du recalage. Ce sont des régions de taille 256×256 pixels extraites des trois images acquises pendant un examen cystoscopique clinique. La taille originale des images est de 720×576 pixels. (a) Image **I** avec une faible illumination et peu de structures visibles. (b) Image **II** bien illuminée avec structures visibles. (c) Image **III** avec peu de structures visibles.

variations du facteur d'échelle, des rotations dans le plan image et des changements du point de vue (perspective). Dans un deuxième temps, nous appliquons aux images des combinaisons de ces transformations élémentaires. Les transformations permettant de contrôler la déformation de l'image sont appelées $T_{théo}$ (transformation théorique).

Les transformations appliquées vont être estimées en recalant l'image de référence transformée par $T_{théo}$ avec l'image de référence. De façon idéale $T_{théo} = T_{calc}^{-1}$ où T_{calc} est la transformation qui superpose l'image transformée par $T_{théo}$ à l'image de référence. Cependant pour estimer la précision du recalage, il est plus simple de comparer des coordonnées calculées et théoriques que des paramètres de transformation calculés et théoriques. Pour ce faire, nous appliquons la transformation théorique $T_{théo}$ aux coordonnées minimales et maximales des images de référence, c'est-à-dire, aux coordonnées $\mathbf{x}_{théo}$ des « coins ». Ensuite, nous effectuons la transformation des coordonnées des coins ainsi transformées avec la matrice T_{calc} obtenue par le recalage :

$$\mathbf{x}_{calc}^i = T_{calc}(T_{théo}(\mathbf{x}_{théo}^i)), \quad (4.1)$$

$\mathbf{x}_{théo}^i = (x_{théo}^i, y_{théo}^i)^T$ sont les coordonnées du $i^{ème}$ coin des images de référence avant transformation et $\mathbf{x}_{calc}^i = (x_{calc}^i, y_{calc}^i)^T$ les coordonnées du $i^{ème}$ coin ré-calculées avec l'équation 4.1. Ce processus est illustré dans la figure 4.2.

Pour calculer les erreurs moyenne err_{moy} et maximale err_{max} , nous utilisons la distance euclidienne d_i de entre les coordonnées de l' $i^{ème}$ coin de l'image de référence et les coordonnées de ces coins après la transformation T_{calc} (issue du recalage) :

$$d_i = \sqrt{(x_{théo}^i - x_{calc}^i)^2 + (y_{théo}^i - y_{calc}^i)^2}, \quad (4.2)$$

avec $i \in [1, 4]$. Avec cette distance nous pouvons calculer les erreurs comme suit :

$$\begin{aligned} err_{moy} &= \frac{1}{4} \sum_i (d_i) \\ err_{max} &= \max(d_i), \end{aligned} \quad (4.3)$$

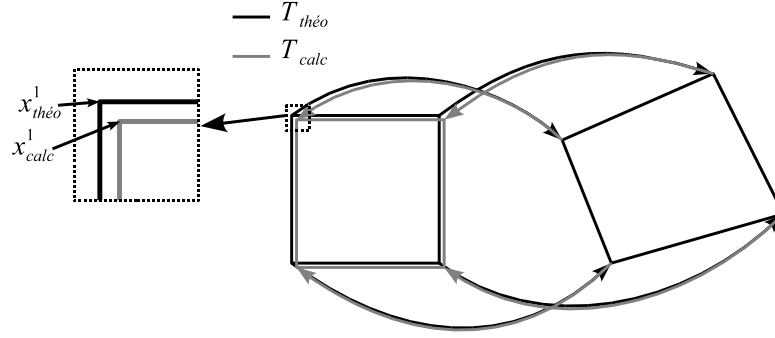


FIG. 4.2 – Illustration du processus de transformation des coordonnées des images pour évaluer la précision du recalage. $T_{théo}$ est la matrice de transformation théorique appliquée initialement à l'image de référence. T_{calc} est la matrice de transformation calculée par notre méthode de recalage. Les distances entre les coordonnées initiales (en noir) et les coordonnées ré-transformées (en gris) sont utilisées pour mesurer la précision du recalage.

$i \in [1, 4]$.

Pour établir les matrices de transformation théoriques $T_{théo}$, nous utilisons le modèle de projection perspective 3D–2D étudié dans la section 2.1.1. Nous allons donc considérer une caméra virtuelle dont la distance focale f est égale à 1 cm. Nous allons aussi considérer que les dimensions des images acquises valent $2 \text{ cm} \times 2 \text{ cm}$. Avec ce modèle, des transformations perspective peuvent être obtenues via des transformations 3D élémentaires appliquées aux images avant projection : des rotations autour de trois axes avec des angles θ , ϕ et ψ , et des translations selon les mêmes trois axes (\vec{x} , \vec{y} et \vec{z}). La figure 4.3 montre le processus de projection 3D–2D.

Les valeurs des paramètres de transformation 3D ont été modifiées jusqu'à trouver les valeurs pour lesquelles la méthode de recalage échouait. Les valeurs maximales des paramètres permettant un recalage correct sont indiquées par une étoile dans les tableaux présentés ci-après. Les paramètres h_{ij} des tableaux correspondent aux 9 paramètres qui composent une matrice de transformation perspective :

$$T_{calc} = \begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{bmatrix}.$$

La signification géométrique des paramètres h_{ij} est donnée dans la section 2.2.3.1 du chapitre 2. Nous utilisons l'algorithme de recalage exposé dans le chapitre 3, avec les étapes de pré-traitement, initialisation par corrélation et recalage itératif. Pour tous les tests, le nombre maximal d'itérations de l'algorithme de recalage a été fixé à 60. Cette valeur a été choisie parce que nous avons observé qu'après ce nombre d'itérations, il n'existe plus d'amélioration des valeurs des paramètres de transformation. Si l'algorithme de recalage n'a pas convergé avant que ce nombre soit atteint, nous allons considérer que le recalage a échoué. Nous présentons d'abord les résultats des tests dans les sections 4.1.1.1 et 4.1.1.2 pour ensuite discuter ces résultats dans la section 4.1.1.3.

4.1.1.1 Test de recalage avec des transformations élémentaires

Le premier test a consisté en l'application de translations selon les axes \vec{x} et \vec{y} (t_y). Les translations (en pixels) qui ont été appliquées aux images sont données dans la deuxième colonne du tableau 4.1. Les valeurs entre parenthèses sont respectivement, des translations selon \vec{x} (t_x) et

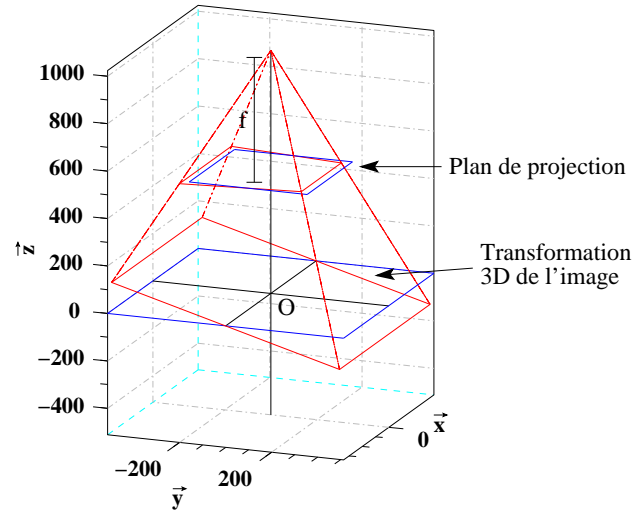


FIG. 4.3 – Modèle de projection 3D-2D que nous utilisons pour obtenir les transformations utilisées dans nos tests.

sur $\vec{y}(t_y)$. Notre algorithme de recalage prend la matrice identité comme transformation initiale. Le tableau montre que les translations maximales pour lesquelles un recalage est possible sont $(t_x, t_y) = (78, 2, -78, 7)$. Cette valeur de translation correspond à un taux de recouvrement de 48% de la taille des deux images à recaler. L'erreur maximale pour ce test pour l'image **I** lors de l'application des translations maximales, est de 1,85 pixels.

Nous avons appliqué des variations du facteur d'échelle sur les images en modifiant la translation selon l'axe \vec{z} (voir figure 4.3 et la deuxième colonne du tableau 4.2 qui fournit les résultats). Les résultats montrent que les valeurs maximales de facteur d'échelle (1,25 et 0,8) autorisant un recalage ont été obtenues pour l'image **II**. Cependant, l'erreur maximale (err_{max} dans l'équation 4.3) pour cette image et pour ce facteur d'échelle est la plus grande (2,86 pixels). Les facteurs d'échelle maximum pour l'image **I**, 2,14 et 0,97, sont les plus faibles des trois images.

Le test suivant a consisté en l'application de rotations. Pour ce test, la valeur de rotation maximum (14°) qui autorise un recalage a été déterminé pour l'image **III**. La valeur minimum de rotation a été calculée avec l'image **II**. L'erreur maximale $err_{max} = 1,045$ pixels est obtenue pour l'image **I**.

La dernière transformation testée a été la perspective. Pour la simuler, il suffit d'appliquer aux images des rotations θ et/ou ϕ selon les axes \vec{x} et \vec{y} respectivement. Pour ce test, nous avons choisi d'effectuer des rotations selon \vec{x} . Le tableau 4.4 donne les résultats pour ce test. La valeur maximum de rotation (26°) a été trouvée pour les images **I** et **II**. La valeur maximale la plus petite (20°) a été trouvée pour l'image **III**. L'image **II** présente l'erreur maximum ($err_{max} = 3,83pixels$) pour ce test.

4.1.1.2 Test de recalage avec des transformations élémentaires combinées

Pour construire les transformations à appliquer aux images **I**, **II** et **III**, nous utilisons également le modèle de projection 3D-2D de la figure 4.3. Les transformations 3D sont contrôlées par 6 paramètres : 3 rotations d'angles θ , ϕ et ψ et 3 translations selon les axes \vec{x} , \vec{y} et \vec{z} (cf. figure 4.3).

Pour ce test, nous procédons de la même façon que pour les tests précédents. Nous avons

TAB. 4.1 – Valeurs des paramètres de T_{calc} obtenues pour les tests de recalage des images montrées dans la figure 4.1. Différentes valeurs de translation selon \vec{x} et \vec{y} ont été appliqués aux images pour construire l'image à recaler avec l'image sans transformation. Les valeurs de la translation, en pixels, sont données par les paramètres h_{13} et h_{23} . Le tableau donne aussi le nombre d'itérations effectuées par l'algorithme qui minimise la différence d'intensité entre les images pour arriver à la solution.

Image	Transl (pixels)	h_{11}	h_{12}	h_{13}	h_{21}	h_{22}	h_{23}	$h_{31} \times 10^6$	$h_{32} \times 10^6$	h_{33}	err_{moy} (pixels)	err_{max} (pixels)	Itér.	Recouv. (%)
I	(78,2, -78,7)*	1,017	3,80e-4	78,055	5,28e-3	1,006	-79,370	55,122	0,617	0,999	0,99	1,85	10	48,1
	(-42,8, 31,6)	1,000	-4,36e-4	-42,775	7,73e-4	0,999	31,568	2,888	-2,990	1,000	0,08	0,16	4	73,00
	(-1,1, 2,6)	1,001	-1,49e-4	-1,076	2,52e-4	1,000	2,558	0,864	-0,641	0,999	0,03	0,05	4	98,56
II	(78,2, -78,7)*	1,010	1,45e-3	77,941	3,52e-3	1,004	-79,099	28,525	4,001	0,999	0,50	1,07	8	48,1
	(-42,8, 31,6)	1,000	-4,24e-5	-42,806	-2,28e-4	1,000	31,594	-0,827	-0,043	1,000	0,04	0,07	4	73,00
	(-1,1, 2,6)	1,000	-5,20e-5	-1,111	4,16e-4	1,000	2,566	1,700	-0,366	0,999	0,04	0,07	4	98,56
III	(78,2, -78,7)*	0,993	-0,011	78,982	-5,47e-4	0,988	-77,819	-4,418	-3,472	0,999	0,10	1,43	7	48,1
	(-42,8, 31,6)	1,000	2,68e-4	-42,854	3,437	1,001	31,558	-0,041	1,829	1,000	0,05	0,07	5	73,00
	(-1,1, 2,6)	0,999	1,65e-4	-1,105	1,888	1,000	2,592	-0,410	0,342	1,000	0,02	0,04	4	98,56

TAB. 4.2 – Résultats du recalage de paires d'images reliées géométriquement par des facteurs d'échelle (voir deuxième colonne). Les valeurs maximales sont indiquées par une étoile. La variation du facteur d'échelle est visible sur les paramètres h_{11} et h_{22} . La dernière colonne fournit le nombre d'itérations qui ont été nécessaires pour que l'algorithme de recalage trouve la solution.

Image	Facteur d'échelle	h_{11}	h_{12}	h_{13}	h_{21}	h_{22}	h_{23}	$h_{31} \times 10^6$	$h_{32} \times 10^6$	h_{33}	err_{moy} (pixels)	err_{max} (pixels)	Itér.
I	1,14*	1,134	-1,98e-3	-0,211	-9,24e-4	1,132	-0,185	-4,064	-8,096	0,997	0,70	0,93	37
	1,1	1,099	-1,34e-3	-0,045	1,39e-5	1,097	-0,077	1,796	-4,538	0,999	0,33	0,47	10
	0,9	0,902	2,82e-4	0,129	1,59e-3	0,901	-0,071	5,524	1,632	0,999	0,38	0,50	14
	0,87*	0,869	5,64e-4	-0,083	1,24e-3	0,869	-0,216	3,622	1,428	0,997	0,21	0,30	36
II	1,25*	1,220	2,40e-3	-2,423	-5,66e-4	1,222	-2,235	-1,638	5,828	0,977	2,72	2,86	26
	1,1	1,099	1,36e-3	-0,452	1,01e-4	1,099	-0,335	0,822	4,265	0,998	0,42	0,51	8
	0,9	0,898	-1,20e-3	-0,025	-5,48e-4	0,896	-0,137	-2,251	-4,092	0,996	0,26	0,38	14
	0,8*	0,783	-1,20e-3	-1,467	-1,64e-5	0,782	-1,556	0,221	-1,999	0,977	2,60	2,83	38
III	1,2*	1,187	-2,03e-3	-1,107	-8,31e-5	1,185	-1,190	-0,777	-5,489	0,990	1,53	1,81	14
	1,1	1,099	-1,16e-3	-0,232	2,57e-4	1,098	-0,270	-0,231	-3,791	0,999	0,39	0,60	7
	0,9	0,899	1,65e-3	-0,079	6,42e-4	0,899	-0,023	2,350	5,061	0,998	0,23	0,42	10
	0,87*	0,868	1,23e-3	-0,101	1,14e-3	0,868	-0,104	5,286	4,409	0,996	0,20	0,28	14

TAB. 4.3 – Paramètres de transformation calculés par le recalage d’images après application des valeurs de rotation données dans la deuxième colonne. Les paramètres h_{12} et h_{21} contrôlent une rotation pure selon l’axe perpendiculaire à l’image (axe \vec{z} de la figure 4.3).

<i>Image</i>	<i>Angle</i> (°)	h_{11}	h_{12}	h_{13}	h_{21}	h_{22}	h_{23}	$h_{31} \times 10^6$	$h_{32} \times 10^6$	h_{33}	<i>err</i> _{moy} (pixels)	<i>err</i> _{max} (pixels)	<i>Itér.</i>
I	11*	0,974	-0,190	-0,095	0,190	0,973	-0,987	2,024	-3,986	0,992	0,99	1,05	25
	8	0,990	-0,140	0,315	0,140	0,988	-0,424	3,531	-3,103	0,999	0,40	0,53	12
	5	0,997	-0,088	0,232	0,088	0,996	-0,194	3,063	-1,634	1,000	0,25	0,30	7
	2	1,000	-0,035	0,083	0,035	0,999	-0,133	2,560	1,797	1,000	0,11	0,16	5
II	10*	0,984	-0,174	0,431	0,173	0,984	-0,272	-1,370	-1,992	1,000	0,48	0,51	19
	8	0,991	-0,139	0,323	0,140	0,991	-0,255	1,378	0,127	1,001	0,41	0,43	10
	5	0,997	-0,087	0,134	0,088	0,997	-0,192	1,311	0,854	1,000	0,25	0,28	6
	2	0,999	-0,035	0,087	0,035	0,999	-0,052	1,891	-1,940	1,000	0,09	0,12	5
III	14*	0,968	-0,242	0,266	0,241	0,970	-0,687	-2,410	5,389	0,998	0,78	0,89	28
	8	0,989	-0,138	0,105	0,139	0,990	-0,423	-1,391	3,626	0,999	0,46	0,52	12
	5	0,996	-0,087	0,141	0,087	0,996	-0,192	-0,930	-0,160	1,000	0,25	0,27	8
	2	0,999	-0,035	0,075	0,035	0,999	-0,085	-0,647	1,093	1,000	0,12	0,18	7

TAB. 4.4 – Paramètres de transformation calculés pour le test impliquant des variations du point de vue (perspective). Les angles testés sont donnés dans la deuxième colonne. Les variations de perspective sont contrôlées principalement par les paramètres h_{31} et h_{32} , cependant les paramètres h_{11} et h_{22} sont aussi modifiés par les changements de perspective.

<i>Image</i>	<i>Angle</i> <i>persp.</i> (°)	h_{11}	h_{12}	h_{13}	h_{21}	h_{22}	h_{23}	$h_{31} \times 10^6$	$h_{32} \times 10^6$	h_{33}	<i>err</i> _{moy} (pixels)	<i>err</i> _{max} (pixels)	<i>Itér.</i>
I	26*	0,991	0,069	-0,126	-3,93e-3	0,962	0,402	-25,944	8,505	0,997	0,42	0,63	50
	20	0,997	0,054	0,083	-2,23e-3	0,993	0,387	-14,697	6,677	1,000	0,45	0,73	41
	10	0,999	0,028	0,016	-1,08e-3	1,012	0,131	-6,780	3,401	1,000	0,25	0,40	13
	5	0,999	0,014	0,026	-8,10e-4	1,010	0,074	-5,296	1,726	1,000	0,14	0,19	8
II	26*	0,983	0,066	-0,825	-2,38e-3	0,953	-0,932	-15,772	8,364	0,986	1,31	1,84	51
	20	0,991	0,053	-0,822	-9,06e-4	0,986	-0,752	-4,005	6,621	0,990	0,78	1,10	33
	10	0,997	0,027	-0,065	-1,43e-3	1,010	-0,072	-6,169	3,360	0,998	0,22	0,43	13
	5	0,998	0,013	0,111	-1,40e-3	1,009	0,051	-5,798	1,678	0,999	0,15	0,30	8
III	20*	0,992	0,054	-0,355	-1,35e-3	0,989	-0,372	-7,207	6,634	0,994	0,50	0,91	44
	10	0,998	0,027	0,084	6,35e-5	1,011	-0,020	-0,605	3,355	0,999	0,18	0,28	11
	5	1,000	0,014	0,0259	-3,33e-5	1,011	-4,43e-4	0,212	1,691	1,000	0,13	0,25	7

effectué ce test en utilisant l'image pour laquelle l'algorithme a eu les moins bons résultats (image **I**) lors des tests précédents de la section 4.1.1.1. Nous appliquons les transformations ainsi calculées aux images. Ensuite, nous effectuons le recalage de l'image transformée avec l'image de référence. Les erreurs dues aux transformations estimées par notre méthode de recalage sont évaluées par la méthode décrite dans la sous-section 4.1.1 (calcul des erreurs moyennes err_{moy} et maximales err_{max} sur la position des coins de l'image recalée, équation 4.3). Le nombre maximum d'itérations de l'optimisation est également fixé à 60.

Nous avons simulé différentes transformations perspectives en fixant des limites sur chacune des transformations indépendantes :

$$\begin{aligned} -10^\circ &\leq \theta &\leq 10^\circ \\ -10^\circ &\leq \phi &\leq 10^\circ \\ -10^\circ &\leq \psi &\leq 10^\circ \\ -50 \text{ pixels} &\leq t_x &\leq 50 \text{ pixels} \\ -50 \text{ pixels} &\leq t_y &\leq 50 \text{ pixels} \\ 0,9 &\leq \text{f. e.} &\leq 1,1 \end{aligned}$$

avec f. e. le facteur d'échelle. Les transformations indépendantes composant les transformations perspective d'un examen cystoscopique réel se trouvent la plupart du temps dans les intervalles choisis. Le facteur d'échelle est à nouveau obtenu par des translations selon l'axe \vec{z} (figure 4.3). Dans un souci de clarté, nous présentons le facteur d'échelle (sans unité) plutôt que sa translation t_z en pixels. Les variations du point de vue (perspective) sont simulées par des rotations θ et ψ autour des axes \vec{x} et \vec{y} .

Puisqu'il est impossible de tester toutes les combinaisons des transformations élémentaires, nous avons défini 12 combinaisons représentatives des mouvements effectués pendant un examen cystoscopique. Plusieurs de ces combinaisons conduisent même à des transformations plus importantes que celles rencontrées lors d'un examen clinique standard. Les transformations combinées utilisées sont données dans le tableau 4.5.

Les différentes combinaisons des transformations élémentaires sont référencées par des numéros allant de 1 à 12. Les résultats du recalage des images par notre méthode sont présentés dans le tableau 4.6.

Les résultats de ce test montrent que notre méthode de recalage a échoué pour les tests 3 et 9, atteignant le nombre maximum d'itérations fixé. Pour le test 3, une fois que nous avons réduit les valeurs des paramètres de translation sur \vec{x} et \vec{y} d'un facteur 3 approximativement (voir test 4), le recalage était possible. Pour le test 9, en faisant passer la valeur du facteur d'échelle de 1,09 à 0,97 (voir test 10), le recalage était à nouveau possible. Cela veut dire que la combinaison de paramètres de transformation près des limites des intervalles (déterminés avec le test des transformations élémentaires) fait échouer la méthode de recalage, particulièrement si l'angle ψ (rotation autour de \vec{z}) est grand (environ 9°). Le recalage est possible même si les autres paramètres sont aux limites des intervalles de transformation élémentaires (test 1 et 7). La rotation dans le plan semble donc une des contraintes limitant la précision de notre méthode de recalage.

4.1.1.3 Discussion sur les limites et la précision de recalage

Les tests impliquant les transformations élémentaires, nous permettent d'établir les limites des transformations qui autorisent le recalage des images. Un recalage d'images ayant en commun au moins 55% de leur surface est possible si la transformation est constituée d'une translation 2D

TAB. 4.5 – Transformations combinées appliquées à l’image **I** pour tester la précision du recalage. Les tests ont été numérotés de 1 à 12. θ , ϕ et ψ sont respectivement des rotations en degrés autour des axes \vec{x} , \vec{y} et \vec{z} . *f.e.* est le facteur d’échelle.

<i>Test</i>	θ (°)	ϕ (°)	ψ (°)	t_x (pixels)	t_y (pixels)	<i>f.e.</i>
1	9	-8	6	-45,0	49,7	1,1
2	-7	4	7	-20,8	25,3	0,9
3	-5	-9	10	43,8	33,4	1,05
4	-5	-9	10	-17,5	10,2	1,05
5	10	10	-6	-12,2	-9,95	0,98
6	4	-5	-2	49,2	5,7	0,95
7	2	-10	3	21,3	40,8	1,08
8	10	-10	10	10,5	3,6	1,01
9	-9	9	-10	4,5	7,1	1,09
10	-9	9	-10	4,5	7,1	0,97
11	3	-4	8	11,4	-12,3	1,02
12	6	3	1,5	-45,1	33,9	1,04

pure. Évidemment, l’erreur maximale err_{max} pour cette amplitude de translation est importante (environ 2 pixels), même si l’erreur moyenne err_{moy} reste inférieure à 1 pixel. Pour les tests impliquant des variations du facteur d’échelle et de la rotation autour de l’axe \vec{z} , la performance de l’algorithme est plus « modeste ». Pour le facteur d’échelle, seules des variations entre 87% et 114% de la surface d’origine assurent une erreur maximale de recalage inférieure à 1 pixel. Pour la rotation dans le plan image, des angles en-dessous de 10° permettent un recalage avec une erreur maximale inférieure à 1 pixel pour les trois images. Du côté des variations du point de vue, des erreurs inférieures à 1 pixel pour les trois images sont obtenues lorsque des rotations avec des angles entre 0° et 10° sont appliquées. Ces tests montrent qu’il existe une corrélation entre l’amplitude de la transformation appliquée et l’amplitude de l’erreur.

Les résultats des tests pour des combinaisons de transformations élémentaires, nous ont appris que l’impact de la rotation autour de \vec{z} sur la performance de l’algorithme est effectivement importante. Si la rotation entre les images est d’environ 10° et les autres transformations sont importantes, le recalage échoue. Si cette rotation est petite ($< 3^\circ$), il est fort probable que l’algorithme de recalage réussisse même si les autres transformations sont importantes. Pour des variations du facteur d’échelle entre 0,9 et 1,1 l’algorithme arrive à recalage les images. Mais, en dehors de cet intervalle, le succès et la précision du recalage est très sensible vis-à-vis des conditions d’illumination et de la « visibilité » des structures (textures) dans les images. D’ailleurs, nous avons observé que pour des taux de recouvrement des images d’environ 70% et pour des transformations dont la rotation dans le plan image est petite, l’algorithme parvient à la solution rapidement et avec une erreur maximale petite.

