

Ecole Doctorale des Sciences pour l'Ingénieur
Université Blaise Pascal

Thèse

N° d'ordre : D. U : 2612
EDSPIC : 714

pour l'obtention du grade de
Docteur de l'université Blaise Pascal

spécialité
Vision par Ordinateur

préparée au sein des laboratoires
Institut Pascal - Université Blaise Pascal - UMR 6602
Image Science For Interventional Techniques - Université d'Auvergne - UMR 6284

Soutenue le 4 Novembre 2015 par
François Chadebecq

**Estimation de l'échelle absolue par vision
passive monofocale et application à la mesure
3D de néoplasies en imagerie coloscopique**

Marie-Odile Berger
Eric Marchand
Fabrice Meriaudeau
Thierry Ponchon
Christophe Tilmant
Adrien Bartoli

INRIA Nancy Grand Est
Université de Rennes 1
Université de Bourgogne
Université de Lyon 1
Université Blaise Pascal
Université d'Auvergne

Présidente
Rapporteur
Rapporteur
Examineur
Co-directeur
Directeur

Absolute Scale Estimation Using Passive Monofocal Vision and its Application to 3D Measurement of Neoplasias in Colonoscopy

ABSTRACT

Vision-based metrology devices generally embed stereoscopic sensors or active measurement systems. Most of the passive 3D reconstruction techniques (*Structure-from-Motion*, *Shape-from-Shading*) adapted to monocular vision suffer from scale ambiguity. Because the processing of image acquisition implies the loss of the depth information, there is an ambiguous relationship between the depth of a scene and the size of an imaged object.

This study deals with the estimation of the absolute scale of a scene using passive monofocal vision. Monofocal vision describes monocular system for which optical parameters are fixed. Such optical system are notably embedded within endoscopic systems used in colonoscopy. This minimally invasive technique allows endoscopists to explore the colon cavity and remove neoplasias (abnormal growths of tissue). Their size is an essential diagnostic criterion for estimating their rate of malignancy. However, it is difficult to estimate and erroneous visual estimations lead to neoplasias surveillance intervals being inappropriately assigned. The need to design a neoplasia measurement system is the core motivation for our study. In the first part of this manuscript, we review state-of-the-art vision-based metrology devices to provide context for our system. We then introduce monofocal optical systems and the specific image formation model used in our study.

The second part deals with the main contribution of our work. We first review in detail state-of-the-art *DfD* (*Depth-from-Defocus*) and *DfF* (*Depth-from-Defocus*) approaches. They are passive computer vision techniques that enable us to resolve scale ambiguity. Our core contribution is introduced in the following chapter. We define the Infocus-Breakpoint (IB) that allows us to resolve scale from a regular video. The IB is the lower limit of the optical system's depth of field. Our system relies on two novel technical modules : Blur-Estimating Tracking (BET) and Blur-

Model Fitting (BMF). BET allows us to simultaneously track an area of interest and estimate the optical blur information. BMF allows us to robustly extract the IB by fitting an optical blur model to the blur measurement estimated by the BET module. For the optical system is monofocal, the IB corresponds to a reference depth that can be calibrated. In the last chapter, we evaluate our method and propose a neoplasia measurement system adapted to the constraints in colonoscopy examination.

The last part of this manuscript is dedicated to a prospect of extension of our method by a generative approach. We present, as a preliminary study, a new *NRSfM (Non-Rigid Structure-from-Motion)* method allowing the scaled Euclidean 3D reconstruction of deformable surfaces. This approach is based on the simultaneous estimation of dense depth maps corresponding to a set of deformations as well as the *in-focus* color map of the flattened surface. We first review state-of-the-art methods for 3D reconstruction of deformable surfaces. We then introduce our new generative model as well as an alternation method allowing us to infer it.

Keywords : Computer Vision, Medical Imaging, Depth-from-Focus, Endoscopy, 3D Measurement.

Acknowledgement

This work has been supported by the French National Research Agency (ANR) through the TecSan program. The authors involved in the papers which were published during this PhD thesis thank Yansys Medical[®], as well as Professor Thierry Ponchon, head of the Gastroenterology department at Edouard Herriot Hospital (Lyon), Doctor Vincent Lepilliez and Doctor Sylvain Beorchia, private gastroenterologists.

Estimation de l'échelle absolue par vision passive monofocale et application à la mesure 3D de néoplasies en imagerie coloscopique

RÉSUMÉ

La majorité des dispositifs de métrologie basés vision sont équipés de systèmes optiques stéréo ou de systèmes de mesure externes dits actifs. Les méthodes de reconstruction tridimensionnelle (*Structure-from-Motion*, *Shape-from-Shading*) applicables à la vision monoculaire souffrent généralement de l'ambiguïté d'échelle. Cette dernière est inhérente au processus d'acquisition d'images qui implique la perte de l'information de profondeur de la scène. La relation entre la taille des objets et la distance de la prise de vue est équivoque.

Cette étude a pour objet l'estimation de l'échelle absolue d'une scène par vision passive monofocale. Elle vise à apporter une solution à l'ambiguïté d'échelle uniquement basée vision, pour un système optique monoculaire dont les paramètres internes sont fixes. Elle se destine plus particulièrement à la mesure des lésions en coloscopie. Cette procédure endoscopique (du grec *endom* : intérieur et *scopie* : vision) permet l'exploration et l'intervention au sein du côlon à l'aide d'un dispositif flexible (coloscope) embarquant généralement un système optique monofocal. Dans ce contexte, la taille des néoplasies (excroissances anormales de tissu) constitue un critère diagnostique essentiel. Cette dernière est cependant difficile à évaluer et les erreurs d'estimations visuelles peuvent conduire à la définition d'intervalles de temps de surveillance inappropriés. La nécessité de concevoir un système d'estimation de la taille des lésions coloniques constitue la motivation majeure de cette étude.

Nous dressons dans la première partie de ce manuscrit un état de l'art synoptique des différents systèmes de mesure basés vision afin de positionner notre étude dans ce contexte. Nous présentons ensuite le modèle de caméra monofocal ainsi que le modèle de formation d'image qui lui a été associé. Ce dernier est la base essentielle des travaux menés dans le cadre de cette thèse.

La seconde partie du manuscrit présente la contribution majeure de notre étude. Nous dressons

tout d'abord un état de l'art détaillé des méthodes de reconstruction 3D basées sur l'analyse de l'information de flou optique (*DfD (Depth-from-Defocus)* et *DfF (Depth-from-Defocus)*). Ces dernières sont des approches passives permettant, sous certaines contraintes d'asservissement de la caméra, de résoudre l'ambiguïté d'échelle. Elles ont directement inspiré le système de mesure par extraction du point de rupture de netteté présenté dans le chapitre suivant. Nous considérons une vidéo correspondant à un mouvement d'approche du système optique face à une région d'intérêt dont on souhaite estimer les dimensions. Notre système de mesure permet d'extraire le point de rupture nette/flou au sein de cette vidéo. Nous démontrons que, dans le cas d'un système optique monofocale, ce point unique correspond à une profondeur de référence pouvant être calibrée. Notre système est composé de deux modules. Le module *BET (Blur Estimating Tracking)* permet le suivi et l'estimation conjointe de l'information de mise au point d'une région d'intérêt au sein d'une vidéo. Le module *BMF (Blur Model Fitting)* permet d'extraire de façon robuste le point de rupture de netteté grâce à l'ajustement d'un modèle de flou optique. Une évaluation de notre système appliqué à l'estimation de la taille des lésions coloniques démontre sa faisabilité.

Le dernier chapitre de ce manuscrit est consacré à une perspective d'extension de notre approche par une méthode générative. Nous présentons, sous la forme d'une étude théorique préliminaire, une méthode *NRSfM (Non-Rigid Structure-from-Motion)* permettant la reconstruction à l'échelle de surfaces déformables. Cette dernière permet l'estimation conjointe de cartes de profondeurs denses ainsi que de l'image de la surface aplanie entièrement mise au point. Nous dressons dans une première partie de ce chapitre un état de l'art des approches de reconstruction 3D de surfaces déformables. Nous introduisons par la suite le modèle génératif proposé ainsi qu'une méthode d'alternation permettant son inférence.

Mots clés : Vision par Ordinateur, imagerie médicale, *Depth-from-Focus*, endoscopie, mesure 3D.

Remerciements

Cette étude doctorale a été financée dans le cadre du programme Technologie pour la Santé et l'autonomie (TecSan) de l'Agence National de la Recherche Française (ANR). Les auteurs des travaux publiés dans le cadre de cette thèse remercient la société Yansys Medical[®] ainsi que le Professeur Thierry Ponchon, chef de service du département d'Hépatogastroentérologie de l'hôpital Edouard Herriot (Lyon), et les Docteurs Vincent Lepilliez et Sylvain Beorchia, gastroentérologistes libéraux.

Table des matières

1	Introduction	13
1.1	Métrologie par vision active	14
1.2	Métrologie par vision passive	21
1.3	Contributions	25
2	Modélisation des systèmes de vision monofocale	27
2.1	Définition et implications d'un système monofocale	28
2.2	Modèles de formation d'images	30
3	Etat de l'art sur l'estimation de la profondeur à partir du flou optique	43
3.1	Depth-from-Focus	45
3.2	Depth-from-Defocus	49
3.3	Conclusion	61
4	Mesurer la profondeur à partir du point de rupture de netteté	63
4.1	Principe	64
4.2	<i>Blur Estimating Tracking (BET)</i> : suivi et estimation conjointe du flou optique	64
4.3	Évaluation de la méthode de suivi et d'estimation conjointe du flou optique	70
4.4	<i>Blur Model Fitting (BMF)</i> : extraction robuste du Point de Rupture de Netteté	75
4.5	Inférence des mesures 3D à partir de mesures 2D	79
4.6	Discussion	80
5	Mesurer la taille des néoplasies en coloscopie	81
5.1	L'examen coloscopique	81
5.2	Néoplasie maligne : critères anatomo-pathologiques prédictifs et ambiguïté d'échelle	84
5.3	Système de mesure des néoplasies : protocole clinique	85
5.4	Validation expérimentale	87
5.5	Discussion	91
6	Perspective d'extension par un modèle génératif	95
6.1	Méthodes géométriques pour la reconstruction 3D de surfaces déformables	97

6.2 Modélisation	104
6.3 <i>A priori</i> du modèle	107
6.4 Fonction de vraisemblance	109
6.5 Inférence du modèle	109
6.6 Conclusion	112
Conclusion	114
Bibliographie	118

Table des figures

1.1.1	Photogrammétrie par insertion de cibles	15
1.1.2	Système de suivi visuel	16
1.1.3	Système de suivi électromagnétique	17
1.1.4	Reconstruction 3D par projection de lumière structurée	18
1.1.5	Système de mesure laser	19
1.1.6	Caméra plénoptique	20
1.2.1	Vision stéréoscopique	21
1.2.2	Mesure par évaluation du flou optique	22
1.2.3	<i>Depth-from-Defocus</i> (DfD) par ouverture codée	24
2.1.1	Système optique à focale double	28
2.1.2	Lentille mince biconvexe	29
2.1.3	Modèle de caméra monofocale	30
2.2.1	Modèle de formation d'image sténopé	31
2.2.2	Modèle de formation d'image par une lentille mince	34
2.2.3	Modèle de formation d'image par une lentille mince : image mise au point	35
2.2.4	Modèle de formation d'image par une lentille mince : image floue	35
2.2.5	Modèle de flou optique	36
2.2.6	Expérience de Fraunhofer : diffraction d'une onde par une ouverture circulaire	38
2.2.7	Fonction d'étalement du point Gaussienne 2D généralisée	40
2.2.8	Fonction d'étalement du point Gaussienne 2D	40
3.0.1	Méthode de reconstruction 3D par analyse du flou optique	44
3.1.1	<i>Depth-from-Focus</i> (DfF)	46
3.2.1	<i>Depth-from-Defocus</i> (DfD)	49
4.1.1	Mesure par extraction du point de rupture de netteté dans un flux vidéo : principes	65
4.2.1	<i>Blur Estimating Tracking</i> (BET) : stratégie	67
4.2.2	<i>Blur-Estimating Tracking</i> (BET) : appariement inter images	69

4.3.1 Images issues d'examens coloscopiques utilisées comme base du jeu de données synthétiques	71
4.3.2 Génération du jeu de données synthétiques	71
4.3.3 Jeu de données synthétiques	72
4.3.4 Evaluation du module <i>Blur-Estimating-Tracking (BET)</i> de suivi et d'estimation du flou optique	74
4.4.1 Mesure du flou optique absolu par intégration du flou relatif	76
4.4.2 Méthode d'extraction du point de rupture de netteté : <i>Blur-Model Fitting (BMF)</i>	77
4.5.1 Mesure 3D interactive d'une lésion colonique	79
5.1.1 Néoplasies coloniques	82
5.1.2 Systeme endoscopique utilisé pour l'examen du côlon : coloscope	83
5.1.3 Coloscopie	83
5.2.1 Critères diagnostiques des lésions néoplasiques	84
5.3.1 Système de mesure des lésions coloniques	86
5.4.1 Modèle fantôme reproduisant l'apparence d'une cavité colonique	88
5.4.2 Evaluation du système de mesure sur modèle fantôme colonique	89
5.4.3 Evaluation du système de mesure sur modèle fantôme colonique : erreur de suivi de la région d'intérêt	90
5.4.4 Calibration et utilisation du système de mesure des lésions coloniques	91
5.4.5 Evaluation du système de mesure des lésions coloniques	92
5.4.6 Estimation de la taille des néoplasies à l'aide du système de mesure par extraction du point de rupture de netteté : évaluation sur données coloscopiques	94
6.0.1 Problèmes <i>Shape-from-Template (SfT)</i> et <i>Non-Rigid Shape-from-Motion (NRSfM)</i>	97
6.2.1 Problème de reconstruction <i>Non-Rigid Structure-from-Motion (NRSfM)</i>	104
6.2.2 Réseau Bayésien modélisant le processus de formation des images	105

Notations

\mathbb{R}^n	Espace vectoriel à n dimensions sur le corps des réels
\mathbb{P}^n	Espace projectif à n dimensions sur le corps des réels
$[a; b]$	Ensemble des entiers compris entre a et b
$\mathbf{Q}^\top = (Q_1, Q_2, Q_3)$	Transposée du point $\mathbf{Q} \in \mathbb{R}^3$
\mathbf{Q}_d	Point \mathbf{Q} situé à la profondeur $Q_3 = d$ de la caméra
$\mathbf{q}^\top = (q_1, q_2)$	Transposée du point $\mathbf{q} \in \mathbb{R}^2$
$I : \Omega \in \mathbb{R}^2 \rightarrow R, G, B \in \{0; \dots; 255\}$	Fonction image
$I(\mathbf{q}) = (R, G, B)$	
(x, y)	Coordonnées d'un point 2D
ROI	Région d'intérêt (<i>Region Of Interest</i>)
I^k	Image d'indice k de la vidéo
I_d	Image du plan situé à la profondeur d de la caméra
$\hat{I} = I_{\hat{d}}$	Image mise au point du plan de netteté situé à la distance \hat{d} de la caméra
\mathcal{I}	Transformée de Fourier de l'image I
$M_{m \times n}$	Matrice M de taille $m \times n$
$\mathbf{t}_{m \times 1}$	Vecteur \mathbf{t} de taille $m \times 1$
Id	Matrice identité
P	Matrice de projection
K	Matrice des paramètres internes du système optique
e	Distance séparant la lentille du capteur image
f	Distance focale
r	Rayon de l'ouverture du système optique
Δ	Opérateur Laplacien
∇	Opérateur gradient
$\nabla \cdot$	Opérateur de divergence
\mathcal{J}	Opérateur Jacobien
$\frac{\partial f}{\partial x}$	Derivée partielle de la fonction f en fonction de la variable x

ζ^k	Différence de flou relative entre les images d'indices k et $k+1$
δ^k	Flou absolu estimé dans l'image d'indice k
$P(E)$	Probabilité de l'événement E
$P(E F)$	Probabilité conditionnelle de l'événement E sachant que l'événement F a été réalisé
$P(E, F)$	Probabilité jointe des événements E et F

1

Introduction

Le développement de systèmes optiques haute définition spécialisés a permis d'étendre leurs champs d'applications. Ils sont employés pour le contrôle qualité ou l'asservissement visuel dans le domaine industriel. Dans le domaine médical, l'endoscopie (du grec *endom* : intérieur et *scopie* : vision) est une procédure de référence permettant l'exploration d'organes ainsi que l'intervention chirurgicale non-invasive. La vision par ordinateur généralement associée à ces systèmes permet le traitement et l'analyse algorithmique des informations visuelles afin d'extraire des éléments de contexte, les propriétés d'une scène ou d'en augmenter le contenu. La vision 3D est une des disciplines de la vision par ordinateur. Elle consiste à estimer la représentation tridimensionnelle d'une scène à l'aide d'un ensemble d'observations. De nombreuses techniques efficaces et éprouvées permettent la reconstruction 3D de scènes rigides (*Structure-from-X (SfX)*). Des solutions plus récentes ont été apportées pour la reconstruction de surfaces déformables en imposant des contraintes sur le type de déformations considéré ou sur les propriétés physiques de l'objet [Bartoli et al., 2015; Chhatkuli et al., 2014a].

Une des limitations majeures des méthodes de vision 3D est l'ambiguïté d'échelle qui complexifie l'inférence de l'échelle absolue d'une scène (échelle Euclidienne). Dans le cas d'un système monoculaire, il existe une relation équivoque entre la distance de l'objet observé et ses dimensions. Les systèmes de métrologie s'appuient généralement sur l'adjonction de dispositifs de mesure externes ou l'utilisation de systèmes optiques multi-caméras calibrés. Les contraintes de conception de ces systèmes peuvent être inadéquates aux dispositifs actuels ou aux applications visées. Cela est notamment le cas dans le domaine de l'endoscopie qui nécessite l'utilisation de dispositifs optiques miniaturisés difficilement modifiables. Les travaux réalisés dans le cadre de cette thèse visent à proposer une solution passive (uniquement basée sur la vision) à la problématique d'estimation de l'échelle des lésions en coloscopie. Cette procédure endoscopique permet l'exploration de la cavité colonique à l'aide d'un endoscope flexible (coloscope) embarquant un système optique monofocale : système optique monoculaire dont les paramètres internes sont fixes. La taille est un

critère de diagnostic essentiel qui détermine notamment les intervalles de temps de surveillance des néoplasies (croissances anormales de tissus) [Chaptini et al., 2014]. L'estimation est rendue complexe de par les propriétés optiques des coloscopes ainsi que l'environnement exploré. Nous proposons une solution passive adaptée à l'estimation de la taille des lésions coloniques pouvant être étendue à tout système optique monofocale.

Ce chapitre introductif dresse un état de l'art des systèmes de métrologie par vision. Nous distinguons les systèmes actifs, nécessitant l'ajout (ou la modification) d'équipements au système de mesure, des systèmes passifs uniquement basés sur la vision. Nous placerons nos travaux en perspective de cet état de l'art afin d'en présenter les principales contributions.

1.1 Métrologie par vision active

Les systèmes de métrologie par vision active reposent sur l'adjonction d'un dispositif externe au système optique (ou à la scène) permettant de résoudre l'ambiguïté d'échelle. Ces systèmes sont génériques mais ils imposent différentes contraintes pouvant être inadaptées à l'application visée (notamment à l'endoscopie). Nous associons à ces derniers les dispositifs stéréos qui nécessiteraient la modification de la majorité des dispositifs endoscopiques monoculaires actuels. A toute fin de clarté, nous considérons la mesure d'objets au sein de scènes rigides. Il est important de noter par ailleurs que la majorité des systèmes de métrologies sus-cités considèrent des surfaces possédant des propriétés de réflexions Lambertiennes [Oren and Nayar, 1995]¹.

1.1.1 Photogrammétrie par insertion de cibles

La méthode de photogrammétrie par insertion de cibles (figure 1.1.1) consiste à insérer dans une scène un ou plusieurs repères². Ces derniers permettent de résoudre l'ambiguïté d'échelle dont souffrent les méthodes *Structure-from-Motion (SfM)* [Hartley and Zisserman, 2004]. Les cibles permettent d'estimer le déplacement absolu de la caméra entre chaque prise de vue. Cette technique de métrologie est notamment utilisée dans le domaine de l'aviation, de l'automobile ou de l'archéologie.

Les précisions de mesures sont inférieures au millimètre [King et al., 2014]. Elles sont particulièrement dépendantes du nombre de repères présents dans la scène. Cette méthode est inadaptée à de nombreuses applications comme l'endoscopie mais peut être utilisée pour des scènes dont un

1. Une surface est Lambertienne si sa réflectance est isotrope, c'est à dire si elle réfléchit la lumière avec la même intensité dans toutes les directions.

2. repère : cible visuelle dont les propriétés géométriques sont connues

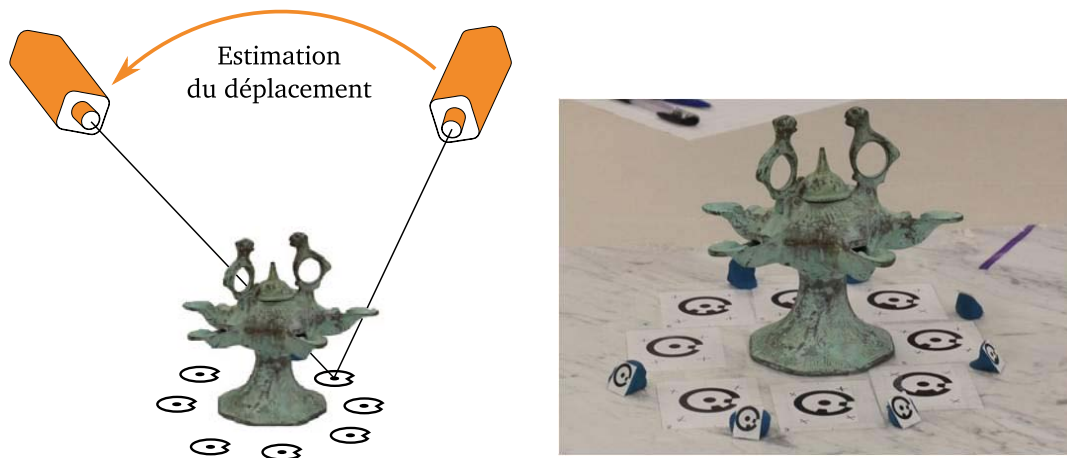


Figure 1.1.1 – Méthode de photogrammétrie par insertion de cibles. La présence d'indices visuels de taille connue dans la scène permet d'estimer le déplacement de la caméra entre chaque prise de vue. National Center for Preservation Technology and Training.

sous-ensemble d'éléments possède des propriétés géométriques connues (par exemple des éléments architecturaux).

1.1.2 Systèmes de suivi

Les systèmes de suivi permettent d'estimer le déplacement d'une caméra opérante à l'aide d'un dispositif externe omniscient. Ce dernier fixe le repère monde de l'environnement considéré.

Systèmes de suivi visuel : Un repère aux motifs discriminants est fixé à la caméra opérante afin d'être suivi par un système de vision externe omniscient (et généralement fixe). En synchronisant les deux systèmes optiques pour chaque prise de vue de la caméra opérante, les déplacements de cette dernière peuvent être estimés. Le suivi visuel est notamment utilisé pour des applications de réalité virtuelle ainsi qu'en chirurgie endoscopique (figure 1.1.2).

En endoscopie, les précisions de mesures obtenues sont de l'ordre de 1,5 mm pour le suivi de l'extrémité distale de l'endoscope et 3 mm pour la reconstruction tridimensionnelle ($\pm 1,9$ mm) [Fusaglia et al., 2012]. Les systèmes de suivi visuel sont peu robustes aux contraintes d'occultation partielles ou totales du repère. La calibration précise requise par cette approche est par ailleurs délicate. Enfin, elle n'est pas adaptée aux systèmes endoscopiques flexibles.

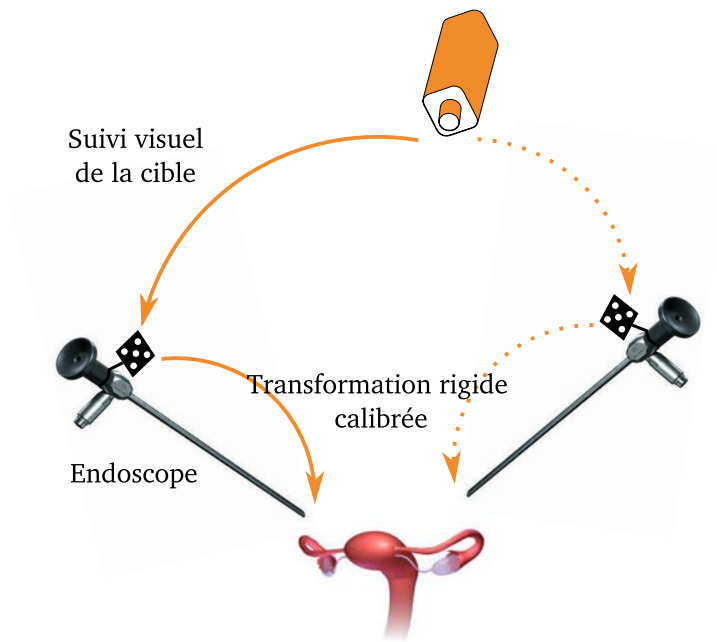


Figure 1.1.2 – Système de suivi visuel d’une caméra endoscopique. Un repère est placé sur l’extrémité proximale de l’endoscope. Ce dernier est suivi par une caméra placée sur le champ opératoire afin d’estimer le déplacement de la caméra endoscopique.

Suivi électromagnétique : Une sonde électromagnétique (émettrice ou émettrice-réceptrice) est intégrée au système optique opérant. Un dispositif externe (récepteur ou émetteur-récepteur), placé à proximité du champ opératoire, traite l’information de position de la sonde (figure 1.1.3). La calibration de la transformation rigide sondes-caméra (dispositif multisondes) permet d’estimer la position de cette dernière et ainsi d’inférer l’échelle de la scène. Cette technique est utilisée en métrologie industrielle (*Coordinate Measuring Machines*) ainsi qu’en endoscopie ou elle peut être intégrée à des endoscopes flexibles (Olympus[®] scopeGuide).

Il n’existe pas, à notre connaissance, d’évaluation pratique de la précision des mesures obtenues à l’aide de ce dispositif. [Reichl et al., 2013] propose une méthode hybride (électromagnétique et visuelle) dont les précisions de suivi sont de l’ordre de 5 mm ($\pm 2,5$ mm) pour la position de l’endoscope et 11,5 degrés (± 6 degrés) pour son orientation. La précision peut être dégradée par la présence de métal dans le champ opératoire. Par ailleurs, ces dispositifs nécessitent de placer l’équipement récepteur à proximité du champ opératoire. Ils peuvent être remplacés par des systèmes basés sur des signaux infrarouges ou ultrasons fonctionnant selon le même principe. Les précisions de suivi sont similaires mais leurs champs d’applications sont fortement contraints

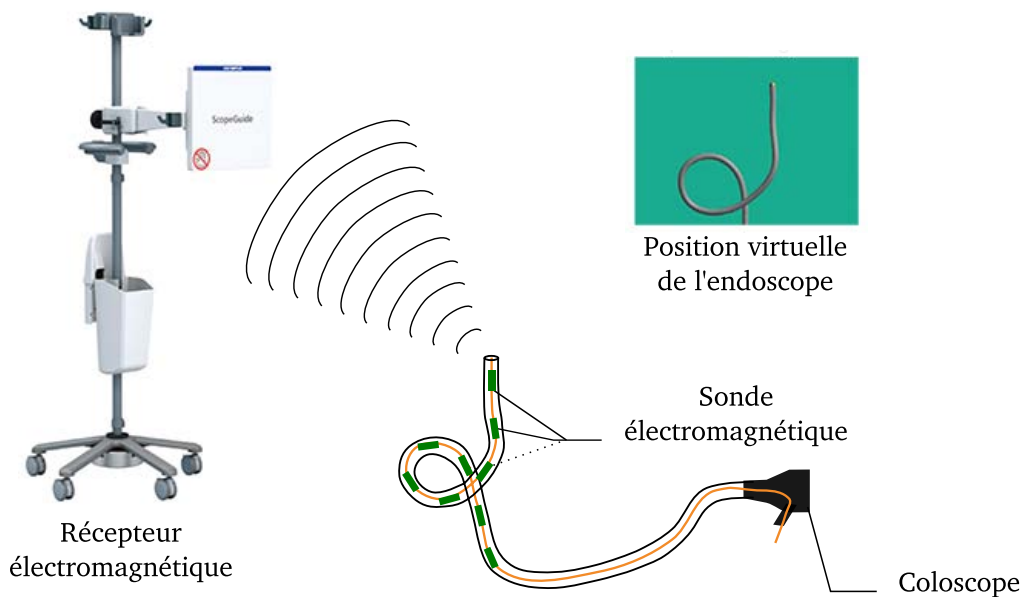


Figure 1.1.3 – Système de suivi électromagnétique intégré à l'équipement endoscopique Olympus[®] scopeguide. Un récepteur placé sur le champ opératoire intègre l'information de position émise par les sondes électromagnétiques. La connaissance de la position de la caméra opérante permet la reconstruction 3D à l'échelle de la scène observée par le praticien.

par l'environnement considéré (problème des occultations).

1.1.3 Projection de lumière structurée

Un motif lumineux de propriétés géométriques connues est projeté dans la scène. La transformation rigide source de lumière-caméra est généralement supposée calibrée. L'analyse de la déformation du motif sur les surfaces observées permet une reconstruction tridimensionnelle à l'échelle de la scène (figure 1.1.4). Cette méthode est particulièrement adaptée à la métrologie industrielle. Elle a également été intégrée à des systèmes endoscopiques rigides [Mirota et al., 2011; Schmalz et al., 2012]. L'acheminement d'une source de lumière structurée au sein d'un endoscope flexible est en pratique complexe.

L'inconvénient majeur de cette méthode est le temps de calcul nécessaire à l'obtention d'une carte de profondeur dense. Les précisions de mesures obtenues sont inférieures à 0,1 mm dans le cas de la reconstruction d'un plan [Wenguo et al., 2013].