Nous pouvons conclure que pour des taux de recouvrement entre les images à recalage supérieurs ou égales 75% de la taille de l’image et pour des rotations dans le plan image inférieures à 8° , un recalage correct avec une erreur maximale inférieure à 1 pixel est systématiquement possible. Cette erreur est acceptable pour la taille des images que nous considérons, puisqu’elle représente à peine 0,4% de la taille de l’image. Les intervalles des paramètres de transformation que nous avons utilisés sont réalistes dans le cadre du recalage d’images de séquences vidéo-cystoscopiques.

TAB. 4.6 – Paramètres de transformation obtenus pour les combinaisons des transformations élémentaires définies dans le tableau 4.5. Le numéro du test est donné par la première colonne. Les tests pour lesquels le recalage a échoué sont marqués par un symbole #. Le nombre d'itérations effectuées par l'algorithme de recalage pour trouver la solution est donné dans la dernière colonne.

Test	h_{11}	h_{12}	h_{13}	h_{21}	h_{22}	h_{23}	$h_{31} \times 10^4$	$h_{32} \times 10^4$	h_{33}	err_{moy} (pixels)	err_{max} (pixels)	Itér.
1	1,070	0,144	-50,506	-0,158	1,114	55,706	-2,571	3,642	1,007	0,48	0,58	9
2	0,888	0,092	-18,743	-0,122	0,871	23,118	-80,346	-2,275	1,002	1,00	1,41	11
3#	1,119	-0,059	9,4357	0,154	0,868	61,998	4,144	-1,260	1,001	91,70	124,05	60#
4	1,041	0,168	-18,962	-0,140	1,016	10,605	3,394	-1,240	0,994	0,47	0,55	21
5	0,935	-0,054	-11,834	0,090	0,985	-10,038	-2,975	3,607	0,998	0,40	0,49	18
6	0,962	-0,024	46,748	0,041	0,960	5,0565	1,638	1,225	0,995	0,45	0,59	23
7	1,088	0,061	22,943	-0,035	1,082	44,163	3,5512	80,574	0,999	1,30	1,50	9
8	1,015	0,208	11,724	-0,185	1,021	5,519	2,672	4,005	1,015	1,62	2,08	31
9#	1,136	-0,267	7,4431	0,115	0,918	13,729	6,563	-8,378	1,008	36,90	68,26	60#
10	0,904	-0,183	3,4774	0,116	0,913	5,270	-3,237	-2,387	0,985	1,54	1,86	23
11	1,016	0,151	11,444	-0,136	1,017	-12,149	1,159	1,194	1,000	0,36	0,43	12
12	1,031	0,044	-47,004	-0,030	1,051	35,297	-1,078	2,073	1,000	1,30	1,34	6

La vitesse d'acquisition (25 images/seconde) fait que les transformations existant entre des images successives de la séquence sont plus petites que les intervalles que nous considérons. Ceci est particulièrement vrai pour la rotation dans le plan image qui est le paramètre auquel notre algorithme est le plus sensible.

Notre méthode de recalage a une performance accrue lorsque la transformation prédominante entre les images à recaler est la translation. Nous avons montré dans la sous-section 3.5.1 que la translation est la transformation qui prédomine dans les séquences vidéo-cystoscopiques. Donc, notre méthode de recalage est bien adaptée à de telles séquences.

4.2 Évaluation de la robustesse du recalage

4.2.1 Méthode

Les tests présentés dans la section précédente ont été choisis pour tester la précision du recalage et les limites de l'algorithme sans tenir compte du bruit. Pour tester la robustesse de notre méthode de recalage vis-à-vis du bruit, nous avons conduit le test suivant. Nous avons répété l'expérience décrite dans la sous-section 4.1.1.2, pour les 10 tests réussis, mais cette fois, les images de référence et à transformer ont été contaminées par un bruit blanc et gaussien de moyenne nulle simulant par exemple un capteur CCD intensifié. La variance σ_b^2 du bruit qui est appliqué aux images est calculée par rapport à la variance des niveaux de gris de l'image « considérée sans bruit » σ_o^2 , avec le ratio :

$$p_b = \frac{\sigma_b^2}{\sigma_o^2}, \quad (4.4)$$

où p_b est le pourcentage de bruit (par rapport à σ_o^2) à appliquer sur l'image originale. Le bruit gaussien $G_{\sigma_b^2}$ de variance σ_b^2 est ajouté à l'image originale $I_o(\mathbf{x})$ pour créer une image bruitée

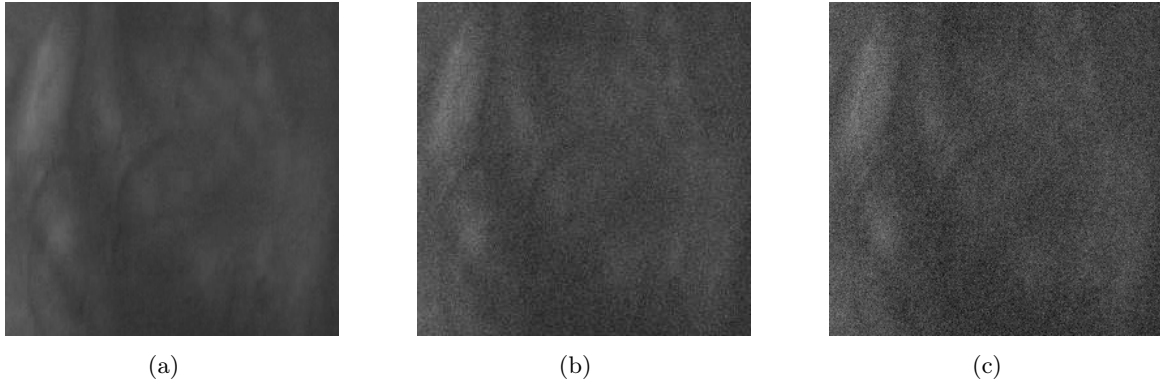


FIG. 4.4 – Images utilisées pour évaluer la robustesse du recalage vis-à-vis du bruit gaussien ajouté aux images. Les images (b) et (c) sont des versions bruitées de l’image (a) avec des variances qui représentent respectivement 10% et 20% de la variance des niveaux d’intensité de l’image originale.

$I_b(\mathbf{x})$ avec laquelle nous allons effectuer les tests :

$$I_b(\mathbf{x}) = G_{\sigma_b^2}(\mathbf{x}) + I_o(\mathbf{x}). \quad (4.5)$$

En réalité, l’image originale $I_o(\mathbf{x})$ étant une image cystoscopique réelle, celle-ci n’est pas dénuée de bruit. Ce dernier est un bruit de référence faible voire négligeable auquel est ajouté un bruit gaussien supplémentaire.

Pour ces tests, nous continuons à utiliser l’image \mathbf{I} donnée dans la figure 4.1. De cette façon, les erreurs obtenues pour cette expérience peuvent être comparées à celles obtenues pour le test décrit dans la sous-section 4.1.1.2. Pour le premier test, nous avons ajouté un bruit dont la variance représente 10% de σ_o^2 ($p_b = 0, 1$) de celle de l’image originale. Pour le deuxième test, nous appliquons un bruit dont la variance vaut 20% de σ_o^2 ($p_b = 0, 2$). L’aspect visuel des images pour les deux niveaux de bruit est présenté dans la figure 4.4.

Les résultats du recalage en utilisant les images bruitées avant et après transformation avec les combinaisons de paramètres du tableau 4.5 sont présentés ci-après (tableau 4.7 pour $p_b = 0, 1$ et tableau 4.8 pour $p_b = 0, 2$).

D’après les résultats des tableaux et par rapport aux résultats obtenus pour l’image sans bruit (tableau 4.6), même si tous les tests ont été réussis, le bruit affecte effectivement les erreurs moyenne err_{moy} et maximales err_{max} de recalage. Si nous considérons le test avec l’image bruitée à 20% de σ_o^2 , nous pouvons constater qu’il existe dans tous les cas, une augmentation du nombre d’itérations nécessaires à la convergence de l’algorithme vers la solution. Les augmentations du nombre d’itérations peuvent atteindre 400% (voir le test 12). Dans tous les cas, il existe aussi une augmentation des erreurs moyenne et maximale, sauf pour le test 8 où l’erreur est légèrement inférieure (environ -0,01 pixel). Les erreurs maximales pour ce test dépassent 2 pixels. L’augmentation de l’erreur la plus importante (passage de 0,47 à 1,43 pixels) est pour le test 4.

Pour les deux tests effectués dans cette section, dans la plupart des expériences, l’erreur augmente avec la première image bruitée (10% de σ_o^2) et avec l’image bruitée à 20% de σ_o^2 . Cependant, il existe une légère diminution de l’erreur (-0,1 pixel) pour le test 1. Pour le test 8, dans les deux tests (pour $p_b = 0, 1$ et $p_b = 0, 2$) le nombre maximum d’itérations a été atteint et nous considérons donc que le recalage a échoué. Ici, les erreurs obtenues pour le test avec l’image bruitée à 10% de σ_o^2 sont environ 2,5 fois plus grandes que pour le test avec l’image bruitée à 20% de σ_o^2 . C’est le même test pour lequel l’erreur a diminuée de l’image sans bruit à l’image bruitée

TAB. 4.7 – Résultat pour le recalage des images bruitées avec une variance de 10% de la variance des niveaux d'intensité des pixels de l'image originale. La première colonne montre le numéro du test (cf. tableau 4.5).

<i>Test</i>	h_{11}	h_{12}	h_{13}	h_{21}	h_{22}	h_{23}	$h_{31} \times 10^4$	$h_{32} \times 10^4$	h_{33}	<i>err</i> _{moy} (pixels)	<i>err</i> _{max} (pixels)	<i>Itér.</i>
1	1,065	0,145	-50,382	-0,159	1,119	55,058	-2,723	3,719	1,006	0,70	1,59	16
2	0,886	0,093	-18,529	-0,122	0,876	22,771	-95,777	-2,244	1,004	1,37	2,20	19
4	1,038	0,170	-18,800	-0,139	1,020	10,098	3,254	-1,212	0,995	0,66	1,60	37
5	0,931	-0,054	-11,526	0,089	0,991	-10,474	-3,107	3,662	0,999	0,86	0,90	28
6	0,959	-0,023	47,055	0,042	0,966	4,558	1,550	1,271	0,996	0,72	1,00	36
7	1,084	0,063	23,105	-0,035	1,087	43,541	3,429	90,874	0,999	1,31	1,78	19
8	1,025	0,212	12,988	-0,186	1,039	5,975	2,600	4,099	1,028	3,22	4,58	60
10	0,908	-0,184	4,451	0,117	0,927	5,477	-3,399	-2,323	0,993	0,78	1,21	38
11	1,013	0,153	11,653	-0,136	1,022	-12,634	1,042	1,243	1,002	0,64	1,55	23
12	1,026	0,045	-46,813	-0,030	1,056	34,762	-1,231	2,130	1,001	1,30	1,72	13

TAB. 4.8 – Paramètres de transformation calculés pour des images bruitées avec une variance à 20% de la variance des niveaux d'intensité des pixels de l'image originale. La première colonne montre le numéro du test (cf. tableau 4.5).

<i>Test</i>	h_{11}	h_{12}	h_{13}	h_{21}	h_{22}	h_{23}	$h_{31} \times 10^4$	$h_{32} \times 10^4$	h_{33}	<i>err</i> _{moy} (pixels)	<i>err</i> _{max} (pixels)	<i>Itér.</i>
1	1,067	0,151	-50,945	-0,163	1,126	55,446	-2,806	4,015	1,005	1,04	1,49	28
2	0,886	0,099	-19,132	-0,126	0,882	22,960	-1,052	-1,910	1,001	1,43	2,44	36
4	1,046	0,178	-19,068	-0,144	1,033	10,984	3,201	-86,804	0,999	1,40	1,98	43
5	0,933	-0,048	-12,101	0,085	0,997	-10,081	-3,222	3,963	0,999	1,15	1,90	44
6	0,959	-0,015	46,446	0,036	0,971	5,080	1,409	1,593	0,995	1,06	1,63	54
7	1,087	0,071	22,544	-0,040	1,097	44,087	3,344	1,257	0,999	1,16	2,07	36
8	1,004	0,215	10,825	-0,189	1,024	5,083	2,354	4,366	1,009	1,59	2,08	60
10	0,912	-0,179	4,184	0,114	0,937	5,954	-3,508	-2,080	0,996	1,10	2,14	54
11	1,016	0,161	11,357	-0,141	1,031	-11,976	90,485	1,600	1,004	1,28	1,73	40
12	1,029	0,051	-47,399	-0,035	1,065	35,175	-1,318	2,460	1,000	1,28	1,99	25

à 20% de σ_0^2 . Une explication possible est que des petites erreurs provoquées par l'addition du bruit soient de signe contraire à l'erreur globale de recalage, « améliorant » d'une certaine façon la position calculée.

4.2.2 Discussion et conclusion

D'après les tests de robustesse du recalage vis-à-vis du bruit gaussien, nous pouvons conclure que notre algorithme de recalage est robuste pour les combinaisons de transformations élémentaires testées et situées dans les intervalles définis dans la section 4.1.1.2. Ajouter un bruit dont la variance est égale à 20%, voire à 10% de la variance des niveaux de gris des images qui sont déjà bruitées revient à se placer dans des conditions de bruit extrêmes. Cependant, même en considérant les images bruitées à 20%, l'erreur de recalage obtenue reste acceptable, car l'erreur maximale obtenue pour cette image (2,44 pixels), ne représente que 0,95% de la taille de l'image. Nous pouvons ainsi affirmer que notre méthode est robuste vis-à-vis du bruit et peut être utilisée de façon fiable dans une méthode de mosaïquage d'images de l'intérieur de la vessie. Le temps de calcul du recalage varie suivant l'amplitude des paramètres de transformation appliqués et le degré de bruit présent dans les images. Cependant, pour 60 itérations, il ne dépasse pas 2,6 secondes pour recalculer une paire d'images, soit environ 0,6 secondes pour 10 itérations. Nous pouvons conclure que notre méthode de recalage est rapide.

4.3 Évaluation de la qualité du mosaïquage

Pour évaluer la précision du mosaïquage nous simulons de façon réaliste les conditions d'acquisition réelles tout en contrôlant (avec une instrumentation adaptée) les mouvements à effectuer pendant l'acquisition. Pour ce faire, nous allons établir un protocole d'acquisition qui reproduit tous les mouvements possibles d'un cystoscope à l'intérieur de la vessie.

4.3.1 Instrumentation

Pour simuler les mouvements d'un cystoscope, nous utilisons système de positionnement qui effectue des translations et de changements d'échelle grâce à des platines de déplacement *Newport UMR8.51* couplées à des moteurs de déplacement linéaires *Newport LTA-HS*. Les translations peuvent être réalisées avec une précision de 0,1 μm . La rotation perpendiculaire à l'axe d'acquisition (rotation dans le plan image) est effectuée avec une platine rotative *Newport PR50CC* qui a une précision de 0,01°. Pour effectuer le changement de point de vue, nous utilisons une platine rotative *Newport URS75CC* qui a aussi une précision de 0,01°. Le contrôle de ces platines est réalisé à travers un contrôleur *Newport XPS* en utilisant une interface programmée en *LabView*. En raison de l'encombrement et du poids du système d'acquisition, les mouvements du cystoscope ont été simulés par des mouvements du fantôme pour lequel nous allons faire les acquisitions. Le fantôme est une photographie d'une vessie de porc qui a été incisée, ouverte et « mise à plat ». Une vessie de porc a, d'un point de vue de la texture, une forte ressemblance avec une vessie. Une illustration de la configuration et une photographie du système de positionnement sont présentées dans la figure 4.5.

Le système d'acquisition d'images est celui décrit dans la section 1.3.5.2. Ce système comporte 3 caméras monochromes *PCO Pixelfly*. La caméra dédiée à l'acquisition vidéo sans filtre est la caméra que nous utilisons pour réaliser les acquisitions des séquences utilisées pour nos tests dans cette section.

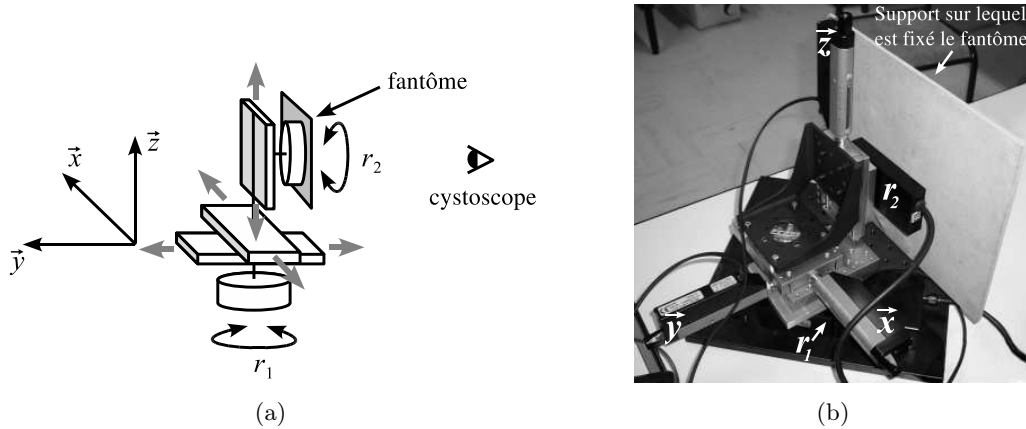


FIG. 4.5 – Matériel utilisé pour l'évaluation de la précision du mosaïquage. (a) Diagramme qui montre les degrés de liberté du système de positionnement. (b) Photographie montrant les composants du système utilisé pour acquérir les images du fantôme permettant d'évaluer la précision du mosaïquage.

4.3.2 Tests effectués

À l'aide du système d'acquisition d'images, nous avons effectué un enregistrement d'une séquence vidéo pendant que le système de positionnement effectue des rotations, des changements d'échelle et des changements de points de vue, suivant un ordre prédéfini. Les acquisitions des séquences d'images ont été effectuées directement par la caméra *sans* le cystoscope. Cette manière de procéder permet de tester la précision inhérente de notre algorithme de mosaïquage, c'est-à-dire sur des données réalistes mais non affectées par les défauts introduits par le cystoscope (distorsion radiale, inhomogénéités d'éclairage).

Le tout premier test effectué a consisté en l'acquisition d'une vidéo d'un parcours constitué uniquement de translations selon \vec{x} et \vec{y} . Le chemin parcouru par le système a la forme d'un carré de 45 mm de côté formant un circuit fermé (la première et la dernière images se trouvent idéalement dans la même position). La séquence acquise contient 1300 images au total. La taille des images acquises est de 640×480 pixels, mais pour notre méthode de mosaïquage nous considérons une région d'intérêt de 400×400 pixels extraite de la partie centrale des images. La translation entre chaque image est approximativement de 2,3 pixels selon chaque axe. Compte-tenu des petites valeurs de translation entre les images, nous avons également testé la reconstruction sur la même séquence, en ne sélectionnant qu'1 image sur 4 et 1 image sur 8. Les images panoramiques calculées sont montrées dans la figure 4.6.

Pour estimer la précision du recalage, il faut comparer les coordonnées des points de la première image avec les coordonnées des points homologues de la deuxième image ramenée dans le repère de l'image panoramique (qui est aussi le repère de la première image). En d'autres termes, dans le cas idéal, la transformation globale pour la dernière image devrait être la même que pour la première image (matrice unité). Les matrices résultantes de ce test pour les trois cas (mosaïquage de toutes les images, recalage d'une image sur quatre et d'une image sur huit) sont présentées dans le tableau 4.9.

L'erreur de mosaïquage err_m (dernière colonne du tableau 4.9) est définie par :

$$err_m = \sqrt{h_{13}^2 + h_{23}^2}/N, \quad (4.6)$$

où N est le nombre de paires d'images de la séquence. Les résultats de ce test montrent que l'erreur

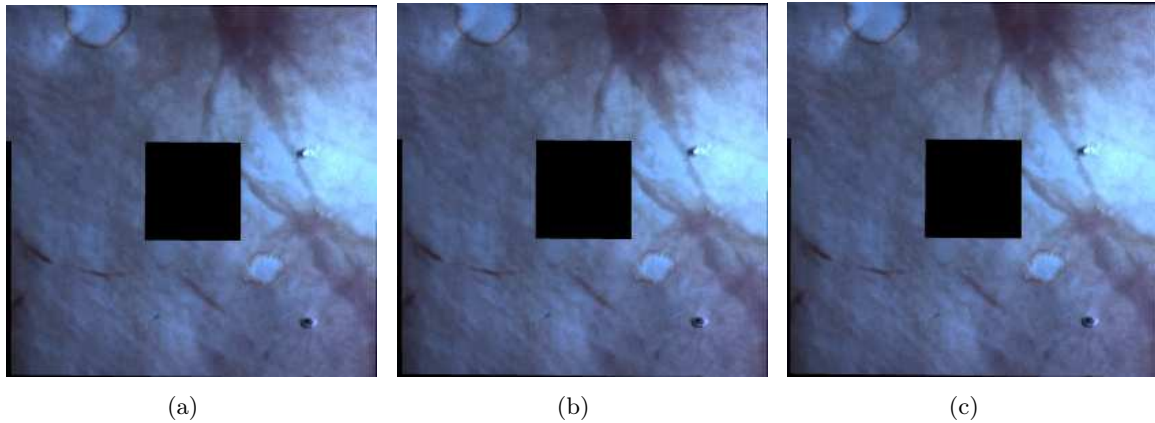


FIG. 4.6 – Images panoramiques obtenues pour le recalage d’une séquence de 1300 images acquises par une caméra *sans* cystoscope. Le lien géométrique entre les images est composé uniquement par des translations sur \vec{x} et \vec{y} . (a) Image panoramique construite en considérant toutes les images (1087×1091 pixels). (b) Image panoramique en prenant 1 image sur 4 (1089×1096 pixels). (c) Image construite en prenant 1 image sur 8 (1089×1098 pixels). Les erreurs de mosaïquage sont visibles à travers les bandes noires.

de mosaïquage augmente avec la distance existant entre les images. Dans tous les cas, cette erreur inhérente reste faible. Les images panoramiques de la figure 4.6 restent toutes visuellement cohérentes.

Le test suivant que nous avons effectué, consiste en l’acquisition d’une séquence d’images en suivant un parcours défini. Ce parcours implique des mouvements de tous les axes qui composent le système de positionnement (3 rotations et 3 translations). La forme globale du parcours est un carré. La première série de mouvements consiste en l’application de translations (amplitude totale 42 mm) selon l’axe \vec{x} combinée avec des translations selon l’axe \vec{z} (amplitude totale 42 mm) (changement d’échelle). Le mouvement suivant est formé par une translation selon \vec{y} (de 0 à 42 mm) associée à une rotation dans le plan image allant de 0 à 20° (ici appelée r_2). Le troisième mouvement consiste en une translation selon \vec{x} (faisant revenir le dispositif à la valeur $y = 0$) et d’amplitude -42 mm qui est combinée à une rotation de 16° hors plan qui provoque des changements de point de vue. Cette rotation hors plan autour de l’axe \vec{z} est appelée r_1 . Le dernier mouvement implique des mouvements selon \vec{y} (changement d’échelle) et \vec{z} et des rotations des angles r_2 et r_1 . À la fin de ce dernier mouvement, le dispositif est revenu dans sa position initiale. Le nombre d’images ainsi acquises est de 169, ce qui correspond à 42 images pour chaque côté du carré plus une image acquise pour la position initiale du dispositif. Une illustration du parcours effectué est donnée dans la figure 4.7.

Ainsi mentionné dans la section 4.3.1, l’image utilisée pour simuler l’aspect visuel de la parois de la vessie humaine est une impression en niveaux de gris d’une photographie haute résolution d’une vessie de porc. Pour ce deuxième test, dans le but de faciliter l’estimation de la précision du mosaïquage, nous avons imprimé sur la photographie de la vessie de porc une grille de points noirs espacés de 2,24 cm selon les axes vertical (\vec{z}) et horizontal (\vec{x}) (voir figure 4.5(a)). Ces points seront utilisés pour obtenir une mesure de l’erreur de la construction d’une image panoramique. Une fois l’acquisition des images effectuée (voir figure 4.8 pour des exemples des images acquises) en suivant le parcours décrit précédemment, nous obtenons une image panoramique avec notre méthode de mosaïquage (voir figure 4.9(b)). La première étape de l’évaluation de la précision de l’image panoramique construite est le recalage par sélection manuelle de 4 points dans la première

TAB. 4.9 – Paramètres de la transformation globale pour la dernière image d’un parcours composé seulement de translations. Les trois dernières lignes du tableau donnent les résultats du mosaïquage en utilisant respectivement toutes les images, 1 image sur 4 et 1 image sur 8. La deuxième ligne donne les valeurs des translations existant effectivement entre la première et la dernière image. Les colonnes h_{13} et h_{23} correspondent respectivement aux translations selon \vec{x} et \vec{y} . La dernière colonne, contient l’erreur totale divisée par le nombre de paires d’images (err_m , erreur moyenne pour la séquence d’images).

	h_{11}	h_{12}	h_{13}	h_{21}	h_{22}	h_{23}	h_{31}	h_{32}	h_{33}	err_m
Valeurs théoriques	1,00	0,00	0,00	0,00	1,00	0,00	0,00	0,00	1,00	—
Toutes	1,019765	-0,004438	-17,39109	0,010603	0,999708	-8,359688	0,000008	-0,000004	0,994196	0,015
1 sur 4	1,018839	-0,004420	-16,75642	0,008911	0,999169	-6,616971	0,000009	-0,000002	0,993020	0,049
1 sur 8	1,014920	-0,002948	-14,275747	0,007819	0,997745	-4,274116	0,000009	-0,000003	0,994068	0,091

image collée sur l’image panoramique avec leurs homologues dans la photographie numérisée. Pour effectuer ce recalage, un modèle rigide (translation 2D, rotation et facteur d’échelle) est utilisé.