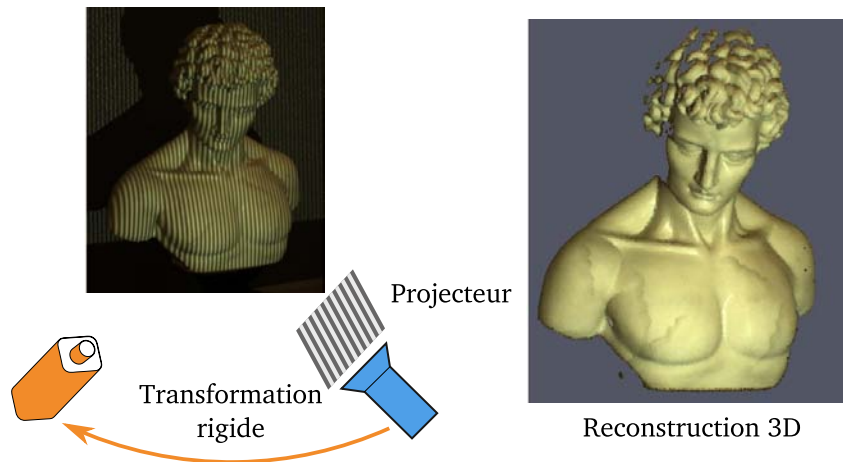


Figure 1.1.4 – Reconstruction 3D par projection de lumière structurée. Un motif de forme géométrique connue (franges, motif circulaire . . .) est projeté dans la scène. L’analyse de la déformation du motif permet une reconstruction 3D à l’échelle de celle-ci. Université Brown, Pr. Gabriel Taubin.

1.1.4 Systèmes de mesure indépendants

Des systèmes de mesure indépendants qualifiés de “temps de vol” (laser, ultra-sons, . . .) sont fréquemment adjoints aux systèmes optiques. Ils sont utilisés en photographie (système autofocus présent sur une majorité d’appareils compacts) ou équipent des systèmes d’assistance visuelle embarqués sur des véhicules (*LiDAR* : *light detection and ranging*, figure 1.1.5). Les systèmes de mesure laser ou ultra-sons, intègrent généralement le temps d’émission-réception d’une onde (lumineuse ou sonore) balayant la scène à reconstruire. Le système ultra-sons présente l’inconvénient d’être sensible aux conditions d’acquisition (température de l’air, . . .) mais contrairement aux systèmes laser il n’est soumis à aucune contrainte sur les propriétés de réflexion des surfaces observées (surfaces translucides, . . .).

Les précisions de mesure des systèmes “temps de vol” sont dépendantes de la technologie employée ainsi que des applications visées. Elles sont généralement de l’ordre du centimètre dans le cas de systèmes infrarouges (pour des objets aux surfaces non spéculaires [Hansard et al., 2012]). Dans le cadre de la reconstruction d’environnements urbains, la précision de mesure d’un système LiDAR est de l’ordre de quelques centimètres [Leberl et al., 2010].

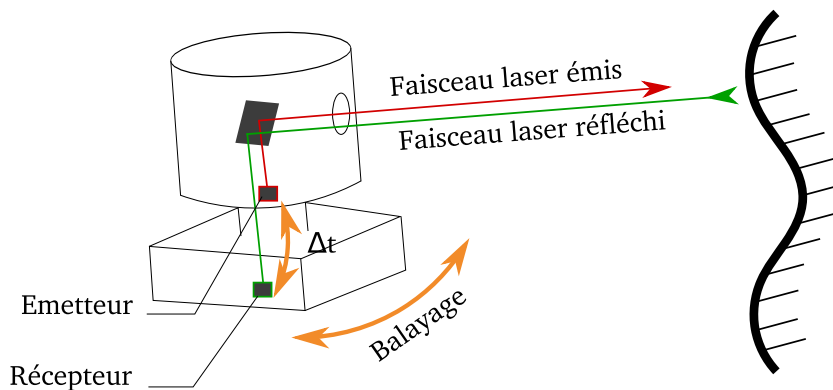


Figure 1.1.5 – Système de mesure laser. Un dispositif émetteur-récepteur intègre le temps de parcours d'un faisceau laser balayant la scène. Une reconstruction dense de la carte de profondeur de cette dernière est alors inférée.

1.1.5 Caméra plénoptique

Le modèle de caméra plénoptique (*light field camera* [Adelson and Wang, 1992]) diffère du modèle de caméra conventionnel par la présence d'un ensemble de microlentilles placées en amont du capteur photosensible (figure 1.1.6 (i)). L'ensemble des rayons lumineux issu d'un point 3D n'est pas directement intégré par une cellule (ou un ensemble de cellules voisines) du capteur mais il est réfracté par les microlentilles pour être intégré en différentes cellules connues de ce dernier (ce qui limite la résolution des images). Grâce à une analyse logicielle post-acquisition, la position et la direction de chaque rayon traversant le système optique peut être estimée. L'image peut alors être mise au point *a posteriori* par intégration des différents rayons issus d'un point 3D (somme cohérente des rayons, figure 1.1.6 (ii,iii)).

Le modèle plénoptique peut également être considéré comme un ensemble de caméras sténopés (voir section 1.2.1). Chaque rayon lumineux provenant d'un point 3D et traversant le système optique correspond à un point de vue différent de la scène (figure 1.1.6 (i)). Le système optique pouvant être calibré, il est alors possible d'obtenir (à l'aide de l'information multi-vues mais également de l'information de flou optique) une reconstruction 3D à l'échelle [Tao et al., 2013]. Cette étude ne présente pas d'évaluation quantitative de la précision des mesures 3D.

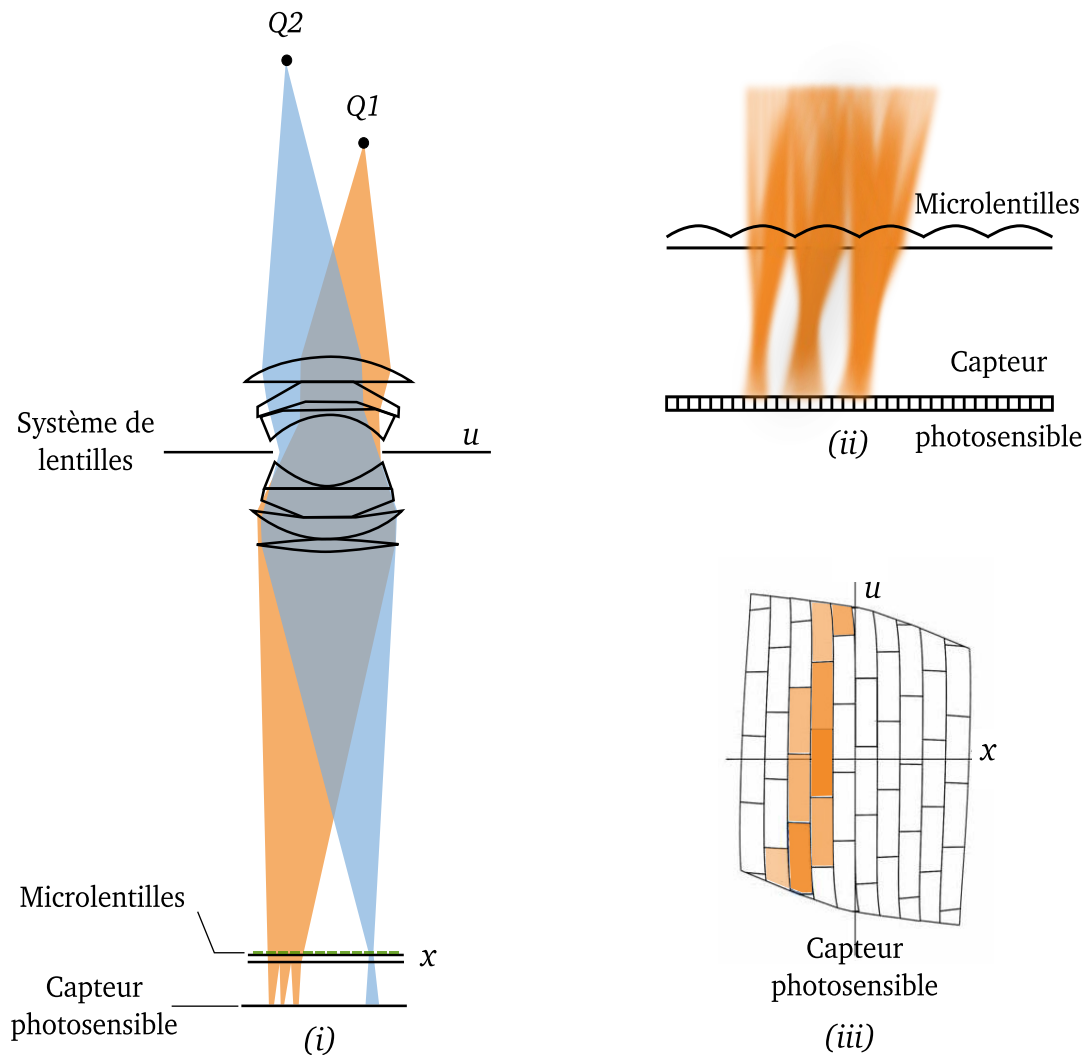


Figure 1.1.6 – Modèle de caméra plénoptique. (i) Un ensemble de microlentilles placées en amont du capteur photosensible réfracte les rayons lumineux traversant le système optique. Les rayons réfractés sont intégrés en différentes cellules du photocapteur. (ii) Réfraction des rayons incidents par l'ensemble des microlentilles. (iii) Les cellules du photocapteur intègrent un sous ensemble des rayons réfractés. Un traitement post-acquisition permet de modifier la mise au point de l'image ou d'effectuer la reconstruction 3D à l'échelle de la scène (chaque rayon issu d'un point 3D pouvant être *a priori* isolé, les systèmes plénoptiques peuvent être assimilés à des systèmes multi vues) [Ren, 2006]

1.2 Métrologie par vision passive

La majorité des dispositifs passifs de métrologie basés vision reposent sur des systèmes optiques stéréos (ou multi-vues calibrés). Les systèmes de métrologie monofocale s'appuient sur la présence de références de taille connue dans la scène ce qui est inadapté à l'endoscopie. Ils peuvent également être basés sur la mesure de flou optique qui est directement liée à la profondeur de la prise de vue dans le cas d'un système optique dont les paramètres internes sont fixes. Cette mesure est cependant délicate et peu robuste au bruit.

1.2.1 Vision stéréoscopique

Ce dispositif permet de résoudre l'ambiguïté d'échelle lorsque la transformation rigide inter-caméras est calibrée. De par son faible coût, il est utilisé dans de nombreuses applications (réalité virtuelle, systèmes de mesure industriels) et notamment en endoscopie lorsque la procédure médicale permet l'utilisation d'un dispositif endoscopique de diamètre suffisant. Afin d'améliorer la précision et la robustesse des systèmes de mesure stéréo, des dispositifs actifs leurs sont fréquemment couplés (figure 1.2.1).

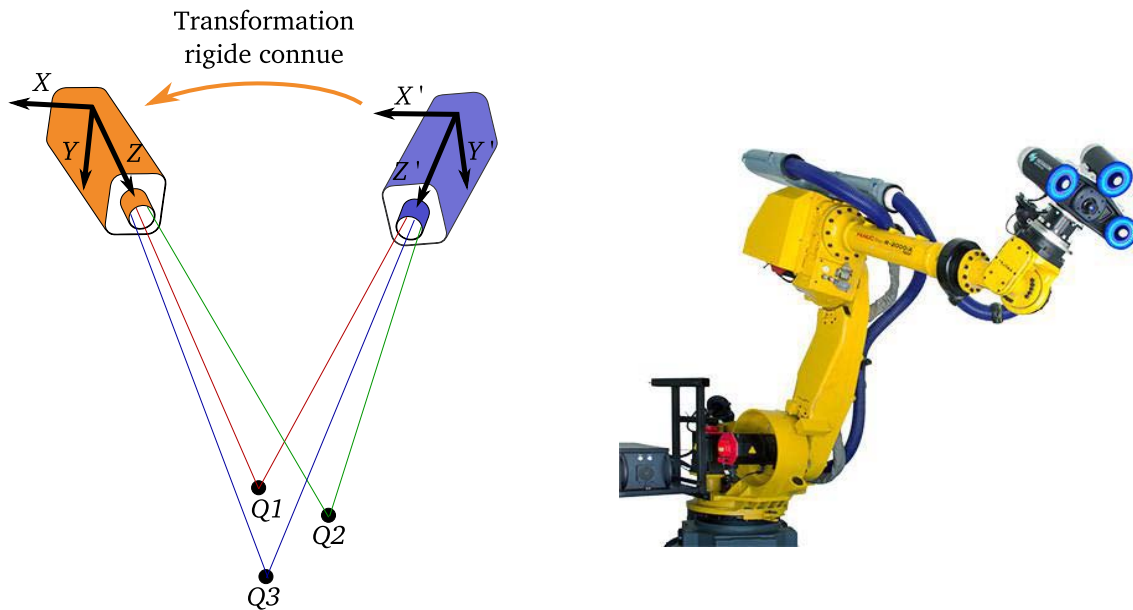


Figure 1.2.1 – Système optique stéréo et système de mesure par stéréovision Hexagon[®] Metrology WLS400A.

Les précisions de reconstruction sont dépendantes du nombre de prises de vues réalisées ainsi

que de la résolution des images. Les dispositifs récents permettent d'atteindre des précisions de mesures inférieures à 0,1 mm (pour l'acquisition d'une paire d'image de résolution 2048×1536 pixels [Hien et al., 2014]).

1.2.2 Mesure par évaluation du flou optique

Considérant un système optique monofocale dont les paramètres internes sont fixes (distance focale, ouverture, ...), le degré de mise au point (ou la quantité de flou optique) d'un élément composant une scène est directement lié à sa distance d'acquisition (voir la figure 1.2.2 ainsi que le chapitre 2). Les méthodes basées sur la mesure de flou optique exploitent cette relation afin de reconstruire la carte de profondeur d'une scène. Elles peuvent également être utilisées pour modifier la mise au point d'une image *a posteriori*. Elles s'appliquent à la vision monoculaire mais nécessitent généralement l'asservissement des paramètres internes (paramètres modifiant la mise au point) ou externes de la caméra (déplacement). Afin de résoudre l'ambiguïté d'échelle, elles requièrent une calibration précise de la caméra permettant d'établir une table de correspondance liant le degré de mise au point d'un objet, sa profondeur d'acquisition et les paramètres internes du système optique (cas des approches *Depth-from-Defocus* (DfD)).



Figure 1.2.2 – Illustration de la relation liant le degré de mise au point d'un objet à sa profondeur d'acquisition. La quantité de flou optique augmente en fonction de la distance d'un point de la surface à la profondeur de champ (plage de distance pour laquelle un objet apparaît visuellement nette).

Nous introduisons dans les sous-sections suivantes les différentes approches de reconstruction tridimensionnelle par mesure du flou optique [Pentland, 1987]. Les travaux présentés dans ce manuscrit s'inscrivent dans cette catégorie d'approche. Un état de l'art exhaustif est détaillé dans le chapitre 3.

Depth-from-Focus (DfF) : Ces approches consistent à acquérir un ensemble d'images (au minimum dix) en contrôlant les déplacements de la caméra (généralement un mouvement de translation fronto-parrallèle à la scène). Les paramètres du système optique sont fixes. Pour chacune des images, une évaluation de la quantité de flou optique est réalisée afin d'extraire les zones de netteté (situées à une distance calibrée de la caméra). La reconstruction de la scène est obtenue en intégrant l'information de déplacement de la caméra et les distances de netteté locales précédemment extraites. Les méthodes *DfF* sont particulièrement adaptées à l'imagerie microscopique [Minhas et al., 2009].

Depth-from-Defocus (DfD) : Ces méthodes nécessitent généralement l'acquisition de deux images réalisées depuis le même point de vue mais en modifiant les paramètres internes du système optique. L'analyse de la variation de netteté entre les deux observations permet d'évaluer l'influence des changements de paramètres de la caméra et d'inférer la carte de profondeur de la scène. Il est généralement nécessaire d'utiliser deux images afin de résoudre les problématiques liées à l'apparence de la surface (apparence par nature floue). Les méthodes *DfD* sont notamment utilisées pour effectuer la mise au point automatique de caméras [Zhang et al., 2014].

Ouverture codées : Les méthodes de reconstruction 3D par ouverture codée peuvent être assimilées à des approches de *DfD* actives. Elles sont introduites dans cette sous-section à des fins de clarté. Des masques obturateurs sont placés sur une des lentilles du dispositif optique afin de changer la géométrie de l'ouverture. L'utilisation de géométries adaptées permet d'améliorer l'estimation du degré de mise au point d'une image [Lin et al., 2013; Zhou et al., 2009] (figure 1.2.3).

Le choix de la géométrie de l'ouverture est un compromis entre qualité de restauration d'une image et précision de reconstruction [Levin et al., 2007]. Les précisions de reconstruction 3D ont été évaluées à 4 mm pour des images texturées (erreur RMS évaluée sur un ensemble d'images synthétiques pour des objets situés entre 0,8 et 1,8 m) et 18 mm pour des images faiblement texturées [Zhou et al., 2009].

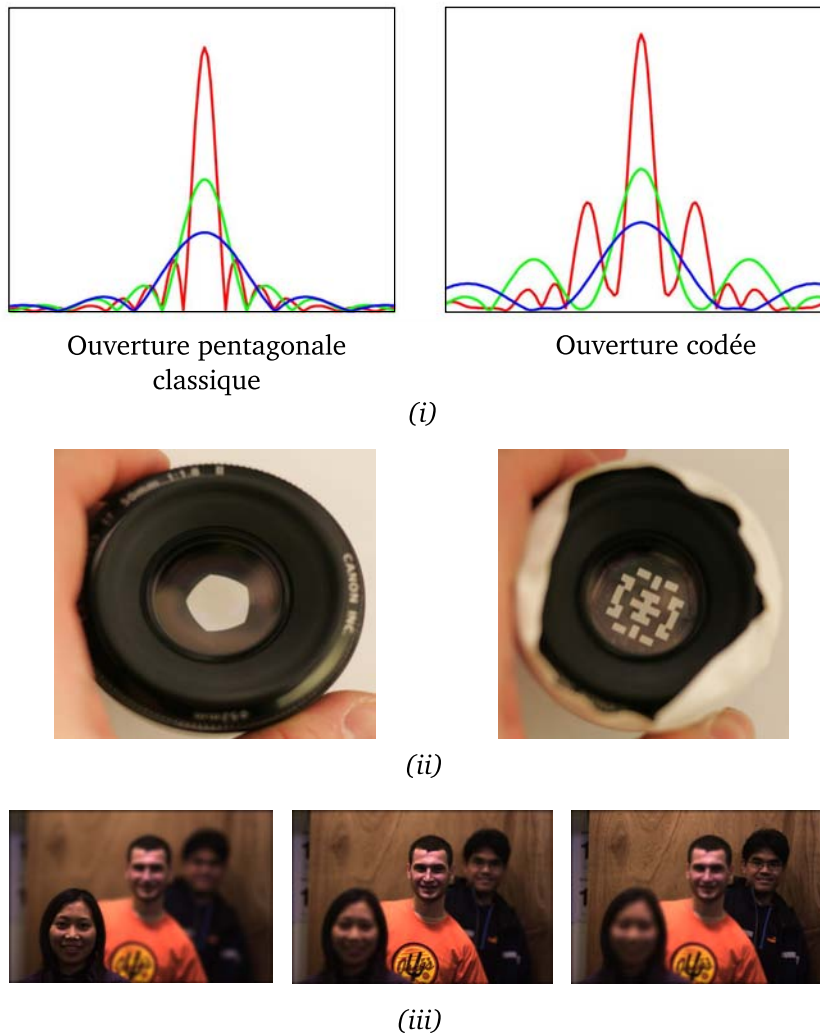


Figure 1.2.3 – Méthode de *Depth-from-Defocus* (DfD) par ouverture codée. (i) Transformée de Fourier 1D du motif d'ouverture pour trois tailles d'ouverture différentes. Le processus de formation des images peut être modélisé par la convolution de l'image mise au point avec la transformée de Fourier du système optique (voir chapitre 2). Pour une ouverture codée adaptée, les valeurs nulles du spectre sont facilement séparables ce qui permet d'améliorer la précision des mesures de flou optique (voir chapitre 3). (ii) Réalisation pratique d'une ouverture codée. (iii) Modification de la mise au point d'une image *a posteriori* basée sur la mesure de flou optique [Levin et al., 2007].

1.3 Contributions

L'étude présentée dans ce manuscrit a pour objet la mesure et l'estimation de l'échelle d'une scène par vision passive monofocale. L'application visée est la mesure endoscopique de la taille des lésions pour laquelle il n'existe pas de solution adaptée aux équipements endoscopiques flexibles. Nous avons plus particulièrement considéré les solutions passives car elles n'impliquent aucune modification des équipements actuels et sont adaptées à de nombreux systèmes de mesure basés vision. L'état de l'art des méthodes de métrologie passives souligne que seules les méthodes basées sur l'analyse du flou optique (et plus particulièrement les méthodes DfF) répondent à notre problématique. Elles imposent cependant l'asservissement des paramètres externes ou internes de la caméra les rendant inadaptées au cadre applicatif considéré.

Notre contribution principale porte sur une méthode discriminative de mesure 3D basée région. Elle a été appliquée à l'estimation de la taille des lésions en coloscopie, examen endoscopique permettant l'observation et l'intervention au sein du côlon. Il s'agit de la première technique de mesure des lésions par vision passive. Elle repose sur l'étude de la redondance de l'information de flou optique observée dans un flux vidéo correspondant à un mouvement d'approche approximatif de la caméra vis à vis d'un objet d'intérêt. La méthode proposée s'affranchit des contraintes d'asservissement de la caméra imposées par les techniques *DfF* et *DfD*. Elle introduit la notion de point de rupture de netteté (*Infocus-Breakpoint* IB) défini prosaïquement comme la borne inférieure de la profondeur de champ. Deux modules originaux nous permettent d'extraire l'IB de façon robuste. Le suivi d'une région d'intérêt et l'estimation conjointe de l'évolution du degré de netteté constitue le premier module. Contrairement aux techniques de l'état de l'art, notre méthode est robuste à d'éventuelles déformations de l'environnement imagé (ce qui est fréquemment le cas dans le contexte de l'imagerie endoscopique). Un second module permet d'extraire de façon robuste l'IB dans la séquence vidéo en s'appuyant sur l'ajustement d'un modèle de flou idoine. Les résultats obtenus en coloscopie démontrent la robustesse de la méthode proposée ainsi que sa précision. L'erreur de mesure est de l'ordre du millimètre.

Nous présentons par ailleurs, sous la forme d'une étude théorique préliminaire, une méthode générative permettant la reconstruction 3D à l'échelle de surfaces déformables. Les méthodes discriminatives reposent sur l'interprétation directe d'observations. Elles cherchent donc à estimer des paramètres ou des probabilités *a posteriori*. Les méthodes génératives reposent sur un modèle de génération des observations nécessitant l'apprentissage de probabilités *a priori*. La solution proposée est basée sur une approche *Non-Rigid Structure-from-Motion (NRsFM)*. Elle s'applique

à la reconstruction dense de surfaces subissant des transformations isométriques³. Nous proposons un modèle de formation d'images original intégrant la modélisation du flou optique. Cette méthode apporte trois contributions à l'état de l'art *NRSfM*. *i)* Elle permet une reconstruction dense et à l'échelle de la surface déformée et s'affranchit des problématiques d'occultations dont souffrent la majorité des méthodes de reconstruction génératives. *ii)* La carte de texture entièrement mise au point de la surface reconstruite est conjointement estimée. *iii)* Après avoir appris les paramètres du modèle, de nouvelles vues de la surface déformée peuvent être efficacement générées. Elle est par ailleurs adaptée à de nombreuses applications de métrologie 3D.

3. Une transformation isométrique conserve localement les distances à la surface de l'objet déformé.

2

Modélisation des systèmes de vision monofocale

Ce chapitre présente les notions et éléments de base de l'optique sur lesquelles s'appuie cette thèse. Nous définissons tout d'abord la notion de système monofocale : un système optique monoculaire dont les paramètres internes sont fixes (distance focale, ouverture, position du capteur image). Nous détaillons par la suite le modèle de formation d'image associé. En s'appuyant sur le modèle optique géométrique, nous mettons plus particulièrement en évidence la relation liant la profondeur de la prise de vue avec le degré de mise au point (flou) de la scène. Pour un système optique monofocale, nous montrons qu'il existe une profondeur théorique unique pour laquelle l'image d'un point 3D est mise au point (nette). Ce modèle est complété par la modélisation optique ondulatoire qui permet de définir de façon plus précise la réponse impulsionnelle du système optique communément appelé *PSF (Point Spread Function)*.

Les systèmes d'acquisition d'images numériques se sont diversifiés et complexifiés afin de répondre à des exigences applicatives spécifiques (imagerie médicale, imagerie microscopique ...). Des modèles de formation d'images simplifiés ont été substitués aux modèles optiques théoriques difficilement exploitables. Ils approximent un sous-ensemble de leurs caractéristiques optiques afin de permettre l'analyse et le traitement efficace des images (le modèle sténopé est par exemple largement utilisé en reconstruction 3D). Il convient ainsi de définir clairement le modèle de caméra considéré ainsi que le modèle de formation d'image associé afin de le placer en perspective de l'application de vision souhaitée.

Afin de répondre aux contraintes de l'examen coloscopique, les coloscopes sont généralement équipés de systèmes optiques monoculaires dont les propriétés optiques sont fixes (distance focale, ouverture, ...). Pour des raisons de clarté, nous utilisons le terme monofocale pour décrire ce

type de système. Le modèle de caméra monofocale et la définition de ses propriétés optiques seront présentés dans la première section. Nous présenterons dans la seconde section le modèle de formation d'image qui lui a été associé afin de répondre à notre problématique : la résolution de l'ambiguïté d'échelle par vision passive monofocale.

2.1 Définition et implications d'un système monofocale

De nombreux termes communs à la vision par ordinateur, la physique optique et la photographie possèdent des définitions différentes et les confusions sont fréquentes. Nous définissons dans cette section les différentes mesures caractéristiques du système optique considéré.

La majorité des caméras actuelles permettent l'ajustement manuel ou automatique des paramètres optiques modifiant par conséquent les propriétés de la prise de vue. Dans le cas d'un système monofocale, tel que nous le définissons, l'ensemble des éléments du système optique est fixé. Ce système, de conception mécanique simple (*a contrario* des systèmes dits autofocus), est fréquemment utilisé dans le cadre d'applications industrielles ou dans le domaine médical pour lequel il est nécessaire de répondre à des contraintes pratiques. Dans le cas de l'endoscopie par exemple, la distance d'observation des organes est faible et le facteur de grossissement ainsi que l'angle de prise de vue est important afin de permettre une observation détaillée des divers organes explorés par le praticien. Par extension, les systèmes dits à focale double (possédant deux configurations optiques fixes) seront assimilés à des caméras monofocales pour les deux configurations (figure 2.1.1).



Figure 2.1.1 – Système optique à focale double embarqué dans la nouvelle gamme de coloscopes Olympus® Evis Lucera Elite (extrait de la brochure). (i) Système dans son mode de fonctionnement dit “normal” utilisé pour l'exploration du côlon (image mise au point dans une plage de distance d'environ 2 mm à 10 cm). (ii) Seconde configuration du système optique permettant l'observation détaillée de la lésion (à une faible distance de l'extrémité distale du coloscope).

La modélisation exhaustive d'un système optique est complexe et requière la connaissance de la configuration mécanique et optique de ce dernier. Cette configuration étant propre à chaque

système, il est commun de considérer un système simplifié. L'ensemble des lentilles est généralement modélisé par une unique lentille mince centrée sur un axe de rotation imaginaire : l'axe optique (symétrie rotationnelle relative à l'axe optique). Cette hypothèse est largement admise dans le cadre des méthodes de reconstruction 3D. L'hypothèse de lentille mince est acceptable si l'épaisseur de la lentille est négligeable face à son diamètre. En d'autres termes, l'épaisseur de la lentille a un effet négligeable sur l'angle de réfraction des rayons incidents contrairement à une lentille dite épaisse. Dans le cadre de l'endoscopie, il est difficile de connaître précisément le système optique utilisé (généralement protégé par un brevet). Nous considérons dans notre étude une lentille mince convergente biconvexe, c'est à dire composée de deux dioptries convexes opposés (figure 2.1.2). Nos expérimentations ont démontrées la validité pratique de cette hypothèse.

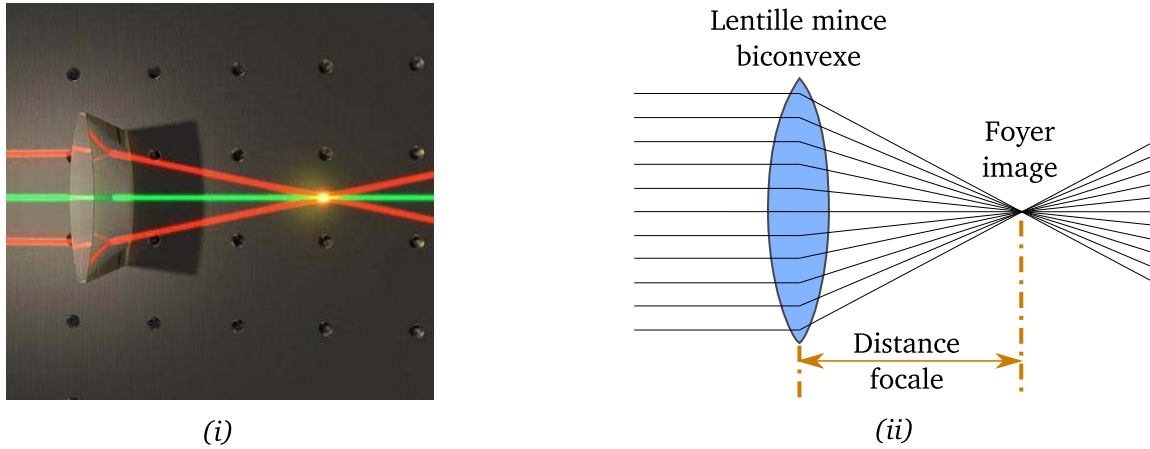


Figure 2.1.2 – Lentille mince biconvexe. (i) Réfraction de 3 rayons incidents parallèles par une lentille biconvexe dite mince Visual photo : David Parker. (ii) Modèle optique géométrique associé.