L’étape suivante est une segmentation manuelle des points de la grille de l’image panoramique (voir figure 4.10(a)). Les 31 points couverts par l’image panoramique ont été tous trouvés par la méthode de segmentation. Ensuite, nous déterminons les coordonnées du barycentre de chaque point segmenté. Ces coordonnées sont finalement comparées aux coordonnées du centre des points imprimés sur la photographie et connus *a priori* (voir figure 4.10(b)). La distance euclidienne entre les coordonnées des points homologues des deux images est la mesure de la précision du mosaïquage. Après ce processus, nous obtenons pour les 31 points segmentés une erreur moyenne de 3,82 pixels et une erreur maximale de 6,73 pixels. La taille de l’image panoramique étant 1947×1187 pixels, l’erreur maximum représente 0,57% de la hauteur de l’image (1187 pixels). Nous estimons que cette erreur est suffisamment petite pour une application de la méthode au mosaïquage d’images endoscopiques. Cette affirmation sera vérifiée sur des données patients. Par ailleurs, les transformations obtenues par l’algorithme de recalage permettent de replacer les points de référence détectés dans l’image panoramique dans le repère de l’image d’origine dans lequel la séquence a été acquise (voir figure 4.10(a)). Le fait que ces points forment visuellement un quadrillage constitué de lignes et colonnes parallèles confirme la cohérence de l’image panoramique et indique que le recalage est précis. Il est à noter que les points de référence forment un quadrillage dans l’image panoramique (voir figure 4.9(b)).

4.4 Application de la méthode de mosaïquage à des données patient

Les tests précédents ont montré que notre approche est adaptée, en théorie à la construction d’images panoramiques. Nous avons aussi appliqué notre méthode de mosaïquage sur des images cystoscopiques cliniques pour démontrer son applicabilité dans une situation réelle (avec des données patients). Les images utilisées pour construire les images panoramiques présentées dans cette section ont été acquises au *Centre Alexis Vautrin (CAV)* qui est un centre de lutte contre

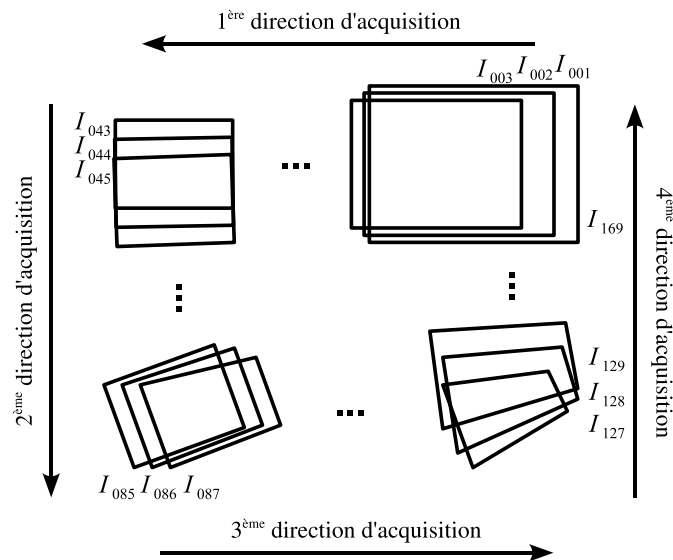


FIG. 4.7 – Illustration du parcours suivi pour l’acquisition des images utilisées pour le deuxième test d’évaluation de la précision du mosaïquage d’images. $I_{001}, I_{002} \dots I_{169}$ sont les images acquises pendant le parcours.

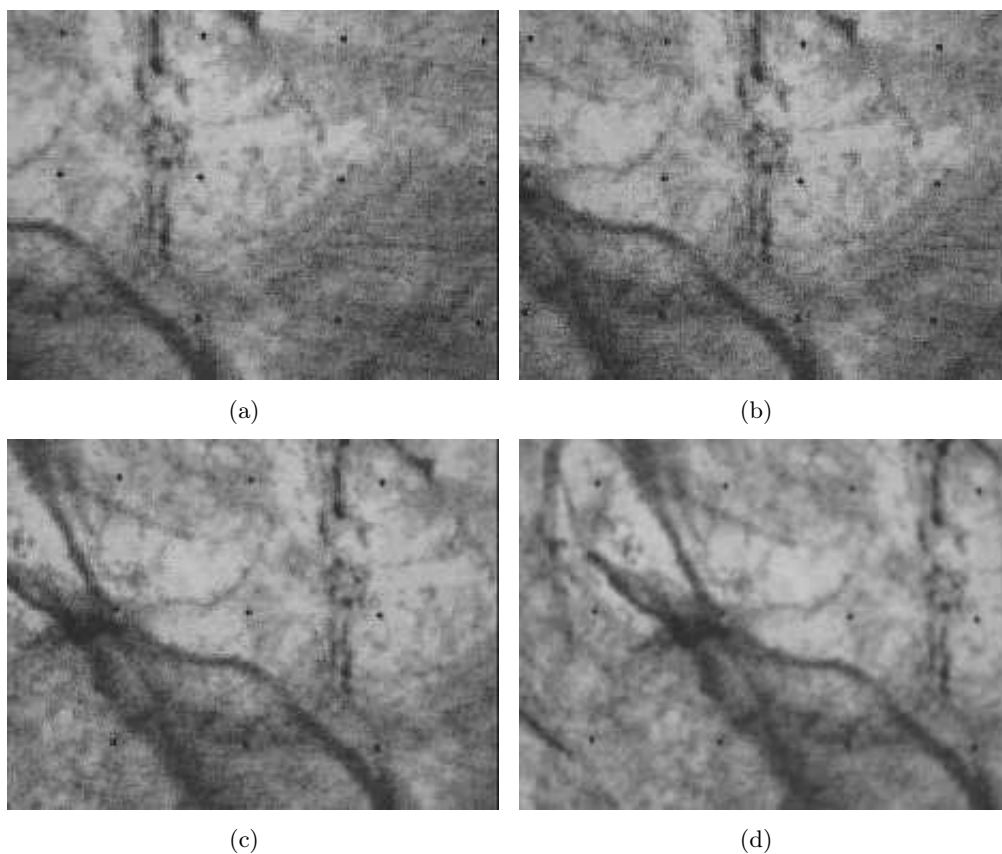
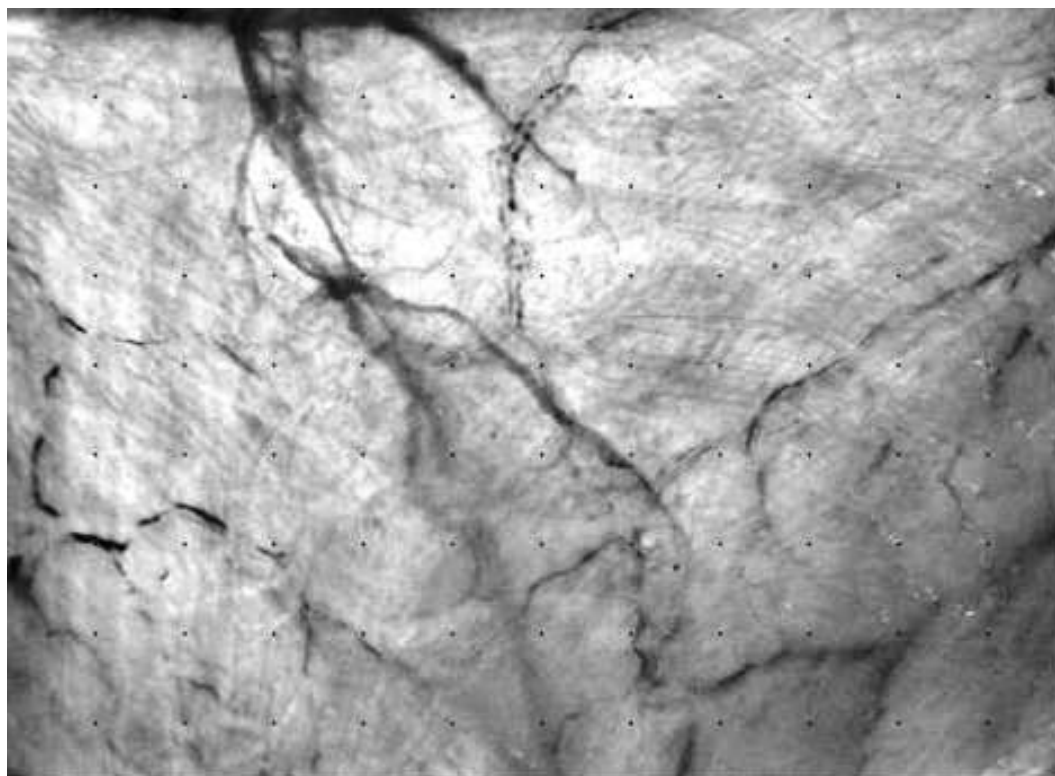
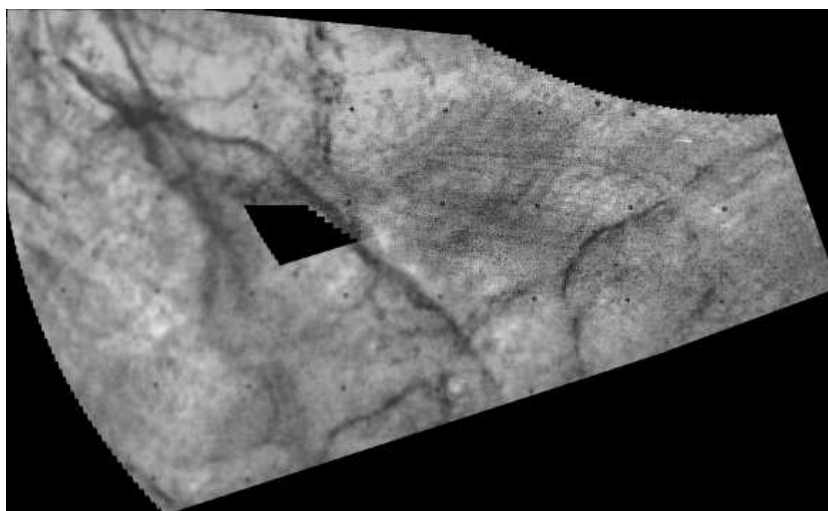


FIG. 4.8 – Les images (a) à (d) sont des exemples des images acquises par la caméra sur une photographie de vessie de porc pour tester la précision du mosaïquage. Nous pouvons observer les points noirs de la grille que nous utilisons pour évaluer la précision du mosaïquage.



(a)



(b)

FIG. 4.9 – (a) Photographie de vessie sur laquelle une grille de points a été placée pour évaluer la précision du mosaïquage. (b) Image panoramique (1947×1187 pixels) construite à partir de 169 images acquises pour la photographie de vessie porc en suivant le parcours illustré dans la figure 4.7. Nous pouvons observer la grille de points noirs utilisé pour l'évaluation de la précision du mosaïquage.

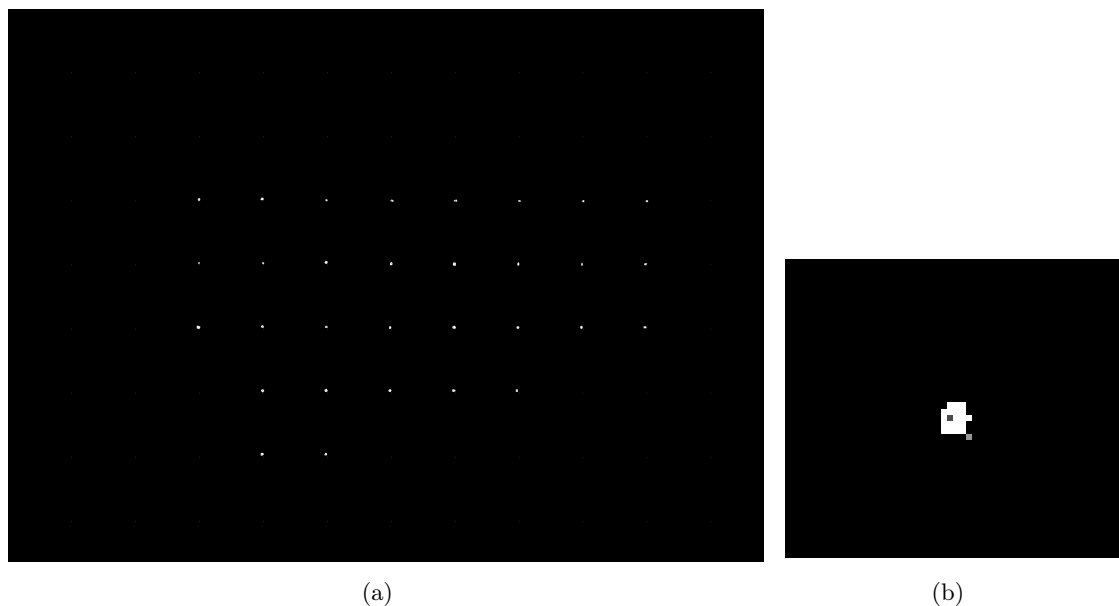


FIG. 4.10 – (a) Image de la figure 4.9(b) ramenée dans le système de coordonnées de l’image de la figure 4.9(a) après la détection des points noirs. Les points de référence de l’image originale et les points de l’image « reconstruite » (barycentre des points détectés) sont placés sur cette image. (b) Zoom qui montre en gris les barycentres des points détectés sur l’image panoramique et en gris clair les points correspondant aux position idéales. La distance moyenne entre les points homologues est de 3,82 pixels.

le cancer¹ situé à Vandœuvre-Lès-Nancy. Des examens ont été réalisés pour 6 patients différents. Une séquence vidéo a été acquise par patient. La durée des séquences varie entre 4 et 11 minutes approximativement. Une colonne endoscopique est utilisée pour effectuer les examens. Les composants du système cystoscopique d’acquisition d’images sont illustrés dans la figure 4.11.

De ces examens, nous avons sélectionné 10 séquences d’images de quelques secondes. Nous avons considéré comme intéressantes les séquences dans lesquelles il est possible de voir les uretères, la bulle d’air, le trigone, des cicatrices ou des polypes. L’intérêt de ces structures anatomiques n’est pas de fournir des structures exploitables par l’algorithme de recalage mais de s’en servir pour pouvoir associer l’image panoramique à un endroit de la vessie. Le nombre d’images des séquences varie entre 150 images (6 secondes) et 1300 images (52 secondes). Il est important de noter que nous n’appliquons pas la correction radiale sur ces images car nous n’avons pas accès au cystoscope utilisé pour les acquisitions. Nous appliquons notre méthode de mosaïquage avec un degré de filtrage 2 (moyen, voir section 3.2.1). Des régions de taille 400×400 , extraites de la partie centrale des images par notre méthode sont utilisées pour recalcr et former les images panoramiques.

Les figures suivantes présentent 5 images panoramiques construites avec 5 des 10 séquences d’images que nous avons utilisées. Les autres 5 images panoramiques construites peuvent être visualisées dans l’annexe B. Différentes caractéristiques particulières y ont été annotées.

Les images panoramiques construites à partir de séquences cliniques, sont visuellement cohérentes et, même si elles n’ont pas la qualité d’une photographie, leur aspect visuel permet

¹Nous remercions le Pr. F. Guillemin et le Dr. M. A. D’Hallewin pour nous avoir fourni les séquences vidéo-cystoscopiques.

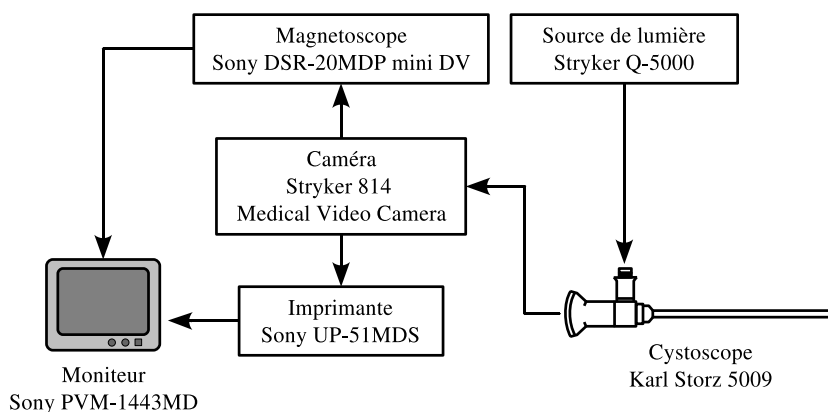


FIG. 4.11 – Composants du système de cystoscopie clinique avec lequel ont été acquises les séquences vidéo que nous utilisons pour tester notre approche de mosaïquage d’images.

d’observer clairement les structures et les détails des parois de la vessie. Dans la figure 4.12, il est possible d’observer en haut de l’image, un fragment de la bulle et en bas à droite de l’image un uretère. Voir, « en un coup d’œil », l’ensemble de ces détails anatomiques sur une image est impossible. Dans la figure 4.13, sur la partie inférieure de l’image panoramique, il est possible d’observer un uretère. Les vaisseaux sanguins de cette partie de la vessie sont clairement visibles. La troisième image construite (voir figure 4.14), montre le parcours effectué par le cystoscope depuis l’endroit où se trouve un polype jusqu’à l’endroit où un autre polype est situé. L’image panoramique suivante (figure 4.15) a été construite à partir d’un parcours du cystoscope tout au long d’une section pariétale, jusqu’à une calcification. Les images utilisées pour construire cette image panoramique sont d’une qualité très inférieure aux images utilisées pour construire les premières images panoramiques (figures 4.12 à 4.14). Malgré cela, l’algorithme de mosaïquage permet de calculer une image panoramique. La calcification qui était à peine visible dans la figure 4.15 est facilement observable dans le champ de vue accru de la figure 4.16.

Le nombre d’images et le temps de calcul pris par notre méthode pour calculer ces 5 images panoramiques présentées sont donnés dans le tableau 4.10.

Les images panoramiques construites permettent de visualiser des régions d’intérêt particulier et ce dans leur totalité à partir d’un support cystoscopique conventionnel (photographique où vidéo). Comme nous l’avons fait remarquer sur les images présentées, avec une vue panoramique il devient possible d’avoir sur un support unique plusieurs repères anatomiques et/ou zones suspectes. Cela facilite la localisation des lésions et la connaissance de leur distribution spatiale.

Étant donné qu’une évaluation qualitative (visuelle) des images panoramiques construites à partir des séquences cliniques est subjective, nous avons décidé de comparer les résultats numériques délivrés par notre méthode de mosaïquage à ceux obtenus par la méthode développée dans une thèse précédente au sein de notre laboratoire. Pour ce faire, nous avons émis l’hypothèse que si les transformations relatives obtenues avec notre méthode coïncident avec celles de la méthode basée sur l’information mutuelle, notre approche est correcte. En effet, même si la méthode basée sur l’information mutuelle est très coûteuse en temps de calcul et est limitée en termes d’amplitude de déplacements entre images, elle s’est avérée être précise (voir [Miranda-Luna *et al.*, 2007]). Pour la comparaison, nous avons tenu compte des translations seules, car elles indiquent la tendance que suit la transformation entre les images (en moyenne, les translations sont les paramètres prépondérants dans les transformations). La figure 4.17 montre les courbes des transformations relatives (entre images recalées) obtenues par les deux méthodes. Nous avons

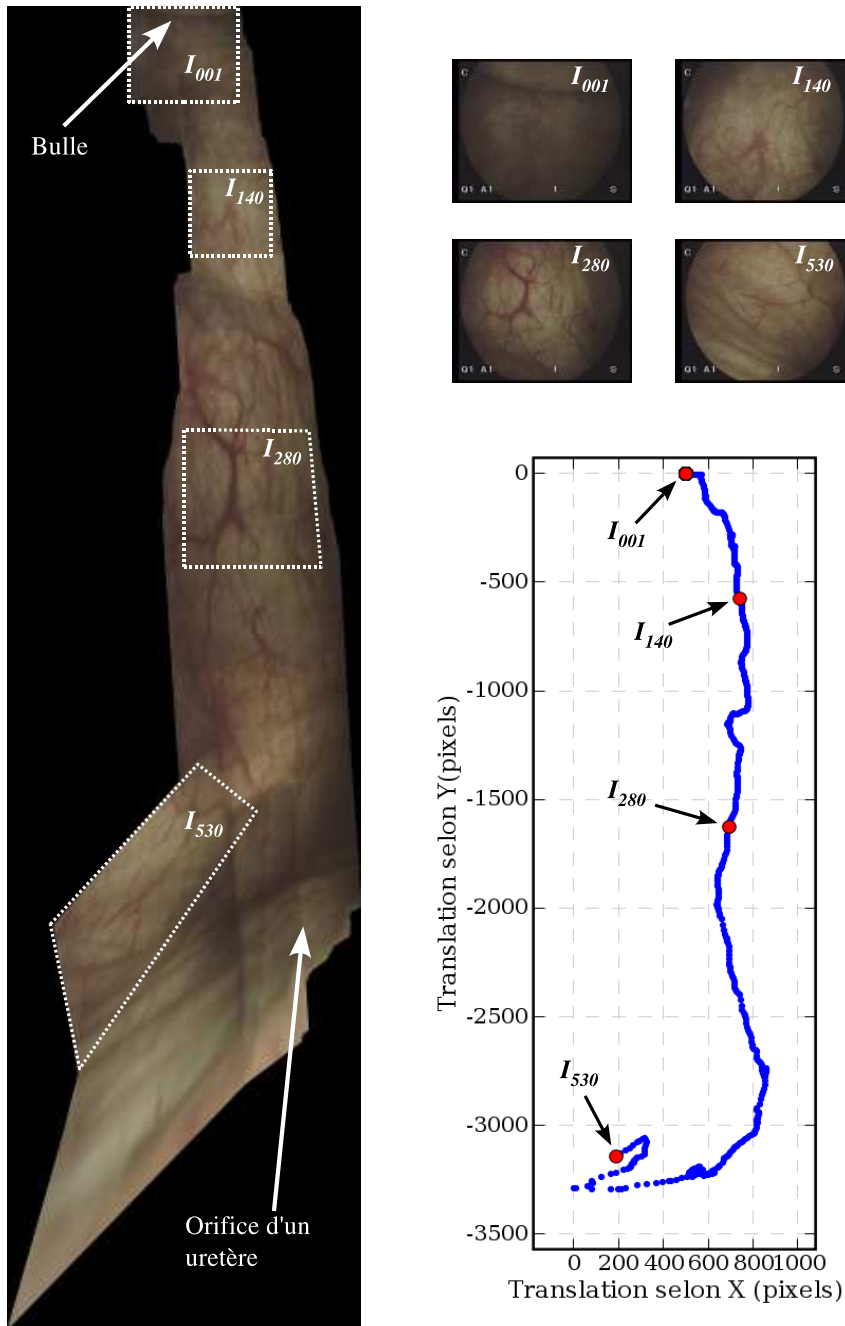


FIG. 4.12 – Image panoramique construite à partir de 530 images. La taille de l'image est de 1458×5471 pixels. Nous pouvons apercevoir la « bulle » dans la partie supérieure de l'image et dans la partie inférieure droite un orifice d'uretère. Voir ces parties de la vessie sur une image unique est impossible. Le graphique de droite montre le parcours suivi par le cystoscope pendant l'acquisition de cette séquence. Ce parcours représente les coordonnées du coin supérieur gauche des images calculées par l'algorithme de mosaïquage.