Le modèle de caméra considéré est présenté dans la figure 2.1.3. Le diaphragme (ou ouverture) permet de restreindre la quantité de rayons lumineux incidents. Dans notre étude, le diamètre de celui-ci est constant. La lentille est principalement caractérisée par sa distance focale (mesure indirecte de son pouvoir de convergence). Le capteur image (confondu avec le plan image) est composé de cellules photosensibles (CCD, CMOS, Foveon . . .) permettant d'intégrer les photons de lumière afin de restituer l'observation. Les propriétés du capteur ne sont pas considérées dans cette étude.

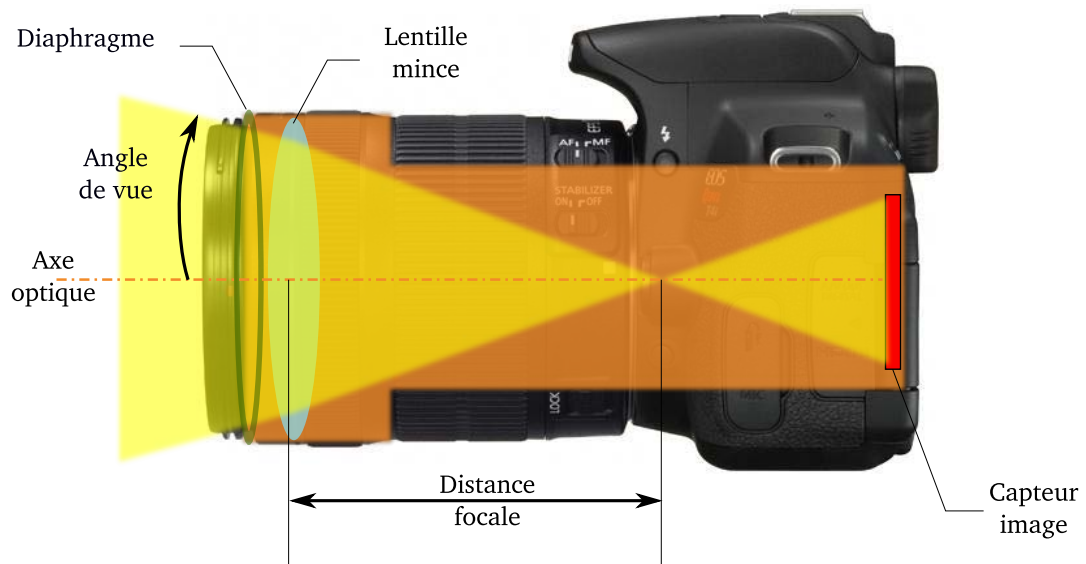


Figure 2.1.3 – Modèle de caméra monofocale considéré dans notre étude. Ce dernier se compose d'une lentille mince centrée sur l'axe optique. Elle est caractérisée par sa distance focale correspondant à la distance séparant le centre de la lentille avec le foyer image (voir la section 2.2).

2.2 Modèles de formation d'images

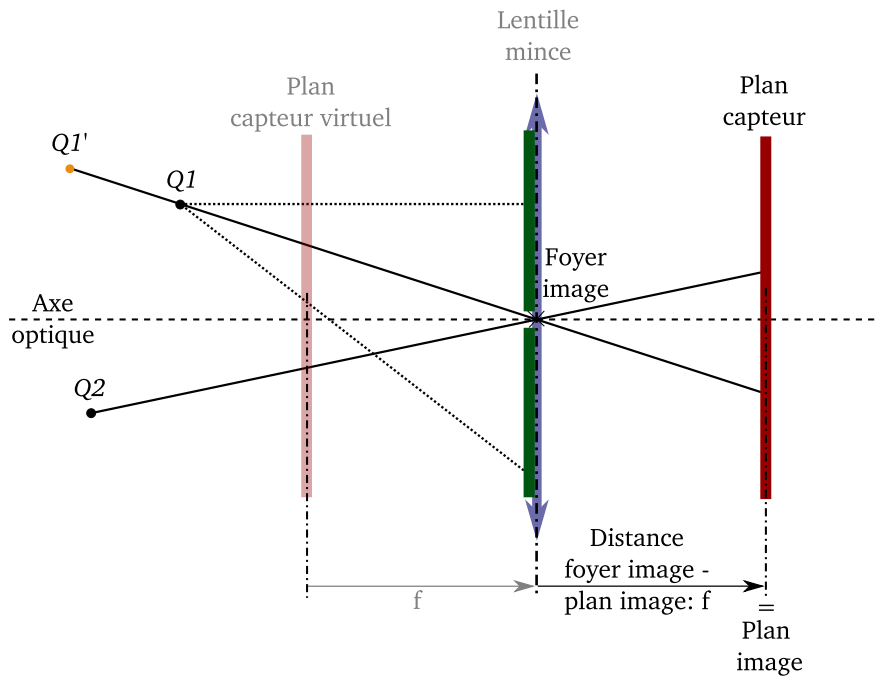
Le modèle de formation d'images peut être considéré suivant différentes approches de l'optique. A chacune d'entre elles correspond une description des propriétés optiques du système de granularité différente.

2.2.1 Modèle optique géométrique

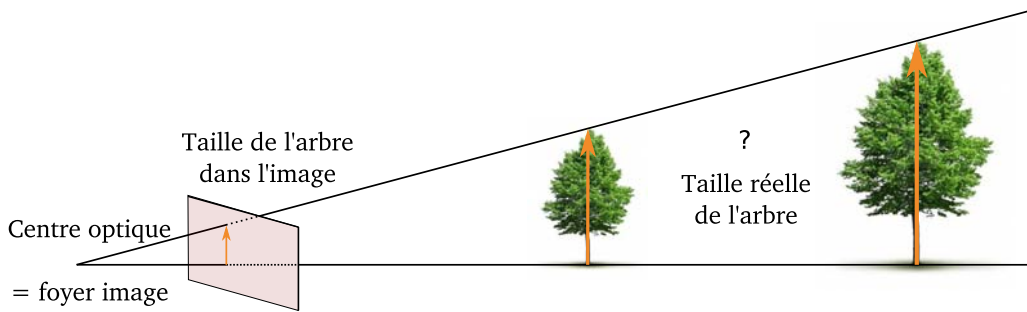
Le modèle optique géométrique (modèle corpusculaire) est basé sur l'hypothèse théorique qu'un faisceau de lumière est constitué d'un ensemble de rayons lumineux.

Modèle optique géométrique simplifié : le modèle sténopé

Le modèle sténopé (ou trou d'épingle) correspond au système d'acquisition d'image le plus élémentaire. Les premières réalisations de *camera obscura* sont généralement attribuées à Ibn al-Haytham dont les recherches ont été publiées dans son traité d'optique en 1021. Ce modèle est une simplification du modèle optique géométrique (ou modèle rayon) détaillé dans la section suivante. Nous considérons la configuration optique illustrée dans la figure 2.2.1.



(i)



(ii)

Figure 2.2.1 – (i) Modèle de formation d’image sténopé. Seuls les rayons incidents passant par le centre de la lentille sont intégrés par le capteur photosensible. Le plan capteur est confondu avec le plan image. Pour faciliter l’analyse mathématique du modèle, le plan capteur est virtuellement placé en amont de la lentille à la distance f de celle-ci (équivalence à une symétrie axiale près). (ii) Illustration de l’ambiguïté d’échelle inhérente à cette modélisation. La distance d’observation étant *a priori* inconnue la taille de l’objet imagé ne peut être inférée.

Dans cette configuration, le diaphragme possède un diamètre de dimension infinitésimale. Il est centré sur l’axe optique. Seuls les rayons lumineux passant par le centre de la lentille sont intégrés par le capteur image. Ceux-ci ne subissant pas d’effet de réfraction, la lentille peut être

éliminée de la représentation (modèle trou d'épingle). La réalisation d'un tel système optique devient triviale.

Le modèle de formation d'images connexe est exprimé par la géométrie projective. Il permet d'associer un point 3D avec sa projection dans le plan image à l'aide de transformations géométriques (approximation linéaire du système optique géométrique). Cela rend son inférence particulièrement efficace et permet d'établir des propriétés remarquables dans le cadre de la géométrie multi-vues [Hartley and Zisserman, 2004]. La matrice de projection $P_{3 \times 4}$ peut être décomposée en trois transformations linéaires permettant le passage (i) du repère monde au repère caméra (déplacement 3D), (ii) du repère caméra au repère rétine 2D (projection) puis (iii) du repère rétine au repère image (similarité).

$$\begin{pmatrix} q_1 \\ q_2 \\ q_3 \end{pmatrix} = \underbrace{\begin{matrix} (iii) & (ii) & (i) \\ \begin{bmatrix} \tau a & s & x_0 \\ 0 & a & y_0 \\ 0 & 0 & 1 \end{bmatrix} & \begin{bmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} & \begin{bmatrix} & & & \\ & R & & t \\ 0 & 0 & 0 & 1 \end{bmatrix} \end{matrix}}_{P} \begin{pmatrix} Q_1 \\ Q_2 \\ Q_3 \\ 1 \end{pmatrix} \quad (2.1)$$

$$P \sim \underbrace{\begin{bmatrix} \tau f a & s & x_0 \\ 0 & f a & y_0 \\ 0 & 0 & 1 \end{bmatrix}}_{K} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} & & & \\ & R & & t \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

où $R_{3 \times 3}$ représente une matrice de rotation et $t_{3 \times 1}$ un vecteur de translation. f est communément appelé distance focale dans la communauté vision. Cette distance correspond en réalité à la distance séparant le foyer image (ou centre optique) du plan image. a est la densité de pixels le long des axes u et v du repère rétine, x_0 et y_0 sont les coordonnées du point principal dans le repère image, τ représente le ratio entre la longueur et la largeur d'un pixel et s représente l'angle entre les axes du système de coordonnées de l'image. Les coordonnées d'un point dans l'image sont obtenues par division suivant la troisième coordonnée soit : $\mathbf{q} = \left(\frac{q_1}{q_3}, \frac{q_2}{q_3} \right)^T$. La matrice de projection est généralement reformulée afin de dissocier la matrice des paramètres internes de la caméra K et la matrice des paramètres externes. K peut être obtenue par calibration [Bouguet, 2008].

Il est important de noter que lors du passage du repère caméra au repère rétine, l'information de profondeur est perdue. Deux points situés sur la même ligne de vue étant confondus, il est impossible de résoudre géométriquement le problème inverse d'estimation de la profondeur d'un point. Ceci constitue l'ambiguïté d'échelle inhérente au modèle sténopé. La profondeur de champ

de ce dernier est infinie. Sa résolution nécessite la connaissance *a priori* de la taille des éléments de la scène ou de la position absolue de la caméra vis à vis de cette dernière.

Les méthodes de reconstruction *SfM* considèrent un ensemble de vues (un minimum de deux vues) d'une scène fixe et s'appuient sur le modèle sténopé pour inférer, par triangulation, la forme d'une scène rigide. Cependant, si le déplacement des caméras ou la position relative de celle-ci n'est pas connue, l'ambiguïté d'échelle ne peut être résolue et la reconstruction est réalisée à un facteur d'échelle près [Hartley and Zisserman, 2004]).

Si le modèle sténopé est largement utilisé en reconstruction 3D de par sa simplicité il ne permet pas la modélisation de nombreux phénomènes tels que les aberrations géométriques ou le flou optique.

Modèle optique géométrique

Contrairement au modèle sténopé, nous considérons que l'ouverture du diaphragme est non négligeable. Le modèle de formation d'image par une lentille mince associé est décrit par les lois de Snell-Descartes [Born and Wolf, 1999] (figure 2.2.2).

Le diaphragme n'est pas représenté dans ce schéma. Il est ici considéré comme lié à la lentille et son ouverture est équivalente au diamètre effectif de cette dernière. La lentille est rotationnellement symétrique et centrée sur l'axe optique (axe virtuel joignant les deux centres de courbures des dioptres convexes de la lentille considérant un système optique centré). Le foyer image, situé sur l'axe optique, est le point de convergence des rayons lumineux incidents perpendiculaires à la lentille. La distance focale f séparant le centre de la lentille du foyer image est la mesure caractéristique du pouvoir de convergence de la lentille. Elle est directement liée à son indice de réfraction. Il est important de noter que la distance focale correspond à $f = e - f$ suivant la définition de f généralement considérée pour le modèle sténopé (voir section précédente). Le foyer objet est le symétrique du foyer image ($f' = -f$). Les rayons issus du foyer objet sont réfractés par la lentille en formant un faisceau de rayons lumineux parallèles. Les plans focaux image et objet sont perpendiculaires à l'axe optique et passent respectivement par le foyer image et le foyer objet. Le plan capteur situé à la distance e de la lentille correspond à la surface photosensible permettant l'acquisition du signal lumineux.

Dans le cas d'un système optique monofocale, le plan capteur est fixe (il est situé à une distance e fixe de la lentille). Le plan de netteté objet est défini comme le plan objet perpendiculaire à l'axe optique pour lequel un point 3D a pour image un point sur le plan capteur. Ce plan est situé à la distance \hat{d} du centre de la lentille. Nous démontrerons que cette distance est unique pour

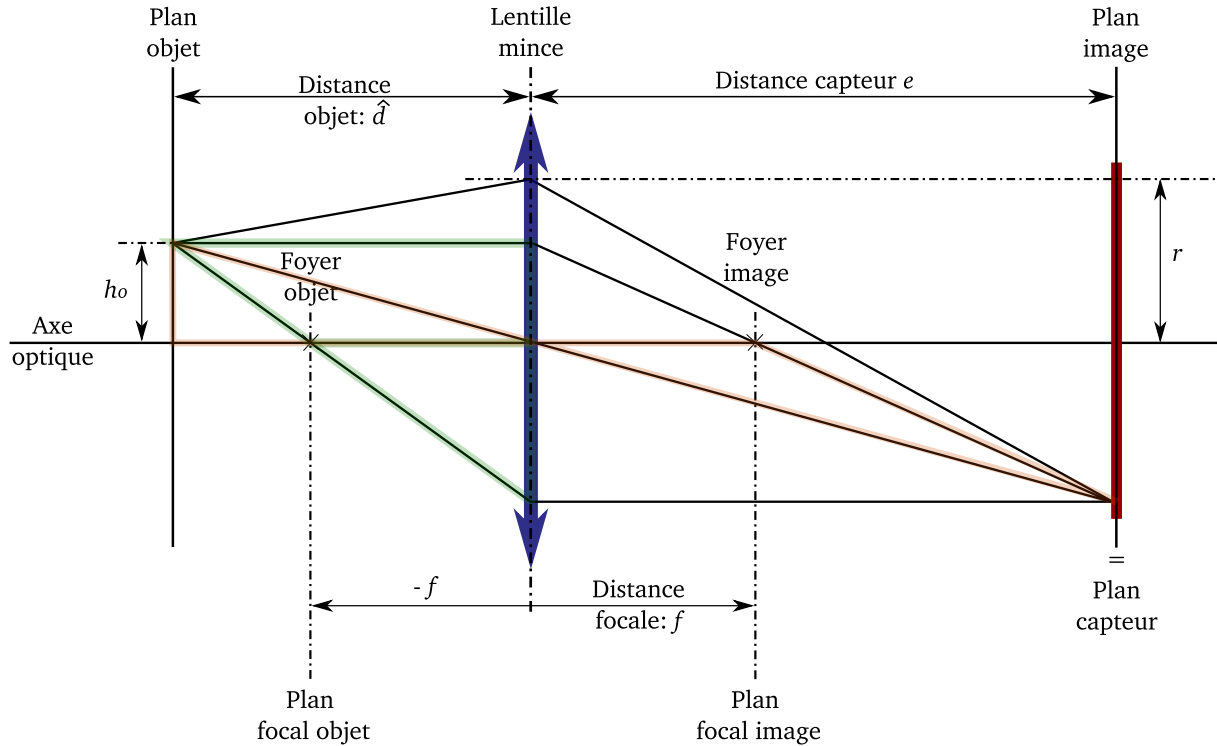


Figure 2.2.2 – Modèle de formation d’image par une lentille mince selon les lois de l’optique géométrique.

un système optique monofocale. Ceci est à différencier des définitions propres au domaine de la photographie pour lequel la notion de profondeur de champ est associée à la plage de profondeur dans laquelle doit se trouver un sujet pour que son image soit considérée comme nette par l’œil humain. Le plan de netteté image peut être assimilé à la définition d’hyperfocale correspondant à la distance minimum à partir de laquelle l’image formée sur le capteur apparaît visuellement nette.

Soit \mathbf{Q}_d un point situé à la distance d de la caméra. L’image de \mathbf{Q}_d est un point si il se trouve sur le plan de netteté objet situé à la distance \hat{d} (figure 1.2.3). Le système optique est dit “mis au point”. Suivant le théorème d’intersection de Thalès (figure 2.2.2), la profondeur de netteté objet \hat{d} est définie par :

$$\left. \begin{array}{l} \frac{h_0}{\hat{d}} = \frac{r}{e} \\ \frac{r}{f} = \frac{r + h_0}{\hat{d}} \end{array} \right\} \Rightarrow \hat{d} = \frac{ef}{e - f} \quad (2.2)$$

où h_0 représente la distance séparant l’axe optique du point \mathbf{Q}_d . La distance de netteté de

l'objet ne dépend pas de la dimension de l'ouverture du dispositif optique, voir figure 2.2.3.

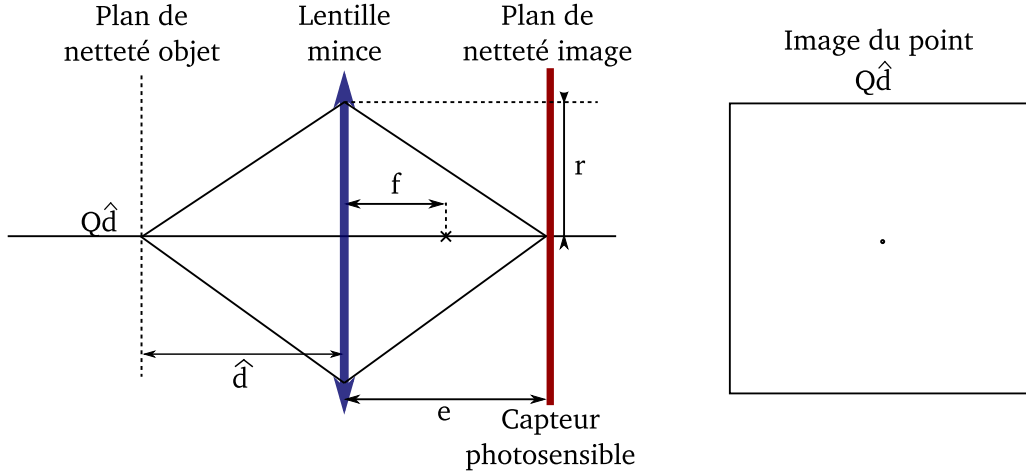


Figure 2.2.3 – Modèle de formation d'une image par une lentille mince. L'image d'un point Q_d est un point si $d = \hat{d}$, ce qui signifie que Q_d se situe sur le plan de netteté objet.

Dans toutes les autres configurations (c'est à dire lorsque l'objet n'est pas situé dans le plan de netteté objet : $d \neq \hat{d}$), l'image de Q_d est un cercle de confusion (figure 2.2.4).

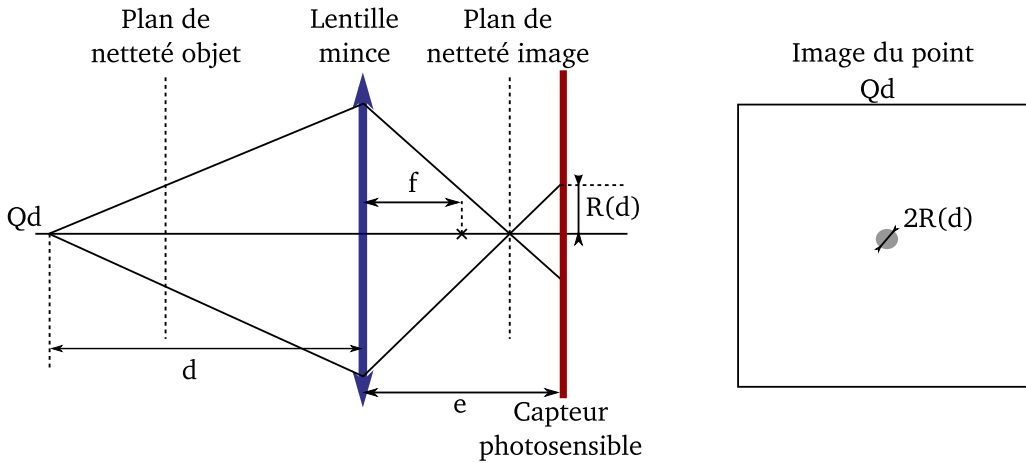


Figure 2.2.4 – Modèle de formation d'une image par une lentille mince. L'image d'un point $Q_{d \neq \hat{d}}$ qui n'est pas situé sur le plan de netteté objet est un cercle de confusion de rayon $R(d)$.

Le rayon R du cercle de confusion est également obtenu par application de la loi de Thalès :

$$\frac{R}{r} = \left| \frac{e - \hat{h}}{\hat{h}} \right| = \left| e \left(\frac{1}{\hat{h}} - \frac{1}{e} \right) \right| \quad (2.3)$$

En s'appuyant sur la formule de lentille mince (2.2), le terme $\frac{1}{\hat{h}}$ peut être substitué :

$$R(d) = r \left| \frac{e}{f} - \frac{e}{d} - 1 \right| \quad (2.4)$$

Il est important de noter que théoriquement $R\left(\hat{d} = \frac{ef}{e-f}\right) = 0$ (l'image est mise au point, voir la figure 2.2.5). Par ailleurs, si l'on considère que le diamètre de l'ouverture est infinitésimal, la profondeur de champ devient infinie (quelle que soit la valeur de d l'image formée par un point sera un point) ce qui est cohérent avec le modèle sténopé présenté à la section précédente.

Considérant un système optique monofocale, seule la distance d séparant le point 3D observé de la lentille est variable. La taille du cercle de confusion est uniquement dépendante de la distance séparant le point \mathbf{Q}_d du plan de netteté objet. La fonction (2.4) n'étant pas bijective, il existe une ambiguïté offrant deux solutions situées de part et d'autre du plan de netteté objet (figure 2.2.5).

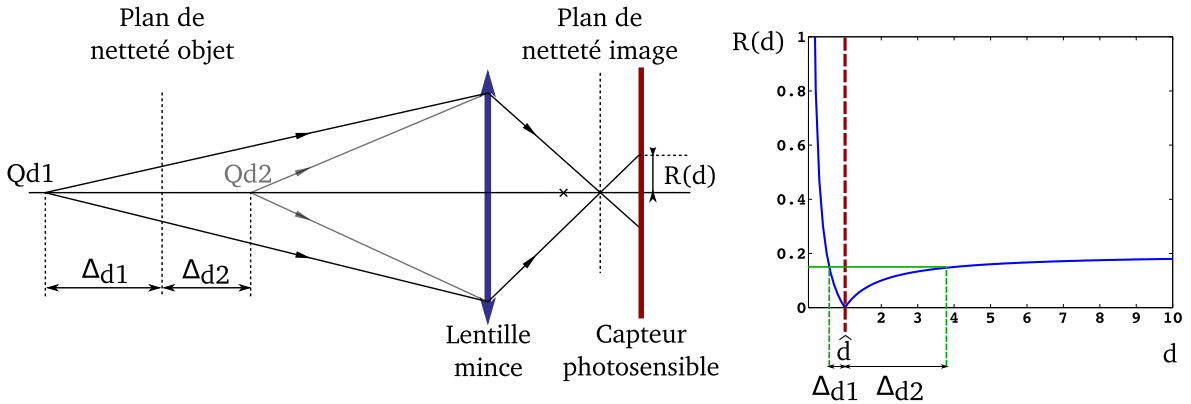


Figure 2.2.5 – Modèle de flou optique. (i) Considérant les lois de l'optique géométrique, deux points situés de part et d'autre du plan de netteté objet possèdent la même image. (ii) Le modèle de flou optique est uniquement dépendant de la distance d'acquisition du point \mathbf{Q}_d .

La **fonction d'étalement du point (PSF : Point Spread Function)** du système optique caractérise sa réponse impulsionnelle. Selon les lois de l'optique géométrique, et supposant un système optique idéal, la distribution de l'intensité de lumière au sein du cercle de confusion est

constante. La *PSF* est alors modélisée par la fonction porte (fonction *pillbox*) :

$$h(x, y) = \begin{cases} \frac{1}{\pi R(d)^2} & \text{si } x^2 + y^2 \leq r^2 \\ 0 & \text{sinon} \end{cases} \quad (2.5)$$

Le modèle optique géométrique permet d'établir un lien entre le degré de flou optique d'un point de l'image et la profondeur Q_3 du point 3D correspondant. Cependant, il ne permet pas de modéliser les phénomènes de diffraction (aberrations géométriques). La PSF défini par ce modèle est imprécise. L'optique ondulatoire complète le modèle optique géométrique et permet d'affiner la fonction d'étalement du point.

2.2.2 Optique ondulatoire

Le modèle optique ondulatoire (ou optique physique) se base sur la représentation de la lumière par une onde électromagnétique. Il complète et valide le modèle géométrique qui en est une simplification basée sur l'hypothèse que tous les éléments composant le système optique sont de grande dimension devant la longueur d'onde de la lumière.

A des fins de concision, nous n'aborderons pas les fondements théoriques de l'optique ondulatoire détaillés dans les ouvrages de référence [Born and Wolf, 1999; Goodman, 2004]. Nous décrirons plus particulièrement le phénomène de diffraction de la lumière par une ouverture circulaire à l'aide des outils de l'optique de Fourier [Goodman, 2004] (approximation de l'optique ondulatoire).

Nous considérons un système optique mis au point. Une source lumineuse monochromatique ponctuelle $S(x, y, z)$ située aux coordonnées $(x, y, 0)^\top$ du plan de netteté objet a pour image le point $I(x, y, \hat{d} + e)$ situé sur le plan de netteté image. L'axe optique est ici confondu avec l'axe de profondeur Z . Le point $I(x, y)$ est défini par la relation suivante :

$$I(x, y) = (h * S)(x, y) \quad (2.6)$$

Cette expression est obtenue par approximation de l'équation de Rayleigh-Sommerfield (approximations de Fresnel et Fraunhofer [Goodman, 2004]) décrivant l'amplitude d'une onde diffractée par une ouverture de forme quelconque. La fonction de transfert optique peut alors être exprimée comme la transformée de Fourier du motif d'ouverture élevée au carré (motif de diffraction de Fraunhofer). Dans le cas d'une ouverture circulaire, le motif obtenu est la tache d'Airy (figure

2.2.6) définie par :

$$\mathcal{A}(\theta) = \mathcal{A}_{max} \left(\frac{2J_1(\kappa r \sin \theta)}{\kappa r \sin \theta} \right)^2 \quad (2.7)$$

où θ désigne un angle d'observation (séparant l'axe optique de la droite joignant le centre de l'ouverture et le point d'observation), \mathcal{A}_{max} correspond à l'intensité maximum de la lumière au centre du motif, J_1 désigne la fonction de Bessel du premier ordre, r est le rayon de l'ouverture et κ est le nombre d'onde égal à $\frac{2\pi}{\lambda}$ (avec λ la longueur d'onde). Le rayon du premier cercle sombre du motif d'Airy (première valeur nulle) peut être approximé par :

$$\sin \theta \approx 1.22 \frac{\lambda}{2r} \quad (2.8)$$

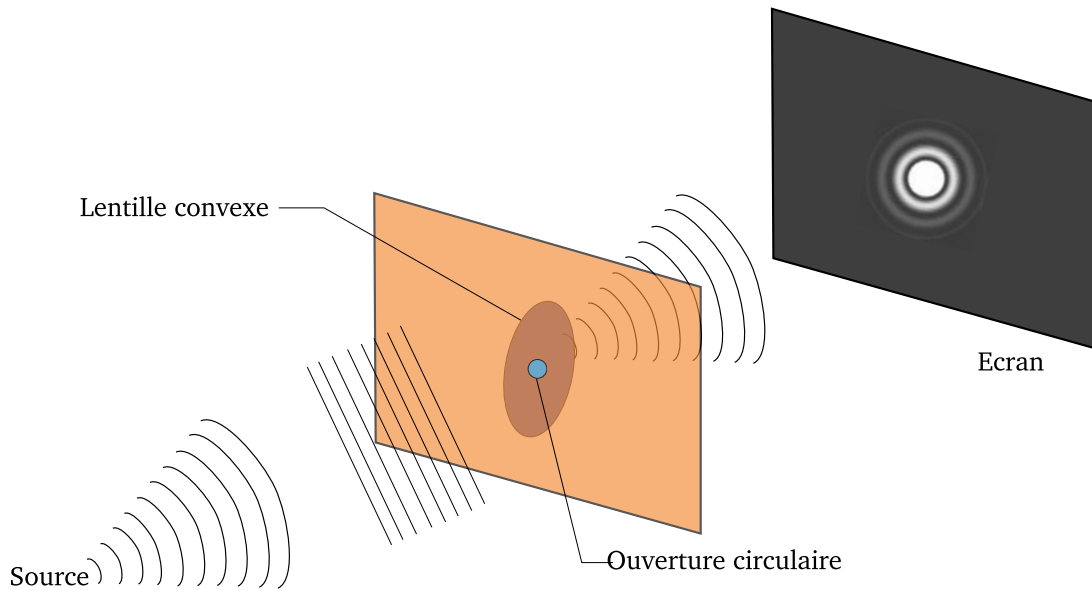


Figure 2.2.6 – Expérience de Fraunhofer réalisée avec une ouverture circulaire. L'onde incidente considérée plane (cas d'une source éloignée du plan obturateur) est diffractée par une ouverture circulaire. Une lentille permet de mettre en évidence le motif de diffraction qui est observé sur un écran confondu avec le plan focal. Dans le cas d'une ouverture circulaire parfaite, le motif de diffraction observé est la tache d'Airy Cambridge in Colour.

Contrairement au modèle optique géométrique, l'image d'un point \mathbf{Q}_d n'est jamais un point. Ceci implique que la résolution maximale d'un capteur photosensible est limitée par la diffraction du système optique (afin d'éviter tout phénomène de recouvrement).

Dans le cas d'une image non mise au point, le flou optique peut être modélisé par une fonc-

tion d'aberration $W(x, y)$ multipliée à la fonction de transfert optique du système mis au point [Claxton, 2007; Saitoh et al., 1997] :

$$W(x, y) = \frac{\pi}{\lambda} \left(\frac{1}{u} + \frac{1}{v} + \frac{1}{f} \right) (x^2 + y^2) \quad (2.9)$$

Les modèles optiques géométriques et ondulatoires soulignent tous deux la relation liant la PSF du système optique et la distance du point imagé. Il est néanmoins important de noter que nous avons ici considéré une lumière monochromatique. Dans le cas d'une lumière polychromatique, il est nécessaire d'intégrer les contributions de chaque longueur d'onde à la fonction d'aberration.

2.2.3 La fonction d'étalement du point : PSF

Si les phénomènes de diffraction doivent être considérés afin de définir de façon précise la PSF d'un système optique, le modèle optique ondulatoire est difficilement exploitable en vision par ordinateur. Cela nécessiterait de considérer une source de lumière polychromatique mais également de modéliser les propriétés d'intégration du capteur photosensible. Une procédure de calibration complexe et minutieuse est par ailleurs nécessaire à l'extraction des constantes du système.

Les méthodes de reconstruction 3D basées sur l'analyse du flou optique s'appuient sur des modèles simplifiés de PSF. [Claxton and Staunton, 2008] montrent que la modélisation Gaussienne généralisée est la plus adaptée aux caméras conventionnelles. Ce modèle permet d'approximer la fonction pillbox ainsi que la fonction Gaussienne en ajustant l'unique paramètre β [Britton, 2006] (figure 2.2.7). La fonction Gaussienne généralisée peut être exprimée par le produit dyadique de deux fonctions Gaussiennes généralisées 1D :

$$g(n_1, n_2) = g1(n_1) \otimes g2(n_2) \quad (2.10)$$

$$g_i(n_i) = e^{-\left(\frac{|n_i - \mu_i|}{\sigma_i \beta}\right)^\alpha}$$

avec $\beta = \left(\frac{\gamma(\frac{1}{\alpha})}{\gamma(\frac{3}{\alpha})}\right)^{\frac{1}{2}}$, γ la fonction gamma), α le paramètre de forme, μ le centre et σ_i l'écart type.

Ce modèle est difficilement exploitable en vision par ordinateur de par le nombre de paramètres à estimer (notamment le paramètre de forme α). Il est par ailleurs sensible au bruit.

La majorité des méthodes de reconstruction par analyse du flou optique s'appuient sur une approximation Gaussienne de la PSF (figure 2.2.8). Cette modélisation a été proposée par [Pentland, 1987] et validée expérimentalement pour une source de lumière polychromatique. La PSF

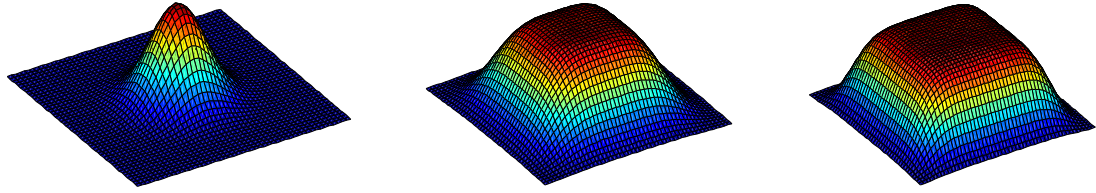


Figure 2.2.7 – Fonction Gaussienne 2D généralisée isotrope d'écart type égal à 2 et de paramètre α égal à 2, 4 et 6 respectivement.

Gaussienne circulaire 2D isotrope et centrée au point $\mu = (x_0, y_0)^\top$ est définie par :

$$g_{x,y,\sigma}(d) = \frac{A}{2\pi\sigma^2} e^{-\frac{(x-x_0)^2+(y-y_0)^2}{2\sigma^2}} \quad (2.11)$$

où A est l'amplitude de la fonction Gaussienne.

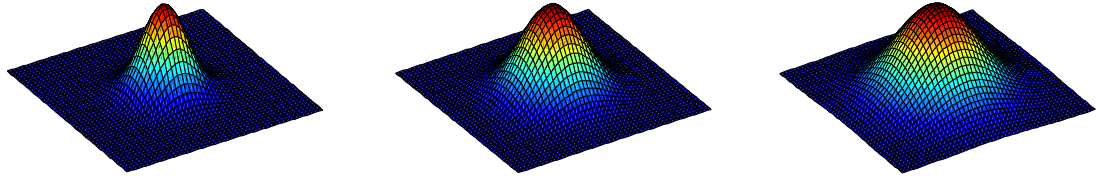


Figure 2.2.8 – Fonction Gaussienne 2D isotrope d'écart type égal à 5, 7 et 9 respectivement.

Nous considérons ce modèle simplifié de PSF qui a prouvé son efficacité en vision par ordinateur. Pour un système optique monofocale, l'écart type $\sigma(d)$ de la PSF Gaussienne est proportionnel à la distance séparant le point 3D \mathbf{Q}_d et le plan de netteté objet [Subbarao, 1988] :

$$\sigma(d) \propto R(d) = r \left| \frac{e}{f} - \frac{e}{\hat{d}} - 1 \right| \quad (2.12)$$

Le coefficient de proportionnalité r dépend directement des paramètres internes de la caméra. Il peut ainsi être calibré.

Nous considérons par ailleurs un système optique parfait (sans aberration et sans perte). L'image I_d d'un plan fronto-parallèle à ce dernier et situé à la profondeur d (suivant l'axe optique) est dérivée de l'équation (2.6) :

$$I_d(x, y) = \left(\hat{I} * g_{\sigma(d)} \right) (x, y) \quad (2.13)$$

où $\hat{I} = I_{\hat{d}}$ correspond à l'image mise au point du plan ou, plus exactement, l'image du plan lorsque ce dernier est situé à la profondeur de netteté objet \hat{d} .

2.2.4 Conclusion

Le modèle de formation d'images considéré suivant les lois de l'optique géométrique et ondulatoire permet de définir un modèle de flou optique liant les paramètres internes de la caméra et la profondeur d'un point 3D à son image. Dans le cas d'un système monofocale, la quantité de flou optique dépend uniquement de la distance séparant le point 3D du plan de netteté objet dont la profondeur de référence \hat{d} peut être calibrée. L'image d'un plan objet situé à la distance d peut être exprimée comme la convolution de l'image mise au point de ce plan par la réponse impulsionnelle du système optique appelé *Point Spread Function (PSF)*. Considérant une source de lumière polychromatique, nous suivons l'hypothèse d'une PSF 2D Gaussienne isotrope dont la variance est proportionnelle à la distance $|d - \hat{d}|$.

$$I_d(x, y) = \left(\hat{I} * g_{\sigma(d)} \right) (x, y) \quad (2.13)$$

$$\sigma(d) \propto R(d) = r \left| \frac{e}{f} - \frac{e}{\hat{d}} - 1 \right| \quad (2.12)$$

Cette modélisation est communément admise par les méthodes de reconstruction 3D par analyse du flou optique [Pentland, 1987]. Ce sont les seules méthodes passives basées vision permettant de résoudre l'ambiguïté d'échelle sans *a priori* forts sur les éléments composant la scène.

3

Etat de l'art sur l'estimation de la profondeur à partir du flou optique

Ce chapitre dresse un état de l'art des techniques de reconstruction 3D par estimation du degré de flou optique. Nous présentons dans une première section l'approche *Depth-from-Focus (DfF)*. Cette dernière repose sur l'acquisition d'un ensemble d'images d'une scène fixe en contrôlant le déplacement de la caméra. L'estimation de l'information de flou optique permet d'extraire les zones de netteté de chaque image et d'inférer la reconstruction 3D de la scène. Nous présentons par la suite l'approche *Depth-from-Defocus (DfD)*. Elle repose généralement sur l'acquisition de deux images depuis le même point de vue mais en contrôlant les paramètres internes de la caméra. L'analyse de la variation du degré de mise au point permet d'inférer la carte de profondeur de la scène. Les travaux menés dans le cadre de cette thèse étendent les méthodes de reconstruction par estimation du flou optique en s'affranchissant des contraintes liées au contrôle des paramètres internes et externes de la caméra. Une étude portant sur la faisabilité de l'estimation de l'échelle des tumeurs en coloscopie par quantification du flou optique a été publiée dans [Chadebecq et al., 2012b]. Cette dernière compare différentes méthodes de mesure du flou optique pour l'estimation de la profondeur absolue d'une scène coloscopique.

La genèse des méthodes d'estimation de la profondeur par analyse du flou optique est communément attribuée à Alex P. Pentland [Pentland, 1982, 1987] malgré l'existence de travaux antérieurs soulignant la relation profondeur d'une scène-degré de mise au point [Grossmann, 1987; Horn, 1968; Jarvis, 1983; Krotkov, 1987; Lighthart and Groen, 1982; Ogle, 1968; Schlag et al., 1983; Subbarao, 1987]. Deux classes de méthodes sont généralement distinguées : les méthodes *Depth-*

from-Focus (DfF) et *Depth-from-Defocus* (DfD) [Crofts, 2007].

Le principe de ces deux approches est rappelé dans l'illustration 3.0.1.

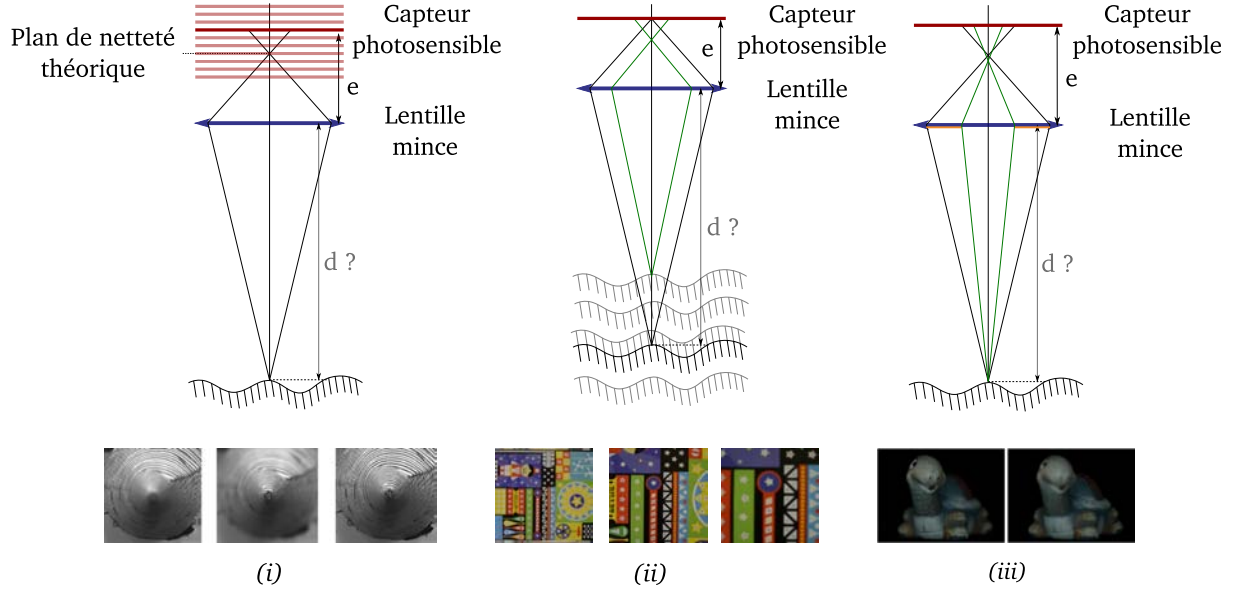


Figure 3.0.1 – Méthodes *Depth-from-Focus* (DfF) et *Depth-from-Defocus* (DfD). L'approche DfF consiste à acquérir un ensemble d'image en contrôlant (i) la distance e séparant la lentille du capteur image [Ahmad and Choi, 2008] ou (ii) la distance d'acquisition d . (iii) L'approche DfD consiste à acquérir deux images de la scène depuis le même point de vue mais en modifiant les paramètres internes de la caméra (par exemple l'ouverture du diaphragme, ici $f = 3.5$ et $f = 16$ respectivement). L'analyse de l'information de flou optique ainsi que la calibration de la PSF du système optique permet une reconstruction 3D de la scène [Li et al., 2013b].

Les méthodes DfF et DfD reposent généralement sur un ensemble d'hypothèses communes :

(i) le système optique est supposé calibré. Au-delà de la calibration géométrique, l'estimation de l'échelle absolue d'une scène nécessite de définir une table de correspondance liant le ou les paramètres du système optique modifiés, la profondeur de l'objet imagé et la mesure de la variation de flou associée. Cette calibration est généralement réalisée en laboratoire dans un environnement contrôlé (système mécanique de translation millimétrique) à l'aide d'une mire plane texturée. Cette procédure est coûteuse et peu adaptée à de nombreux cas pratiques tel que l'imagerie médicale.

(ii) La majorité des méthodes considèrent la reconstruction de surfaces planes fronto-parallèles au système optique. Cette hypothèse d'équifocalité est implicitement admise par les méthodes basées région.

(iii) Les surfaces observées sont Lambertiennes et la scène est rigide.

(iv) Les éléments de la scène sont texturés ou possèdent des contours saillants. Le modèle de flou optique Gaussien considéré (équation (2.13)) correspond à un filtrage passe-bas de l'image mise au point. La fréquence de coupure de ce filtre est liée à l'écart type de la PSF. L'augmentation du degré de flou correspond à une dégradation des composantes haute fréquence de l'image. Ces dernières portent l'information de contour (et les détails de l'image) permettant l'estimation de l'information de mise au point. Les zones homogènes, à l'inverse, ne subissent aucune dégradation et ne peuvent être exploitées.

(v) Dans le cas des méthodes DfD , le paramètre optique généralement modifié est la dimension de l'ouverture. La taille du diaphragme influe sur le degré de netteté de l'observation mais elle n'implique aucune modification spatiale de l'image. La variation des autres paramètres internes du système optique ou des paramètres externes de la prise de vue implique de réaliser le recalage des images.

Un état de l'art des principales contributions DfF et DfD est présenté dans les sous-sections suivantes.

3.1 Depth-from-Focus

Nous considérons plus particulièrement les méthodes basées sur le contrôle des paramètres externes de la caméra (figure 3.1.1). Les observations sont supposées recalées bien que cette étape puisse s'avérer délicate dans de nombreux cas (par exemple en endoscopie [Alcantarilla et al., 2013; Kaufman and Wang, 2008]).

Les méthodes DfF consistent à extraire, pour chaque acquisition, les zones de netteté (équation (2.2)). A l'aide des données de calibration, la profondeur de la scène peut être inférée par substitution. Cela suppose que chaque élément de la scène est observé mis au point dans une des images acquises. L'estimation dense d'une carte de profondeur nécessite de plus un nombre d'image important impactant les temps de calcul. Les méthodes DfF ont principalement été associées à des systèmes de mise au point automatique ainsi qu'à des méthodes de reconstruction 3D pour l'imagerie microscopique.

Les approches DfF reposent essentiellement sur le choix d'un opérateur de mesure du flou optique efficace. Une étude comparative des différents opérateurs est présentée dans [Pertuz et al., 2013a,b]. Les méthodes de fusion d'images "multi-focus" visant à reconstruire l'image entièrement mise au point d'une scène s'appuient sur des opérateurs de mesure de flou similaires [Anish and Jebaseeli, 2012; Li and Yang, 2008]. Deux catégories peuvent être distinguées : les opérateurs s'appliquant dans le domaine spatial (directement appliqués à l'image) et ceux s'appliquant dans

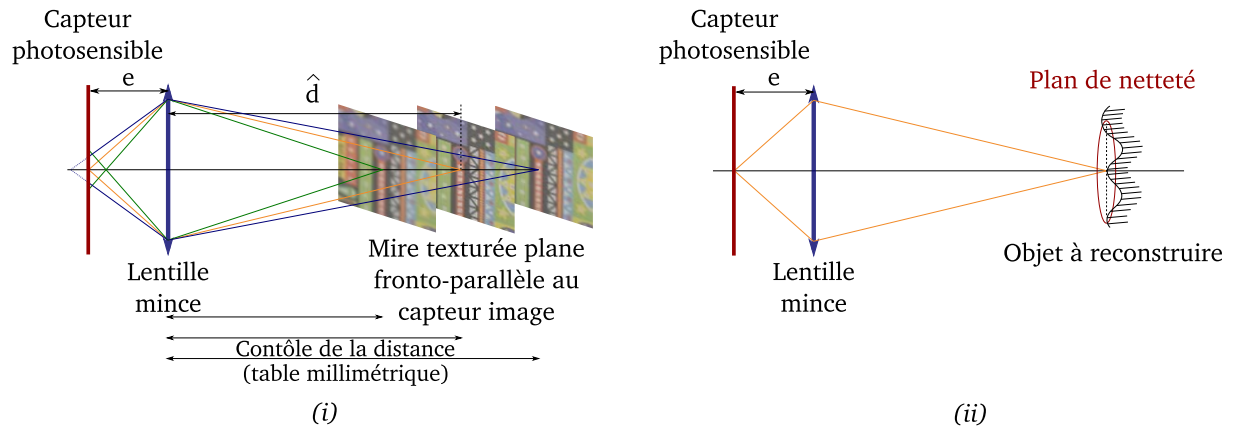


Figure 3.1.1 – Méthodes *Depth-from-Focus* (DfF). (i) La phase de calibration consiste à définir la profondeur de référence \hat{d} correspondant au plan de netteté objet du système optique. Une mire plane texturée est déplacée fronto parallèlement au système optique. La mesure de l'information de flou permet d'extraire \hat{d} . (ii) Durant la phase de mesure un ensemble d'images est acquise en modifiant la distance séparant la caméra de la scène imagée. Pour chaque acquisition, une mesure de flou est réalisée afin d'extraire les zones de netteté.

le domaine fréquentiel.

Parmi les opérateurs spatiaux, les opérateurs dérivatifs à l'ordre un (opérateur de différence finie et opérateur de Tenengrad [Malik and Choi, 2008; Minhas et al., 2009]) et deux (opérateur Laplacien [Nayar and Nakagawa, 1994]) permettent une estimation efficace de l'information de flou. Les opérateurs dits statistiques sont basés sur l'analyse du contraste [Helmlí and Scherer, 2001] ou de la variance locale de la luminance (au voisinage d'un pixel) [Gaganov and Ignatenko, 2009; Huang and Jing, 2007; Wee and Paramesran, 2007]. Ces opérateurs sont souvent associés à des approches fenêtrées (utilisant notamment des fenêtres glissantes recouvrantes) afin de limiter leurs sensibilités au bruit.

Les opérateurs fréquentiels basés sur la décomposition en cosinus discret [Sang-Yong et al., 2008] (systèmes de mise au point automatique) permettent une estimation globale de la mesure de flou. Les opérateur temps-échelle basés sur la décomposition en ondelette des images [Lin et al., 2004; Minhas et al., 2011; Xie et al., 2006] permettent une analyse locale du flou optique.

Les expériences menées dans l'étude [Pertuz et al., 2013a] montrent que dans des conditions d'imagerie idéales (peu sujettes au bruit, aux changements d'illuminations ...), les opérateurs basés Laplacien obtiennent les meilleures performances. Suivant les conditions de prise de vue et le niveau de bruit, il est cependant difficile de choisir l'opérateur idoine. Les opérateurs basés sur la décomposition en ondelette s'avèrent généralement plus robustes aux changements de condition d'acquisition. Nous présentons plus particulièrement ces deux opérateurs de mesure de flou dans

les sous-sections suivantes.

Les méthodes *DfF* sont par ailleurs sensibles au mode d'acquisition des images [Muhammad and Choi, 2012] ainsi qu'au choix de la taille de la fenêtre de mesure considérée [Florczak and Petko, 2014]. Afin de s'affranchir de ces contraintes, ainsi que de la contrainte d'équifocalité, les méthodes nommées *Focused Image Surface* (FIS) [Choi and Yun, 2000; Rahmat et al., 2013; Shim and Choi, 2010; Subbarao and Choi, 1995] considèrent uniquement la surface formée par l'ensemble des points de netteté observés dans l'ensemble d'acquisition.

3.1.1 Opérateur spatial de mesure de flou optique Laplacien modifié

L'opérateur Laplacien est généralement utilisé pour approximer la dérivée seconde d'un signal. Il est défini par [Ahmad and Choi, 2007] :

$$\Delta I = \frac{\partial^2 I}{\partial x^2} + \frac{\partial^2 I}{\partial y^2} \quad (3.1)$$

Il peut également être interprété comme un filtre passe-haut [Nayar and Nakagawa, 1994]. L'extraction de l'image la plus nette d'une sous région au sein d'un ensemble d'images correspond ainsi à la recherche de l'argument maximum de son Laplacien. L'inconvénient de cet opérateur est que les dérivées partielles suivant les coordonnées x et y peuvent s'annuler. Ceci est notamment le cas pour des images texturées ce qui rend la mesure de l'information de flou instable. [Nayar and Nakagawa, 1994] proposent une version modifiée de l'opérateur Laplacien afin de palier à cette problématique :

$$\tilde{\Delta} I = \left| \frac{\partial^2 I}{\partial x^2} \right| + \left| \frac{\partial^2 I}{\partial y^2} \right| \quad (3.2)$$

Pratiquement, le calcul du Laplacien modifié est réalisé à l'aide de deux filtres de convolution 1D :

$$\tilde{\Delta} I(x, y) = |(\mathcal{L}_x * I)(x, y)| + |(\mathcal{L}_y * I)(x, y)| \quad (3.3)$$

avec $\mathcal{L}_x = \mathcal{L}_y^\top = [-1 \ 2 \ -1]$. La dérivée seconde est estimée par différence finie. [Thelen et al., 2009] proposent de compléter cette mesure en considérant les axes diagonaux du filtre. L'équation (2.13) (voir le chapitre 2) devient :

$$(\tilde{\Delta} * I_d)(x, y) = \left(\left(\tilde{\Delta} * \hat{I} \right) * g_{\sigma(d)} \right) (x, y) \quad (3.4)$$

Différentes adaptations de l'opérateur Laplacien modifié ont été proposées dans la littérature.

[An et al., 2008] s'appuie sur un sous ensemble d'images successives afin de tirer parti de la redondance de l'information de flou. Cette méthode de reconstruction 3D offre une meilleure robustesse au bruit.

Les opérateurs spatiaux d'estimation du flou optique sont particulièrement rapides à évaluer. Ils ont notamment été intégrés à des systèmes de mise au point automatique. Ils sont cependant peu adaptés à l'estimation de l'échelle d'une scène. La mesure est particulièrement dépendante du contenu de cette dernière et elle est peu robuste au bruit.

3.1.2 Opérateur temps-échelle de mesure du flou optique par décomposition en ondelette

Contrairement à la transformée de Fourier qui permet uniquement une analyse en fréquence [Granlund and Knutsson, 1995], la décomposition en ondelette permet une analyse temps-échelle. Nous n'aborderons pas les fondements mathématiques de la décomposition en ondelette qui sont détaillés dans [Mallat, 2009]. Cette méthode est basée sur une mesure de similarité (produit scalaire) entre une fonction généralement oscillante et de moyenne nulle appelée ondelette (correspondant à une petite oscillation) et l'image. Au premier niveau de décomposition, la mesure implique l'ondelette mère. Au niveau supérieur, la mesure implique les ondelettes filles correspondant à l'ondelette mère dilatée. Cette méthode permet ainsi de dissocier les petites échelles (premier niveau correspondant aux détails de l'image) des grandes échelles. Chaque niveau est composé des coefficients de similarité calculés suivant les trois directions principales de l'image (W_{LH1} , W_{HL1} , W_{HH1}) et du reste de la décomposition (W_{LL1} correspondant aux grandes échelles).

Les coefficients extraits aux premiers niveaux de décomposition fournissent ainsi une mesure indirecte du flou optique [Xie et al., 2006; Yang and Nelson, 2003]. Les premières approches se sont appuyées sur la base de décomposition orthogonale de Daubechies [Daubechies, 1992]. Différentes mesures ont alors été proposées comme la somme des coefficients suivant les trois directions principales de l'image ou l'analyse de variances [Yang and Nelson, 2003]. A des fins de robustesse, la mesure peut impliquer plusieurs niveaux de décomposition [Xie et al., 2006]. Des opérateurs de mesure du flou optique basés curvelet (généralisation des ondelettes permettant une décomposition suivant différents angles) ont également été proposés suivant des stratégies similaires [Minhas et al., 2011]. L'utilisation d'une base de décomposition adaptée au contenu de l'image permet d'améliorer la précision des mesures de flou.

Cette approche permet la reconstruction 3D ou la mise au point automatique de systèmes

optiques mais elle n'est pas adaptée à l'estimation de l'échelle d'une scène. Elle est peu résistante au bruit et la valeur des coefficients ne peut être associée de façon robuste à la profondeur d'une scène.

3.2 Depth-from-Defocus

Les méthodes *DfD* sont intrinsèquement inadaptées à l'imagerie monofocale. Elles requièrent la modification des paramètres internes du système optique afin d'évaluer la variation de la mise au point d'une scène rigide (figure 3.2.1). Il est néanmoins pertinent de les introduire dans notre étude car elles impliquent une mesure indirecte de l'information de flou. Les méthodes multi-vues sont présentées dans une première section. Nous présenterons dans une seconde section différentes adaptations des méthodes *DfD* au cas d'une image unique. Nous détaillerons par la suite les approches *DfD* basées sur une estimation conjointe du flou relatif et d'un appariement inter-image permettant de s'affranchir des contraintes d'acquisition de la scène. Le tableau 3.2.1 présenté à la suite de cette section récapitule les principales contributions aux méthodes *DfD* citées dans ce document. Il actualise l'état de l'art présenté dans [Claxton, 2007; Crofts, 2007; Raj, 2009].

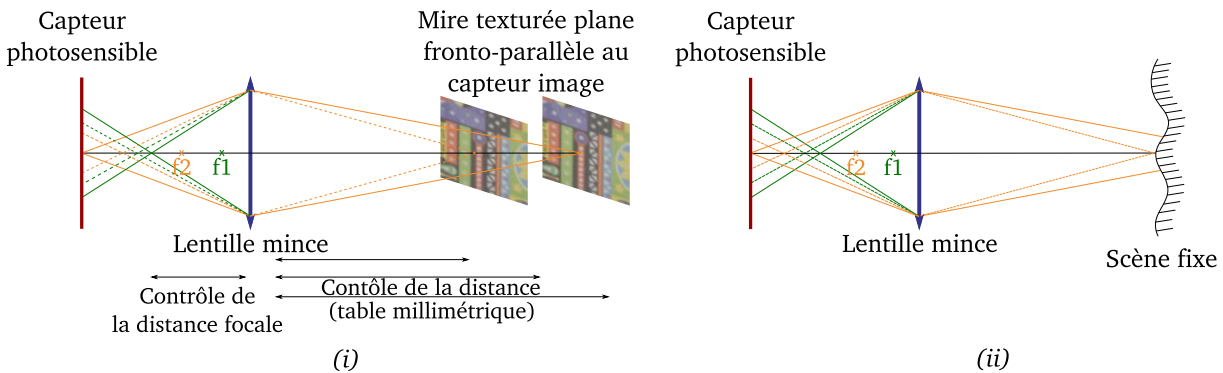


Figure 3.2.1 – *Depth-from-Defocus* (DfD). (i) L'étape de calibration consiste généralement à acquérir deux (ou plusieurs) images d'une mire plane texturée à différentes profondeurs. Chaque paire d'image correspond à des paramètres internes différents et connus du système optique. La variation de flou est mesurée et une table de correspondance est construite. (ii) Durant la phase de mesure, deux images sont acquises en modifiant les paramètres internes idoines du système optique. La mesure de la variation de flou relative permet d'inférer la reconstruction 3D à l'échelle de la scène.

3.2.1 Méthodes multi-vues

Nous distinguons les méthodes basées sur une mesure du flou optique réalisée dans le domaine fréquentielle et celles pour lesquelles la mesure est réalisée dans le domaine spatial.

Méthodes fréquentielles

Transformée de Fourier et décomposition en ondelette : Dans ses travaux fondateurs, [Pentland, 1987] définit l’approche *DfD* comme un problème de filtrage inverse. Il propose une solution basée région considérant deux images d’une scène réalisées en utilisant les ouvertures extremums du système optique (aucun appariement n’est nécessaire). La région j de l’image k notée $I^{k,j}$ est exprimée en coordonnées polaires par l’équation suivante :

$$I^{k,j}(r, \theta) = I^k(x_0 + r \cos \theta, y_0 + r \sin \theta) \quad (3.5)$$

où x_0, y_0 représentent les coordonnées du centre de la région j et (r, θ) représente la position d’un point en coordonnées polaires (rayon et angle respectivement).

Afin d’estimer la variation de flou entre deux régions similaires, il est nécessaire d’éliminer de l’équation (2.13) la fonction correspondant à l’image de la région entièrement mise au point. Cela est réalisé dans le domaine de Fourier en considérant le ratio :

$$\frac{\mathcal{I}^{1,j}(\lambda, \Theta)}{\mathcal{I}^{2,j}(\lambda, \Theta)} = \frac{\widehat{\mathcal{I}}(\lambda, \Theta) \times \mathcal{G}(\mu_1, \sigma_1)}{\widehat{\mathcal{I}}(\lambda, \Theta) \times \mathcal{G}(\mu_2, \sigma_2)} = \frac{\mathcal{G}(\mu_1, \sigma_1)}{\mathcal{G}(\mu_2, \sigma_2)} \quad (3.6)$$

où (λ, Θ) représentent respectivement une fréquence spatiale et une fréquence angulaire, $\mathcal{G}(\mu, \sigma)$ représente la transformée de Fourier du modèle de PSF Gaussien isotrope et centré considéré.

Le développement de l’équation (3.6) conduit à l’expression suivante :

$$\ln \mathcal{I}^{1,j}(\lambda) - \ln \mathcal{I}^{2,j}(\lambda) = \ln \frac{\sigma_1^2}{\sigma_2^2} + \lambda^2 2\pi^2 (\sigma_1^2 - \sigma_2^2) \quad (3.7)$$

Cette approche considère un système optique dont les paramètres internes ainsi que le facteur de proportionnalité sont calibrés (équation (2.12)). La profondeur d’acquisition de chaque région est obtenue en substituant les valeurs de σ_1 et σ_2 dans l’équation (2.4). Elle est cependant peu résistante au bruit et nécessite un changement important des paramètres de l’ouverture.