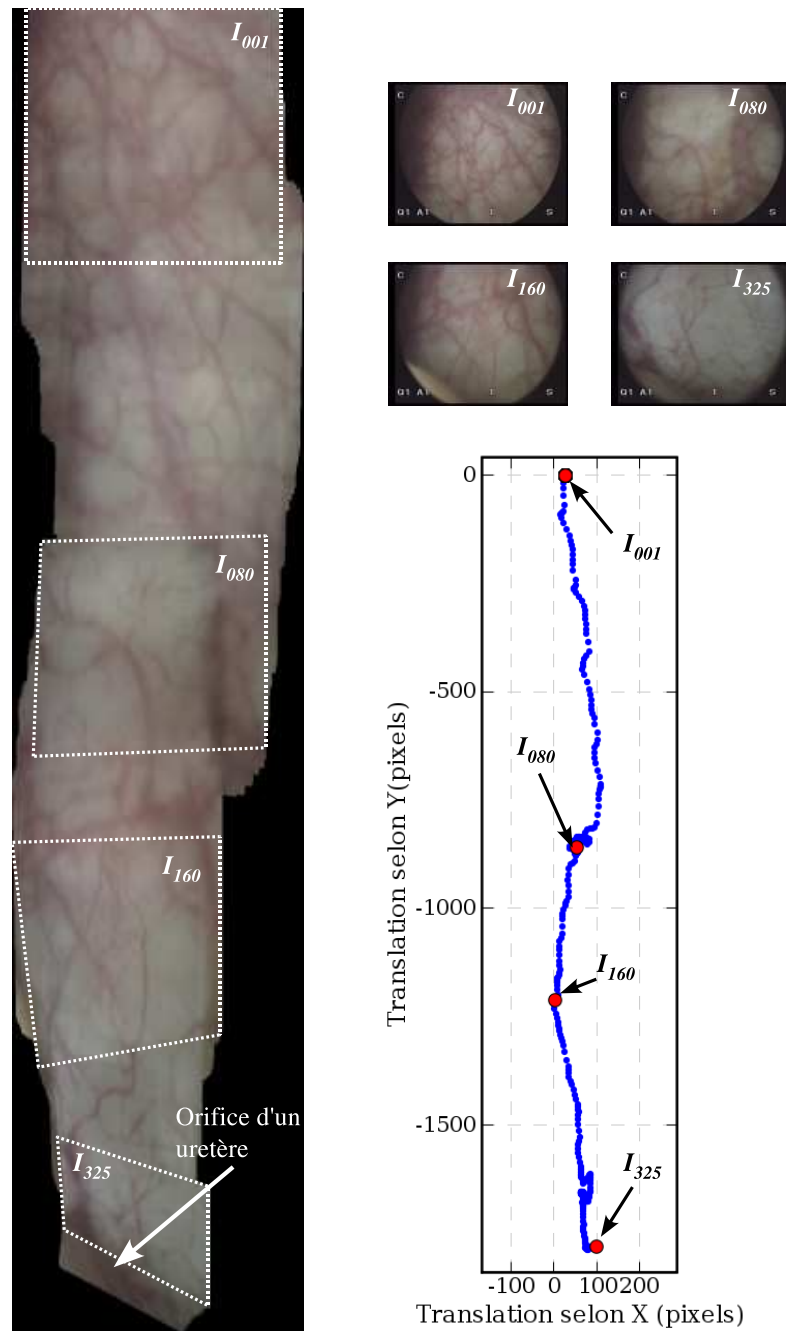


FIG. 4.13 – Image panoramique de taille 462×2104 pixels, calculée avec 325 images. Dans la partie inférieure gauche nous pouvons observer un orifice d’uretère. A gauche, près de l’uretère, une partie de la paroi peut être observée. Le graphique à droite représente la trajectoire de l’endoscope (coordonnées du coin supérieur gauche des images dans l’image panoramique).

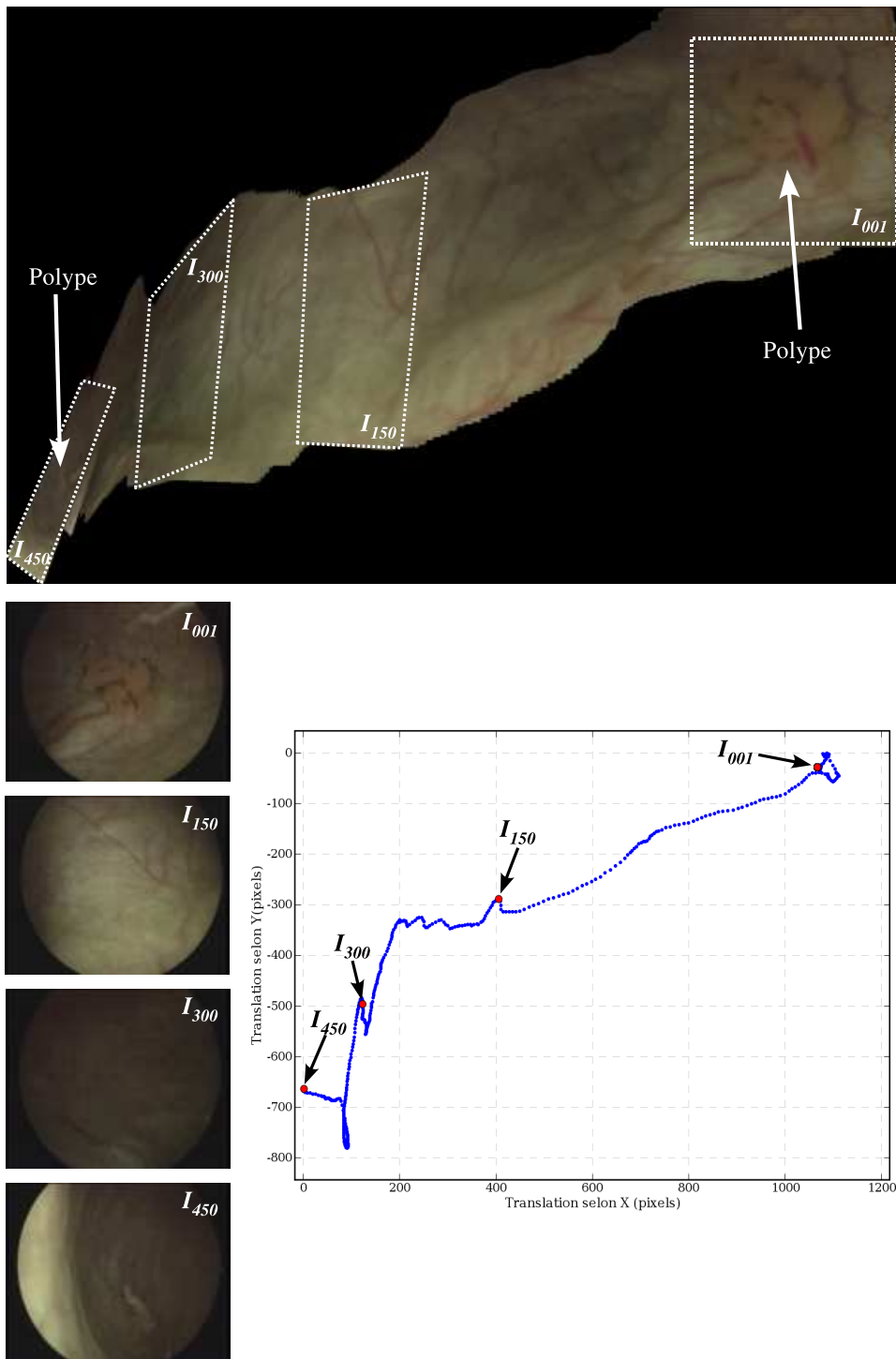


FIG. 4.14 – 450 images ont été utilisées pour construire cette image panoramique de 1479×1049 pixels. Dans cette image il existe deux polypes, l'un dans la partie supérieure droite et l'autre dans la partie inférieure gauche de l'image. Cette image panoramique donne une idée plus précise de la situation d'un polype par rapport à l'autre. En observant les 4 échantillons qui montrent le type d'images utilisées pour construire les images panoramiques, nous remarquons la grande variabilité de l'illumination qui existe parmi les images de la séquence.

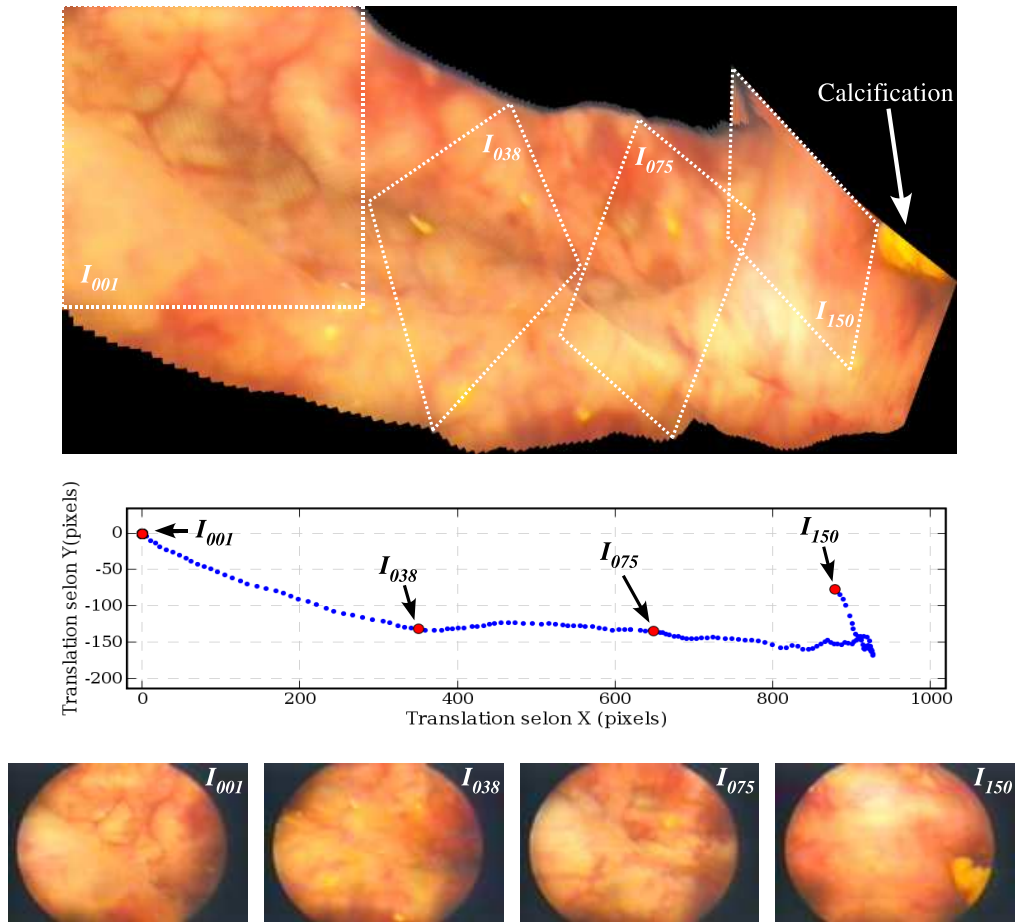


FIG. 4.15 – 150 images ont été utilisées pour construire cette image de taille 1178×588 pixels. Il est possible de suivre le parcours que le cystoscope a effectué le long de la paroi vésicale jusqu'à arriver au début d'une zone de calcification.

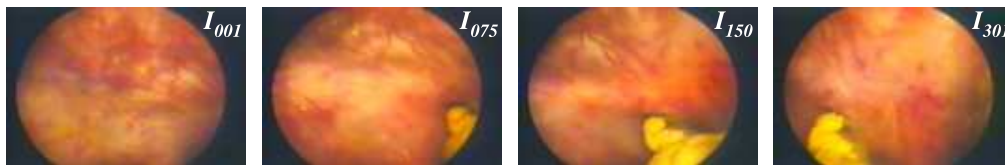
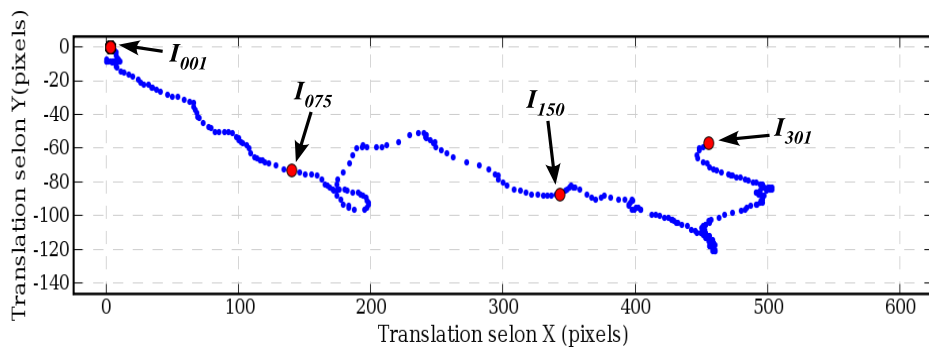
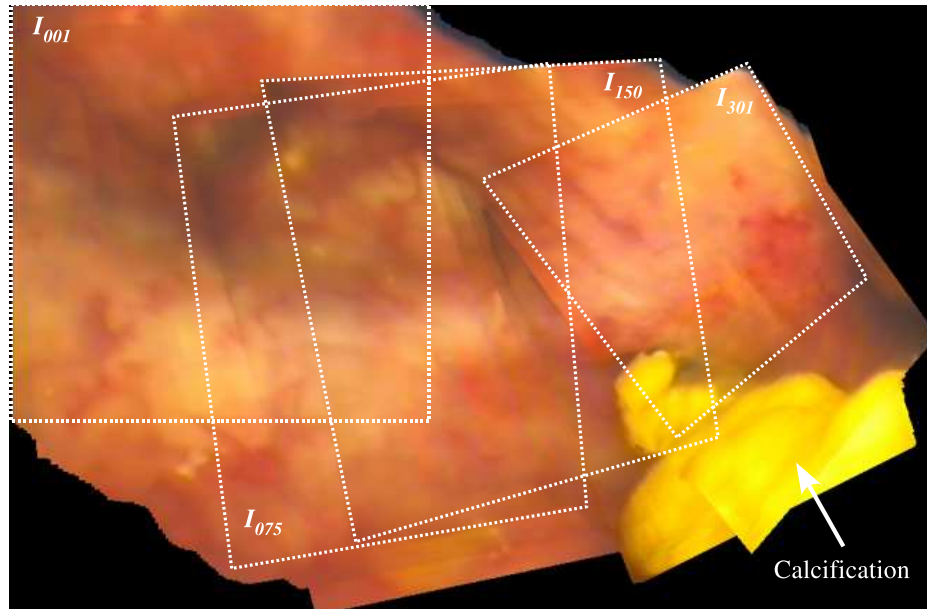


FIG. 4.16 – Cette image panoramique de 889×583 pixels et calculée avec 301 images est le complément de l'image de la figure 4.15. Ici, la calcification est plus facilement observable.

TAB. 4.10 – Taille, nombre d'images et temps de calcul pour les dix images panoramiques fabriquées à partir de séquences cystoscopiques cliniques. Il faut noter que pour les images 6 et 7, la taille de la région d'intérêt est de 256×256 pixels et par conséquent, le temps de calcul par paire d'images est réduit d'environ un facteur 4 par rapport à ceux des autres séquences dans lesquelles des régions de 400×400 pixels ont été extraites.

<i>Image panoramique</i>	<i>Taille (pixels)</i>	<i>Nombre d'images</i>	<i>Temps de calcul du recalage (minutes)</i>	<i>Temps de calcul du collage (minutes)</i>	<i>Temps de calcul total (minutes)</i>	<i>Temps de calcul par paire d'images (secondes)</i>
1	1458×5471	530	9,03	3,65	12,69	1,44
2	462×2104	325	6,26	0,61	6,88	1,27
3	1479×1049	450	8,66	0,72	9,39	1,25
4	1178×588	150	3,40	0,32	3,72	1,5
5	889×583	301	5,97	0,77	6,74	1,35
6*	1319×2647	1350	6,77	2,04	8,82	0,39
7*	1073×827	1150	6,20	0,96	7,16	0,37
8	859×1484	194	4,14	0,63	4,77	1,48
9	2072×1755	390	6,63	1,57	8,20	1,26
10	2544×4362	301	4,03	3,36	9,31	1,86

effectué le recalage des dix séquences d'images cystoscopiques cliniques utilisées précédemment, cependant nous présentons seulement les résultats pour les trois séquences où la différence était plus importante. L'erreur moyenne entre les deux courbes pour toutes les séquences est de 1,31 pixels selon \vec{x} et 1,54 selon \vec{y} . La différence maximale entre les translations obtenues par les deux méthodes est de 59,12 pixels. L'erreur maximale est pour la translation selon \vec{y} dans la séquence 7, où l'amplitude des translations est importante et nous pensons qu'elle dépasse les limites de la méthode de recalage basée sur l'information mutuelle. En effet, les images panoramiques construites avec notre méthode ne présentent pas, de façon visuelle, des déviations de position d'une amplitude égale à l'erreur entre les courbes. Il est important de noter que le nombre maximum d'images pouvant être recalées par la méthode de recalage par information mutuelle est de 511. C'est pour cette raison que le nombre maximum de paires d'images pour les courbes est de 510.

Les courbes montrent que pour la plupart des paires d'images, les valeurs de translation sont corrélées (proches en valeur). Cette comparaison n'est pas une évaluation quantitative absolue, mais, puisque la méthode basée sur l'information mutuelle a été déjà validée [Miranda-Luna *et al.*, 2007], nous pouvons dire que la tendance des valeurs est correcte.

4.5 Sélection d'images pour éviter le surplus de calcul

Dans les séquences vidéo, il est très commun de trouver des zones où le cystoscope reste presque immobile ou avance très lentement. Dans ce cas, la méthode va recalculer inutilement des images très similaires (peu ou pas transformées) plusieurs fois. En considérant que les changements de paramètres les plus importants dans les matrices de transformation des séquences cystoscopiques correspondent aux translations, nous avons pris les estimations délivrées par l'étape de corrélation des images en tant qu'un indicateur de la distance en pixels entre les images. À partir de la caractérisation de la distribution des translations sur les 10 séquences cystoscopiques que nous avons utilisées dans le test précédent (cf. sous-section 3.3.1.2), nous avons ajouté une

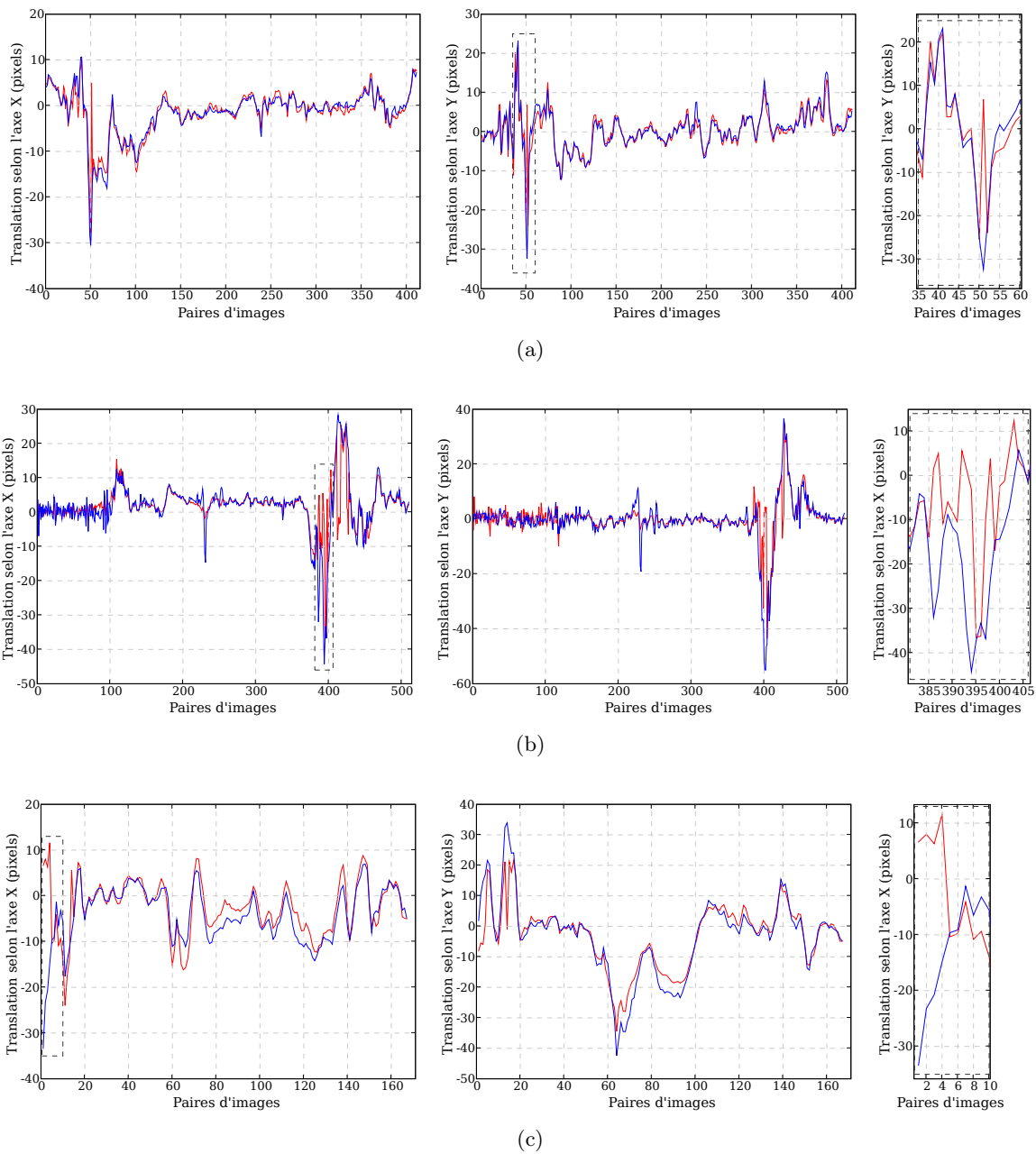


FIG. 4.17 – Comparaison des valeurs des translation obtenues par l’algorithme de recalage basé sur la différence d’intensité entre les images (trait bleu) et celles obtenues par l’algorithme de recalage basé sur l’information mutuelle (trait rouge) pour des images consécutives de trois séquences cystoscopiques cliniques. Les petites fenêtres à droite sont des grossissements des zones indiquées par les rectangles en trait pointillé. (a) Valeurs de translation pour la séquence 9. Erreur maximale selon \vec{x} : 30,13 pixels; erreur maximale selon \vec{y} : 39,19 pixels. (b) Courbes des translations entre paires d’images de la séquence 7. L’erreur maximale selon \vec{x} et \vec{y} est respectivement, 41,13 et 59,13 pixels. (c) Translations obtenues pour la séquence 8. Nous avons une erreur maximale de 40,00 pixels sur \vec{x} et de 34,96 pixels sur \vec{y} entre les deux courbes.

étape conditionnelle à notre algorithme : lorsque la translation entre les images est inférieure à 9 pixels, le recalage itératif des images et le calcul des matrices de transformation globales n'est pas effectué (voir figure 4.18). Dès que ce seuil est dépassé, le recalage entre l'image actuelle et l'image précédente, ainsi que l'actualisation de la matrice de transformation globale sont réalisés.

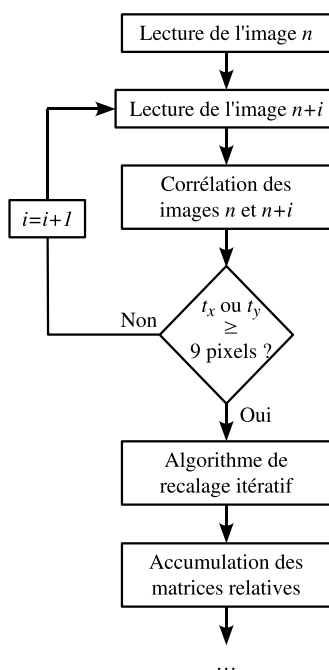


FIG. 4.18 – Organigramme de l'algorithme de sélection des images basé sur les valeurs de translations (t_x, t_y) délivrées par l'algorithme de corrélation. Le seuil de 9 pixels est choisi d'après l'étude des translations que nous avons effectuée dans le chapitre 2.

Puisque la partie la plus coûteuse au niveau du temps de calcul est l'algorithme de recalage itératif basé sur la différence d'intensité entre les images, nous obtenons un gain substantiel de temps de calcul en faisant le recalage itératif seulement lorsque le seuil a été dépassé. Cette technique va éviter le recalage entre des images « trop » proches. Pour tester cette approche, nous reprenons les séquences d'images cliniques utilisées lors du test précédent. Le temps gagné varie selon les séquences, mais en général nous obtenons des diminutions situées autour de 40% du temps nécessaire au mosaïquage de l'ensemble des images et ce, sans perte de qualité de l'image construite. Les images 4.19, 4.20 et 4.21 sont des exemples d'images panoramiques construites avec la méthode de sélection d'images. Nous pouvons observer que même s'il existe de petites variations au niveau de la forme de quelques régions des images panoramiques, l'allure générale de l'image est la même avec et sans sélection des images. Le tableau 4.11 donne les temps de calcul avec la sélection des images et le nombre d'images utilisées. Avec la limite de 9 pixels que nous avons fixée, le recalage de la séquence d'images 7 échoue (la taille de l'image panoramique devient trop importante).