[Subbarao, 1987, 1988; Subbarao and Gurumoorthy, 1988] étendent [Pentland, 1987] en proposant des méthodes basées sur la modification des différents paramètres internes du système

optique. [Subbarao, 1989] propose une formalisation théorique des problèmes de filtrage inverse en considérant différents types de surface ainsi que l’extension de la fonction d’étalement du point aux lignes (*Edge Spread Function, ESF*) afin d’améliorer la robustesse et la précision de reconstruction 3D de cette méthode.

[Favaro and Duci, 2008] proposent un modèle de formation d’images original reposant la modélisation de la PSF du système optique par une fonction Gaussienne 3D et de la scène par une densité de volume $W : \mathbb{R}^3 \rightarrow [0, \infty)$. L’information photométrique est donnée par l’intensité de W et la géométrie est donnée par la position de chaque point (équation (2.6)). Ce modèle permet d’introduire des *a priori* géométriques ou radiométriques relatifs à la nature de la surface. Deux solutions sont proposées. La première méthode est basée sur un algorithme de déconvolution et requière un nombre important d’images (au moins 100). La seconde tire partie des propriétés de linéarité du modèle de formation d’images afin d’estimer la profondeur de la scène. Elle a été évaluée pour un ensemble de 5 images. Seule une évaluation qualitative de la méthode a été réalisée afin de valider le modèle théorique introduit dans cette étude. Les méthodes proposées ne permettent pas d’inférer simplement l’échelle de la scène.

Afin de s’affranchir de l’hypothèse d’équifocalité imposée par l’analyse fréquentielle, des techniques *DfD* temps-échelle basées sur la décomposition en ondelette ont été proposées [Asif et al., 2005; Rooms, 2002]. Ces dernières reposent sur l’analyse de la variation des coefficients de décomposition. Malgré leur efficacité, elles ne permettent pas d’inférer l’échelle d’une scène et ont été principalement appliquées à la mise au point post-acquisition d’images (déconvolution aveugle).

Filtrage à l’aide d’opérateurs rationnels : Les méthodes *DfD* par filtrage ont initialement été introduites afin de pallier aux problématiques inhérentes aux méthodes fréquentielles fenêtrées (choix de la dimension de la fenêtre, . . .) [Xiong and Shafer, 1997]. Elles sont cependant coûteuses et nécessitent l’utilisation d’un grand nombre de filtres afin de réaliser une reconstruction 3D dense. [Watanabe and Nayar, 1998] proposent une approche *DfD* basée sur un ensemble de filtres rationnels. Elle repose sur l’expression du modèle de flou relatif normalisé entre deux images comme une fraction rationnelle de fonctions de base :

$$\frac{M(u, v; \alpha)}{P(u, v; \alpha)} = \frac{\mathcal{I}^2(u, v; \alpha) - \mathcal{I}^1(u, v; \alpha)}{\mathcal{I}^2(u, v; \alpha) + \mathcal{I}^1(u, v; \alpha)} = \frac{\sum_{i=1}^{n_P} c_{P_i}(u, v) b_{P_i}(\beta)}{\sum_{i=1}^{n_M} c_{M_i}(u, v) b_{M_i}(\beta)} \quad (3.8)$$

où α correspond à la profondeur normalisée. La profondeur de l’élément de la scène le plus distant est égale à -1 pour la première image. Inversement, la profondeur de l’élément le plus proche de la camera est égale à 1 pour la seconde image. Ceci implique que l’ensemble des éléments de la scène

est inclus dans cette plage de distance. β est la distance estimée. $b_{P_i}(\beta)$ ($i = 1, \dots, n_P$) et $b_{M_i}(\beta)$ ($i = 1, \dots, n_M$) représentent les fonctions de bases. $c_{P_i}(u, v)$ et $c_{M_i}(u, v)$ sont des coefficients (poids) dépendants des fréquences spatiales (u, v) . Les auteurs s'appuient sur trois fonctions de bases ($n_P = 2, n_M = 1$) : $b_{P1}(\beta) = \beta$, $b_{P2}(\beta) = \beta^3$ et $b_{M1}(\beta) = 1$. L'erreur d'estimation du ratio $\frac{M}{P}$ est minimisée à l'aide de la méthode de Newton-Raphson afin d'estimer les coefficients du modèle [Bonnans et al., 2006]. Cette méthode fenêtrée nécessite un pré-traitement des images mais elle utilise une base de filtres rationnels compacte. [Li et al., 2013a] l'étendent en proposant des opérateurs basés sur le modèle de PSF Gaussien et Gaussien généralisé. Un état de l'art des méthodes *DfD* basées sur des opérateurs rationnels est présenté dans ces travaux.

Equation de diffusion de la chaleur : L'équation aux dérivés partielles de diffusion de la chaleur isotrope est donnée par l'expression :

$$\frac{\partial I(x, y; t)}{\partial t} = b \left(\frac{\partial^2 I(x, y; t)}{\partial x^2} + \frac{\partial^2 I(x, y; t)}{\partial y^2} \right) \quad (3.9)$$

où b est une constante de diffusivité thermique caractérisant la capacité d'un matériau à diffuser la chaleur en son sein. Cette équation modélise également le processus de changement de mise au point d'une image $I(x, y; 0)$ [Namboodiri and Chaudhuri, 2007; Sochen et al., 2001]. La solution à l'équation de la chaleur est donnée par la convolution de cette dernière avec une fonction Gaussienne dont l'écart type évolue en fonction du temps : $\sigma^2 = 2bt$.

[Namboodiri and Chaudhuri, 2007] s'appuient sur le modèle de diffusion de la chaleur pour expliciter l'évolution du degré de mise au point entre deux images $I(x, y; t_1)$ et $I(x, y; t_2)$ avec $\sigma_1 < \sigma_2$. Cette méthode nécessite l'utilisation d'une mesure de mise au point afin d'extraire la profondeur de la scène. Les deux auteurs proposeront par la suite une approche statistique adaptée au cas d'une seule vue [Namboodiri and Chaudhuri, 2008]. Elle est cependant imprécise et ne permet pas d'estimer l'échelle d'une scène.

Les méthodes *DfD* basées sur l'équation de diffusion de la chaleur seront significativement étendues dans le domaine spatial [Favaro et al., 2008] (voir section suivante).

Méthodes spatiales

Méthodes algébriques [Ens and Lawrence, 1993] soulignent les inconvénients des méthodes fréquentielles fenêtrées. Ils proposent différentes approches matricielles basées sur l'estimation

directe de la variation de flou $h_3(x, y)$ dans le domaine spatial :

$$(h_1 * h_3)(x, y) = h_2(x, y) \quad (3.10)$$

où $h_1(x, y)$ et $h_2(x, y)$ correspondent respectivement à la valeur de la PSF de la première et de la seconde image. La variation de flou est obtenue, pour une région de l'image, à l'aide d'une recherche itérative (dans une table de correspondance) du paramètre de variance de la fonction Gaussienne $h_3(x, y)$ minimisant la fonction suivante :

$$\sigma_{ROI} = \underset{\sigma_3}{\operatorname{argmin}} \sum_{x=0}^{N-j} \sum_{y=0}^{N-j} ((I^1 * h_3(\sigma_3))(x, y) - I^2(x, y))^2 \quad (3.11)$$

où N est la taille d'une région de la scène, j est la taille du filtre Gaussien $h_3(x, y)$. Cette méthode est particulièrement coûteuse. L'erreur RMS d'estimation de la profondeur est de 1,3 % (pour des images avec un faible niveau de bruit). Elle sera améliorée dans [Claxton and Staunton, 2008] en considérant la modélisation de l'ESF par une fonction Gaussienne généralisée.

[Subbarao and Surya, 1994] proposent une méthode basée sur une transformation spatiale de l'image (*Spatial Transform Method (STM)*). L'image mise au point est approximée par une fonction cubique. Considérant une PSF rotationnellement symétrique, le développement en série de Taylor de l'équation (2.13) permet de formuler la STM inverse :

$$\hat{I}(x, y) = I^1(x, y) - \frac{\sigma_1^2}{4} \Delta I^1(x, y) \quad (3.12)$$

La profondeur de la scène est approximée en comparant localement (pour des régions de taille 9×9 pixels) deux images. La STM généralisée permet de s'affranchir de la modélisation de l'image par un polynôme cubique en s'appuyant sur l'expression de ses dérivées par sa convolution avec une base de filtres orthogonaux [Meer and Weiss, 1990]. La méthode STM est efficace mais les précisions de reconstruction sont médiocres (notamment dépendante de la forme et de l'apparence de la surface). Elle se destine à des systèmes de mise au point automatique.

[Ziou, 1998; Ziou and Deschênes, 2001] améliorent STM en s'appuyant sur la décomposition des régions de l'image par une polynôme Hermitien. Les auteurs démontrent que l'acquisition la plus floue d'une paire d'image peut être exprimée par les dérivées partielles de la seconde image et de la différence de netteté. La méthode proposée améliore significativement les précisions de reconstruction (erreur de 2,21 % pour un objet plan situé dans la plage de distance 115-125 cm

contre une erreur de 4,22 % pour la méthode STM). [Deschênes et al., 2004; Deschênes et al., 2008] étendent ce modèle afin de prendre en considération les mouvements de la camera entre deux acquisitions (déplacement spatial 2D et 3D ainsi que le facteur de grossissement, voir section 3.2.3).

[Favaro and Soatto, 2005] propose une méthode basée région s'appuyant sur le calcul d'opérateurs de projection orthogonaux. Elle permet d'estimer simultanément la topologie d'une surface et son apparence. Si la PSF du système optique est connue les opérateurs sont directement estimés à l'aide d'une décomposition en vecteurs singuliers. Dans le cas contraire, un processus d'apprentissage est nécessaire. Cette méthode est particulièrement rapide. Elle s'avère néanmoins difficilement adaptable à l'estimation de l'échelle d'une scène.

Méthodes statistiques : [Rajagopalan and Chaudhuri, 1999] modélisent les paramètres de mise au point (paramètres relatifs à la PSF ainsi qu'à l'image entièrement mise au point) par des champs de Markov. Une méthode de recuit simulé permet d'optimiser l'équation du maximum *a posteriori* définie par :

$$\max_{\zeta, \hat{I}} \frac{P(I^1, I^2, \dots, I^n | \zeta, \hat{I}) P(\zeta) P(\hat{I})}{P(I^1, I^2, \dots, I^n)} \quad (3.13)$$

où ζ dénote le champ aléatoire correspondant aux paramètres de flou et \hat{I} au champ aléatoire représentant l'image mise au point. Plusieurs observations peuvent être utilisées afin d'améliorer la précision des mesures. Cette solution s'avère néanmoins coûteuse.

[Bhasin and Chaudhuri, 2001] étendent cette méthode afin de considérer les occultations. Considérant deux plans fronto-parallèles au système optique formant un échelon, des phénomènes d'occultations partielles peuvent biaiser la mesure de l'information de netteté au voisinage de la discontinuité. Une étape de détection des contours est considérée. L'information de flou est réajustée à l'aide d'une méthode itérative (considérant l'estimation de profondeur courante). Seules les discontinuités linéaires ont été considérées. Aucune évaluation sur des données réelles n'a été considérée.

Récemment, [Li et al., 2013b] ont combiné une approche statistique *DfD* à une approche SfS [Drouot et al., 2008]. Le processus de formation des images est modélisé par un réseau Bayésien dont les nœuds suivent des probabilités Markoviennes. L'approche consiste à minimiser la probabilité :

$$\tilde{\mathcal{D}} = \underset{\mathcal{D}}{\operatorname{argmax}} P(\mathcal{D} | I^1, I^2) = \underset{\mathcal{D}}{\operatorname{argmax}} P(I^1, I^2) P(\mathcal{D}) \quad (3.14)$$

où $\tilde{\mathcal{D}}$ correspond à la carte de profondeur estimée. Cette dernière est contrainte par l'information

d'ombrage présente dans la scène contrairement à la majorité des méthodes de l'état de l'art s'appuyant sur des contraintes de lissage. Les résultats obtenus pour des données synthétiques démontrent le gain de précision de cette méthode. Aucune expérimentation sur données réelles n'a été réalisée.

Equation de diffusion de la chaleur : [Favaro et al., 2008] proposent une méthode *DfD* améliorant de façon significative les précédentes approches basées sur l'équation de diffusion de la chaleur [Namboodiri and Chaudhuri, 2007]. Cette dernière s'appuie sur l'équation de diffusion homogène afin de considérer la nature spatialement variante du flou (estimation dense) :

$$\begin{cases} \frac{\partial I}{\partial t}(x, y; t) = \nabla \cdot (c(x, y)\nabla I(x, y; t)) \text{ avec } t \in (0, \infty) \\ \frac{\partial I}{\partial t}(x, y; t) = r(x, y) \end{cases} \quad (3.15)$$

Les paramètres de cette fonction sont explicités dans l'équation (3.9). La solution $s : \mathbb{R}^2 \times [0, \infty) \mapsto [0, \infty)$ au temps $t = \tau$ est l'image $I(x', y') = I(x, y; \tau)$. Dans le cas où l'équation de diffusion est appliquée à la modélisation du flou relatif entre deux images, l'équation (3.15) peut être reformulée par :

$$\begin{cases} \frac{\partial I}{\partial t}(x, y; t) = \nabla \cdot (c(x, y)\nabla I(x, y; t)) \text{ avec } t \in (0, \infty) \\ I(x, y; 0) = I^1(x, y) \quad \forall (x, y) \in \Omega \\ (c(x, y)\nabla I(x, y; t)) \cdot \mathbf{n}(x, y) = 0 \quad \forall (x, y) \in \partial\Omega \end{cases} \quad (3.16)$$

avec Ω le domaine de l'image, $I(x, y; \Delta t) = I^2(x, y) \quad \forall (x, y) \in \Omega$, \mathbf{n} est le vecteur normal aux bornes de $\partial\Omega$. La troisième contrainte permet de garantir l'unicité de la solution. La résolution du système impose également que l'image I^1 soit plus nette que l'image I^2 (*forward diffusion*) ou qu'une partition de l'image soit réalisée afin de reformuler l'équation (3.16). Pratiquement la solution du système est donnée par un algorithme itératif permettant la minimisation de :

$$\hat{\mathcal{D}} = \underset{\mathcal{D}}{\operatorname{argmin}} \int_{\Omega} (I(x, y; \delta t) - I^2(x, y))^2 dx dy \quad (3.17)$$

Cette méthode élégante est néanmoins coûteuse et aucune évaluation de la précision d'estimation de l'échelle n'a été réalisée au profit d'une analyse qualitative de la méthode.

[Wei et al., 2009, 2012] étendent [Favaro et al., 2008] au cas d'une camera monofocale. La seconde image est acquise en déplaçant le système optique. L'image I^2 est exprimée en fonction de I^1 et de la variation de flou relative. Dans le cas où le déplacement du système optique n'est

pas connu, un algorithme itératif est proposé afin d’actualiser l’estimation courante de la carte de profondeur en fonction de la différence de flou relative. Une évaluation de cette méthode sur des images microscopiques de nano-grilles a été réalisée. L’erreur moyenne de la mesure est de 2,3%. La problématique d’appariement des images n’a pas été traitée dans cette approche.

3.2.2 Méthodes basée sur une vue unique

[Saxena et al., 2008] proposent une approche basée sur un apprentissage supervisé de différents environnements (forêt, routes, immeubles, ...). L’image est divisée en régions. Pour chacune d’entre elles un descripteur de profondeur absolue ainsi qu’un descripteur relatif inter-région (pour un voisinage donné) est calculé. Ce dernier est basé sur des indices locaux : la variation et les gradients de texture ainsi que la couleur. Le descripteur de profondeur absolue est lui calculé à partir d’une approche multi-échelles permettant de capturer, pour différentes résolutions, l’apparence d’une région. Cette approche s’appuie sur l’hypothèse qu’un objet imagé à différentes profondeurs présente une apparence singulière pour différentes résolutions. La relation inter région est modélisée par un champ de Markov aléatoire multi-échelle et hiérarchique. Différents *a priori* sont imposés au modèle. Deux pixels voisins de couleurs identiques sont par exemple associés à une même profondeur. Cette méthode a notamment été combinée à une méthode de reconstruction stéréo dans le cas de la détection d’objets mobiles. Les performances sont peu adaptées à une application de métrologie. Plus récemment, une approche probabiliste basée sur la décomposition de l’image par une base de filtre de Gabor a été introduite [Zhu et al., 2013]. Chaque filtre correspond à un sous-ensemble de fréquences spatiales. [Chakrabarti et al., 2010] démontrent une relation probabiliste entre les coefficients associés à ces filtres et l’écart type de la PSF. Cette méthode a été appliquée à l’estimation de cartes de profondeurs hiérarchiques. Elle s’avère difficilement adaptable à l’estimation d’une carte de profondeur dense (sans augmenter de façon importante la dimension de la base).

L’approche [Zhuo and Sim, 2009] permet l’inférence de la carte de profondeur hiérarchique d’une scène en mesurant l’influence du filtrage Gaussien des contours saillants d’une image. La mesure est réalisée à l’aide d’un opérateur gradient. Après avoir estimé localement la profondeur de la scène, une carte de profondeur est approximée en minimisant une fonction d’énergie basée sur un champ de Markov aléatoire (chaque pixel de l’image défini un nœud du champ Markovien). Les contours saillants sont *a priori* associés à des discontinuités de profondeur et les zones homogènes sont supposés lisses. Seulement trois niveaux de profondeur ont été considérés dans l’étude. [Zhuo and Sim, 2011] proposent de remplacer la fonction d’énergie par une fonction basée sur un lissage

Laplacien de l'image [Levin et al., 2008]. L'extension de cette approche à l'inférence d'une carte de profondeur absolue est complexe.

3.2.3 Méthodes d'appariement et d'estimation du flou conjointe

[Deschênes et al., 2008] étendent les méthodes [Deschênes et al., 2004; Myles and da Vitoria Lobo, 1998]. Ils proposent une approche permettant de réaliser l'appariement de deux images (approche directe basée sur l'intensité de l'image) et l'estimation conjointe de la variation de flou. Les auteurs considèrent des transformations inter-images rigides (déformation affine de régions planes). L'observation la plus floue de la paire d'image (ou ses dérivées partielles) est exprimée par :

$$I^{2(n,m)}(\mathbf{q}') = \left(I^{1(n,m)} * g_\zeta \right) (\mathbf{q}) \quad (3.18)$$

où $\mathbf{q} = \mathbf{A}\mathbf{q}' + \mathbf{t}$ correspond aux coordonnées d'un point de la première image, \mathbf{q}' correspond au point correspondant dans la seconde image, $\mathbf{A}_{2 \times 2}$ est une transformation linéaire (composée d'un facteur d'échelle anisotrope, d'un facteur de cisaillement et d'une rotation) et \mathbf{t} désigne un vecteur de translation, $\zeta^2 = \sigma_2^2 - \sigma_1^2$ représente la différence de flou relative entre les deux images, et n et m sont les ordres des dérivées partielles : $I^{k(n,m)}(x,y) = \frac{\partial^{n+m} I^k(x,y)}{\partial x^n \partial y^m}$. Le développement en série de Taylor de cette formulation permet d'exprimer une image en fonction des dérivées partielles des deux images, de la différence de flou relative et des paramètres de la transformation affine. La résolution d'un système d'équations non linéaires (basé sur une homotopie [Martínez, 1994]) permet d'estimer ces paramètres. Différentes évaluations ont permis de démontrer la robustesse de cette méthode. Cependant, aucune évaluation de la précision des mesures de profondeur n'a été réalisée. Cette méthode est particulièrement sensible au bruit ainsi qu'aux occultations.

[Yang and Schonfeld, 2010] proposent une méthode reposant sur les outils de l'analyse de Fourier. Une image de référence est extraite d'un flux vidéo ainsi qu'un ensemble d'images voisines. L'analyse fréquentielle (étude de la phase et du ratio fréquentiel) permet d'obtenir les paramètres de la transformation rigide inter images (transformation affine) ainsi que la mesure du flou optique dans l'image de référence. Cette approche tire parti de la redondance de l'information de flou ce qui lui confère une robustesse face au bruit. Elle a été principalement proposée afin de reconstruire une vidéo entièrement mise au point de la scène. Aucune évaluation sur la précision de la reconstruction 3D n'a été menée.

[Ben-Ari, 2014] souligne l'influence de l'appariement des images sur la précision des mesures de flou. Considérant un mouvement de caméra perpendiculaire à l'axe optique (similarité), l'auteur

propose de minimiser la fonction d'énergie suivante :

$$[\hat{\zeta}(\mathbf{q}), \hat{T}] = \underset{\zeta, T}{\operatorname{argmin}} E_d[\zeta, T] + \alpha \int_{\Omega} \|\nabla \zeta\| d\mathbf{q} \quad (3.19)$$

où \hat{T} correspond à l'estimation des paramètres de la similarité, $\alpha \int_{\Omega} \|\nabla \zeta\|$ est un terme de lissage (variation totale) pondéré par le poids α et $E_d[\zeta, T]$ est le terme d'attache aux données défini par :

$$E_d[\zeta, T] = \int_{\Omega} \|(g_{\zeta} * I^1)(\mathbf{q}) - I^2(T\mathbf{q})\|_1 H_{\zeta} + \|(g_{-\zeta} * I^2)(T\mathbf{q}) - I^1(\mathbf{q})\|_1 (1 - H_{\zeta}) d\mathbf{q} \quad (3.20)$$

où la pseudo norme de Huber est définie par $\|f\|_1 \simeq \sqrt{f^2 + \nu^2}$ (ν désigne une constante de petite taille), $g_{\zeta}(\mathbf{q})$ désigne la fonction Gaussienne 2D isotrope ayant pour variance ζ et H_{ζ} correspond à une fonction de signe définie par $H_{\zeta} = 1$ si $\zeta \geq 0$ et 0 sinon.

Le problème de minimisation (équation (3.19)) est résolu à l'aide d'une approche variationnelle numérique similaire à [Favaro, 2010]. Cette méthode peut être implémentée efficacement sur GPU (estimation de cartes de profondeurs denses pour des images de 1 MPixel en 2 secondes). Les expérimentations démontrent la contribution de cette méthode à la problématique d'appariement. Aucune évaluation quantitative de la mesure de l'échelle d'une scène n'a été menée.

Le tableau suivant récapitule les principales contributions à l'approche DfD ([Claxton, 2007; Crofts, 2007; Raj, 2009]).

Référence	Méthode	Modèle PSF	Intervalle d'évaluation	Erreur
Pentland [1982, 1987]	Fréquentielle fenêtrée (analyse du degré de saillance des contours).	Gaussien 2D isotrope	0,75 m	1,9%
Ens and Lawrence [1993]	Matricielle.	Adaptatif	0,015 à 1 m	1,3%
Subbarao and Surya [1994]	<i>Spatial Transform Method</i> (STM) (approximation par un polynôme cubique de l'image mise au point).	Adaptatif	≈ 4 m	2,3% à 0,6 m et 20% à 5 m
Watanabe and Nayar [1998]	Ensemble de filtres rationnels.	Pillbox	0,5 m à 0,9 m	0,5% à 1,2% de la distance objet camera
Ziou [1998]	Résolution d'un système d'équations polynomiales hermitiennes.	Gaussien 2D isotrope	–	–
Rajagopalan and Chaudhuri [1999]	Modélisation probabiliste du processus de flou par un champ de Markov. Estimation du maximum <i>a posteriori</i> .	Gaussien 2D isotrope	plan ≈ 1 m	4%
Bhasin and Chaudhuri [2001]	Etend [Rajagopalan and Chaudhuri, 1999] en considérant les auto occultations.	Gaussien 2D isotrope	–	–
Favaro and Soatto [2005]	Apprentissage d'une base d'opérateur de projection orthogonaux.	adaptatif	0,5 à 0,85 m	27 mm
Namboodiri and Chaudhuri [2007]	Modélisation du processus de flou par l'équation de diffusion de la chaleur.	Gaussien 2D isotrope	–	–
Deschênes et al. [2008]; Myles and da Vitoria Lobo [1998]	Estimation conjointe du flou relatif et de l'appariement d'un ensemble d'images au sein d'une vidéo.	Gaussien 2D isotrope	–	–
Favaro et al. [2008]	Modélisation du processus de flou par l'équation de diffusion de la chaleur.	Gaussien 2D isotrope	–	–
Favaro and Duci [2008]	Fréquentielle, basée sur l'analyse de Fourier.	adaptatif	–	–
Saxena et al. [2008]	Modèle probabiliste multi échelles basé sur un apprentissage supervisé.	–	Jusqu'à 81 m	Erreur multiplicative de 20% à 35%
Zhuo and Sim [2009, 2011]	Evaluation de l'influence d'un filtre Gaussien passe-haut sur le contenu d'une l'image.	Gaussien 2D isotrope	–	–

Référence	Méthode	Modèle PSF	Intervalle d'évaluation	Erreur
Yang and Schonfeld [2010]	Estimation de la profondeur d'une image de référence extraite d'une vidéo.	Pillbox, Gaussien ou polynomial	synthétique 1,3 m à 0,8 m	5%
Wei et al. [2009, 2012]	Modélisation du processus de flou par l'équation de diffusion de la chaleur.	Gaussien 2D isotrope	–	2,3%
Li et al. [2013a]	Utilisation de filtres rationnels Gaussien et Gaussien généralisés.	Gaussien 2D isotrope ou Gaussien 2D isotrope généralisée	0,05 à 0,8 m	inférieur à 7 mm
Zhu et al. [2013]	Méthode probabiliste basée sur une analyse fréquentielle basée région.	Adaptatif	–	–
Li et al. [2013b]	Méthode probabiliste combinant estimation du flou (DFD) et contraintes d'ombrage (SFS).	Gaussien 2D isotrope	–	–
Ben-Ari [2014]	Estimation conjointe du flou relatif et de l'appariement pour une paire d'images.	Gaussien 2D isotrope	–	–

3.3 Conclusion

Si la majorité des méthodes de reconstruction par analyse du flou permettent de résoudre l’ambiguïté d’échelle, peu d’entre elles s’appliquent à la métrologie. Elles imposent le contrôle précis des paramètres internes ou externes de la caméras les rendant inadaptées aux systèmes optiques monofocales utilisés en endoscopie (et plus particulièrement en coloscopie). Les méthodes DfF nécessiteraient l’adjonction d’un dispositif actif de suivi de l’endoscope. Les méthodes DfD basées sur une vue unique permettent généralement une reconstruction hiérarchique pour différentes profondeurs relatives prédéfinies. Elles sont par ailleurs particulièrement sensibles au bruit. Ceci est également le cas des méthodes basées sur l’estimation conjointe de l’appariement inter-images et du flou optique qui sont plus particulièrement destinées à améliorer la robustesse de l’appariement.

Afin d’estimer l’échelle d’une scène par vision passive monofocale, Il est nécessaire d’apporter une solution robuste à la mesure de l’information de flou optique. Ceci est plus particulièrement le cas en endoscopie qui nécessite une mesure millimétrique. Il est également nécessaire de s’affranchir des contraintes liées aux méthodes DfF et DfD qui ne peuvent être appliquées aux endoscopes flexibles utilisés en coloscopie.

4

Mesurer la profondeur à partir du point de rupture de netteté

Ce chapitre présente la contribution majeure des travaux menés dans le cadre de cette thèse. Nous proposons une nouvelle méthode permettant d'estimer le point de rupture de netteté appelé *Infocus-Breakpoint (IB)* au sein d'une vidéo. Considérant un système monofocale, l'*IB* correspond au point de rupture nette/flou qui intervient lorsque la caméra suit un mouvement d'approche de vitesse approximativement constante face à une région d'intérêt. Notre système est composé de deux modules dénommés *Blur-Estimating Tracking (BET)* et *Blur-Model Fitting (BMF)*. Le module *BET* permet le suivi et l'estimation conjointe du flou relatif entre deux images successives de la séquence vidéo. Nous détaillons ce module dans une première section et comparons, à l'aide d'une évaluation sur données synthétiques, un module *BET* basé sur un appariement inter-images affine avec un module basé sur un appariement déformable représenté par une *TPS (Thin-Plate Spline)*. Cette évaluation souligne la robustesse de la méthode TPS-BET face aux artefacts communément rencontrés dans le cadre d'une application endoscopique (notamment le bruit et les flous de mouvement). Le module *BMF* permet d'extraire de façon robuste l'*IB*. Le modèle de flou optique, présenté dans le second chapitre de ce manuscrit, est ajusté aux mesures de flou obtenues à l'aide du module *BET*. Nous proposons une stratégie permettant de s'affranchir des erreurs d'estimations proches du point de rupture de netteté (notamment liées à la proximité du système optique avec l'objet d'intérêt). Cette étude a fait l'objet des publications suivantes [Chadebecq et al., 2012a, 2013a,b, 2015; D'Orazio et al., 2014].

4.1 Principe

Nous proposons une méthode de mesure passive ne requérant aucun contrôle des paramètres internes ou externes du système optique [Chadebecq et al., 2015]. Cette dernière est basée sur l’acquisition d’une courte vidéo correspondant à un mouvement d’approche (ou de recul) approximativement constant de la caméra vis à vis d’un objet d’intérêt (figure 4.1.1). Elle exploite la redondance de l’information de flou optique afin d’estimer la mesure 3D d’un objet d’intérêt. Notre système s’appuie sur un premier module d’appariement et d’estimation conjointe du flou optique basé région nommé *Blur Estimating Tracking (BET)*. L’estimation est réalisée entre chaque paire d’images successives de la séquence. Un second module nommé *Blur Model Fitting (BMF)* permet d’extraire de façon robuste le point de rupture de netteté (*Infocus-Breakpoint (IB)*) en s’appuyant sur l’ajustement de l’information de flou estimée à un modèle de flou optique (voir section 2.2). Les expériences réalisées sur modèle synthétique ainsi que l’application de notre système à la mesure de lésions ont démontrés la faisabilité de notre approche. Notre système s’est montré robuste aux artefacts ponctuels tels que les flous de mouvement ou le bruit qui ne sont généralement pas considérés par les méthodes de l’état de l’art.

Pratiquement, le système de mesure comporte deux étapes. La phase de calibration consiste à acquérir une séquence vidéo correspondant à un mouvement d’approche du système optique face à une mire plane de calibration. Elle permet l’obtention des paramètres internes de la caméra ainsi que de la profondeur de référence \hat{d} correspondant à l’*IB*. Cette dernière est extraite à l’aide des modules *BET* et *BMF*. La seconde phase correspond à l’acquisition de la vidéo. Les modules *BET* et *BMF* permettent la mesure 3D interactive. Ces deux étapes seront plus particulièrement détaillées dans le cadre de la mesure de lésions en coloscopie (voir chapitre 6).

4.2 *Blur Estimating Tracking (BET)* : suivi et estimation conjointe du flou optique

L’évaluation du degré de mise au point d’une région d’intérêt au sein d’une video nécessite d’effectuer son suivi. Nous supposons que l’objet d’intérêt possède des contours saillants et une apparence discriminante.

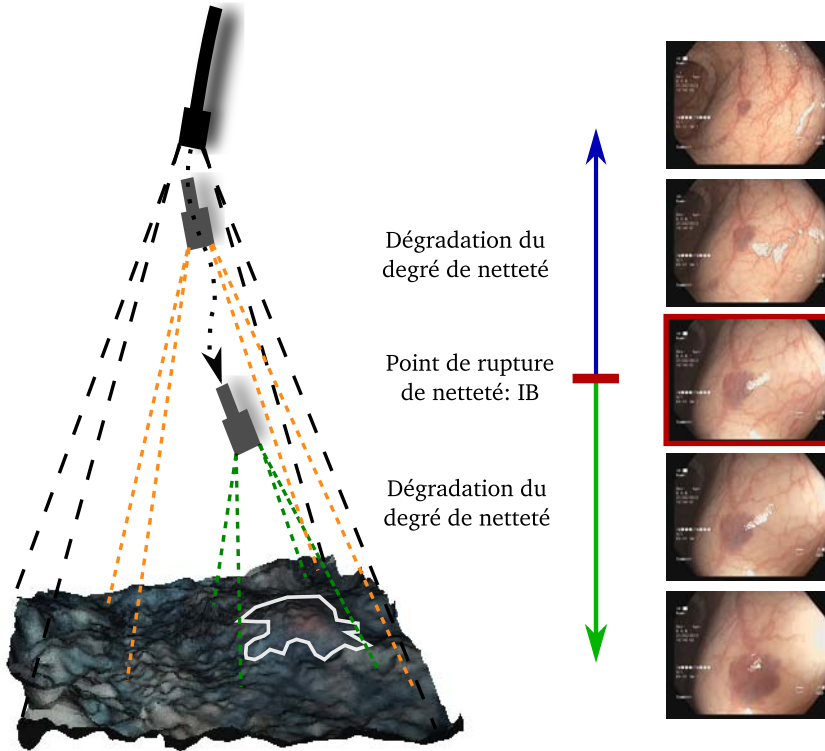


Figure 4.1.1 – Mesure par extraction du point de rupture de netteté dans un flux vidéo : principe.

4.2.1 Modèle géométrique de déformation des images

Nous avons utilisé deux modèles de transformation inter-images. Le premier est une simple transformation affine adaptée à de faibles déplacements du système optique vis à vis d'une scène supposée rigide (voir section 3.2.3) :

$$\mathcal{W}_{\text{AFF}}(\mathbf{q}) = \mathbf{A}\mathbf{q} + \mathbf{t} \quad (4.1)$$

Le second modèle de transformation permet de prendre en considération les déformations de la scène. Il est exprimé par une fonction Thin-Plate Spline 2D (TPS-Warp) [Bookstein, 1989]. Cette dernière est obtenue en empilant deux fonctions TPS ($\mathbb{R}^2 \rightarrow \mathbb{R}$) partageant leurs centres de déformation ainsi que leurs poids de régularisation internes [Bartoli et al., 2010] :

$$\mathcal{W}_{\text{TPS}}(\mathbf{q}) = \mathcal{M}\ell_{\mathbf{q}} \text{ avec } \mathcal{M}^{\text{T}} = \mathbf{E}_{\lambda}\mathbf{P}' \quad (4.2)$$

où $\mathbf{c}'_j \in \mathbb{R}^{2 \times 1}$ (avec $j \in [1; l]$) correspond aux l centres de déformation de l'image cible. Ils sont cumulés dans la matrice $\mathbf{P}' \in \mathbb{R}^{l \times 2}$. La fonction noyau de la TPS permettant le calcul de la distance des points de l'image aux centres de déformation est donnée par $\rho(d^2) = d^2 \log d^2$. Elle permet de définir $\ell_{\mathbf{q}}^\top = (\rho(d^2(\mathbf{q}, \mathbf{c}_1)) \cdots \rho(d^2(\mathbf{q}, \mathbf{c}_l)) \mathbf{q}^\top \mathbf{1})$. Enfin, \mathbf{E}_λ est une matrice fixe construite à l'aide des centres de déformation de l'image source $\mathbf{c}_j \in \mathbb{R}^{2 \times 1}$ (avec $j \in [1; l]$) et du poids de régularisation interne $\lambda = 10^{-5}$.

Le modèle de déformation affine est global et rigide alors que le modèle de déformation TPS est local et flexible. Le choix du modèle TPS est motivé par la volonté de rendre notre méthode robuste aux déformations de l'environnement imagé (ceci est notamment le cas des environnements explorés au cours de nombreux examens endoscopiques) ainsi qu'aux artefacts ponctuels (flous de mouvements, acquisitions bruitées, ...). Les performances des modules AFF-BET et TPS-BET sont comparées dans la section 4.3.2.

4.2.2

Les méthodes de suivi basées sur la détection et l'appariement de points d'intérêt ([Lowe, 2004], ...) sont peu adaptées aux scènes faiblement texturées. Ceci est notamment le cas en coloscopie (et dans de nombreux cas d'examen endoscopiques) où certaines sections du côlon possèdent une texture homogène [Alcantarilla et al., 2013; Kaufman and Wang, 2008].

Stratégie : Nous proposons une méthode de suivi dite directe basée sur l'appariement de régions d'intensité similaire. Dans le cas d'une vidéo, l'appariement peut être réalisé entre une image de référence (image source) et l'ensemble des images voisines. Cependant, cette stratégie est peu adaptée à notre méthode de par l'important changement d'échelle des objets lié au mouvement d'approche de la caméra. Nous avons donc choisi de réaliser l'appariement entre deux images successives I^k et I^{k+1} de la vidéo. La différence de flou relative $\delta^k = s^k - t^k$, où s^k et t^k correspondent respectivement au degré de mise au point des images sources et cibles, est calculée simultanément pour chaque paire d'image. Elle s'apparente, selon nos hypothèses, à la dérivée de la mesure de flou absolu à la distance d'acquisition d de la région d'intérêt correspondant à l'image cible. le module *BMF* permet l'intégration des mesures de flou relatif (à une constante d'intégration près) afin d'obtenir la mesure de flou absolu (voir section 5.3). La stratégie employée par le module *BET* est illustrée par la figure 4.2.1 et résumée par l'algorithme 4.2.1.

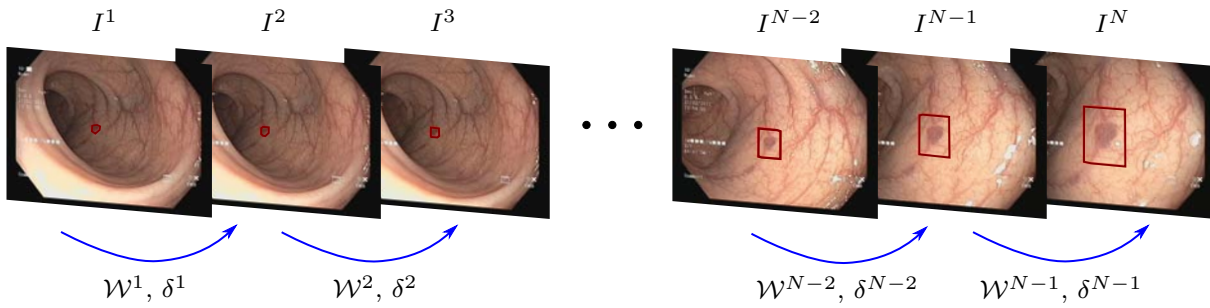


Figure 4.2.1 – Stratégie employée par le module de suivi et d’estimation conjointe du flou optique : *Blur Estimating Tracking (BET)*. Cette dernière s’appuie sur l’estimation simultanée d’une déformation géométrique \mathcal{W}^k et d’un flou optique relatif δ^k pour chaque paire d’images consécutives de la vidéo (I^k, I^{k+1}) avec $k \in [1; N - 1]$.

Blur-Estimating Tracking (BET)

Entrées :

- Vidéo (I^1, \dots, I^N)
- ROI initiale (ROI^1)

Algorithme :

Pour $k = 1$ à $N - 1$ faire

suivi et estimation conjointe du flou relatif : $(\mathcal{W}^k, \delta^k) \leftarrow \text{Appariement}(I^k, I^{k+1}, ROI^k)$
mise à jour de la région d’intérêt : $ROI^{k+1} \leftarrow \mathcal{W}^k(ROI^k)$

Fin Pour

Sorties :

- Flou relatif $(\delta^1, \dots, \delta^{N-1})$
 - (optionel) Paramètres de la déformation géométrique $(\mathcal{W}^1, \dots, \mathcal{W}^{N-1})$
 - (optionel) ROIs (ROI^1, \dots, ROI^N)
-

Table 4.2.1 – Algorithme *Blur Estimating Tracking (BET)* : suivi d’une région d’intérêt au sein d’une vidéo. Le module *BET* estime simultanément une déformation géométrique et l’information de flou relatif entre deux images successives du flux vidéo.

4.2.3 Appariement d’une paire d’images

L’équation (2.13) présentée dans la section 2.2.3 permet de définir le système d’équations suivant :

$$\begin{cases} I^k(\mathbf{q}) = (\hat{I} * g_{\sigma(d^k)})(\mathbf{q}) \\ I^{k+1}(\mathbf{q}') = (\hat{I} * g_{\sigma(d^{k+1})})(\mathbf{q}') \end{cases} \quad (4.3)$$

où d^k et d^{k+1} représente la profondeur du point 3D \mathbf{Q} imagé au pixel \mathbf{q} dans l'image I^k et \mathbf{q}' dans l'image I^{k+1} .

L'image nette \hat{I} étant inconnue, elle doit être éliminée des équations. Nous utilisons l'image la plus nette de la paire comme "génératrice" de la seconde. Ceci est déterminé grâce au signe de $\Delta^k = \sigma(d^k) - \sigma(d^{k+1})$ ou Δ^k représente la différence de flou absolu : si $\Delta^k > 0$, I^k est l'image la plus nette. Nous reformulons ainsi l'équation (4.3) :

$$\begin{cases} \left(I^k * g_{\sigma(d^{k+1})} \right) (\mathbf{q}) & = \left(\left(\hat{I} * g_{\sigma(d^k)} \right) * g_{\sigma(d^{k+1})} \right) (\mathbf{q}) \\ \left(I^{k+1} * g_{\sigma(d^k)} \right) (\mathbf{q}') & = \left(\left(\hat{I} * g_{\sigma(d^{k+1})} \right) * g_{\sigma(d^k)} \right) (\mathbf{q}') \end{cases} \quad (4.4)$$

La convolution avec une fonction Gaussienne étant commutative, il en résulte l'expression suivante :

$$\left(I^k * g_{\sigma(d^{k+1})} \right) (\mathbf{q}) = \left(I^{k+1} * g_{\sigma(d^k)} \right) (\mathbf{q}') \quad (4.5)$$

Le flou relatif, mesuré pour chaque image, peut alors être estimé à une constante g_κ (fonction Gaussienne 2D isotrope d'écart type κ non nul). Celle-ci ne peut être négligée, particulièrement en présence de bruit ou de flou de mouvement dans les images. Pour $\kappa > 0$, l'équation (4.5) devient :

$$\left(I^k * \left(\underbrace{g_{\sigma(d^{k+1})} * g_\kappa}_{g_{t^k}} \right) \right) (\mathbf{q}) = \left(I^{k+1} * \left(\underbrace{g_{\sigma(d^k)} * g_\kappa}_{g_{s^k}} \right) \right) (\mathbf{q}') \quad (4.6)$$

où g_{s^k} et g_{t^k} correspondent à des mesures de flou relatif dans les images source et cible I^k et I^{k+1} . L'appariement inter-images, illustré dans la figure 4.2.1, nécessite de résoudre le problème de minimisation des moindres carrés non linéaires suivant :

$$\min_{(\zeta^k, s^k, t^k, \mathcal{W}^k)} \sum_{\mathbf{q} \in ROI} \rho^k(\mathbf{q}) \left[\left(I^k * g_{t^k} \right) (\mathbf{q}) - \zeta^k \left(\left(I^{k+1} * g_{s^k} \right) (\mathcal{W}^k(\mathbf{q})) \right) \right]^2 + \lambda \min(s^k, t^k)^2 \quad (4.7)$$

Le masque ρ^k permet d'ignorer les zones saturées de l'image. Il est obtenu par un seuillage basé sur l'intensité des pixels. Le gain ζ^k permet d'ajuster les variations d'intensité globales entre les images sources et cibles. La pénalité $\min(s^k, t^k)^2$ prévient la sur-estimation de l'information de flou optique. Elle garantit que $\kappa \approx 0$ et ainsi $t^k \approx \sigma(d^{k+1})$ et $s^k \approx \sigma(d^k)$. La valeur de λ a été expérimentalement fixée à 100 (pour l'ensemble des évaluations présentées dans ce document). L'évaluation de l'influence de ce paramètre n'a pas été réalisée au profit d'une évaluation globale portant sur la robustesse de la méthode BET (elle est envisagée dans la perspective d'une évalua-

tion clinique de nos travaux portant sur la mesure des lésions en coloscopie). La fonction de coût proposée correspond trivialement à la mesure d'une carte de disparité (différence d'intensité) entre l'image source floutée et l'image cible déformée et floutée.

Nous résolvons le problème (4.7) à l'aide de la méthode Levenberg-Marquardt [Levenberg, 1944] en utilisant les valeurs d'initialisation suivante : $\zeta^k = 1$, $s^k = t^k = \varepsilon$ et $\mathcal{W}^k = Id$ (aucune transformation géométrique). Le calcul des dérivées partielles de la fonction de coût a été réalisé numériquement.

La distorsion géométrique n'est pas considérée dans le module *BET*. La correction de cette aberration pourrait en effet corrompre la cohérence des mesures de flou optique (la correction étant dépendante de la position spatiale de l'objet dans l'image [Brauer-Burchardt et al., 2006]). Elle sera uniquement corrigée dans l'image correspondant à l'*IB* afin de réaliser la mesure interactive d'un objet (module *BMF*). Expérimentalement, nous n'avons constaté aucun impact significatif sur le suivi et l'estimation du flou, même lorsque les objets à mesurer n'étaient pas centrés sur l'axe optique de la caméra.

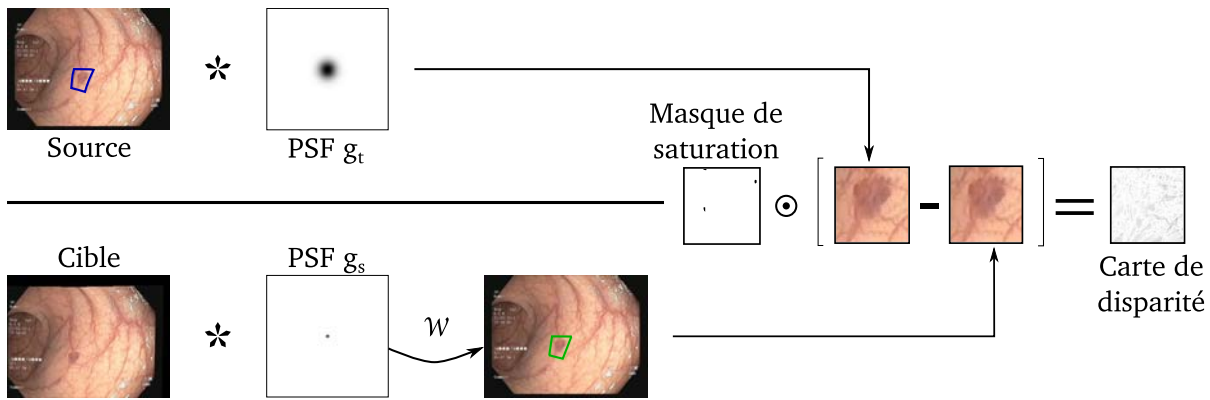


Figure 4.2.2 – *Blur-Estimating Tracking (BET)* appliqué à une paire d'images coloscopiques. L'image source est plus nette que l'image cible. *BET* minimise la carte de disparité obtenue en comparant l'image source floutée avec l'image cible déformée et floutée.

Entrées :

- Paire d'images (I^k, I^{k+1})
- ROI dans l'image I^k (ROI^k)

Algorithme :

Calculer $(\zeta^k, s^k, t^k, \mathcal{W}^k)$ en résolvant l'équation (4.7)
Calculer $\delta^k \leftarrow s^k - t^k$

Sorties :

- Déformation géométrique \mathcal{W}^k
 - Flou relatif δ^k
-

Table 4.2.2 – Algorithme *Blur Estimating Tracking (BET)* : appariement d'une paire d'images. Les paramètres de déformation géométrique et de flou relatif entre une image source et une image cible sont conjointement estimés.

4.3 Évaluation de la méthode de suivi et d'estimation conjointe du flou optique

4.3.1 Données synthétiques

Afin d'évaluer les performances du module *BET*, nous avons créé une base de données synthétiques. La méthode proposée se destinant à la coloscopie, nous avons cherché à reproduire les déformations inter-images observées lors de cet examen. L'environnement exploré peut subir de légères déformations dues aux mouvements naturels des organes ou à des manipulations chirurgicales.

Dix images comportant des néoplasies ont été extraites d'examens coloscopiques (figure 4.3.1). Ce choix a été motivé par leurs différentes apparences (contours saillants, présence de texture ...). Des déformations aléatoires leurs ont été appliquées afin d'évaluer la robustesse des modules AFF-BET et TPS-BET.

Le processus de génération des images synthétiques est basé sur les étapes suivantes (figure 4.3.2) :

(i) nous créons une sphère virtuelle dont le centre, situé sur l'axe optique d'une caméra endoscopique simulée, se trouve à 4 cm du centre de projection de cette dernière. Ce choix correspond à une distance d'observation réaliste considérant les différents diamètres du côlon. Un sous-ensemble de points 3D de la sphère est sélectionné aléatoirement puis projeté dans l'image cible. Les points projetés définissent les centres de déformation 2D dans cette image.

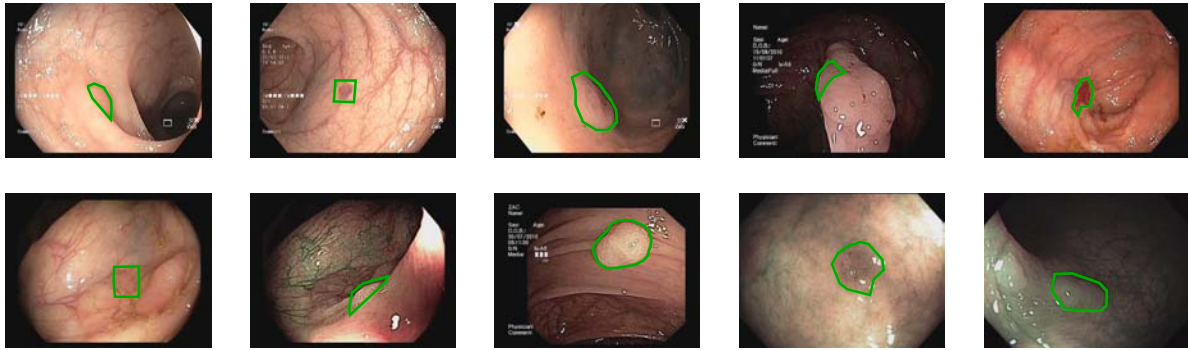


Figure 4.3.1 – Différentes images de néoplasies extraites d'examens coloscopiques ont été utilisées pour générer un jeu de données synthétique. Pour chacun des paramètres expérimentaux (bruit additif Gaussien, flou optique, amplitude de déformation), 100 transformations aléatoires ont été appliquées à chacune des images de références pour générer un ensemble de 1000 nouvelles images.

(ii) Nous appliquons à la sphère une rotation et une translation 3D aléatoire. Le sous-ensemble de point 3D est projeté dans l'image source afin de définir des centres de déformation 2D dans cette image.

(iii) Une déformation basée sur une TPS est calculée à partir des centres de déformation précédemment générés. Celle-ci est ensuite appliquée à l'une des 10 images de référence.

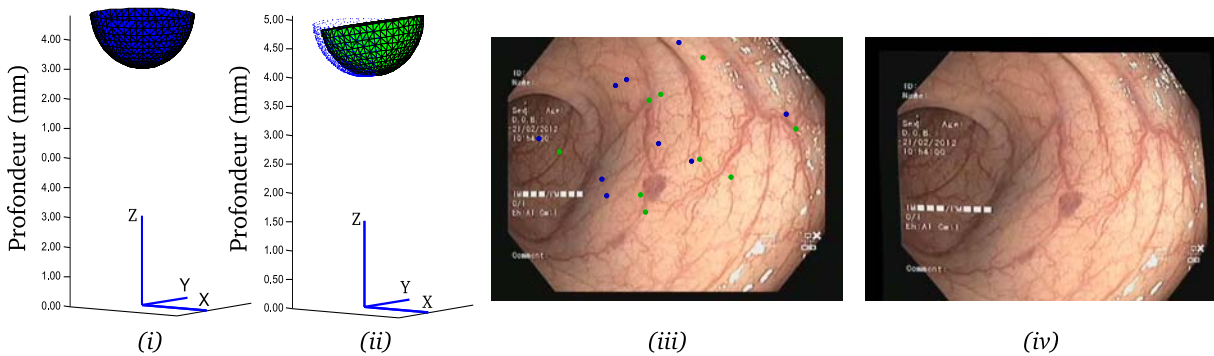


Figure 4.3.2 – Génération du jeu de données synthétiques. (i) et (ii) Une sphère centrée sur l'axe optique d'un endoscope simulé subit une translation et une rotation 3D. (iii) Un sous-ensemble des points 3D de la sphère de référence et de la sphère cible est projeté sur le plan image de l'endoscope virtuel afin de générer des points de contrôle. Ces derniers permettent de calculer une déformation basée TPS. (iv) Une nouvelle image est générée en appliquant la déformation précédemment calculée à l'une des 10 images de référence extraite d'examens coloscopiques.

Trois critères de robustesse ont été évalués : le bruit (additif Gaussien), le flou optique et

l'amplitude de déformation (figure 4.3.3). Ce dernier est exprimé indirectement par l'amplitude de déplacement de la sphère 3D. Il peut également être assimilé à la vitesse de déplacement de la caméra. La rotation aléatoire appliquée à la sphère a été limitée afin de générer des déformations réalistes de l'image de référence. Cela permet également de ne pas biaiser l'étude en favorisant l'algorithme TPS-BET. Pour chaque valeur de paramètre évaluée, 100 images cibles ont été générées.

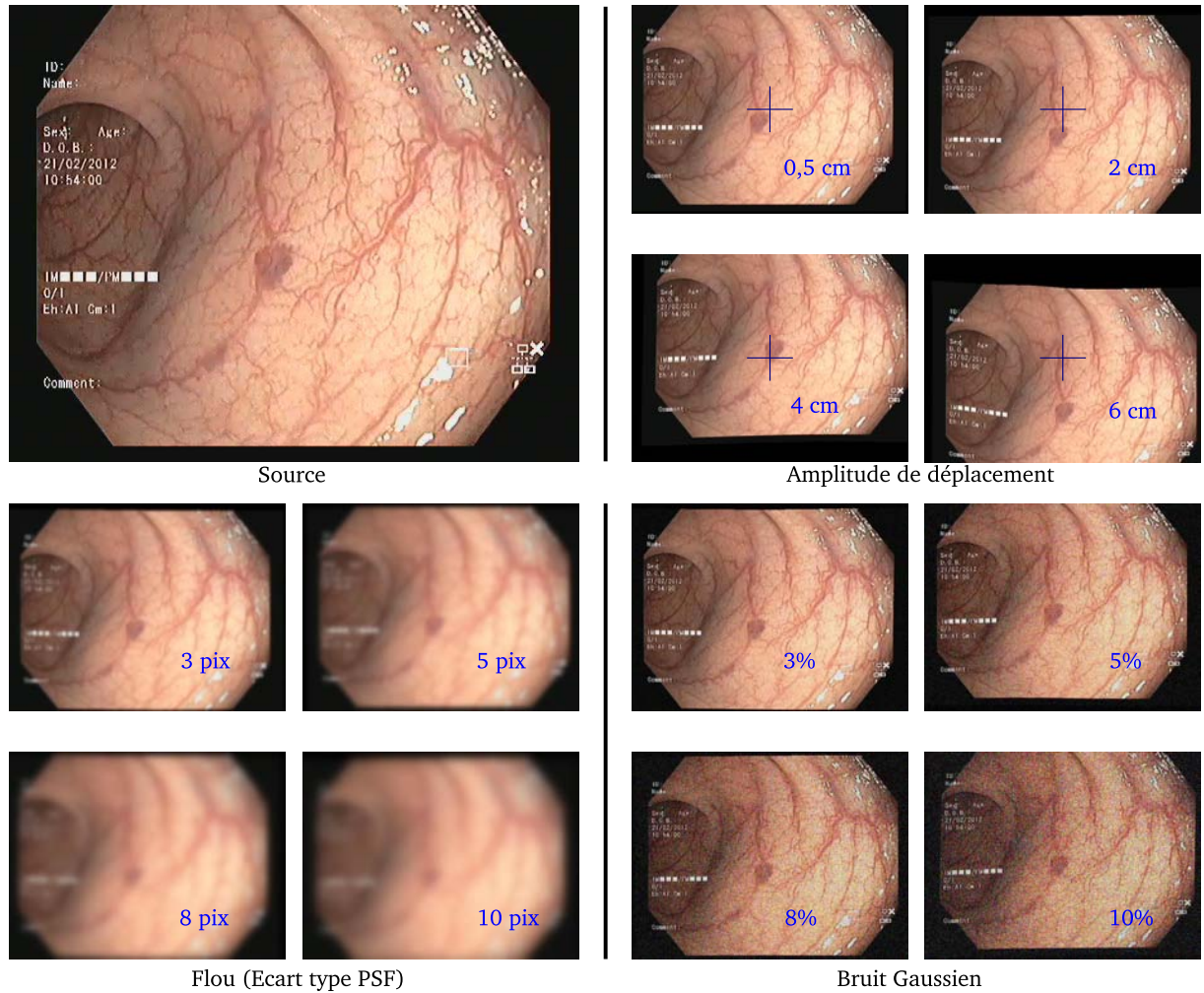


Figure 4.3.3 – Jeu de données synthétiques permettant d'évaluer le module de suivi et d'estimation conjointe du flou optique *Blur Estimating Tracking (BET)*. Cette illustration est basée sur l'image de référence présentée dans la vignette supérieure gauche. Trois critères de robustesse ont été évalués : l'amplitude de déplacement (vignette supérieur droite), le flou optique (vignette inférieur gauche) et le bruit (vignette inférieur droite).

Les intervalles d'évaluation des critères de robustesse sont les suivants :

amplitude de déplacement de la sphère 3D : $[0; 6]$ cm. Ce choix correspond pratiquement à un déplacement des centres de déformation compris dans l'intervalle $[0; 70]$ pixels. L'amplitude de rotation de la sphère est définie aléatoirement dans l'intervalle $[\frac{\pi}{20}; \frac{\pi}{15}]$. L'écart type du flou Gaussien appliqué à l'image source est de 1 pixel et le bruit 1%.

Flou optique Gaussien : $[0; 10]$ pixels. L'amplitude de déplacement de la sphère est fixée à 1 cm (ce qui correspond à un déplacement des centres de déformation d'environ 10 pixels) et le bruit à 1%.

Bruit additif Gaussien : $[0; 10\%]$. L'amplitude de déplacement de la sphère est fixée à 1 cm et l'écart type du flou Gaussien à 1 pixel.

4.3.2 Analyse statistique

Les résultats sont présentés dans la figure 4.3.4. L'algorithme TPS-BET se montre plus robuste aux déformations géométriques que l'algorithme AFF-BET. TPS-BET permet l'appariement d'images ainsi qu'une estimation correcte du flou optique pour des déplacements de la sphère allant jusqu'à 3,5 cm alors que AFF-BET n'est performant que pour des déplacements inférieurs à 2 cm. Cela représente une amélioration significative si l'on considère les difficultés de déplacement d'un endoscope au sein de la cavité colonique. Lorsque la lésion (ou la région d'intérêt) est proche de la caméra, un léger déplacement de cette dernière peut entraîner un déplacement important de la lésion dans l'image. Ces résultats démontrent également que les erreurs ou dérives de suivi inter-images entraînent des erreurs significatives d'estimation du flou optique. Celles-ci peuvent générer des ambiguïtés d'extraction de l'*IB* malgré l'utilisation d'une vidéo cohérente et du module *BMF*.

TPS-BET et AFF-BET ont démontré des performances similaires en présence de bruit et de flou optique. L'erreur de suivi est inférieure à 0,5 pixel même en présence d'un flou optique important. L'erreur d'estimation de l'écart type de la PSF est inférieure à 0,5 pixel lorsque le flou réellement appliqué à l'image de référence est inférieur à 6 pixels. Au delà de cette valeur, l'erreur d'estimation du flou s'accroît linéairement pour les deux algorithmes évalués. Pratiquement, une telle magnitude de flou n'a pas été observée sur les vidéos endoscopiques dont nous disposons.