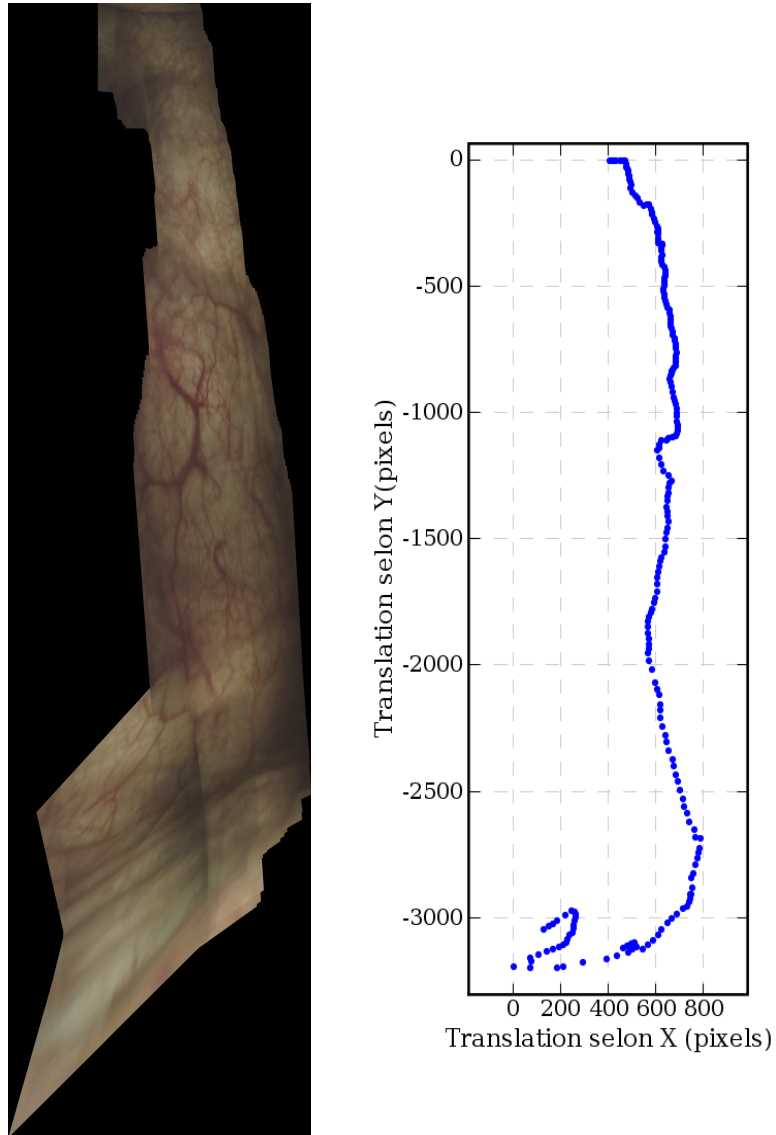


FIG. 4.19 – Image panoramique (1366×5138 pixels) construite avec notre méthode de sélection d'images en utilisant 239 images sur 530 disponibles. Nous avons utilisé la même séquence que pour la construction de l'image 4.12. L'aspect des deux images est très similaire.

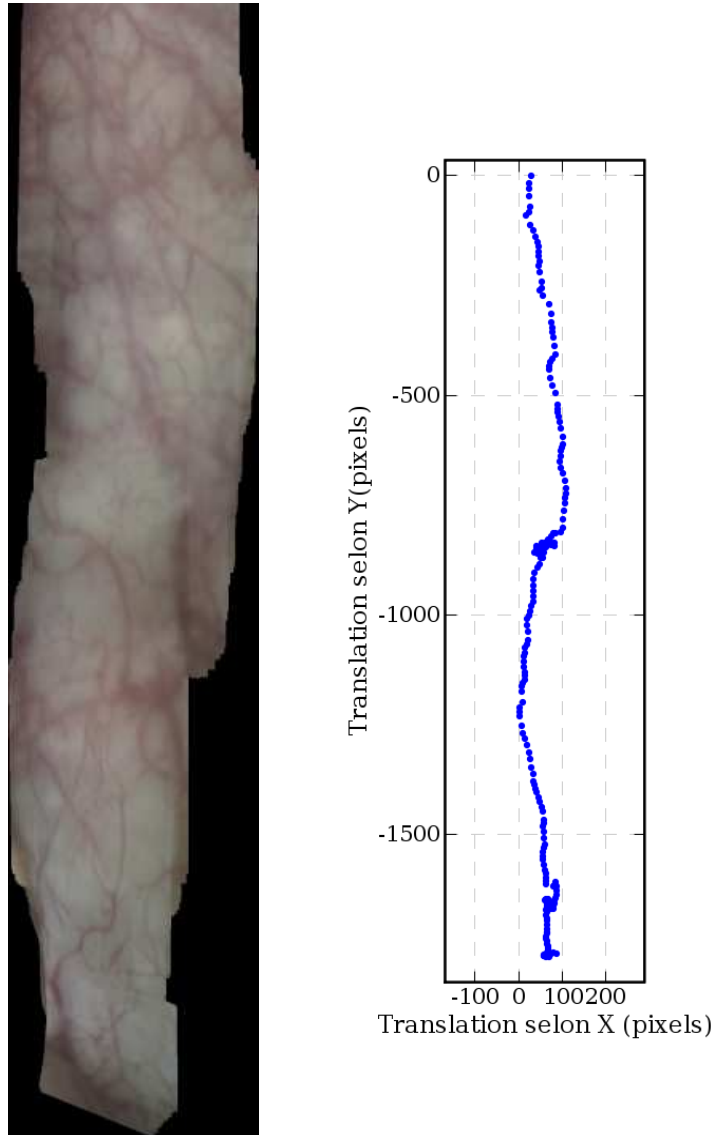


FIG. 4.20 – Image panoramique construite avec la méthode de sélection d'images correspondant à l'image 4.13 construite avec la méthode sans sélection. La taille de l'image est de 463×2092 pixels et a été construite avec 202 images sur les 450 images disponibles.

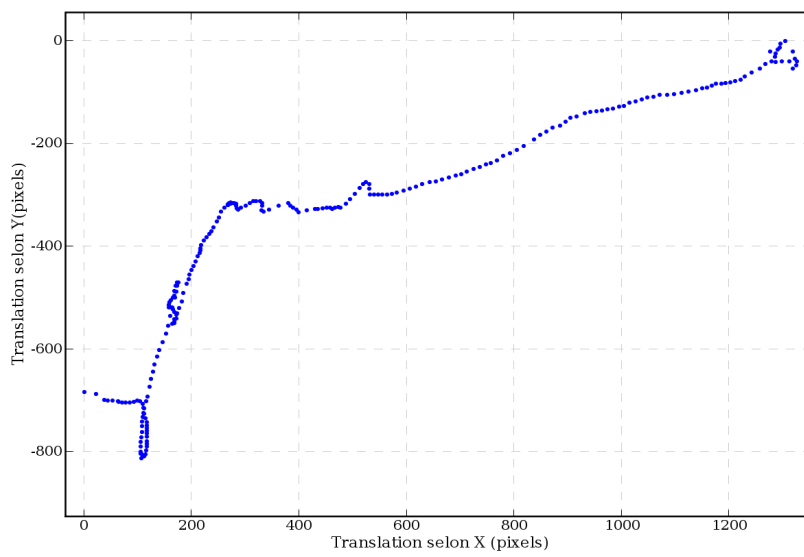


FIG. 4.21 – Image formée à partir de 236 images avec notre méthode de mosaïquage avec sélection des images. La taille de l'image est de 1685×1089 pixels. Elle est très similaire à l'image 4.14 construite avec 450 images.

TAB. 4.11 – Temps de calcul pour la méthodes de mosaïquage avec sélection d’images. L’utilisation de la méthode avec sélection d’images fait échouer le recalage de la séquence 7. Pour la séquence 6, il n’est pas possible d’utiliser les 1350 images parce que les transformations globales deviennent trop importantes et la taille de l’image panoramique augmente à une taille impossible à gérer. Si nous utilisons 1150 images le mosaïquage est possible. Néanmoins, pour le reste des séquences, la méthode réussit à diminuer significativement le nombre d’images utilisées et le temps de calcul nécessaire (pour comparaison avec le tableau 4.10).

<i>Image panoramique</i>	<i>Taille (pixels)</i>	<i>Nombre d'images utilisées / total d'images</i>	<i>Temps de calcul du recalage (minutes)</i>	<i>Temps de calcul du collage (minutes)</i>	<i>Temps de calcul total avec / sans sélection (minutes)</i>	<i>Temps de calcul par paire d'images (secondes)</i>
1	1366 × 5138	239 / 530	5,23	1,47	6,70 / 12,69	1,68
2	463 × 2092	202 / 325	4,48	0,40	4,88 / 6,88	1,45
3	1685 × 1089	236 / 450	6,41	0,50	6,91 / 9,39	1,50
4	1242 × 616	97 / 150	2,87	0,36	3,23 / 3,72	1,94
5	905 × 622	157 / 301	2,45	0,20	2,65 / 6,74	1,01
6* (1150)	2996 × 3218	368 / 1150	8,57	1,09	9,66 / 8,82	1,57
7*	—	—	—	—	—	—
8	856 × 1490	109 / 194	2,90	0,34	3,25 / 4,77	1,79
9	2013 × 1420	132 / 390	3,36	0,60	3,95 / 8,20	1,80
10	2705 × 4929	157 / 301	3,21	1,53	4,74 / 9,31	1,81

4.6 Combinaison des informations spatiales et fonctionnelles

Les examens cystoscopiques en lumière blanche permettent au clinicien de repérer des lésions cancéreuses au niveau macroscopique à un stade où elles ont généré des modifications tissulaires visibles à l’œil nu. Pour des lésions pré-cancéreuses ou cancéreuses précoces infra-cliniques, cette méthode de détection en lumière blanche est insuffisante. De nombreux travaux de recherche se concentrent actuellement sur l’exploitation de la fluorescence des tissus sous excitation UV. Notre approche d’imagerie panoramique en cystoscopie trouve tout son intérêt applicatif dans l’idée de combiner sur un support panoramique unique, les informations spatiales données par les images acquises en lumière blanche et les informations données par les images acquises sous excitation UV. Sur la séquence acquise en lumière blanche, il est possible d’appliquer notre méthode de mosaïquage. Les matrices de transformation globales sont alors enregistrées dans des fichiers texte pendant le processus de recalage. Les transformations globales ainsi enregistrées, sont utilisées pour construire la seconde image panoramique avec les images de fluorescence. De cette façon nous pouvons combiner les informations spatiales fournies par l’image panoramique en lumière blanche avec les informations de détection spectrale fournies par les images acquises en fluorescence.

4.7 Discussion et conclusion

Dans la première partie du chapitre nous avons testé la précision inhérente et la robustesse de notre méthode de recalage. Les résultats obtenus montrent que cette méthode est bien adaptée à notre application sur des séquences vidéo-cystoscopiques. Nous avons observé que notre méthode de recalage est très performante si la transformation prédominante entre les images à recalier est la translation. Cependant si des rotations perpendiculaires au plan d’acquisition supérieures à 10° sont présentes, la précision du recalage diminue. Les variations du facteur d’échelle supérieures

à 10% peuvent aussi affecter la précision du recalage des images. Les changements du point de vue affectent moins la précision du recalage. Néanmoins, sur une séquence vidéo, du fait de la vitesse d'acquisition (25–30 images par seconde), l'amplitude des transformations reliant géométriquement les images est en général, inférieure aux limites des transformations qui assurent un recalage précis des images. D'autre part, nous avons prouvé que notre méthode est robuste à des niveaux de bruit gaussien dont la variance varie jusqu'à 20% de celle de la distribution des niveaux d'intensité de l'image sans bruit. L'erreur maximale pour le recalage des images bruitées à 20% a été de 2,44 pixels pour des images de taille 256×256 . Cette erreur reste acceptable, car ce déplacement de la position réelle ne représente que 0,95% de la taille de l'image. Nous avons conclu que notre méthode de recalage peut être utilisée effectivement dans une méthode de mosaïquage applicable à la cystoscopie. En terme de précision de recalage d'images cystoscopiques, la méthode de recalage développé par [Miranda Luna, 2005] obtient une erreur maximale de 3,05 pixels pour le recalage d'images bruitées à 20% de la distribution des niveaux de gris de l'image originale. Nous obtenons une erreur maximale de 2,44 pixels pour une image similaire, mais pour une image bruitée à 10% l'erreur maximale est de 4,58 pixels. Cependant, ces erreurs sont à relativiser, car les transformations et les images utilisés pour tester les deux méthodes de recalage ne sont pas les mêmes.

Concernant les tests effectués pour évaluer la précision du mosaïquage, nous avons vu que l'erreur maximale obtenue permet de construire des images panoramiques représentant de façon précise les scènes acquises par une caméra. Pour ces tests, nous avons volontairement voulu éliminer les contraintes liées au positionnement précis d'un cystoscope. En effet, le seul cystoscope auquel nous avons accès est un cystoscope *Karl Storz 27005*, lequel possède un angle de vue de 30° . Le fait de positionner de façon parallèle le cystoscope au plan d'acquisition n'est pas un problème trivial. Étant donné que notre objectif était de tester la précision de notre méthode de mosaïquage dans la construction d'images panoramiques, nous avons fait les acquisitions d'images en supposant que le dispositif d'acquisition était parallèle à l'axe d'acquisition et que la distorsion radiale sur les images avait été corrigée. C'est-à-dire, les acquisitions des images ont été effectuées directement par une caméra. Le test effectué avec les 1300 images nous a montré qu'il existe bien une accumulation de l'erreur liée au calcul des matrices globales des images (par l'accumulation des matrices relatives). Néanmoins, cette erreur cumulée est petite (0,015 pixels pour des images translatées d'environ 2,3 pixels), mais augmente de façon proportionnelle à la distance existant entre les images avant le recalage (0,091 pour des images translatées d'environ 18,4 pixels). Le deuxième test nous a permis d'évaluer la précision de recalage pour une combinaison de transformations : des translations, des rotations dans le plan, des changements d'échelle et des changements du point de vue. L'erreur maximale estimée par la différence de position des points de référence a été de 6,73 pixels. Cette erreur représente moins d'un pourcent de la taille de l'image. Ce test nous a montré que notre méthode est tout à fait adaptée à la construction d'images panoramiques à partir d'images déformées par un modèle projectif acquises par une caméra. Cette erreur est comparable à celle rapporté par [Miranda Luna, 2005] (8,3 pixels) après l'application de sa méthode de correction de l'erreur de recalage cumulée. Au niveau de la vitesse de construction de l'image panoramique, nous arrivons à construire une image panoramique à partir d'une séquence de 511 images en 3,74 minutes. La méthode basée sur l'information mutuelle met environ 8 heures 7 minutes pour construire une image panoramique avec la même séquence.

Les images construites à partir des séquences acquises en clinique démontrent clairement que notre méthode est bien adaptée à cette application. Les images construites représentent de façon fidèle les scènes vues par le cystoscope. La qualité visuelle des images construites permet une observation aisée de l'anatomie de la région de la vessie qu'elles représentent. Ces images permettent d'avoir des vues qui ne peuvent pas être obtenues par les moyens conventionnels

d'enregistrement et de support visuel utilisés actuellement en cystoscopie. Par exemple, sur l'image 4.12 il est possible de voir sur une seule image un orifice d'un uretère et la bulle d'air.

Nous avons observé qu'aucune des séquences cliniques utilisées pour nos tests n'est constituée de parcours fermés. C'est pour cette raison, que nous n'appliquons pas la méthode de correction de l'erreur de recalage cumulée. Pour une application de cette méthode se référer à [Miranda-Luna *et al.*, 2007].

Cependant, notre méthode de mosaïquage n'est pas infaillible et peut, dans le cas de transformations dépassant certaines limites caractérisées, placer de façon imprécise les images d'une séquence sur le support panoramique. Nous pensons que le fait d'utiliser des images sans correction radiale affecte aussi le comportement de notre méthode. Un exemple de ce comportement a été observé sur l'image montrée dans la figure B.2 (annexe B) où, lorsque le cystoscope fait un mouvement circulaire rapide, l'erreur de recalage entre les images est visuellement grande.

La comparaison entre les paramètres de translation obtenus par notre méthode et ceux obtenus par la méthode de recalage basée sur l'information mutuelle développée précédemment au laboratoire, montre une concordance entre des résultats des deux méthodes. Cette comparaison a permis de démontrer que notre méthode arrive à recalibrer des images même lorsque la transformation est trop importante pour être trouvée par la méthode de recalage précédente.

Conclusion et perspectives

Le cœur de notre méthode de mosaïquage est un algorithme de recalage basé sur le travail de [Baker and Matthews, 2002]. Cet algorithme utilise comme mesure de similarité la différence des valeurs d'intensité entre les pixels des images. L'optimisation de cette mesure est effectuée par une méthode de minimisation du deuxième ordre (Gauss-Newton). L'approche retenue permet de calculer en une seule fois certaines valeurs (notamment le Hessien) qui sont ré-utilisées à chaque itération de l'optimisation. Cela permet de réduire le nombre d'opérations à effectuer et par conséquent, le temps de calcul. Pour des images de taille 400×400 pixels et pour une séquence d'une trentaine de secondes (750 images), notre méthode va calculer une image panoramique en environ 18 minutes sur un ordinateur *Pentium IV* à 2,4 GHz. Une séquence d'une trentaine de secondes est en général suffisante pour acquérir une zone d'intéressante du point de vue médical. Si nous utilisons la méthode de sélection d'images, le temps est réduit d'environ 40% (ce qui ramène le temps de calcul de 18 à 11 minutes). Ce temps de calcul est inférieur à la durée d'un examen cystoscopique clinique standard (dont la durée est de 15 minutes en moyenne). Il est donc possible de fournir au clinicien, directement après l'examen, une image panoramique 2D calculée à partir d'une séquence d'une trentaine de secondes. Pour des images de taille 256×256 pixels, le temps de calcul est réduit d'un facteur d'environ 4. Cela permet de construire les images panoramiques à une cadence proche de 2 images/seconde.

Nous avons montré que notre méthode peut gérer la grande variabilité de texture, de couleur et d'illumination des images cystoscopiques, ainsi que les changements de point de vue entre des images successives dus aux mouvements du cystoscope à l'intérieur de la vessie. Cela est possible grâce au pré-traitement des images que nous avons développé pour éliminer les inhomogénéités d'illumination et le pattern de fibres optiques (d'un fibroscope). Cette étape est indispensable pour le fonctionnement de l'algorithme de recalage. La méthode de pré-traitement est basée sur une approche pyramidale dans le domaine spatial. Nous proposons une plage réduite de quatre niveaux de filtrage (aucun, faible, moyen et fort) pour simplifier le choix à l'utilisateur. Pour les séquences que nous avons testées, les valeurs que nous proposons suffisent pour traiter tous les cas.

Les résultats de l'évaluation du recalage ont montré que l'erreur de recalage obtenue représente toujours moins de 1% de la taille de l'image, même en présence de bruit gaussien fort. Des expériences nous ont permis d'établir les limites pour lesquelles notre méthode fonctionne correctement (en termes de précision et de bruit). L'évaluation de l'erreur de mosaïquage a montré que l'erreur maximale (6,73 pixels pour une image de taille 1947×1187 pixels) permet de construire avec précision des images panoramiques de l'intérieur de la vessie. Cette erreur est similaire à celle rapporté dans [Miranda Luna, 2005] (8,33 pixels).

Notre méthode de mosaïquage n'est pas infaillible. Si la transformation géométrique liant deux images successives se trouve en dehors des limites caractérisées de notre méthode, le recalage sera imprécis ou échouera. Cependant, ces limites sont au delà des transformations géométriques existant dans les examens standards.

À ce jour, le développement d'une méthode de mosaïquage automatique d'images cystoscopiques est en elle-même une originalité car, à notre connaissance, si le mosaïquage d'images est utilisé avec succès dans divers domaines de la médecine (notamment dans l'ophtalmologie), son application aux organes creux était inexplorée. L'automatisation du processus de mosaïquage fait partie de cette contribution. Notre méthode ne nécessite aucune intervention de l'utilisateur pendant son fonctionnement. Elle est complètement automatique (le réglage du niveau de filtrage des fibres optiques peut être réglé une fois pour toutes pour un cystoscope donné). Cela réduit le temps de construction des images panoramiques et élimine la nécessité d'avoir des connaissances techniques ou de devoir, via des interventions humaines, localiser des points homologues dans les images de la séquence.

La caractérisation de la distribution des valeurs de translations présentée dans le chapitre 3 est une étude originale qui nous a permis de démontrer que, du fait du taux élevé de recouvrement des images, une méthode de recalage basée sur l'utilisation directe des niveaux d'intensité des pixels des images est bien adaptée aux séquences vidéo-cystoscopiques. Les résultats de cette étude ont été utilisés pour le choix d'un seuil qui nous a permis d'améliorer la vitesse de calcul des images panoramiques en évitant de recalculer des images trop proches en termes de translations.

Une autre contribution importante est la possibilité de combiner des informations spatiales et fonctionnelles sur un support panoramique unique. Cela peut améliorer l'analyse et le diagnostic de lésions vésicales cancéreuses par la combinaison des informations données par la fluorescence (et invisibles dans la modalité de lumière blanche) avec les informations des repères anatomiques et de morphologie des tissus visibles avec la lumière blanche.

Nous avons donc développé une méthode rapide pour la construction automatique et robuste d'images panoramiques de l'intérieur de la vessie à partir de séquences d'images issues d'exams cystoscopiques cliniques standard. Les images ainsi construites, représentent de façon fidèle des régions de la paroi interne de ces vessies. L'algorithme proposé ne modifie en rien le protocole d'acquisition clinique. Ces images panoramiques constituent des vues globales d'une zone qui ne peuvent pas être obtenues par les supports visuels et d'enregistrement utilisés conventionnellement en cystoscopie. Elles peuvent être utilisées pour un archivage de données non redondant mais surtout pour la détection/localisation de lésions, le guidage des biopsies et le suivi des lésions (exams ultérieurs, observations postopératoires). L'intérêt applicatif supplémentaire de cette approche réside dans la combinaison d'informations spatiales et fonctionnelles sur un support panoramique unique à l'aide d'un dispositif d'acquisition spécifique. Nous pouvons utiliser un prototype développé au laboratoire permettant d'acquérir de façon simultanée des séquences d'images en lumière blanche et en fluorescence sous excitation dans le proche UV. Nous avons démontré qu'avec les images acquises en lumière blanche il est possible de construire des images panoramiques. Les transformations géométriques trouvées pendant la construction de cette image permettent d'obtenir une seconde image panoramique de fluorescence. Les informations extraites de la modalité de fluorescence peuvent être « superposées » sur l'image panoramique en lumière blanche.

Perspectives

En termes de perspectives, notre travail pourrait être poursuivi selon trois principaux axes. Un premier axe concerne l'amélioration de l'outil développé. Nous pensons que des techniques de prédiction des positions des images, basées sur les informations de translation existant entre des images consécutives d'une séquence, pourraient être développées. La construction d'images panoramiques plus « étendues » à partir de plusieurs images panoramiques est un axe devant être exploré. Par exemple, nous avons mis en évidence plusieurs exemples qui montrent clairement le

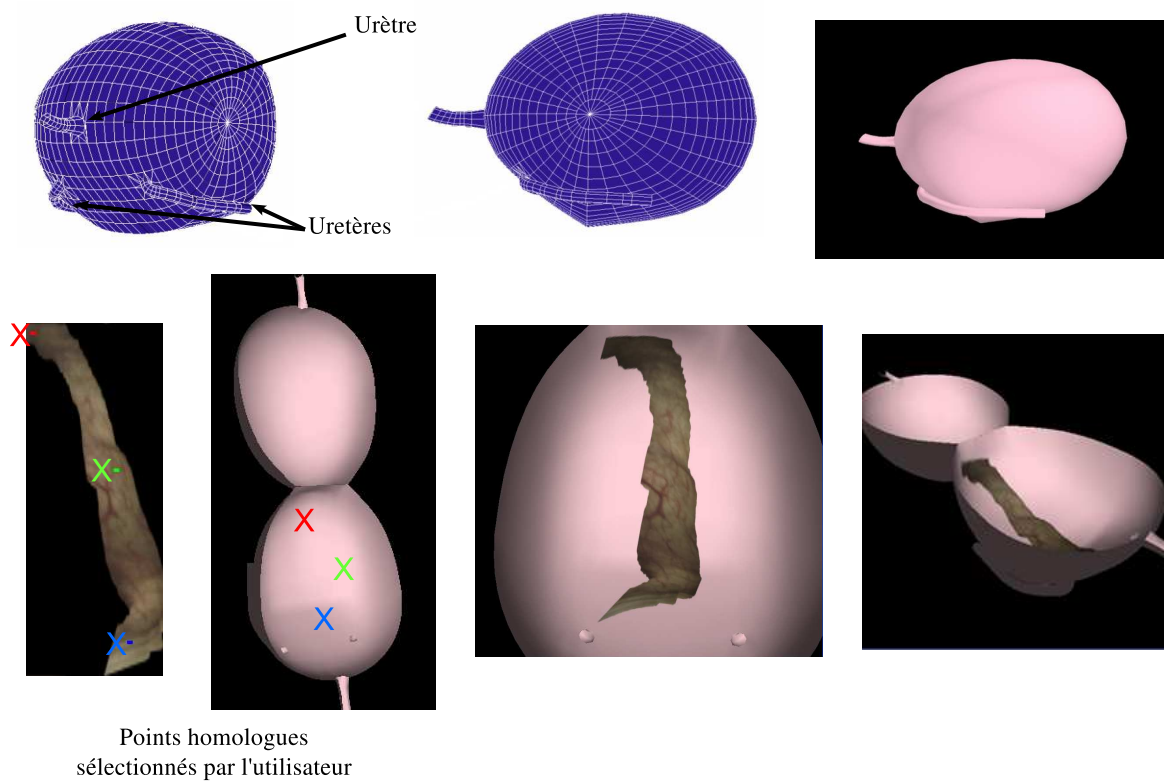
déplacement linéaire du cystoscope (déplacement du cystoscope en zig-zag ou sur des trajectoires parallèles). Nous pouvons ré-mosaïquer ces « bandes » pour former une image plus grande.

Du côté du temps de calcul, notre méthode est encore loin de la cadence vidéo, mais nous pensons qu'une optimisation du code et l'utilisation de techniques multi-résolution pourrait réduire le temps de calcul des images panoramiques (recalage d'images sous-échantillonnées). D'autre part, le test que nous avons effectué avec le détecteur de Harris sur les images cystoscopiques n'était pas exhaustif. Peut-être, qu'une étude plus approfondie sur l'application de cette méthode ou des méthodes similaires est nécessaire pour écarter (ou adopter) l'utilisation de primitives images dans notre domaine d'application. Une méthode basée sur des primitives image, permettrait de recalculer des images beaucoup plus distantes dans une séquence vidéo avec, en conséquence, une réduction du temps de calcul.

Un deuxième axe est l'adaptation de l'outil pour une utilisation clinique. Un travail d'adaptation du matériel et des logiciels développées permettant d'effectuer le transfert direct du support vidéo utilisé en clinique vers un support numérisé directement utilisable par notre algorithme, est aussi nécessaire. Cela serait un pas supplémentaire vers l'adaptation de notre solution à une utilisation en clinique. Nous avons démontré clairement qu'il est possible de combiner des informations spatiales et fonctionnelles dans un support panoramique unique par les tests que nous avons effectués sur fantôme. Il faut maintenant prouver son utilité clinique par des tests sur des images acquises *in vivo*.

Le dernier axe concerne la visualisation 3D des images panoramiques et la possibilité d'obtenir des informations de la forme 3D de la vessie à partir du recalage effectué entre les images 2D. Cette visualisation en 3D permettrait de réajuster les images conformément à un modèle « réaliste » de vessie et de réorienter les images (cf. figure B.5 qui est probablement "de travers"). Concernant la visualisation des images panoramiques, nous sommes parvenu dans ce travail à présenter les images panoramiques sur des plans 2D. Nous pensons que ce n'est pas la façon optimale de présenter ces informations, car au cours d'un examen, le clinicien se fait une représentation mentale de la forme 3D de la vessie. Nous pouvons imaginer présenter les images panoramiques en les projetant sur un volume représentant la vessie du patient. D'ailleurs, les premiers tests réalisés dans notre équipe² avec un volume « générique » de vessie semblent avoir convaincu le clinicien de l'utilité d'une représentation 3D des parois de la vessie. Actuellement, pour construire cette représentation 3D, l'utilisateur indique trois points sur l'image panoramique en se servant par exemple, des repères anatomiques. L'utilisateur indique aussi sur le volume générique les trois points correspondant à ceux marqués sur l'image panoramique. Même si l'approche n'est pas automatique, les résultats sont encourageants car, ainsi confirmé par des cliniciens, cette visualisation de la vessie rejoint la représentation 3D que se fait mentalement l'opérateur. Ce processus de marquage et les images 3D résultantes sont illustrés dans la figure (a). L'étape suivante serait d'automatiser le processus de projection sur le volume. Mais, nous pensons que la présentation au clinicien des données 3D devrait se faire par la construction du volume en utilisant l'information fournie uniquement par un cystoscope. Des techniques de reconstruction basées sur la lumière structurée ou sur l'obtention de volumes par des vues multiples pourraient être explorées. Un travail dont le but est notamment de compléter le cystoscope par une optique qui donne, en plus de l'image 2D, une information de profondeur, va être démarré au laboratoire.

²Tests réalisés par Yannick Rémy et Sophie Olijnyk.



Points homologues
sélectionnés par l'utilisateur

FIG. (a) – Processus de projection manuelle d’une image panoramique sur un volume générique de vessie. Le clinicien utilise des repères anatomiques (urètre, orifices des uretères, bulle) pour marquer des points sur la forme générique 3D de la vessie ainsi que leurs homologues (approximatifs) dans l’image panoramique. Les points homologues sont ensuite utilisés pour plaquer l’image panoramique sur le volume « ouvert » de la vessie.



Développement de l'algorithme de Lucas et Kanade

Cette annexe présente un développement plus complet de l'algorithme de Lucas et Kanade que celui présenté dans le chapitre 3. Pour ce faire, nous allons compléter les équations du chapitre 3 par des équations intermédiaires. Nous utilisons la même notation : $I_R(\mathbf{x})$ est l'image de référence (destination) et $I_D(\mathbf{x})$ est l'image à recaler (source) avec $I_R(\mathbf{x})$. $\mathbf{x} = (x, y)^T$ est un vecteur colonne contenant les coordonnées des pixels. La transformation des coordonnées est notée $T(\mathbf{x}; \mathbf{p})$, où $\mathbf{p} = (p_1, \dots, p_n)^T$ est le vecteur contenant les paramètres de transformation. $n = 8$ pour une transformation perspective. L'algorithme de Lucas et Kanade utilise le carré des différences des niveaux de gris des pixels homologues pour recaler deux images :

$$\sum_{\mathbf{x}} [I_D(T(\mathbf{x}; \mathbf{p})) - I_R(\mathbf{x})]^2. \quad (\text{A.1})$$

L'algorithme de [Lucas and Kanade, 1981] considère que les paramètres de transformation \mathbf{p} initiaux sont connus et cherche ensuite les incréments $\Delta\mathbf{p}$ qui, ajoutés ou soustraits à ces paramètres, feront que l'image I_D devienne de plus en plus similaire à l'image I_R :

$$\sum_{\mathbf{x}} [I_D(T(\mathbf{x}; \mathbf{p} + \Delta\mathbf{p})) - I_R(\mathbf{x})]^2. \quad (\text{A.2})$$

Cette fonction coût est minimisée par rapport à $\Delta\mathbf{p}$ et les paramètres de transformation \mathbf{p} sont actualisés à chaque itération avec les valeurs de $\Delta\mathbf{p}$ ainsi calculées (voir équation A.3) jusqu'à la convergence de \mathbf{p} vers la solution :

$$\mathbf{p} \leftarrow \mathbf{p} + \Delta\mathbf{p}. \quad (\text{A.3})$$

Généralement, il est admis que l'algorithme a convergé lorsque le module du vecteur $\Delta\mathbf{p}$ est en-dessous d'un seuil ϵ pré-défini, c'est-à-dire $\|\Delta\mathbf{p}\| \leq \epsilon$.

Les incréments de $\Delta\mathbf{p}$ sont obtenus par une optimisation non-linéaire du deuxième ordre (Gauss-Newton). D'abord, la mesure de similarité est linéarisée par un développement en série de Taylor du premier degré de $I_D(T(\mathbf{x}; \mathbf{p} + \Delta\mathbf{p}))$ selon $\Delta\mathbf{p}$. Le développement en série de Taylor d'une fonction de forme $f(x + \Delta x)$ est donné par :

$$f(x + \Delta x) = f(x) + \Delta x f'(x) + \frac{\Delta x^2}{2!} f''(x) + \dots + \frac{\Delta x^n}{n!} f^n(x),$$

Nous avons donc dans notre cas :

$$I_D(T(\mathbf{x}; \mathbf{p} + \Delta\mathbf{p})) \approx I_D(T(\mathbf{x}; \mathbf{p})) + \Delta\mathbf{p} \frac{dI_D(T(\mathbf{x}; \mathbf{p}))}{d\mathbf{p}}, \quad (\text{A.4})$$

et par la règle de dérivation des fonctions composées :

$$\frac{dI_D(T(\mathbf{x}; \mathbf{p}))}{d\mathbf{p}} = \frac{\partial I_D}{\partial \mathbf{x}} \frac{\partial T}{\partial \mathbf{p}} \quad (\text{A.5})$$

avec

$$\frac{\partial I_D}{\partial \mathbf{x}} = \left(\frac{\partial I_D}{\partial x}, \frac{\partial I_D}{\partial y} \right)^T = \nabla I_D, \quad (\text{A.6})$$

∇I_D est le gradient de l'image I_D défini par ses composantes selon \vec{x} et \vec{y} . Puisque $T(\mathbf{x}; \mathbf{p}) = (T_x(\mathbf{x}; \mathbf{p}), T_y(\mathbf{x}; \mathbf{p}))^T$, la dérivée première des coordonnées transformées par rapport aux paramètres de transformation (le *Jacobien*) est :

$$\frac{\partial T}{\partial \mathbf{p}} = \begin{pmatrix} \frac{\partial T_x}{\partial p_1} & \dots & \frac{\partial T_x}{\partial p_n} \\ \frac{\partial T_y}{\partial p_1} & \dots & \frac{\partial T_y}{\partial p_n} \end{pmatrix}. \quad (\text{A.7})$$

Pour un modèle de transformation perspective (8 paramètres), les coordonnées transformées sont $(x'$ et y' sont les coordonnées homogènes et w le facteur de normalisation en coordonnées standard) :

$$T(\mathbf{x}; \mathbf{p}) = \begin{bmatrix} p_1 & p_2 & p_3 \\ p_4 & p_5 & p_6 \\ p_7 & p_8 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \begin{bmatrix} p_1x + p_2y + p_3 \\ p_4x + p_5y + p_6 \\ p_7x + p_8y + 1 \end{bmatrix} = \begin{bmatrix} x' \\ y' \\ w \end{bmatrix}. \quad (\text{A.8})$$

La conversion des coordonnées homogènes en coordonnées standard est effectuée comme suit :

$$\begin{bmatrix} \frac{x'}{w} \\ \frac{y'}{w} \end{bmatrix} = \begin{bmatrix} \frac{p_1x + p_2y + p_3}{p_7x + p_8y + 1} \\ \frac{p_4x + p_5y + p_6}{p_7x + p_8y + 1} \end{bmatrix}. \quad (\text{A.9})$$

La dérivée des coordonnées transformées par rapport aux paramètres de transformation peuvent être obtenues comme suit. En sachant que :

$$\frac{d \left[\frac{f(x)}{g(x)} \right]}{dx} = \frac{g(x)f'(x) - f(x)g'(x)}{[g(x)]^2}.$$

Les dérivées des coordonnées x par rapport aux paramètres sont :

$$\frac{\partial T_x}{\partial p_1} = \frac{(p_7x + p_8y + 1)x - (p_1x + p_2y + p_3) \cdot (0)}{(p_7x + p_8y + 1)^2} = \frac{x}{p_7x + p_8y + 1},$$

$$\frac{\partial T_x}{\partial p_2} = \frac{(p_7x + p_8y + 1)y - (p_1x + p_2y + p_3) \cdot (0)}{(p_7x + p_8y + 1)^2} = \frac{y}{p_7x + p_8y + 1},$$

$$\frac{\partial T_x}{\partial p_3} = \frac{(p_7x + p_8y + 1)(1) - (p_1x + p_2y + p_3) \cdot (0)}{(p_7x + p_8y + 1)^2} = \frac{1}{p_7x + p_8y + 1},$$

$$\frac{\partial T_x}{\partial p_4} = 0, \quad \frac{\partial T_x}{\partial p_5} = 0, \quad \frac{\partial T_x}{\partial p_6} = 0,$$

$$\frac{\partial T_x}{\partial p_7} = \frac{(p_7x + p_8y + 1) \cdot (0) - (p_1x + p_2y + p_3)x}{(p_7x + p_8y + 1)^2} = \frac{-x(p_1x + p_2y + p_3)}{(p_7x + p_8y + 1)^2},$$

$$\frac{\partial T_x}{\partial p_8} = \frac{(p_7x + p_8y + 1) \cdot (0) - (p_1x + p_2y + p_3)y}{(p_7x + p_8y + 1)^2} = \frac{-y(p_1x + p_2y + p_3)}{(p_7x + p_8y + 1)^2},$$

et les dérivées des coordonnées y par rapport aux paramètres de transformation :

$$\frac{\partial T_y}{\partial p_1} = 0, \quad \frac{\partial T_y}{\partial p_2} = 0, \quad \frac{\partial T_y}{\partial p_3} = 0,$$

$$\frac{\partial T_y}{\partial p_4} = \frac{(p_7x + p_8y + 1)x - (p_4x + p_5y + p_6) \cdot (0)}{(p_7x + p_8y + 1)^2} = \frac{x}{p_7x + p_8y + 1},$$

$$\frac{\partial T_y}{\partial p_5} = \frac{(p_7x + p_8y + 1)y - (p_4x + p_5y + p_6) \cdot (0)}{(p_7x + p_8y + 1)^2} = \frac{y}{p_7x + p_8y + 1},$$

$$\frac{\partial T_y}{\partial p_6} = \frac{(p_7x + p_8y + 1)(1) - (p_4x + p_5y + p_6) \cdot (0)}{(p_7x + p_8y + 1)^2} = \frac{1}{p_7x + p_8y + 1},$$

$$\frac{\partial T_y}{\partial p_7} = \frac{(p_7x + p_8y + 1) \cdot (0) - (p_4x + p_5y + p_6)x}{(p_7x + p_8y + 1)^2} = \frac{-x(p_4x + p_5y + p_6)}{(p_7x + p_8y + 1)^2},$$

$$\frac{\partial T_y}{\partial p_8} = \frac{(p_7x + p_8y + 1) \cdot (0) - (p_4x + p_5y + p_6)y}{(p_7x + p_8y + 1)^2} = \frac{-y(p_4x + p_5y + p_6)}{(p_7x + p_8y + 1)^2}.$$

La forme matricielle des dérivées partielles par rapport aux paramètres de transformation est obtenue en injectant les relations précédentes dans l'équation A.7 :

$$\frac{\partial T}{\partial \mathbf{p}} = \frac{1}{p_7x + p_8y + 1} \begin{bmatrix} x & y & 1 & 0 & 0 & 0 & \frac{-x(p_1x + p_2y + p_3)}{p_7x + p_8y + 1} & \frac{-y(p_1x + p_2y + p_3)}{p_7x + p_8y + 1} \\ 0 & 0 & 0 & x & y & 1 & \frac{-x(p_4x + p_5y + p_6)}{p_7x + p_8y + 1} & \frac{-y(p_4x + p_5y + p_6)}{p_7x + p_8y + 1} \end{bmatrix} \quad (\text{A.10})$$

Pour minimiser la fonction coût, nous linéarisons l'équation A.2 en utilisant l'équation A.4 pour obtenir :

$$\sum_{\mathbf{x}} \left[I_D(T(\mathbf{x}; \mathbf{p})) + \nabla I_D \frac{\partial T}{\partial \mathbf{p}} \Delta \mathbf{p} - I_R(\mathbf{x}) \right]^2, \quad (\text{A.11})$$

Cette minimisation conduit à un problème de moindres carrés dont la solution peut être obtenue en calculant la dérivée première de l'équation A.11 selon $\Delta \mathbf{p}$:

$$2 \sum_{\mathbf{x}} \left[\nabla I_D \frac{\partial T}{\partial \mathbf{p}} \right]^T \left[I_D(T(\mathbf{x}; \mathbf{p})) + \nabla I_D \frac{\partial T}{\partial \mathbf{p}} \Delta \mathbf{p} - I_R(\mathbf{x}) \right]. \quad (\text{A.12})$$

En ne tenant pas compte de la multiplication par 2, nous pouvons obtenir les incréments $\Delta \mathbf{p}$ à ajouter aux paramètres \mathbf{p} pour minimiser l'équation A.2 en développant et en annulant l'équation A.12 :

$$\sum_{\mathbf{x}} \left[\left[\nabla I_D \frac{\partial T}{\partial \mathbf{p}} \right]^T I_D(T(\mathbf{x}; \mathbf{p})) + \left[\nabla I_D \frac{\partial T}{\partial \mathbf{p}} \right]^T \left[\nabla I_D \frac{\partial T}{\partial \mathbf{p}} \right] \Delta \mathbf{p} - \left[\nabla I_D \frac{\partial T}{\partial \mathbf{p}} \right]^T I_R(\mathbf{x}) \right] = 0.$$

Ensuite, nous isolons $\Delta \mathbf{p}$:

$$\Delta \mathbf{p} = \sum_{\mathbf{x}} \frac{\left[\nabla I_D \frac{\partial T}{\partial \mathbf{p}} \right]^T I_R(\mathbf{x}) - \left[\nabla I_D \frac{\partial T}{\partial \mathbf{p}} \right]^T I_D(T(\mathbf{x}; \mathbf{p}))}{\left[\nabla I_D \frac{\partial T}{\partial \mathbf{p}} \right]^T \left[\nabla I_D \frac{\partial T}{\partial \mathbf{p}} \right]},$$

$$\Delta \mathbf{p} = H^{-1} \sum_{\mathbf{x}} \left[\nabla I_D \frac{\partial T}{\partial \mathbf{p}} \right]^T [I_R(\mathbf{x}) - I_D(T(\mathbf{x}; \mathbf{p}))] \quad (\text{A.13})$$

avec

$$H = \sum_{\mathbf{x}} \left[\nabla I_D \frac{\partial T}{\partial \mathbf{p}} \right]^T \left[\nabla I_D \frac{\partial T}{\partial \mathbf{p}} \right]. \quad (\text{A.14})$$

H est l'approximation que l'algorithme de Gauss-Newton fait du Hessien (matrice des dérivées secondes). H va contrôler les pas effectués par l'algorithme à chaque itération.

De cette façon nous allons actualiser à chaque itération les paramètres de transformation minimisant l'équation A.2.

B

Autres images panoramiques

Cette annexe contient les figures fournissant le reste des images panoramiques construites à partir d'images cystoscopiques cliniques.

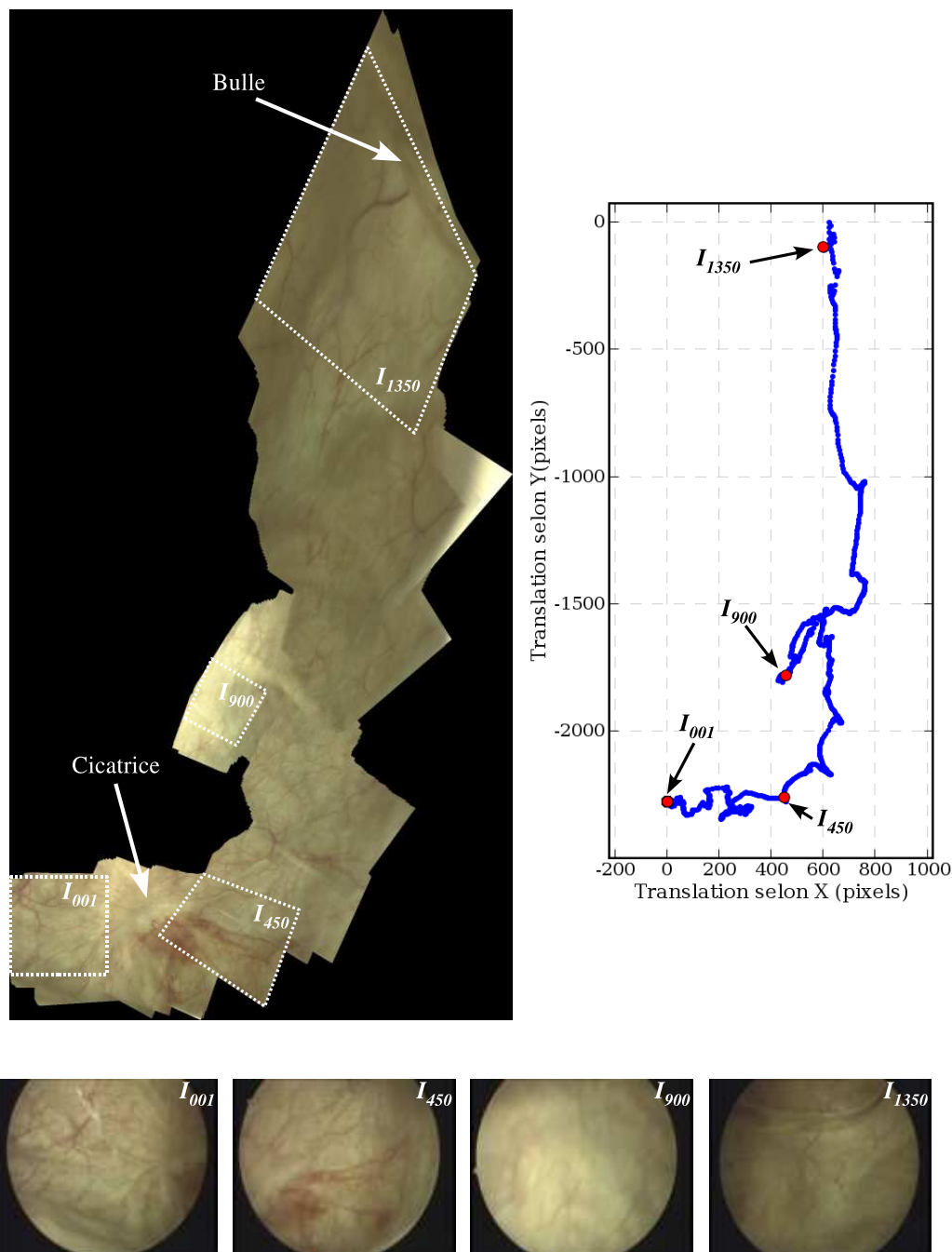


FIG. B.1 – Image panoramique construite à partir de 1350 images. La taille de l'image est de 1319×2647 pixels. Nous pouvons apercevoir une « bulle » dans la partie supérieure de l'image et dans la partie inférieure droite un orifice d'uretère. Cette vue ne serait pas possible pendant un examen cystoscopique. À droite, le graphique qui montre le parcours suivi par le cystoscope pendant l'acquisition de cette séquence (trajectoire du coins supérieur gauche des images recalées et placées dans le repère de la mosaïque).

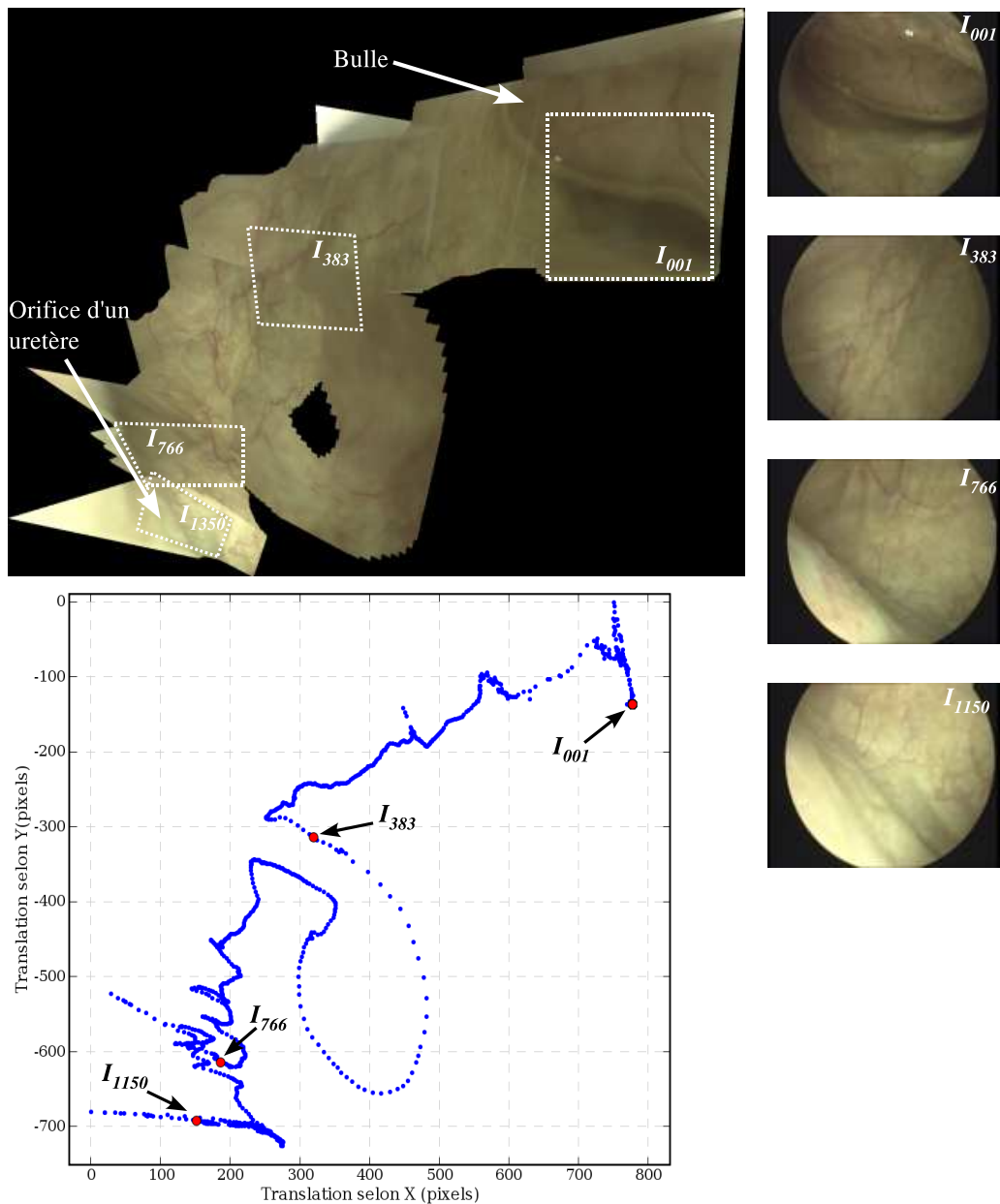


FIG. B.2 – Image panoramique de taille 1073×827 pixels, formée de 1150 images. Dans la partie inférieure gauche de l'image nous pouvons observer un orifice d'uretère. A gauche de l'image, une partie de la paroi peut être observée. Sur cette image, autour de l'image I_{383} il est possible d'observer des imprécision de mosaïquage des images. Cela est dû aux mouvements trop rapides du cystoscope.