Le bruit a une légère influence sur l'erreur d'appariement lorsque celui-ci est inférieur à 6%. Pour des valeurs plus importantes, cette erreur s'accroît rapidement bien que l'algorithme TPS-BET se montre plus robuste. L'estimation du flou est particulièrement affectée par la présence de bruit. L'erreur est inférieure à 1 pixel lorsque le bruit est inférieur à 3%. Au delà de cette

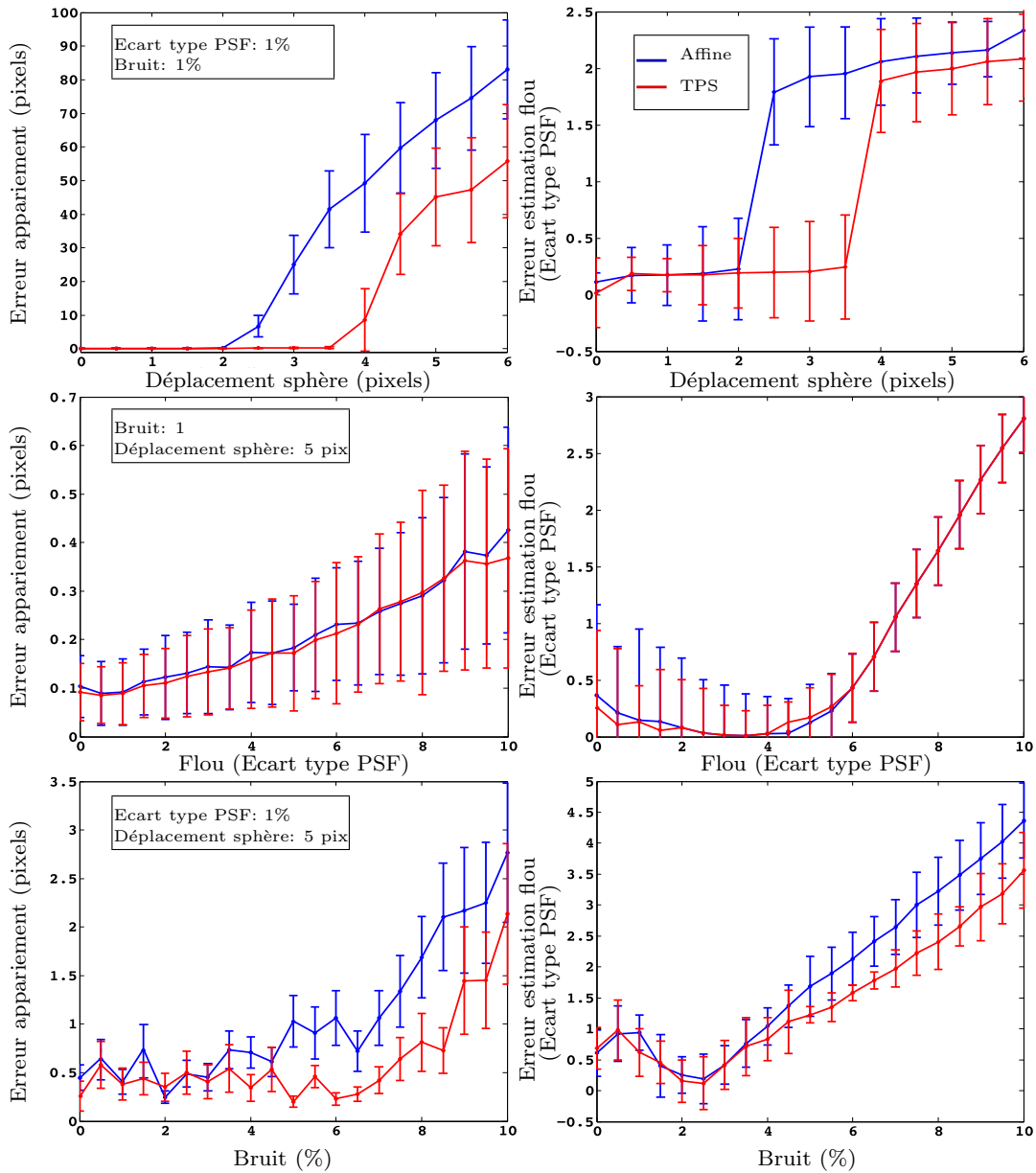


Figure 4.3.4 – Evaluation du module *Blur-Estimating-Tracking (BET)* de suivi et d'estimation du flou optique. Les graphiques situés à gauche de la figure présentent les erreurs d'appariements relatifs à l'algorithme *BET* Affine (bleue) et TPS (rouge) en fonction de la vitesse de déplacement de la caméra, du flou optique et du bruit. Les graphiques situés à droite présentent l'erreur moyenne d'estimation du flou optique. Chaque ligne correspond à l'un des paramètres expérimentaux évalués.

valeur, l'erreur d'estimation du flou s'accroît linéairement ce qui peut entraîner des ambiguïtés d'extraction de l'*IB*.

Cette évaluation statistique souligne la robustesse de la méthode TPS-BET face à la méthode AFF-BET. Elle se montre notamment plus adaptée à des déplacements importants de la caméra, aux déformations de l'environnement ainsi qu'aux bruit d'acquisition (considérant un bruit additif Gaussien). Nous constatons également que l'estimation du flou optique est biaisée lorsque ce dernier est important mais qu'elle reste cohérente. Cette évaluation du module *BET* est complétée par une évaluation de la méthode (module *BET* et *BMF*) sur un ensemble de vidéos coloscopiques (chapitre 5). Elle démontre également que la méthode TPS-BET apporte un gain de robustesse et de précision sur la méthode TPS-BET.

4.4 *Blur Model Fitting (BMF) : extraction robuste du Point de Rupture de Netteté*

Mesure du flou optique par intégration du flou relatif : L'extraction robuste de l'*IB* basée sur l'ajustement du modèle de flou optique géométrique (équation (2.5)) nécessite la mesure absolue du flou. Le module *BET* ne réalise qu'une mesure relative de ce dernier pour chaque paire d'images d'une vidéo. L'approche naturelle consiste à intégrer les valeurs de δ^k puis à extraire l'*IB* grâce à l'information de changement de signe. La constante d'intégration v est alors définie afin que le flou absolu correspondant à l'*IB* soit nul. Cette méthode est néanmoins sensible aux bruit ainsi qu'aux erreurs d'estimation qui induisent des ambiguïtés d'extraction de l'*IB*, comme cela est illustré dans la figure 4.4.1. Ceci est notamment le cas lorsque le système optique est proche de l'objet d'intérêt à mesurer. La présence de flous de mouvement biaise l'approche d'intégration directe et implique une imprécision d'estimation de l'*IB*. Cela est également le cas lorsque la vitesse de déplacement du système optique est particulièrement variable. Nous proposons une stratégie alternative permettant l'intégration robuste des valeurs de flou relatif.

Méthode d'extraction robuste du Point de Rupture de Netteté : A des fins d'exemple, nous considérons une séquence vidéo correspondant à un mouvement d'approche du système optique vis à vis d'un objet d'intérêt. Lorsque $d > \hat{d}$, l'image cible est plus nette que l'image source. Ainsi, $\Delta^k < 0$ et s^k représente la différence de flou absolu (alors que t^k est théoriquement égal à 0). De même, lorsque $d \leq \hat{d}$, $\Delta^k > 0$; l'image source est plus nette que l'image cible et ainsi t^k représente la différence de flou absolu (s^k est théoriquement égal à 0). Nous considérons

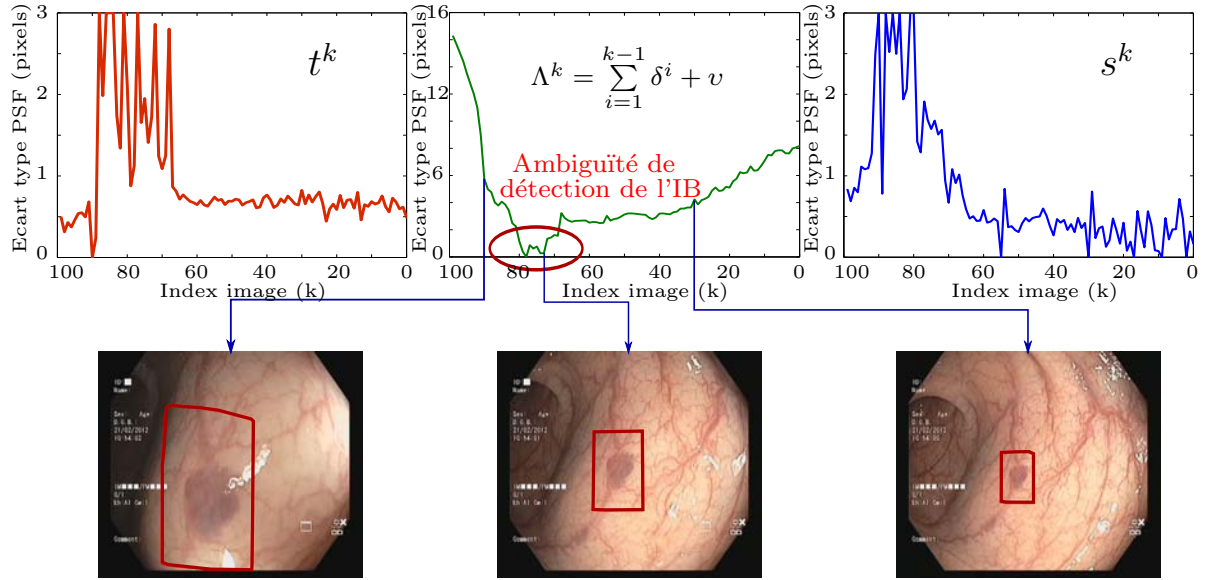


Figure 4.4.1 – La mesure du flou optique absolu par intégration directe du flou relatif implique des ambiguïtés d'extraction du point de rupture de netteté. La présence de flous de mouvements ou une vitesse de déplacement du système optique inconstante rendent la méthode d'intégration directe inadaptée.

les relations suivantes :

$$\Delta^k \approx \begin{cases} -s^k & \text{pour } d > \hat{d} \\ t^k & \text{pour } d \leq \hat{d} \end{cases} \quad (4.8)$$

Il est ainsi nécessaire de définir une approximation initiale de l'*IB* afin d'intégrer la mesure de flou absolu. Cette dernière est obtenue à l'aide de la méthode d'intégration directe. Elle est affinée grâce à l'ajustement du modèle de flou optique (2.12). Nous choisissons l'index $k_0 \in [1; N - 1]$ correspondant à l'argument minimum de $\Lambda^k = \sum_{i=1}^{k-1} \delta^i + v$ comme première approximation de l'*IB*. La constante d'intégration v correspond au flou absolu mesuré dans la première image de la vidéo. La valeur de l'écart type de la PSF Gaussienne étant théoriquement nulle pour l'*IB*, nous définissons $v = \left| \min_{j=2, \dots, N} \sum_{l=1}^{j-1} \delta^l \right|$. La majorité des artefacts apparaissent lorsque le système optique est proche de l'objet d'intérêt. En cas d'ambiguïté, nous choisissons l'index d'image le plus important comme première approximation de l'*IB*. La mesure de flou absolu Ψ^k est ensuite réalisée en intégrant les valeurs de flou relatifs (dérivées de l'écart type de la PSF, voir figure

4.4.2) :

$$\Psi^k \stackrel{\text{def}}{=} \sigma(d^k) \approx \begin{cases} -\sum_{i=0}^k (s^i) + v & \text{pour } k < IB \text{ (} d > \hat{d} \text{)} \\ \Psi^{IB-1} + \sum_{i=k_{IB}}^N (t^i) + v & \text{pour } k \geq IB \text{ (} d \leq \hat{d} \text{)} \end{cases} \quad (4.9)$$

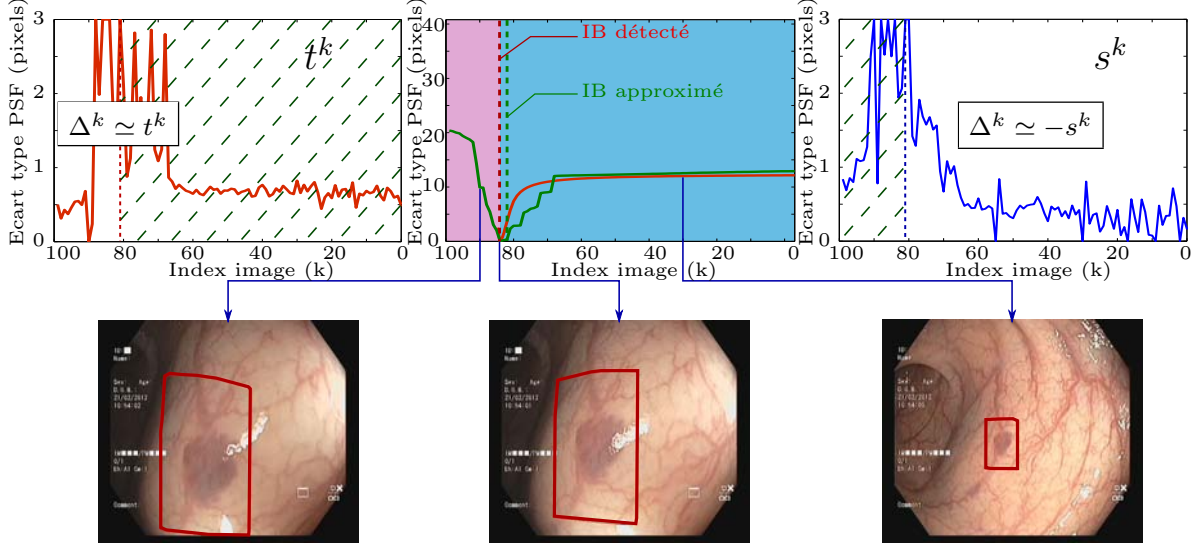


Figure 4.4.2 – Blur-Model Fitting (BMF). Les graphiques gauche et droite correspondent aux mesures de flou relatif respectivement dans les images source et cible de la vidéo. Le graphique central correspond à l'estimation du flou absolu obtenu par intégration des mesures de flou relatif source et cible de part et d'autre de l'approximation initiale du point de rupture de netteté (IB). Le modèle de flou optique, ici représenté par la courbe rouge, est ensuite ajusté à l'estimation de flou absolu afin d'extraire l'*Infocus-Breakpoint* (IB) de façon robuste. Les images positionnées au dessous des graphiques présentent la vidéo coloscopique considérée ainsi que le suivi de la lésion (réalisé par le module *Blur-Estimating Tracking* (BET)).

Le modèle paramétrique (équation (2.12)) peut ensuite être ajusté aux mesures de flou absolu en supposant la vitesse de déplacement de la caméra β (pixels par image) constante :

$$\min_{r,e,f,\beta} \sum_{k=1}^{k_0} \left[\Psi^k - \left(r \left| \frac{e}{f} - 1 - \frac{e}{\beta k} \right| \right) \right]^2 \quad (4.10)$$

Ce problème de minimisation non linéaire de moindres carrés est résolu en utilisant la méthode de Levenberg-Marquardt. Afin d'estimer la dérivé de l'équation (4.10), nous remplaçons la valeur absolue du modèle paramétrique (2.12) par la pseudo-norme de Huber [Huber,

1964] : $L\left(\left|\frac{e}{f} - 1 - \frac{e}{\beta k}\right|\right) = \sqrt{\left(\frac{e}{f} - 1 - \frac{e}{\beta k}\right)^2 + \epsilon}$, avec ϵ un réel positif proche de 0 (par exemple $\epsilon = 10^{-6}$). Les paramètres initiaux sont issus de la calibration géométrique de la caméra.

L'ajustement du modèle théorique est uniquement réalisé pour les images d'index compris entre $(1 \dots k_0)$. Lorsque la caméra est très proche de l'objet d'intérêt, la mesure du flou optique est imprécise et biaisée (voir section 4.3.2). Par ailleurs, l'hypothèse de lentille mince n'est plus valide. Seules les profondeurs supérieures ou égales à la profondeur de l'*IB* sont ainsi considérées. L'algorithme 4.4.1 présente la méthode *BMF* d'extraction de l'*IB*.

Blur-Model Fitting (BMF)

Entrées :

- Flou relatif $(\delta^1, \dots, \delta^{N-1})$

Algorithme :

résidu, *IB* \leftarrow 0

Calcul de l'*IB* prédit : $k_0 \leftarrow \underset{i}{\operatorname{argmin}} \sum_{i=1}^{N-1} \delta^i$

Pour $k_{IB} = k_0 - 10$ à $k_0 + 10$

 Calcul du flou absolu relatif à l'index k_{IB} en utilisant l'équation (4.9)

 Obtention de la valeur courante de l'*IB* ainsi que du résidu en résolvant l'équation (4.10)

 Si résidu courant < résidu

IB, résidu \leftarrow *IB* courant, résidu courant

 Fin si

Fin pour

Sorties :

- *IB*
 - (optionel) erreur résiduel *BMF*
-

Table 4.4.1 – Algorithme *Blur Model Fitting (BMF)*. L'intégration des mesures de flou relatif obtenues à l'aide du module *Blur-Estimating Tracking (BET)* permet d'inférer la mesure de flou absolu ainsi qu'une première approximation du point de rupture de netteté (*Infocus-Breakpoint (IB)*). Le modèle de flou optique (équation (4.9)) peut ensuite être ajusté à l'estimation du flou absolu afin d'extraire de façon robuste l'*IB*.

Le module *BMF* permet d'affiner de façon précise la première approximation de l'*IB*. Il est par ailleurs robuste aux déplacements approximativement constants de la caméra. Comme l'*IB* correspond à une profondeur de référence unique et calibrée, la taille de l'objet d'intérêt peut être inférée dans l'image correspondant à $I^{IB} \approx \hat{I}$ (voir section 2.2.3). Afin de réaliser la mesure, la distorsion de l'image I^{IB} est corrigée.

4.5 Inférence des mesures 3D à partir de mesures 2D

Une fois l'*IB* détecté au sein de la vidéo, l'image I^{IB} la plus proche de ce point peut être extraite. La distorsion est corrigée à l'aide des paramètres de calibration. L'hypothèse de planarité de la région d'intérêt ainsi que la calibration de la profondeur de référence correspondant à l'*IB* permet d'inférer la mesure d_{3d} de cette dernière en mm depuis sa mesure d_{2D} effectuée dans l'image de référence en pixels :

$$d_{3D} = \frac{\hat{d}}{\alpha f} d_{2D} \quad (4.11)$$

où α est la densité du capteur de la caméra en pixels par mm. Il est important de noter que la calibration géométrique permet d'obtenir le produit αf .

La mesure interactive d'une lésion est illustrée dans la figure 4.5.1. Tout autre image du flux vidéo pourrait être utilisée pour réaliser la mesure interactive. L'objet d'intérêt ayant été suivi au cours de la vidéo, la sélection d'une région d'intérêt peut être reportée dans l'image de référence correspondant à l'*IB*.

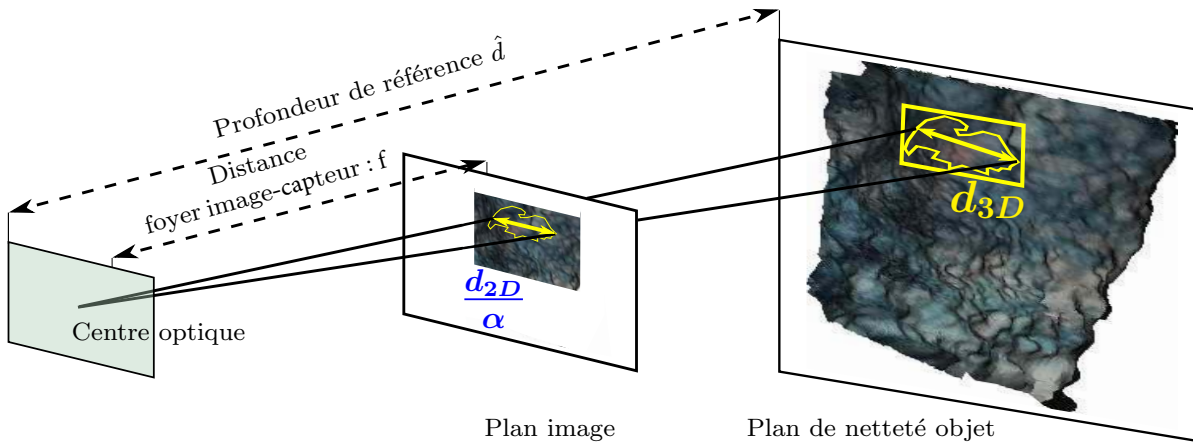


Figure 4.5.1 – Mesure 3D interactive d'une lésion colique.

Lorsque l'hypothèse d'équifocalité n'est pas vérifiée, la quantité de flou optique varie spatialement, notamment aux contours de l'objet suivi. Nous mesurons alors une profondeur moyenne bornée par la profondeur minimum et maximum de la région d'intérêt. Les expériences réalisées en coloscopie ont démontré que l'hypothèse d'équifocalité n'avait pas d'impact significatif sur la qualité des mesures considérant l'échelle de l'environnement colique exploré (voir chapitre 5).

4.6 Discussion

Nous avons présenté dans ce chapitre un système de mesure de profondeur basé sur l'extraction du point de rupture de netteté au sein d'une vidéo. Cette dernière correspondant à un mouvement d'approche, à une vitesse approximativement constante, de la caméra face à une région d'intérêt. Notre méthode est composée de deux modules. Le module *Blur-Estimating Tracking (BET)* permet l'estimation conjointe de l'appariement inter-images et de la mesure de flou relative. Le module *Blur Model Fitting (BMF)* permet l'intégration des mesures de flou et l'estimation robuste du point de rupture de netteté dénommé *Infocus-Breakpoint (IB)*. Notre système permet de s'affranchir des contraintes d'asservissement des paramètres externes et internes de la caméra respectivement imposées par les méthodes *Depth-from-Focus (DfF)* et *Depth-from-Defocus (DfD)*. Il a été utilisé à des fins de calibration pré-opératoire (extraction du point de rupture de netteté) ainsi que des mesures per-opératoire. Pour ces deux étapes, la stratégie employée est identique. Les évaluations menées ont démontrées que l'étape de calibration pouvait être réalisée efficacement en déplaçant une mire plane texturée face à la caméra. Cette phase de calibration est détaillée dans le chapitre suivant.

Les performances du module *BET* ont été évaluées sur un jeu de données synthétiques reproduisant les transformations inter-images observables lors d'examens coloscopiques. Ces expériences ont démontré que la méthode TPS-BET basée sur un appariement inter-images déformable (représenté par une *Thin-Plate Spline*) est plus performante que la méthode AFF-BET basée sur un appariement rigide (transformation affine). Elle s'avère plus robuste aux déplacements importants de la caméra (liés à sa vitesse d'approche) ainsi qu'au bruit. Ce gain de performance est significatif dans le contexte de la coloscopie. La manipulation du coloscope au sein de la cavité colonique est délicate, notamment lorsque ce dernier est proche d'une lésion ou de la paroi du côlon. Cela peut entraîner d'importants changements d'échelle de la néoplasie dans les images ainsi que des flous de mouvement. Nous avons par ailleurs proposé une méthode d'extraction robuste de l'*IB* permettant de s'affranchir des erreurs de mesures de flou lorsque le coloscope est proche de la profondeur de référence. L'application de notre système à la mesure 3D de néoplasies en coloscopie est présentée dans le chapitre suivant.

5

Mesurer la taille des néoplasies en coloscopie

La coloscopie est une procédure endoscopique permettant l'exploration et l'intervention au sein de la cavité colique. Parmi les critères diagnostics des néoplasies (croissances anormales de tissu) la taille est essentielle. Les erreurs de mesure peuvent notamment entraîner des erreurs d'évaluation des intervalles de surveillance des lésions. Comme il est impossible de connaître la taille réelle (vérité terrain) d'une néoplasie au cours d'une coloscopie, nous présentons une évaluation qualitative sur modèle fantôme. Cette dernière démontre la robustesse de la méthode proposée face aux artefacts communément rencontrés en coloscopie. L'erreur de mesure est de l'ordre du millimètre. Cette étude est complétée par une évaluation sur des cas de coloscopies réelles. Les mesures sont comparées à une estimation visuelle. Cette évaluation démontre la pertinence de la méthode de mesure des néoplasies par extraction du point de rupture de netteté. Cette étude a été publiée dans [Chadebecq et al., 2015]. Le système de mesure des néoplasies a par ailleurs été intégré avec succès à la méthode de reconstruction 3D de l'environnement colique présentée dans [Alcantarilla et al., 2013].

5.1 L'examen coloscopique

Le cancer colorectal est la quatrième cause mondiale de décès par cancer selon les derniers rapports de l'Organisation Mondiale de la Santé [World Health Organization, 2008] (694 000 décès en 2012). Il est en France la seconde cause de décès par cancer chez l'homme (9 275 décès en 2012 et 23 266 nouveaux cas détectés) et la troisième chez la femme (8 447 décès en 2012 et 18 926 cas détectés) [Institut National du Cancer, 2014]. Les campagnes de dépistage, et plus

particulièrement les tests immunologiques, sont aujourd’hui généralisés à toutes personnes âgées de 50 à 74 ans. Ces derniers permettent de détecter la présence de sang occulte dans les selles. En cas de test positif, un examen coloproctal est prescrit au patient afin de confirmer ou d’infirmar la présence de lésions cancéreuses ou de néoplasies, excroissances anormales de tissu pouvant dégénérer en tumeurs malignes (figure 5.1.1).

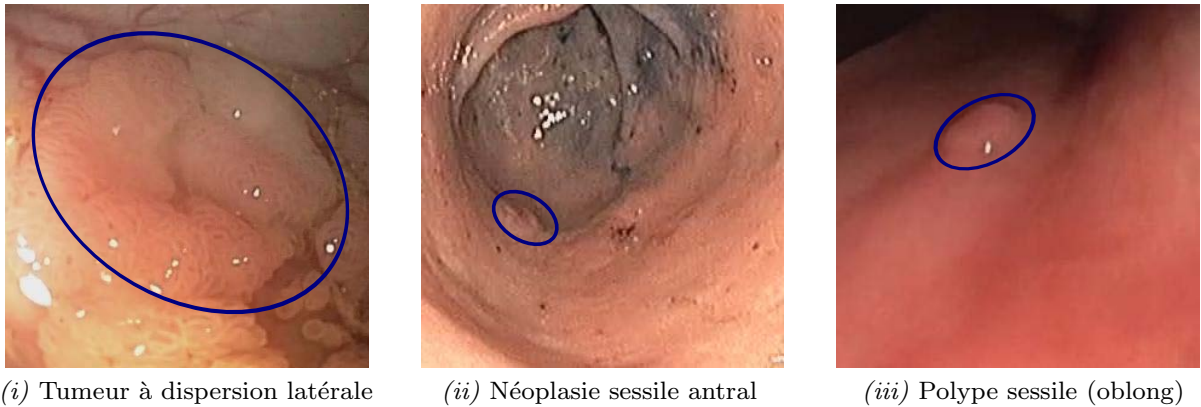


Figure 5.1.1 – Néoplasies coloniques. (i) Néoplasie suspecte présentant un risque de mutation ou étant maligne. (ii) et (iii) néoplasies bénignes. Ces trois exemples illustrent la difficulté d’analyse des images coloscopiques. La paroi colonique ainsi que les néoplasies sont peu texturées. Les méthodes de suivi basées sur l’appariement de points d’intérêt sont inadaptées [Kaufman and Wang, 2008].

L’examen coloproctal de référence est la coloscopie [Kim et al., 2007]. Cette procédure endoscopique est qualifiée de peu invasive car elle ne nécessite aucune incision d’organes internes contrairement à une chirurgie conventionnelle. Elle consiste à explorer le côlon à l’aide d’un endoscope flexible appelé coloscope (figure 5.1.2). Un gaz (CO_2) est insufflé au sein du côlon afin de le gonfler et ainsi faciliter le déplacement du coloscope. Cela permet également de déployer certains replis du côlon. Un médecin gastro-entérologue (spécialiste des maladies de l’appareil digestif) introduit alors le coloscope par voie basse dans le corps du patient et pratique un examen exploratoire en s’appuyant sur les images diffusées sur un écran placé sur le champ opératoire (figure 5.1.3 (i)). Cet examen permet l’exploration du rectum et de l’intégralité du côlon jusqu’à sa jonction avec l’intestin grêle (figure 5.1.3 (ii)).

Malgré le développement des techniques de coloscopie virtuelle (*Computed Tomography colonography* ou encore coloscanner) ainsi que des Pill Cam[®] la coloscopie est l’examen privilégié. Il permet notamment l’excision et la biopsie des néoplasies durant l’exploration du côlon. Le coloscope est équipé d’un canal opérateur permettant d’acheminer des instruments chirurgicaux

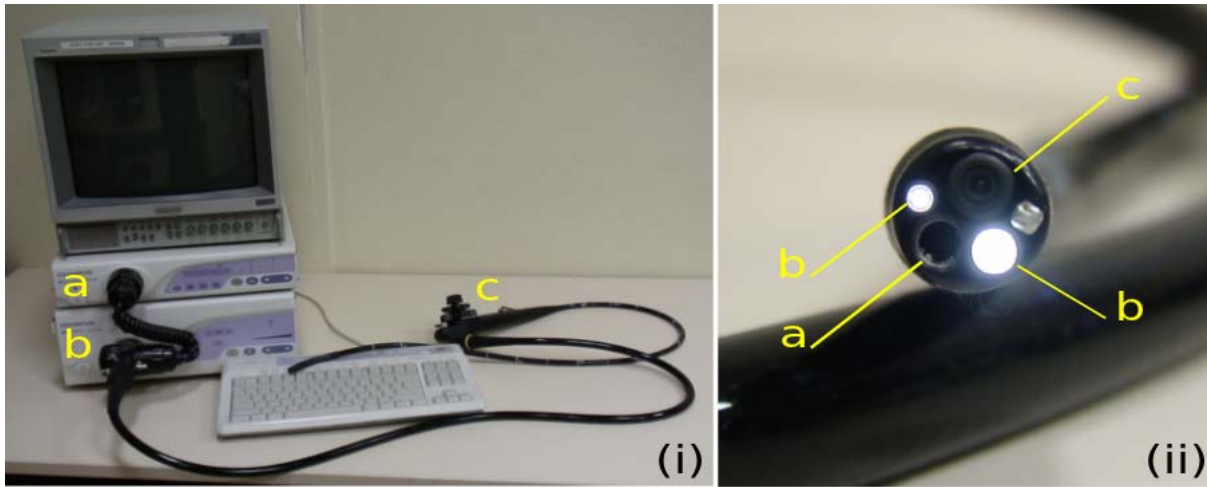


Figure 5.1.2 – Coloscope Olympus® Exera CV-160 - (i) colonne d'endoscopie connectée à son coloscope (a : module de traitement d'images, b : source de lumière, c : commande du coloscope) - (ii) Extrémité distale du coloscope (a : canal opérateur, b : source d'illumination, c : système optique).