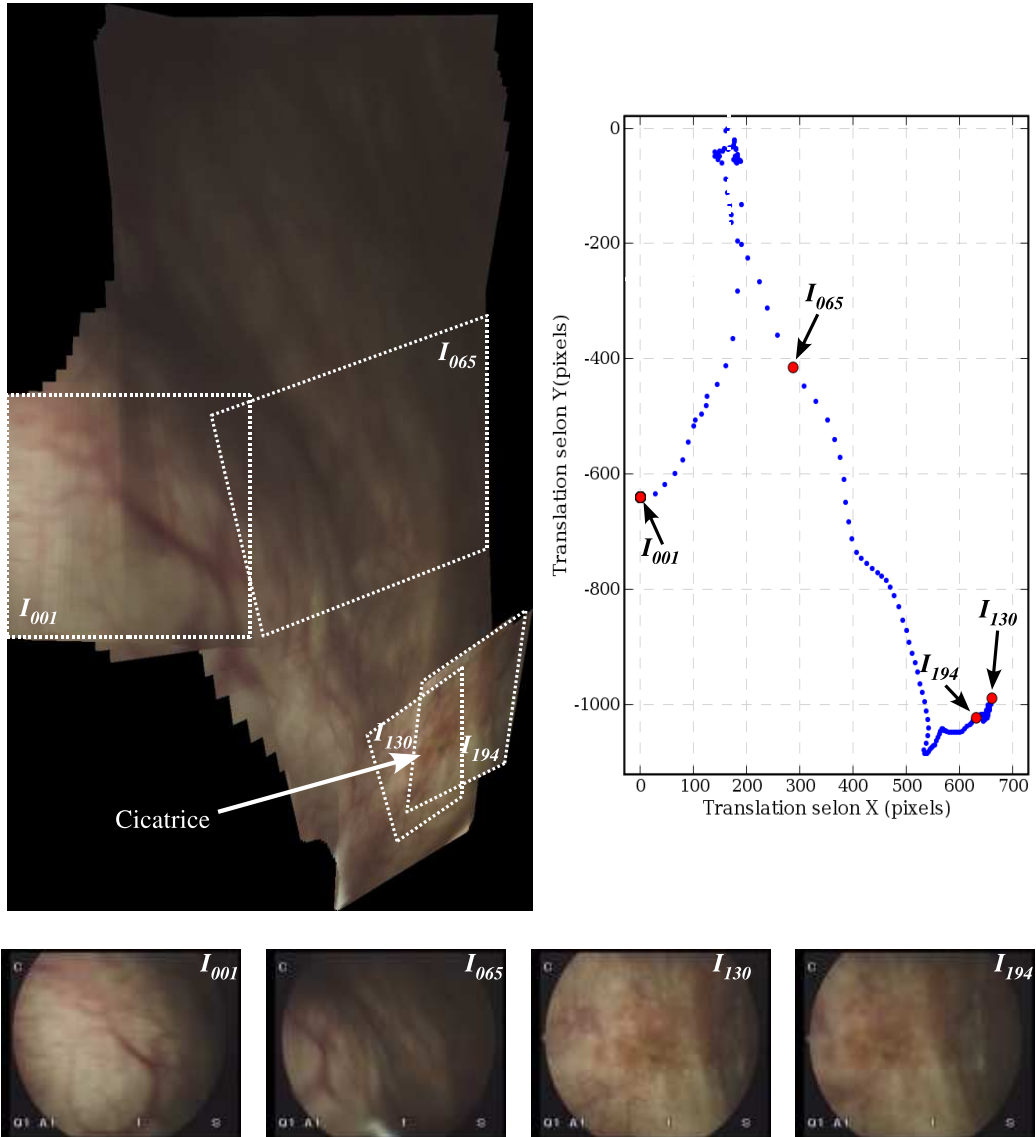


FIG. B.3 – 194 images ont été utilisées pour construire cette image panoramique (859×1484 pixels). Dans cette image, en bas à droite, il est possible de visualiser une cicatrice.

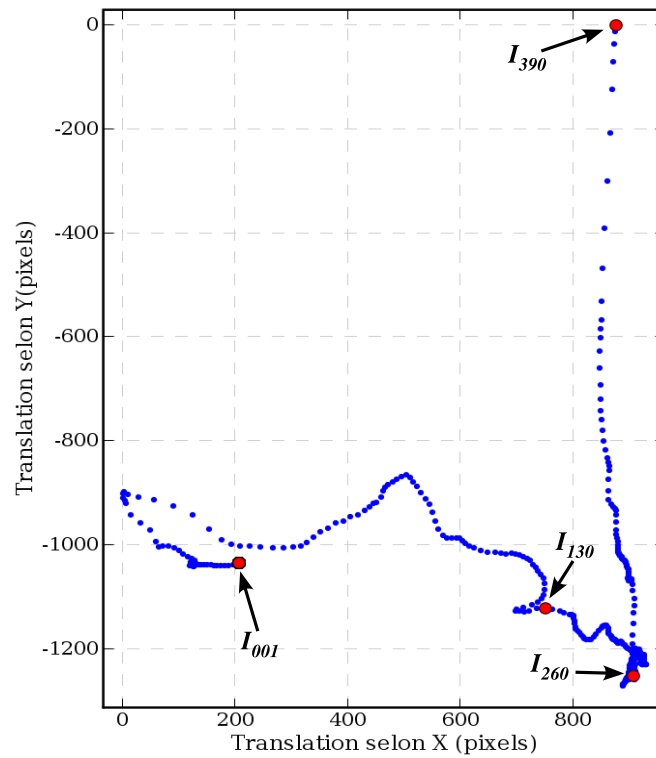
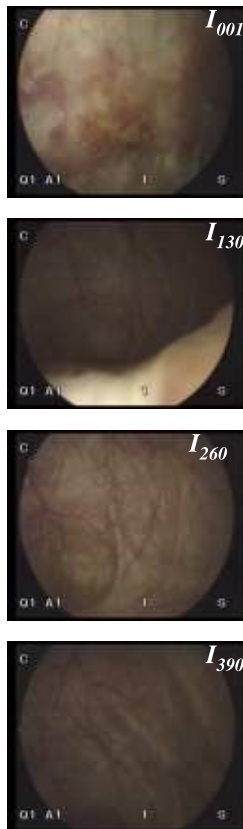
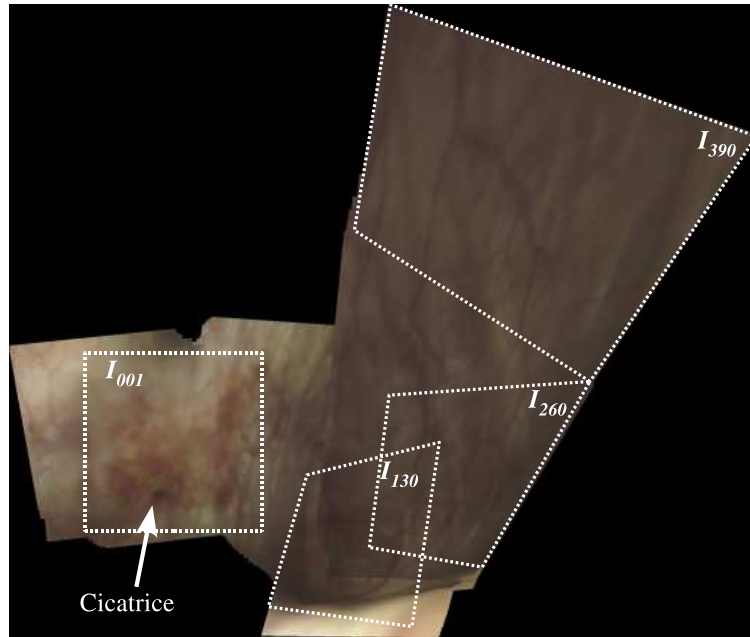


FIG. B.4 – Pour construire cette image de taille 2072×1755 pixels, 390 images ont été utilisées. Cette image est complémentaire à l'image B.3. Ici, nous observons que la cicatrice est située près d'une paroi de la vessie.

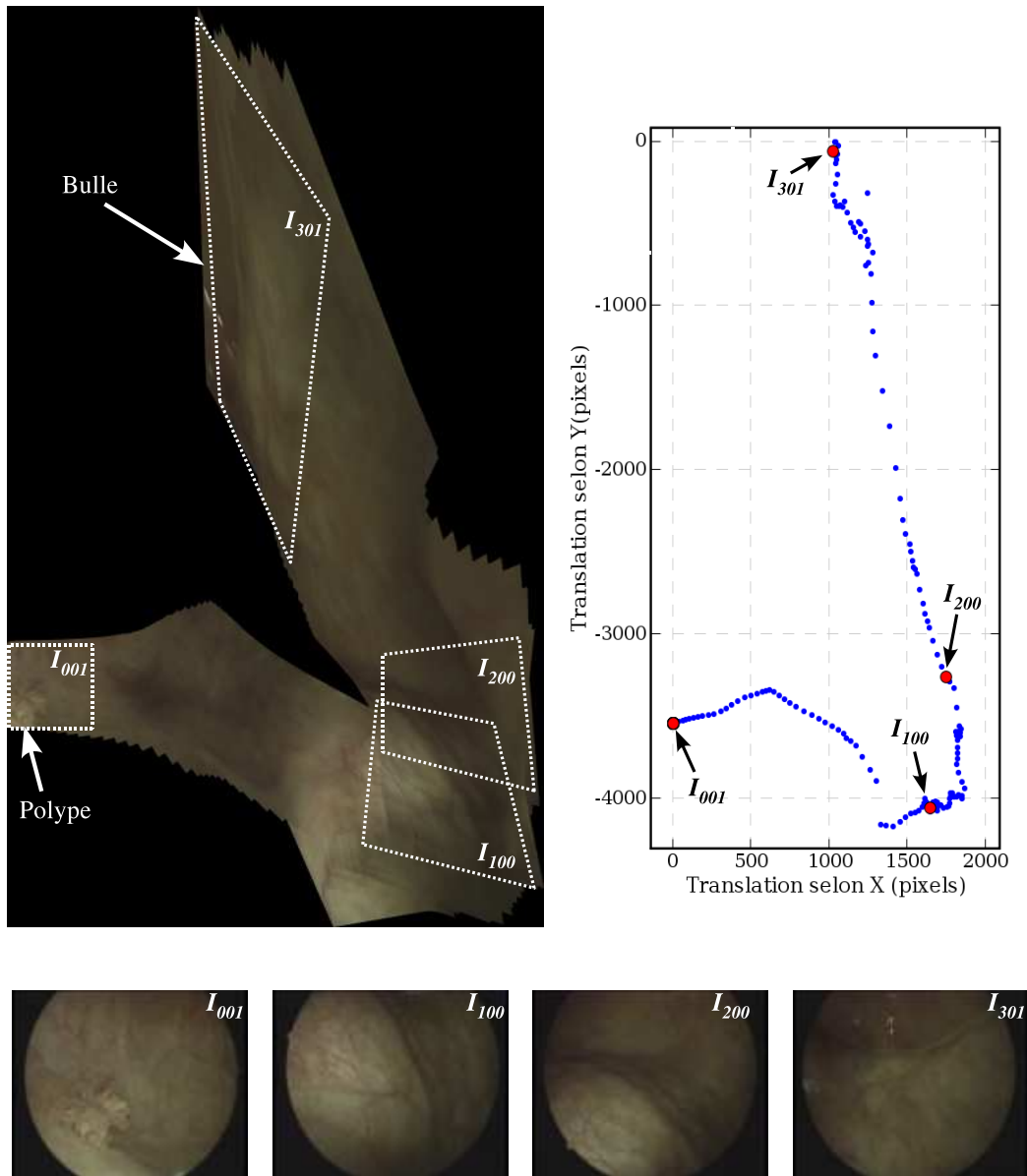


FIG. B.5 – Dans cette image de taille 2544×4362 pixels et construite avec 301 images, il est possible de situer un polype (visible à gauche en bas de l'image) par rapport à la bulle visible dans la partie supérieure.

Bibliographie

- [Abelev and Kopnin, 2000] G. I. Abelev and B. P. Kopnin. Biochemistry of cancer : An introduction to this issue. *Biochemistry (Moscow)–Translated from Biokhimiya*, 65(1) :3–4, 2000.
- [Alterovitz *et al.*, 2004] R. Alterovitz, K. Goldberg, J. Kurhanewicz, J. Pouliot, and I. C. Hsu. Image registration for prostate MR spectroscopy using biomechanical modeling and optimization of force and stiffness parameters. In *Proceedings of the 26th International Conference of the IEEE Engineering in Medicine and Biology Society (EMBS)*, volume 1, pages 1722–1725, 2004.
- [Asari *et al.*, 1999] K. V. Asari, S. Kumar, and D. Radhakrishnan. Technique of distortion correction in endoscopic images using a polynomial expansion. *Medical & Biological Engineering & Computing*, 37(1) :8 – 12, January 1999.
- [Baker and Matthews, 2002] S. Baker and I. Matthews. Lucas-Kanade 20 years on : A unifying framework : Part 1. Technical Report CMU-RI-TR-02-16, Robotics Institute, Carnegie Mellon University, Pittsburgh, PA, July 2002.
- [Baudisch *et al.*, 2005] P. Baudisch, D. Tan, D. Steedly, E. Rudolph, M. Uyttendaele, C. Pal, and R. Szeliski. Panoramic viewfinder : providing a real-time preview to help users avoid flaws in panoramic pictures. In *OZCHI '05 : Proceedings of the 19th Conference of the computer-human interaction special interest group (CHISIG) of Australia on Computer-human interaction*, pages 1–10, 2005.
- [Bay *et al.*, 2006] H. Bay, T. Tuytelaars, and L. Van Gool. SURF : Speeded up robust features. In *Proceedings of the 9th European Conference on Computer Vision (ECCV2006)*, volume 3951, pages 404–417, May 2006.
- [Black and Anandan, 1991] M. J. Black and P. Anandan. Robust dynamic motion estimation over time. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'91)*, volume 31, pages 296–302, 1991.
- [Blondel *et al.*, 2005] W. Blondel, C. Daul, S. Villette, R. Miranda-Luna, Y. Hernandez-Mier, G. Bourg-Heckly, and D. Wolf. Autofluorescence endoscopic spectro-imaging and 2D-cartography for *in situ* localisation and diagnosis of cancerous lesions. *Traitement du signal*, 23 :213–233, 2005.
- [Bourg-Heckly *et al.*, 2000] G. Bourg-Heckly, J. Blais, J. J. Padilla, O. Bourdon, J. Etienne, F. Guillemain, and L. Lafay. Endoscopic ultraviolet-induced autofluorescence spectroscopy of the esophagus : tissue characterization and potential for early cancer diagnosis. *Endoscopy*, 32 :756–765, 2000.
- [Boyle and Ferlay, 2005] P. Boyle and J. Ferlay. Cancer incidence and mortality in europe, 2004. *Annals of Oncology*, 16 :481–488, 2005.

- [Brown and Lowe, 2003] M. Brown and D. G. Lowe. Recognising panoramas. In *Proceedings of the 9th IEEE International Conference on Computer Vision (ICCV2003)*, volume 2, pages 1218–1227, October 2003.
- [Brown, 1992] L. G. Brown. A survey of image registration techniques. *ACM Computing Surveys*, 24(4) :325–376, 1992.
- [Burt and Adelson, 1983] P. J. Burt and E. H. Adelson. The laplacian pyramid as a compact image code. *IEEE Transactions on Communications*, 31(4) :532–540, April 1983.
- [Calhoun *et al.*, 1998] V. Calhoun, T. Adali, and G. Pearlson. A frequency-space approach for motion correction in functional MRI. In *Proceedings of the 10th Image and MultiDimensional Signal Processing (IMDSP) Annual Meeting*, page 229, 1998.
- [Can *et al.*, 2002] A. Can, C. V. Stewart, B. Roysam, and H. L. Tanenbaum. A feature-based, robust, hierarchical algorithm for registering pairs of images of the curved human retina. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(3) :347–364, March 2002.
- [Canny, 1986] J. Canny. A computational approach to edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 8(6) :679–698, 1986.
- [Capel and Zisserman, 2003] D. Capel and A. Zisserman. Computer vision applied to super-resolution. *IEEE Signal Processing Magazine*, 20(3) :75–86, 2003.
- [Capel, 2001] D. P. Capel. *Image Mosaicing and super-resolution*. PhD thesis, Robotics Research Group. Department of Engineering Science. University of Oxford, 2001.
- [Cheeseman *et al.*, 1994] P. Cheeseman, B. Kanefsky, Richard Kraft, J. Stutz, and R. Hanson. Super-resolved surface reconstruction from multiple images. Technical Report FIA-94-12, NASA Artificial Intelligence Research Branch, Moffett Field, CA 94035, USA, Décembre 1994.
- [Chmielewski *et al.*, 2002] L. Chmielewski, P. F. Kukulowicz, P. Gut, and A. Dabrowski. Assessment of the quality of radiotherapy with the use of portal and simulation images - the method and the software. *Journal of Medical Informatics & Technologies*, 3 :MI-171 – MI-179, November 2002.
- [Chopin and Gattengo, 2001] D. K. Chopin and B. Gattengo. *Résumé du rapport 2001 de l'association française d'urologie (AFU) sur les tumeurs superficielles de la vessie*, 2001.
- [Cuzol *et al.*, 2005] A. Cuzol, P. Hellier, and E. Mémin. A novel parametric method for non-rigid image registration. In *Proceedings of the 19th International Conference on Information Processing in Medical Imaging, IPMI'2005*, volume 3565, pages 456–467, 2005.
- [Davatzikos, 1996] C. Davatzikos. Nonlinear registration of brain images using deformable models. In *Proceedings of the IEEE Workshop on Mathematical Methods in Biomedical Image Analysis*, pages 94–103, June 1996.
- [De Castro and Morandi, 1987] E. De Castro and C. Morandi. Registration of translated and rotated images using finite Fourier transforms. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 9(5) :700–703, 1987.
- [D'Hallewin *et al.*, 2002] M.-A. D'Hallewin, L. Bezdetnayab, and F. Guillemin. Fluorescence detection of bladder cancer : A review. *European Urology*, 42 :417–425, 2002.
- [Douze, 2004] M. Douze. *Estimation d'homographies inter-images. Cas des mosaïques et du suivi en temps réel. Applications en réalité augmentée*. PhD thesis, Institut National Polytechnique de Toulouse, 2004.
- [DREES, 2006] Direction de la recherche, des études, de l'évaluation et des statistiques DREES. Ministère de l'emploi, de la cohésion sociale et du logement et ministère de la santé et des

solidarités. *Données sur la situation sanitaire et sociale en France en 2005. Annexe 1 au projet de loi de financement de la sécurité sociale pour 2006*, 2006.

- [Dydenko *et al.*, 2003] I. Dydenko, D. Friboulet, and I. E. Magnin. A variational framework for affine registration and segmentation with shape prior : application in echocardiographic imaging. In *Proceedings of the Ninth IEEE International Conference on Computer Vision (ICCV)*, pages 926–938, 2003.
- [Ericsson and Åström, 2003] A. Ericsson and K. Åström. Minimizing the description length using steepest descent. In *Proceedings of the British Machine Vision Conference, Norwich, United Kingdom*, volume 2, pages 93–102, 2003.
- [Fedorov *et al.*, 2002] D. Fedorov, L. M. G. Fonseca, C. Kenney, and B. S. Manjunath. Automatic registration and mosaicking system for remotely sensed imagery. In *In Proceedings of the 9th International Symposium on Remote Sensing (SPIE)*, September 2002.
- [Fischler and Bolles, 1981] M. A. Fischler and R. C. Bolles. Random sample consensus : A paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, 24(6) :381–395, June 1981.
- [Fitzgibbon, 2001] A. W. Fitzgibbon. Robust registration of 2D and 3D point sets. In *Proceedings of the British Machine Vision Conference 2001, BMVC 2001*, page 10, September 2001.
- [Forsyth and Ponce, 2003] D. A. Forsyth and J. Ponce. *Computer Vision : A Modern Approach*. Prentice Hall Professional Technical Reference, 2003.
- [Gattengo and Chopin, 2001] B. Gattengo and D. Chopin. *Progrès en Urologie*, volume 11, chapter Diagnostic, pages 993–1019. Urofrance, 2001.
- [Gay-Bellile *et al.*, 2006] V. Gay-Bellile, M. Perriollat, A. Bartoli, and P. Sayd. Image registration by combining thin-plate splines with a 3D morphable model. In *Proceedings of the International Conference on Image Processing (ICIP '06)*, 2006.
- [Gee *et al.*, 2003] A. H. Gee, G. M. Treece, R. W. Prager, C. J. C. Cash, and L. Berman. Rapid registration for wide field-of-view freehand three-dimensional ultrasound. *IEEE Transactions on Medical Imaging*, 22(11) :1344–1357, 2003.
- [Goujon *et al.*, 2003] D. Goujon, M. Zellweger, A. Radu, P. Grosjean, B. C. Weber, H. van Den Bergh, P. Monnier, and G. Wagnieres. *In vivo* autofluorescence imaging of early cancers in the human tracheobronchial tree with a spectrally optimised system. *Journal of Biomedical Optics*, 8(1) :17–25, 2003.
- [Gracias and Negahdaripour, 2005] N. Gracias and S. Negahdaripour. Underwater mosaic creation using video sequences from different altitudes. In *In Proceedings of the MTS/IEEE OCEANS '05*, volume 1, pages 1234–1240, 2005.
- [Gracias *et al.*, 2002] N. Gracias, S. van der Zwaan, A. Bernardino, and J. Santos-Victor. Results on underwater mosaic-based navigation. In *Proceedings of the MTS/IEEE OCEANS 2002*, page 1234, 2002.
- [Guillén González *et al.*, 1998] M. Guillén González, P. Holifield, and M. Varley. Improved video mosaic construction by accumulated alignment error distribution. In *Proceedings of the British Machine Vision Conference 1998, BMVC 1998*, volume 2, pages 377–387, 1998.
- [Harris and Stephens, 1988] C. Harris and M. J. Stephens. A combined corner and edge detector. In *Proceedings of the Alvey Vision Conference*, pages 147–152, 1988.
- [Hartley and Zisserman, 2004] R. I. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, 2004.

- [Heikkilä and Pietikäinen, 2005] M. Heikkilä and M. Pietikäinen. An image mosaicing module for wide-area surveillance. In *Proceedings of the third ACM International Workshop on Video surveillance & Sensor Networks (VSSN '05)*, pages 11–18, 2005.
- [Heikkilä, 2000] J. Heikkilä. Geometric camera calibration using circular control points. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(10) :1066–1077, 2000.
- [Heintzelman *et al.*, 2000] L. D. Heintzelman, U. Utzinger, H. Fuchs, A. Zuluaga, K. Gossage, A. Gillenwater, R. Jacob, B. Kemp, and R. Richards-Kortum. Optimal excitation wavelengths for in vivo detection of oral neoplasia using fluorescence spectroscopy. *Photochemistry and Photobiology*, 72(1) :103–113, 2000.
- [Hernandez-Mier *et al.*, 2005] Y. Hernandez-Mier, R. Miranda-Luna, W. C. P. M. Blondel, C. Daul, and D. Wolf. Fast construction of mosaics from video-endoscopic images using phase correlation registration. In *Proceedings of the Journées Sciences, Technologies et Imagerie pour la Médecine, JSTIM*, page 3, March 2005.
- [Hernández-Mier *et al.*, 2005] Y. Hernández-Mier, W. C. P. M. Blondel, C. Daul, D. Wolf, and G. Bourg-Heckly. Constructing panoramic views of internal walls of a bladder from cystoscopic image sequences. In *Proceedings of the European Biomedical Engineering Conference, EMBECE'05*, page 4, November 2005.
- [Hernández-Mier *et al.*, 2006] Y. Hernández-Mier, W. C. P. M. Blondel, C. Daul, and D. Wolf. 2-D panoramas from cystoscopic image sequences and potential application to fluorescence imaging. In *Proceedings of the Modelling and Control in Biomedical Systems, MCBMS'06*, page 4, September 2006.
- [Hoeksma *et al.*, 2005] M. R. Hoeksma, J. L. Kenemans, C. Kemner, and H. van Engeland. Variability in spatial normalization of pediatric and adult brain images. *Clinical Neurophysiology*, 166(5) :1188–1194, May 2005.
- [Hoge *et al.*, 2003] W. S. Hoge, H. Mamata, and S. E. Maier. Efficient construction of histology slide mosaics via phase correlation registration of high resolution tiles. In *Proceedings of the IEEE International Conference on Image Processing ICIP 2003*, September 2003.
- [Huang *et al.*, 2005] C. N. H. Huang, J. D. Lee, and J. H. Huang. Registration of CT image and facial surface data using adaptive genetic algorithm. *Biomedical Engineering - Applications, Basis and Communications*, 17(4) :201–206, 2005.
- [Huntsberger *et al.*, 1999] T. L. Huntsberger, E. T. Baumgartner, H. Aghazariana, Y. Chenga, P. S. Schenker, P. C. Legerb, K. D. Iagnemmac, and S. Dubowsky. Sensor-fused autonomous guidance of a mobile robot and applications to march sample return operations. In *In Proceedings of the International Society for Optical Engineering (SPIE)*, volume 3839, pages 2–8, 1999.
- [Hurskainen and Pellikka, 2005] P. Hurskainen and P. Pellikka. Change detection of informal settlements using multi-temporal aerial photographs – the case of voi, se-kenya. In *In Proceedings of the 5th AARSE conference (African Association of Remote Sensing of the Environment)*, October 2005.
- [Häußinger *et al.*, 2005] K. Häußinger, F. Stanzel H. Becker, A. Kreuzer, B. Schmidt, J. Strausz, S. Cavaliere, F. Herth, M. Kohlhaufl, K.-M. Muller, R.-M. Huber, U. Pichlmeier, and Ch. T. Bolliger. Autofluorescence bronchoscopy with white light bronchoscopy compared with white light bronchoscopy alone for the detection of precancerous lesions : a European randomised controlled multicentre trial. *Thorax*, 60 :496–503, 2005.

-
- [Irani *et al.*, 1996] M. Irani, P. Anandan, J. Bergen, R. Kumar, and S. Hsu. Efficient representations of video sequences and their applications. *Signal processing. Image communication*, 8(4) :327–351, May 1996.
- [Jang *et al.*, 2006] T. W. Jang, C. H. Oak, B. K. Chun, and M. H. Jung. Detection of pre-invasive endobronchial tumors with D-light/autofluorescence system. *Korean Journal of Medical Sciences*, 21 :242–246, 2006.
- [Johansson, 2006] M. Johansson. Image registration with simulated annealing and genetic algorithms. Master’s thesis, School of Computer Science and Engineering Royal Institute of Technology, Stockholm, Sweden, 2006.
- [Kara *et al.*, 2004] M. Kara, R. S. DaCosta, B. C. Wilson, N. E. Marcon, and J. Bergman. Autofluorescence-based detection of early neoplasia in patients with Barrett’s esophagus. *Digestive Diseases*, 22 :134–141, 2004.
- [Kirkpatrick *et al.*, 1983] S. Kirkpatrick, C. D. Gelatt, and M. P. Vecchi. Optimization by simulated annealing. *Science*, 220(4598) :671–680, 1983.
- [Kotsas, 2005] P. D. Kotsas. 2D rigid registration of MR scans using the 1d binary projections. *Enformatika ; Transactions on Engineering, Computing and Technology*, 9 :157–161, November 2005.
- [Kuglin and Hines, 1975] C. D. Kuglin and D. C. Hines. The phase correlation image alignment method. In *Proceedings of the 1975 IEEE International Conference Cybernetics Society*, pages 163–165, 1975.
- [Leong *et al.*, 2003] F. J. W.-M. Leong, M. Brady, and J. O’D. McGee. Correction of uneven illumination (vignetting) in digital microscopy images. *Journal of Clinical Pathology*, 56 :619–621, 2003.
- [Lin *et al.*, 2001] C.Y. Lin, M. Wu, J. A. Bloom, I. J. Cox, M. L. Miller, and Y. M. Lui. Rotation, scale, and translation resilient watermarking for images. *IEEE Transactions on Image Processing*, 10(5) :767–782, May 2001.
- [Litvin *et al.*, 2003] A. Litvin, J. Konrad, and W. C. Karl. Probabilistic video stabilization using kalman filtering and mosaicking. In *Proceedings of the IS&T/SPIE Symposium on Electronic Imaging, Image and Video*, January 2003.
- [Lomonosov *et al.*, 2006] E. Lomonosov, D. Chetverikov, and A. Ekárt. Pre-registration of arbitrarily oriented 3D surfaces using a genetic algorithm. *Pattern Recognition Letters*, 27(11) :1201–1208, 2006.
- [Lowe, 2004] D. G. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2) :91–110, 2004.
- [Lucas and Kanade, 1981] B. D. Lucas and T. Kanade. An iterative image registration technique with an application to stereo vision (darpa). In *Proceedings of the 1981 DARPA Image Understanding Workshop*, pages 121–130, April 1981.
- [Maes *et al.*, 1999] F. Maes, D. Vandermeulen, and P. Suetens. Comparative evaluation of multiresolution optimization strategies for multimodality image registration by maximization of mutual information. *Medical Image Analysis*, 3(4) :373–386, 1999.
- [Maintz and Viergever, 1998] J. B. A. Maintz and M. A. Viergever. A survey of medical image registration. *Medical Image Analysis*, 2 :1–36, 1998.
- [Matsushita *et al.*, 2005] Y. Matsushita, E. Ofek, X. Tang, and H.-Y. Shum. Full-frame video stabilization. In *Proceedings of 2005 IEEE International Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2005.

- [Matthies and Okutomi, 1989] L. Matthies and M. Okutomi. A bayesian foundation for active stereo vision. In *Proceedings of SPIE Conference 1198, Sensor Fusion II : Human and Machine Strategies*, pages 296–302, November 1989.
- [Mikolajczyk and Schmid, 2004] K. Mikolajczyk and C. Schmid. Scale & affine invariant interest point detectors. *International Journal of Computer Vision*, 60(1) :63–86, 2004.
- [Miranda-Luna *et al.*, 2004a] R. Miranda-Luna, W.C.P.M. Blondel, Ch. Daul, Y. Hernandez-Mier, R. Posada, and D. Wolf. A simplified method of endoscopic image distortion correction based on grey level registration. In *Proceedings of the 2004 International Conference on Image Processing (ICIP '04)*, volume 5, pages 3383–3386, 2004.
- [Miranda-Luna *et al.*, 2004b] R. Miranda-Luna, Y. Hernandez-Mier, Ch. Daul, W.C.P.M. Blondel, and D. Wolf. Mosaicing of medical video-endoscopic images : data quality improvement and algorithm testing. In *Proceedings of the 1st International Conference on Electrical and Electronics Engineering, 2004 (ICEEE)*, pages 530–535, 2004.
- [Miranda-Luna *et al.*, 2005] R. Miranda-Luna, Y. Hernandez-Mier, W. C. P. M. Blondel, Ch. Daul, and D. Wolf. Mosaicing of video-endoscopic images and evaluation of the resulting cartography. In *Proceedings of the Imaging for medical and life sciences (IMVIE2) Symposium*, page 4, March 2005.
- [Miranda-Luna *et al.*, 2007] R. Miranda-Luna, C. Daul, W. C. P. M. Blondel, Y. Hernandez-Mier, and D. Wolf. Image mosaicing of endoscopic video-sequences : registration algorithm and endoscope calibration. *IEEE Transactions on Biomedical Engineering*, Accepted, 2007.
- [Miranda Luna, 2005] R. Miranda Luna. *Contribution au mosaïquage d'images endoscopiques d'organes creux pour l'aide au diagnostic du cancer en phase précoce*. PhD thesis, Institut National Polytechnique de Lorraine, 2005.
- [Moravec, 1979] H. P. Moravec. Visual mapping by a robot rover. In *Proceedings of the International Joint Conference on Artificial Intelligence*, pages 598–600, 1979.
- [Moreno Velo *et al.*, 2001] F. J. Moreno Velo, I. Baturone, S. Sánchez Solano, and A. Barriga. XFSL : a tool for supervised learning of fuzzy systems. In *Proceedings of the 2001 European Symposium on Intelligent Technologies, Hybrid Systems and their implementation on Smart Adaptive Systems (EUNITE 2001)*, pages 58–59, 2001.
- [Moshir, 2005] M. Moshir. *Spitzer space telescope data processing and algorithmic complexity*. In *Astronomical Data Analysis Software and Systems XIV. ASP Conference Series*, volume 347, pages 434–443, 2005.
- [Mäkelä *et al.*, 2001] T. Mäkelä, P. Clarysse, J. Lötjönen, O. Sipilä, K. Lauerma, H. Hänninen, E.-P. Pyökkimies, J. Nenonen, J. Knuuti, T. Katila, and I. E. Magnin. A new method for the registration of cardiac PET and MR images using deformable model based segmentation of the main thorax structures. In *Proceedings of the 4th International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI'01)*, volume 2208, pages 557–564, 2001.
- [Palm *et al.*, 1998] Ch. Palm, A. Pelkmann T. Lehmann, and K. Spitzer. Distortion correction of videobased laryngoscopic images. In *Proceedings of the 3rd International Workshop in Advances in Quantitative Laryngoscopy, Voice and Speech Research*, pages 117–126, 1998.
- [Palmer *et al.*, 2003] G. M. Palmer, C. Zhu, T. M. Breslin, F. Xu, K. W. Gilchrist, and N. Ramanujam. Comparison of multiexcitation fluorescence and diffuse reflectance spectroscopy for the diagnosis of breast cancer (march 2003). *IEEE Transactions on Biomedical Engineering*, 50(11) :1233–1242, 2003.

-
- [Pirrone and Gambino, 2006] E. Ardizzone R. Pirrone and O. Gambino. Illumination correction on mr images. *Journal of Clinical Monitoring and Computing*, 20 :391–398, 2006.
- [Pluim *et al.*, 2003] J. P. W. Pluim, J. B. A. Maintz, and M. A. Viergever. Mutual-information-based registration of medical images : A survey. *IEEE Transactions on Medical Imaging*, 22(8) :986–1004, 2003.
- [Posada *et al.*, 2004] R. Posada, C. Daul, D. Wolf, P. Aletti, and R. Miranda. Towards a fractionated treatment in conformational radiotherapy using 3d-multimodal data registration. In *Proceedings of the IEEE International Conference on Image Processing (ICIP 2004)*, October 2004.
- [Posada *et al.*, 2007] R. Posada, C. Daul, D. Wolf, and P. Aletti. Towards a non invasive intracranial tumor irradiation using 3–D optical imaging and multimodal data registration. *International Journal of Biomedical Imaging, Special Issue on Multimodality Imaging and Hybrid Scanners*, 2007 :14, 2007.
- [Press *et al.*, 1992] William H. Press, Brian P. Flannery, Saul A. Teukolsky, and William T. Vetterling. *Numerical Recipes : The Art of Scientific Computing*. Cambridge University Press, Cambridge (UK) and New York, 2nd edition, 1992.
- [Prosst and Gahlen, 2002] R. L. Prosst and J. Gahlen. Fluorescence diagnosis of colorectal neoplasms : a review of clinical applications. *International Journal of Colorectal Diseases*, 17 :1–10, 2002.
- [Rankov *et al.*, 2005] V. Rankov, R. J. Locke, R. J. Edens, P. R. Barber, and B. Vojnovic. An algorithm for image stitching and blending. In *Proceedings of SPIE 2005*, March 2005.
- [Reddy and Chatterji, 1996] B. S. Reddy and B. N. Chatterji. An FFT-based technique for translation, rotation, and scale-invariant image registration. *IEEE Transactions on Image Processing*, 5(8) :1266–1271, 1996.
- [Rosten and Drummond, 2005] E. Rosten and T. Drummond. Fusing points and lines for high performance tracking. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV2005)*, volume 2, pages 1508–1511, October 2005.
- [Rzhanov *et al.*, 2002] Y. Rzhanov, L. Huff, and R.G. Cutter Jr. Seafloor video mapping : modeling, algorithms, apparatus. In *In Proceedings of the International Conference on Image Processing (ICIP2002)*, page 4, 2002.
- [Sato *et al.*, 2004] T. Sato, S. Ikeda, M. Kanbara, A. Iketani, N. Nakajima, N. Yokoya, and K. Yamada. High-resolution video mosaicing for documents and photos by estimating camera motion. In *Proceedings of the SPIE 2004*, pages 246–253, 2004.
- [Schmid *et al.*, 2000] C. Schmid, R. Mohr, and C. Bauckhage. Evaluation of interest point detectors. *International Journal of Computer Vision*, 37(2) :151–172, 2000.
- [Shi and Tomasi, 1994] J. Shi and C. Tomasi. Good features to track. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR’94)*, pages 593–600, June 1994.
- [Shum and Szeliski, 1997] H.-Y. Shum and R. Szeliski. Panoramic image mosaics. Technical Report MSR-TR-97-23, Microsoft Research, Microsoft Corporation, Redmond, WA 98052, December 1997.
- [Smith and Brady, 1997] S. M. Smith and J. M. Brady. SUSAN – A new approach to low level image processing. *International Journal of Computer Vision*, 23(1) :45–78, 1997.

- [Soman *et al.*, 2003] S. Soman, A. C. S. Chung, W. E. L. Grimson, and W. M. Wells III. Rigid registration of echoplanar and conventional magnetic resonance images by minimizing the kullback-leibler distance. In *Proceedings of the 2nd International Workshop on Biomedical Image Registration*, volume 2717, pages 181–190, 2003.
- [Steadly *et al.*, 2005] D. Steedly, C. Pal, and R. Szeliski. Efficiently registering video into panoramic mosaics. In *In Proceedings of the IEEE International Conference on Computer Vision (ICCV 2005)*, pages 15–21, October 2005.
- [Svistun *et al.*, 2004] E. Svistun, R. Alizadeh-Naderi, A. El-Naggar, R. Jacob, A. Gillenwater, and R. Richards-Kortum. Vision enhancement system for detection of oral cavity neoplasia based on autofluorescence. *Head Neck*, 26 :205–215, 2004.
- [Svoboda, 2005] T. Svoboda. A software for complete calibration of multicamera systems. In *SPIE-IS&T Electronic Imaging, SPIE 2005*, volume 5658, pages 115–128, 2005.
- [Szeliski and Shum, 1997] R. Szeliski and H.-Y. Shum. Creating full view panoramic image mosaics and environment maps. In *SIGGRAPH '97 : Proceedings of the 24th annual conference on Computer graphics and interactive techniques*, volume 31, pages 251–258, 1997.
- [Szeliski, 1996] R. Szeliski. Video mosaics for virtual environments. *IEEE Computer Graphics and Applications*, 16(2) :22–30, March 1996.
- [Telenczuk *et al.*, 2006] B. Telenczuk, M. J. Ledesma-Carbato, J. A. Velazquez-Muriel, C. O. S. Sorzano, J.-M. Carazo, and A. Santos. Molecular image registration using mutual information and differential evolution optimization. In *Proceedings of the 3rd IEEE International Symposium on Biomedical Imaging*, pages 844–847, April 2006.
- [Thompson and Toga, 1996] P. M. Thompson and A. W. Toga. A surface-based technique for warping three-dimensional images of the brain. *IEEE Transactions on Medical Imaging*, 15(4) :402–417, 1996.
- [Triggs *et al.*, 2000] B. Triggs, P. McLauchlan, R. Hartley, and A. Fitzgibbon. *Vision Algorithms : Theory and Practice*, volume 11 of *Lecture Notes in Computer Science*, chapter Bundle Adjustment – A Modern Synthesis, pages 298–372. Springer-Verlag, 2000.
- [Tsai, 1987] R. Y. Tsai. A versatile camera calibration technique for high-accuracy 3d machine vision metrology using off-the-shelf tv cameras and lenses. *IEEE Journal of Robotics and Automation*, 3(4) :323–344, August 1987.
- [Tyczynski and Parkin, 2003] J. E. Tyczynski and D. M. Parkin. Bladder cancer in Europe. *ENCR Cancer Fact Sheets. European Network of Cancer Registries. International Agency for Research on Cancer*, 3, September 2003.
- [Vercauteren *et al.*, 2005] T. Vercauteren, A. Perchant, X. Pennec, and N. Ayache. Mosaicing of confocal microscopic in vivo soft tissue video sequences. In *Proceedings of the 8th International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI2005)*, pages 753–760, October 2005.
- [Verhey *et al.*, 2005] J. F. Verhey, J. Wisserb, T. Kellerb, C. F. Westinc, and R. Kikinis. Rigid overlay of volume sonography and mr image data of the female pelvic floor using a fiducial based alignment-feasibility due to a case series. *Computerized Medical Imaging and Graphics*, 29(4) :243–249, June 2005.
- [Viola and Wells III, 1997] P. Viola and W. M. Wells III. Alignment by maximization of mutual information. *International Journal of Computer Vision*, 24(2) :137–154, 1997.

-
- [Vovk *et al.*, 2007] U. Vovk, F. Pernus, and B. Likar. A review of methods for correction of intensity inhomogeneity in mri. *IEEE Transactions on Medical Imaging*, 26(3) :405–421, March 2007.
- [Wagnières *et al.*, 1998] G. A. Wagnières, W. M. Star, and B. C. Wilson. *In Vivo* fluorescence spectroscopy and imaging for oncological applications. *Photochemistry and Photobiology*, 68(5) :603–632, 1998.
- [Wang and Brady, 1995] H. Wang and M. Brady. Real-time corner detection algorithm for motion estimation. *Image and Vision Computing*, 13(9) :695–703, 1995.
- [Wang and Tian, 2005] X. Wang and J. Tian. Image registration based on maximization of gradient code mutual information. *Image Analysis & Stereology*, 24 :1–7, March 2005.
- [WHO, 2006] WHO. World health organization. *Fact sheet N° 297*. WHO Media Center, Février 2006.
- [Winter *et al.*, 2006] C. Winter, S. Rupp, M. Elter, C. Münzenmayer, H. Gerhäuser, and T. Wittenberg. Technique of distortion correction in endoscopic images using a polynomial expansion. *IEEE Transactions on Biomedical Engineering*, 53(10) :2035–2046, October 2006.
- [Woods *et al.*, 1993] R. P. Woods, J. C. Mazziotta, and S. R. Cherry. MRI-PET registration with automated algorithm. *Journal of Computer Assisted Tomography*, 17 :536–546, 1993.
- [Wu *et al.*, 2006] M. Wu, C. Rosano, M. Butters, E. Whyte, M. Nable, R. Crooks, C. C. Meltzer, C. F. Reynolds III, and H. J. Aizenstein. A fully automated method for quantifying and localizing white matter hyperintensities on MR images. *Psychiatry Research : Neuroimaging*, 148(2-3) :133–142, Décembre 2006.
- [Xie *et al.*, 2003] H. Xie, N. Hicks, G. R. Keller, H. Huang, and V. Kreinovich. An IDL/ENVI implementation of the FFT-based algorithm for automatic image registration. *Computers & Geosciences*, 29 :1045–1055, March 2003.
- [Yaniv and Joskowicz, 2004] Z. Yaniv and L. Joskowicz. Long bone panoramas from fluoroscopic x-ray images. *IEEE Transactions on Medical Imaging*, 23(1) :26–35, 2004.
- [Zaak *et al.*, 2005] D. Zaak, A. Karl, R. Knüchel, H. Stepp, A. Hartmann, O. Reich, A. Bachmann, M. Siebels, G. Popken, and C. Stief. Diagnosis of urothelial carcinoma of the bladder using fluorescence endoscopy. *British Journal of Urology*, 96 :217–222, 2005.
- [Zeng *et al.*, 2004] H. Zeng, M. Petek, M. T. Zorman, A. McWilliams, B. Palcic, and S. Lam. Integrated endoscopy system for simultaneous imaging and spectroscopy for early lung cancer detection. *Optics Letters*, 29(6) :587–589, 2004.
- [Zhang and Rangarajan, 2004] J. Zhang and A. Rangarajan. A unified feature-based registration method for multimodality images. In *Proceedings of the 2004 IEEE International Symposium on Biomedical Imaging (ISBI)*, pages 724–727, 2004.
- [Zhang *et al.*, 2000] C. Zhang, J. P. Helferty, G. McLennan, and W. E. Higgins. Nonlinear distortion correction in endoscopic video images. In *Proceedings of the IEEE International Conference on Image Processing (ICIP 2000)*, volume 2, pages 439–442, 2000.
- [Zhang, 1998] Z. Zhang. A flexible new technique for camera calibration. Technical Report MSR-TR-98-71, Microsoft Research, Microsoft Corporation, Redmond, WA 98052, December 1998.
- [Zhu *et al.*, 2001] Z. Zhu, A R. Hanson, H. S. Bassali, H. Schultz, and E. M. Riseman. Generating seamless stereo mosaics from aerial video. In *In Proceedings of the ASPRS 18th Biennial Workshop on Color Photography & Videography in Resource Assessment*, May 2001.

- [Zitová and Flusser, 2003] B. Zitová and J. Flusser. Image registration methods : a survey. *Image and Vision Computing*, 21(11) :977–1000, October 2003.
- [Zoghlami *et al.*, 1997] I. Zoghlami, O. Faugeras, and R. Deriche. Using geometric corners to build a 2D mosaic from a set of images. In *Proceedings of the International Conference on Computer Vision and Pattern Recognition (CVPR'97)*, pages 420–425, June 1997.
- [Zomet and Peleg, 2000] A. Zomet and S. Peleg. Efficient super-resolution and applications to mosaics. In *In Proceedings of the 15th International Conference on Pattern Recognition (IC-PR'00)*, volume 1, pages 579–583, 2000.

Résumé

Cette thèse propose un algorithme de mosaïquage pour la construction d'images panoramiques des parois internes de la vessie. Le temps de construction de ces images correspondant aux parties intéressantes de la vessie doit être inférieur à la durée d'un examen clinique standard. La méthode de mosaïquage doit aussi être robuste vis-à-vis des variabilités inter-examens liées aux patients et aux instruments. Ces images panoramiques pourront être utilisées par le clinicien comme référence pour guider des examens ultérieurs, pour l'archivage des données et pour suivre l'évolution des lésions. La première étape de l'algorithme est le pré-traitement des images cystoscopiques consistant en l'atténuation des inhomogénéités d'illumination et du motif de fibres optiques visible dans les images acquises par un fibroscope. La deuxième étape est le recalage des images. La solution retenue consiste en la corrélation par les transformées de Fourier des images qui fournit des translations initiales à un algorithme itératif basé sur la différence d'intensité entre les images. Ce dernier délivre les paramètres de la transformation perspective reliant deux images successives de la séquence. Dans la troisième étape nous projetons les images dans un repère commun en utilisant des transformations globales calculées avec les résultats des recalages. Nous utilisons un moyennage pondéré des intensités des pixels pour atténuer les bords visibles lors de la projection. Les résultats quantitatifs obtenus avec un fantôme et des résultats qualitatifs calculés pour des séquences réelles montrent que notre approche automatique de mosaïquage est robuste et rapide (temps compatible avec la durée d'un examen cystoscopique clinique). Nos tests ont également prouvé que l'algorithme de recalage fonctionne pour des transformations géométriques plus grandes que celles rencontrées typiquement entre images d'une séquence vidéo (90% de recouvrement entre images consécutives pour ces dernières).

Mots-clés: Mosaïquage d'images, images panoramiques, cystoscopie, vessie, applicable en cancérologie.

Abstract

This work describes a mosaicing algorithm for constructing panoramic images of internal walls of the bladder. Time relating to the construction of panoramic images including the interesting parts must be shorter than that required by a standard cystoscopic examination. The mosaicing algorithm must be robust against lighting conditions, morphologic and texture variations relating to instruments and patient anatomy. These panoramic images could be used by a clinician for guiding further exams, storing non-redundant data and following-up evolution of lesions. The preprocessing of cystoscopic images is the first stage of the algorithm. Preprocessing consists of shading correction and fiber optics pattern attenuation occurring in fibroscope acquired images. The second stage is image registration. The chosen solution consist of cross-correlating images (using their Fourier transforms) in order to have initial translations for an iterative registration algorithm based on the sum of squared differences of images. In the third stage, images are projected in the coordinate system of the panoramic image using global transformations computed with matrices given by the iterative registration. We use a weighted average of pixel intensities to blend visible borders of images produced in the projection process. Numerical results obtained with a phantom and qualitative results obtained with real sequences show that our automatic approach is robust and allows for a fast construction of panoramic images in a period of time that is shorter than the duration of a clinical cystoscopic examination. Our experiments showed that the registration algorithm can handle geometric transformations that are larger than those existing typically in a video-sequence (90% of superposition between successive images in this case).

Keywords: Image mosaicing, panoramic images, cystoscopy, bladder, useful for cancer research.