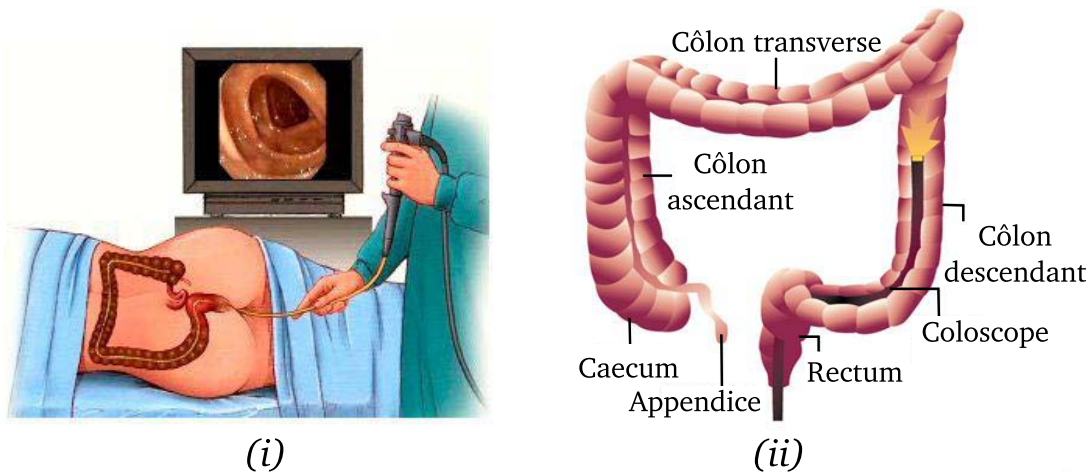


Figure 5.1.3 – La coloscopie : (i) procédure, (ii) sections du côlon explorées lors de l'examen [Schoen, 2002].

telles qu'une pince à biopsie ou un bistouri électrique en forme de lasso. Les Pill Cam® sont de petites capsules (de la taille d'une gélule) contenant une caméra. Elles sont ingérées par le patient et à usage unique. Le temps de transit intestinal est long et l'analyse de la vidéo est fastidieuse. De plus, il est complexe de contrôler l'orientation et le déplacement de la capsule dans le corps. La coloscopie virtuelle utilise la tomodensitométrie pour reconstituer une image en 3D du côlon. La détection des tumeurs inférieures à 5 mm est délicate. L'efficacité de cette

méthode (taux de détection des lésions) fait l’objet d’études contradictoires. Le principal avantage de ces deux techniques et qu’elles ne nécessitent pas d’anesthésie et sont moins invasives que la coloscopie. Elles sont généralement proposées comme alternative dans des cas de contre indications médicales à la coloscopie.

5.2 Néoplasie maligne : critères anatomo-pathologiques prédictifs et ambiguïté d’échelle

Les néoplasies sont classées suivant leurs potentiels de malignité. Le diagnostic est établi à l’aide de critères visuels ainsi que par les biopsies pratiquées au cours de la coloscopie. Parmi les indices visuels, l’apparence, la forme ainsi que la taille de la lésion sont essentielles à l’évaluation de son potentiel de développement au sein du côlon.

Le critère d’apparence correspond aux motifs formés à la surface des néoplasies. Ces motifs sont communément classés suivant la “pit pattern classification” ou “classification de Kudo” [Kudo et al., 1994]. Les critères de forme et de taille sont définis dans l’article de référence : “the Paris classification of superficial neoplastic lesions” [Group, 2005] (figure 5.2.1). Généralement, une néoplasie dont la taille est supérieure à 10 mm a un fort potentiel de malignité alors qu’une néoplasie dont la taille est comprise entre 5 et 10 mm nécessite un suivi [Talbot, 1996]. La taille est également essentielle pour déterminer les intervalles de temps de surveillance (les pratiques médicales peuvent varier en fonction du pays). Si l’analyse de l’apparence et de la forme d’une néoplasie peut être réalisée à l’aide des seules images coloscopiques, il est difficile d’évaluer précisément sa taille visuellement.

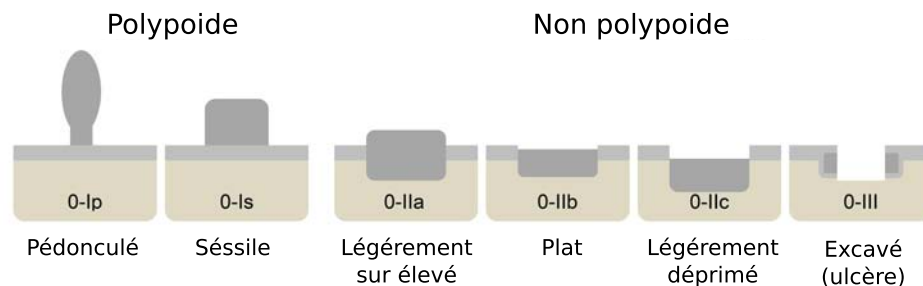


Figure 5.2.1 – Critères diagnostiques des lésions néoplasiques : classification de Paris.

La coloscopie est une modalité d’imagerie monofocale qui ne permet pas d’inférer *a priori* l’échelle des scènes coloscopiques. Les gastro-entérologues réalisent une estimation visuelle basée

sur leurs connaissances anatomiques *a priori* ainsi que sur leurs expertises. En cas d'ambiguïtés, un outil chirurgical est placé en contact avec la lésion comme référence de mesure. Plus rarement, des règles endoscopiques ou des outils marqués (sur lesquels sont placés des marques de référence) peuvent également être employés [Hyun et al., 2011]. De nombreuses études ont démontré que l'estimation visuelle de la taille des néoplasies est biaisée [Margulies et al., 1994; Moug et al., 2010; Schoen et al., 1997]. Récemment, [Chaptini et al., 2014] ont évalué la reproductibilité inter-opérateurs des mesures ainsi que l'influence des erreurs d'estimation sur les intervalles de temps de surveillance après une polypectomie (excision d'une néoplasie). Cette étude révèle que l'erreur d'estimation de la taille des néoplasies est supérieure à 20% dans 52% des cas (2812 cas étudiés). Cela conduit dans 10% des cas à la définition d'intervalles de temps de surveillance inappropriés. Le manque de reproductibilité intra et inter-opérateurs compromet le diagnostic per-opératoire et éventuellement post-opératoire [Rubio et al., 2010]. Par ailleurs, les techniques de mesure employées par les praticiens requièrent un temps de manipulation des instruments chirurgicaux non négligeable. Ceci est plus particulièrement le cas pour des interventions telles que la polypectomie pouvant nécessiter l'utilisation d'instruments de résection telles que le bistouri électrique en forme de lasso. Dans ce cas, l'instrument doit être remplacé par une règle endoscopique avant de pouvoir être à nouveau inséré dans le canal operateur.

La nécessité d'un système d'estimation de la taille des néoplasies permettant une mesure reproductible ainsi que son potentiel impact clinique constituent la motivation première de cette étude. Ce système peut être étendu à de nombreuses procédures endoscopiques et être utilisé à des fins de formation [Chang et al., 2010]. Il est une application directe de la méthode d'extraction du point de rupture de netteté présentée dans le chapitre 4.

La section suivante décrit le protocole clinique pré-opératoire de calibration du système ainsi que le protocole per opératoire de mesure de la taille d'une lésion colonique. L'évaluation du système sur modèle fantôme ainsi que sur des cas réels de coloscopies est présentée dans une seconde partie.

5.3 Système de mesure des néoplasies : protocole clinique

Le système proposé comporte deux étapes illustrées dans la figure 5.3.1. La phase de calibration préopératoire nécessite moins d'une minute. Elle est réalisée une seule fois pour un coloscope. La phase de mesure per-opératoire permet l'évaluation la taille d'une néoplasie. La mesure n'est pas réalisée en temps réel mais les résultats sont obtenus en quelques secondes (implémentation Matlab).

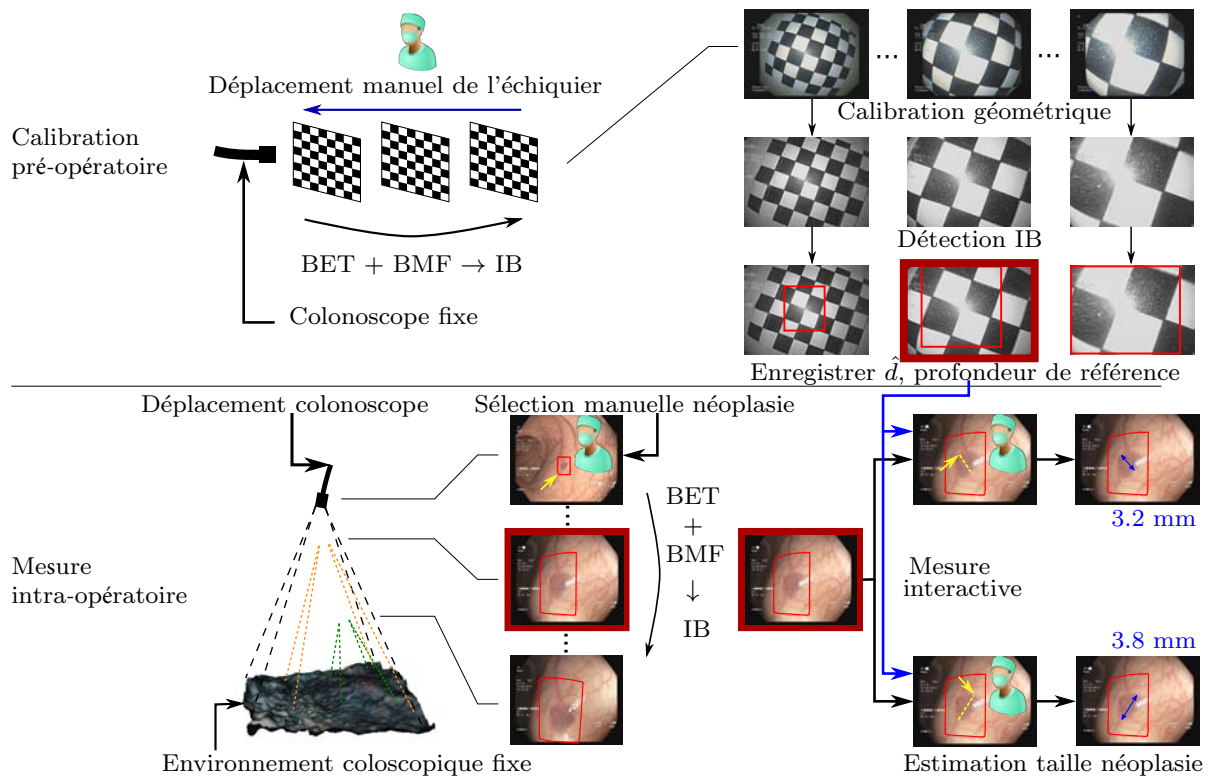


Figure 5.3.1 – Vue d'ensemble du système de mesure des lésions coloniques : protocole de calibration et protocole de mesure. *BET* *Blur-Estimating Tracking* (*BET*) est le module de suivi et d'estimation du flux optique dans la vidéo coloscopique. *Blur-Model Fitting* (*BMF*) est le module permettant l'extraction robuste du *point de rupture de netteté* (*Infocus Breakpoint* (*IB*)). La détection de l'*IB*, dont la profondeur de référence \hat{d} est obtenue par calibration, permet la mesure 3D de la néoplasie.

Calibration pré-opératoire : La calibration n'est pas requise avant chaque intervention. Une mire d'étalonnage ou une image plane texturée est positionné fronto-parallèlement au coloscope à une distance d'environ 10 cm. Il est ensuite déplacé en direction de ce dernier jusqu'à approcher la partie distale du coloscope (distance inférieure au centimètre). La vidéo est enregistrée et la calibration géométrique ainsi que la calibration de l'*IB* sont réalisées conjointement.

La calibration géométrique est nécessaire à l'obtention des paramètres internes de la caméra (paramètres de la matrice K) ainsi que de la fonction de distorsion. Elle a été réalisée à l'aide du module Matlab[®] [Bouguet, 2008]. Elle permet notamment d'obtenir la densité de pixel α du capteur permettant la conversion pixels vers millimètres. La calibration de la profondeur de référence correspondant à l'*IB* est basée sur les modules *BET* et *BMF*. Pour chaque image de la

vidéo, une région de l'échiquier (ou de la mire texturée) est suivie et le flou optique est estimé à l'aide du module *BET*. Le module *BMF* permet ensuite de détecter l'*IB* et la profondeur de référence \hat{d} est enregistrée. L'ensemble du protocole de calibration pré-opérateur nécessite généralement moins d'une minute.

Mesure interactive per-opérateur : Lors de la phase per-opérateur, le gastro-entérologue détecte une lésion colonique dont il souhaite effectuer la mesure. Il positionne alors le coloscope de façon à ce que cette dernière soit entièrement visible et apparaisse visuellement nette. Il sélectionne manuellement dans l'image une zone polygonale contenant la néoplasie. Il déplace ensuite le coloscope en direction de celle-ci jusqu'à la toucher. La vidéo est enregistrée et traitée afin de résoudre l'ambiguïté d'échelle. Le module *BET* permet le suivi automatique de la néoplasie dans la vidéo ainsi que l'estimation conjointe du flou optique au sein de la région polygonale. L'*IB* est extrait grâce au module *BMF*. Une image voisine de l'*IB* est affichée afin de permettre au praticien de réaliser des mesures interactives de longueurs de la lésion, en plaçant des points de mesure, ou de son aire en positionnant une ellipse. Le traitement hors ligne nécessite moins d'une minute.

5.4 Validation expérimentale

Notre système de mesure a été évalué sur des jeux de données :

- *ex-vivo* : un modèle fantôme a été conçu afin de reproduire les conditions d'une coloscopie tout en possédant une vérité terrain (la mesure réelle d'objets d'intérêt). Les vidéos fantômes ont été réalisées à l'aide d'un coloscope Olympus[®] GIF type Q160.
- *In-vivo* : un jeu de données composé d'exams coloscopiques a été réalisé afin de comparer les performances de notre système avec l'estimation visuelle réalisée par les gastro-entérologues. La vérité terrain ne peut être connue, cependant, les néoplasies composant le jeu de données sont de petites tailles afin de minimiser l'erreur d'estimation visuelle (tailles proches de celles des instruments chirurgicaux). Les coloscopies ont été réalisées à l'aide de deux coloscopes haute définition Olympus[®] GIF type N180 et XP190N.

A des fins d'évaluation, nous avons utilisé les deux versions de l'algorithme *BET* : AFF-BET et TPS-BET (chapitre 4). Le module TPS-BET a été paramétré par quatre centres de déformation. Ce choix a été motivé par la nécessité de rendre la TPS robuste aux légères déformations de l'environnement colonique tout en conservant des propriétés de rigidité [Pizarro et al., 2013].

Les centres de déformation sont initialement placés aux angles du rectangle englobant la région d'intérêt à mesurer.

5.4.1 Évaluation sur modèle fantôme

Le modèle fantôme utilisé se compose d'un tube dans lequel un côlon de porc est inséré afin de reproduire l'apparence d'une cavité colonique [Park et al., 2007]. Des objets de tailles comprise entre 1,6 à 15 mm sont enveloppés dans du tissu colonique et placés au sein de cette cavité (figure 5.4.1).

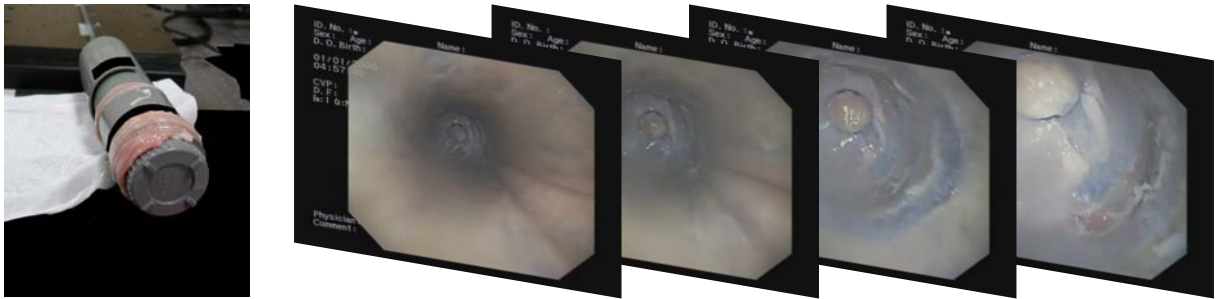


Figure 5.4.1 – Le modèle fantôme est composé d'un tube dans lequel un côlon de porc a été inséré afin de reproduire l'apparence d'une cavité colonique. Des objets de tailles connues ont été enveloppés dans du tissu colonique et placés au sein de la cavité.

Motivation : Il est pratiquement complexe d'obtenir la taille réelle d'une néoplasie *in vivo*. Son excision entraîne par ailleurs une rétraction qui ne permet pas la mesure *ex vivo*. Le modèle fantôme permet de reproduire des conditions similaires à celle d'une coloscopie à l'exception des déformations de l'environnement (flou de mouvement lié à la manipulation du coloscope, saturations lumineuses, ...). Le mouvement du coloscope a été contrôlé manuellement (8 vidéos) ainsi qu'à l'aide d'un système mécanique de translation millimétrique (7 vidéos). Ceci permet d'évaluer de façon pratique la robustesse du système de mesure face aux flous de mouvements ainsi qu'à une vitesse de déplacement irrégulière du coloscope. La figure 5.4.2 présente les erreurs de mesures obtenues (erreur au sens des moindres carrés, *Root Mean Square (RMS)*).

Résultats : Les résultats démontrent que le système basé sur le module TPS-BET est plus performant que celui basé sur le module AFF-BET. Il permet l'obtention de mesures précises et s'avère plus robuste aux artefacts. Lorsque le coloscope est contrôlé manuellement, l'erreur moyenne de mesure est de 6,9% pour la méthode TPS-BET soit environ 0,8 mm alors que

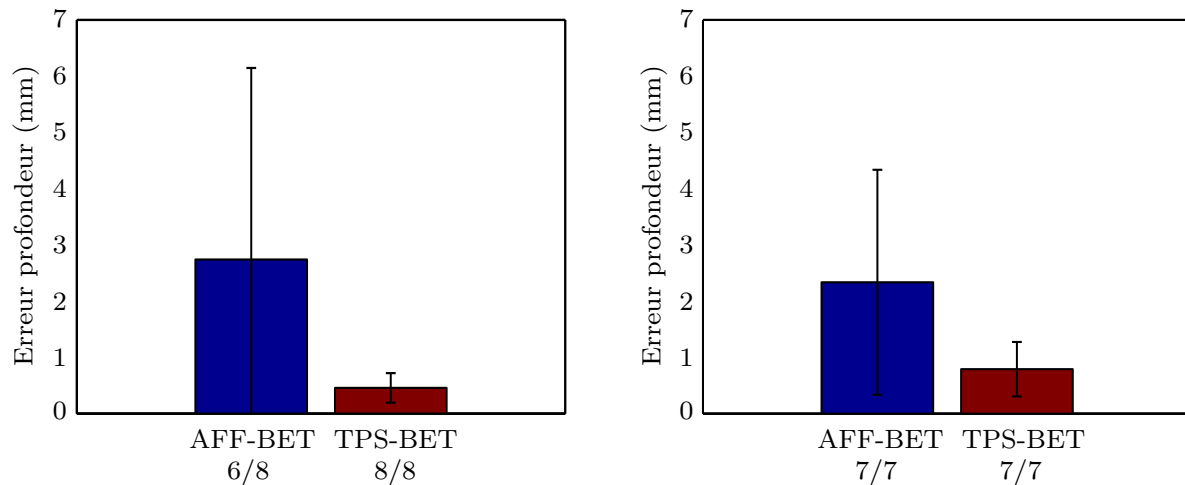


Figure 5.4.2 – Evaluation du système de mesure sur modèle fantôme colonique. Le système de mesure a été évalué sur 8 vidéos réalisées manuellement (*i*) ainsi que sur 7 vidéos réalisées à l'aide d'un système mécanique de translation (*ii*). Les modules AFF-BET (*Blur Estimating Tracking affine Affine*) et TPS-BET (basé sur une TPS) ont été intégrés au système afin de comparer leurs influences sur la précision des mesures. Parmi les vidéos réalisées manuellement (*i*), seules 6 séquences ont pu être traitées à l'aide de la méthode AFF-BET, notamment de par la présence de flous de mouvements.

L'erreur moyenne de mesure est de 26% pour la méthode AFF-BET soit environ 2,8 mm. Ceci est notamment dû à des erreurs de suivi pouvant impliquer des ambiguïtés d'extraction de l'*IB* (figure 5.4.3). Deux vidéos issues de ce jeu de données n'ont pu être traitées à l'aide de la méthode AFF-BET de part la présence de flous de mouvements. Elles ont été correctement traitées à l'aide de la méthode TPS-BET malgré une dégradation de la précision des mesures.

Concernant les vidéos réalisées à l'aide du système mécanique de translation, l'erreur moyenne de mesure est de 24% (environ 2,4 mm) pour la méthode AFF-BET. Ces derniers ont été dégradés par la présence d'humidité sur l'extrémité distale du coloscope. La méthode TPS-BET permet d'obtenir une erreur moyenne de mesure de 4,9% (environ 0,5 mm). Elle a par ailleurs un comportement plus robuste aux flous de mouvements ainsi qu'aux artefacts inhérents aux examens coloscopiques (humidité, saturations, ...). Il est par ailleurs important de noter que l'hypothèse d'équifocalité a un impact peu significatif sur la précision des mesures considérant l'échelle de la scène observée ainsi que les contraintes diagnostiques portant sur la taille des néoplasies. Cette étude qualitative est à mettre en perspective des travaux de [Chaptini et al., 2014] qui révèlent que l'erreur d'estimation visuelle de la taille des néoplasies est supérieure à 20% dans 52% des cas (2812 cas étudiés).

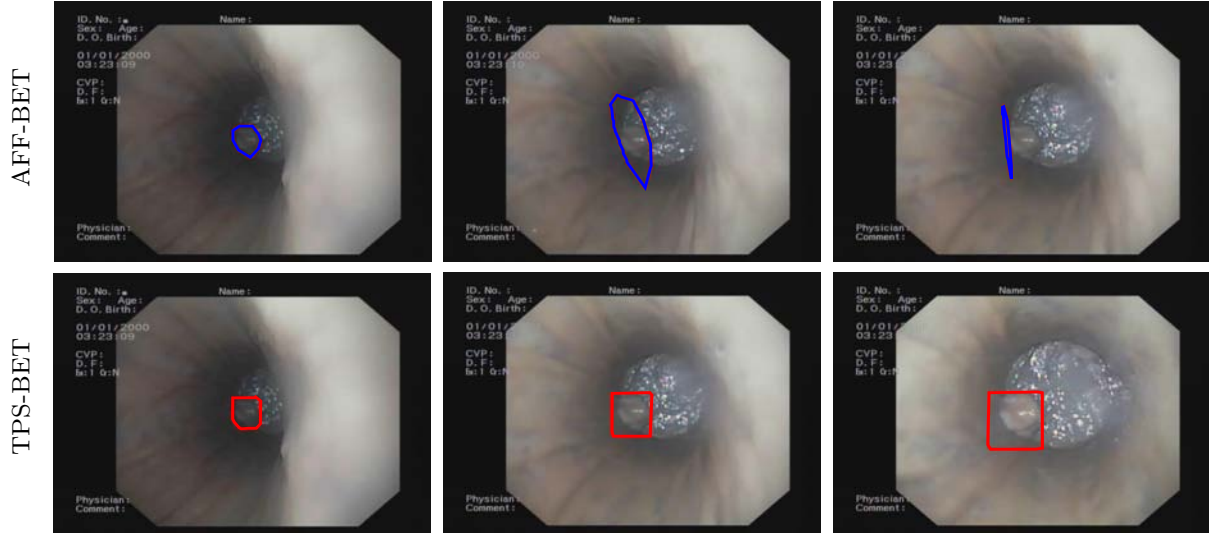


Figure 5.4.3 – Erreur de suivi d’une région d’intérêt lors de l’évaluation du système de mesure des néoplasies sur modèle fantôme colonique. Le contrôle manuel du coloscope engendre des flous de mouvements. Ces derniers font échouer le module AFF-BET (*Blur Estimating Tracking* affine, ligne supérieure) alors que le module TPS-BET (*BET* basé sur une TPS) permet le suivi de la région d’intérêt.

5.4.2 Evaluation en coloscopie

Cinq vidéos ont été acquises par deux gastro-entérologues lors de coloscopies. Ces dernières ont été réalisées suivant le protocole précédemment proposé (section 5.3). La calibration pré-opératoire ainsi que la mesure per-opératoire d’une néoplasie sont illustrées dans la figure 5.4.4.

De façon à minimiser l’influence des flous de mouvements et s’assurer que les néoplasies soient entièrement visibles au sein de la vidéo, il a été demandé aux gastro-entérologues d’appliquer au coloscope un mouvement linéaire et constant. Les résultats obtenus avec AFF-BET et TPS-BET ont été comparés avec les estimations visuelles réalisées par ces derniers. Afin d’obtenir une évaluation visuelle fiable nous avons volontairement considéré la mesure de petites lésions dont la taille était comparable à celle d’un instrument chirurgical. La figure 5.4.5 présente l’erreur RMS moyenne pour les deux algorithmes *BET*. Une planche récapitulative des résultats est présentée dans la figure 5.4.6.

Resultats : Les mesures obtenues par les systèmes AFF-BET et TPS-BET sont similaires. L’erreur RMS moyenne est de 4,5% (environ 0,2 mm en comparaison des estimations visuelles) pour l’algorithme TPS-BET et de 6.5% (environ 0.3 mm) pour l’algorithme AFF-BET. De façon

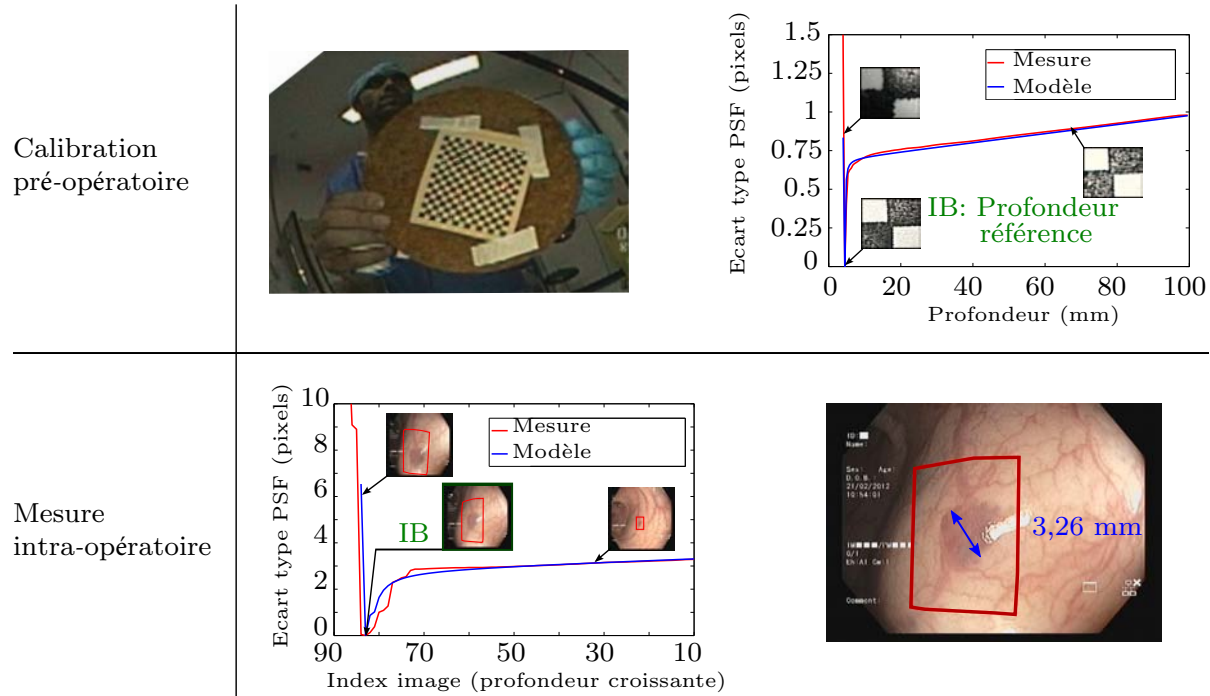


Figure 5.4.4 – Calibration et utilisation du système de mesure des lésions coloniques. La première ligne du tableau illustre la calibration pré-opératoire de la profondeur de calibration d'un coloscope. La ligne inférieure présente la mesure per-opératoire d'une néoplasie basée sur l'extraction du point de rupture de netteté (*Infocus Breakpoint (IB)*).

similaire à l'évaluation sur modèle fantôme, le module AFF-BET a échoué pour l'une des vidéos du jeu de données. La présence de flous de mouvements a compromis le suivi et la mesure de flou s'est avérée trop imprécise pour extraire de façon robuste l'*IB* (malgré le module *BMF*). Cette étude confirme le gain de robustesse de l'algorithme TPS-BET. Il est par ailleurs important de noter que les mesures obtenues par notre système sont reproductibles et de nouvelles mesures peuvent être réalisées post-opérativement.

5.5 Discussion

Les résultats expérimentaux, et plus particulièrement ceux obtenus à l'aide du modèle fantôme, valident la méthode proposée de mesure interactive des lésions coloniques par détection du point de rupture de netteté. Cette méthode s'appuie sur un protocole clinique simple qui a pu être appliqué sans difficulté par les deux gastro-entérologues impliqués dans notre étude. Le module

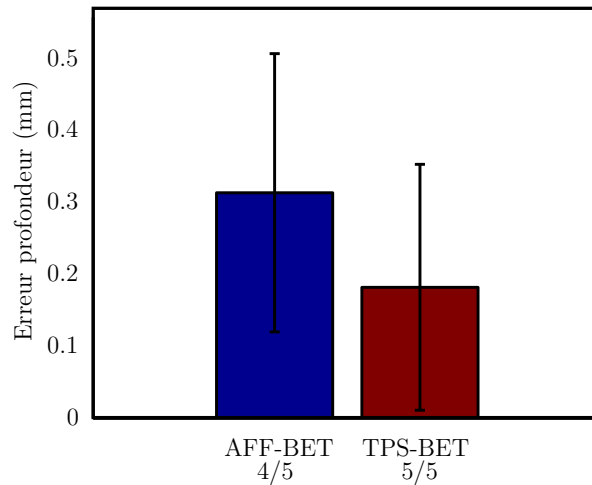


Figure 5.4.5 – Evaluation du système de mesure des lésions coloniques. L’influence des modules AFF-BET (*Blur Estimating Tracking Affine*) et TPS-BET (*BET basée sur une TPS*) sur la précision de mesure des lésions coloniques est évaluée pour 5 vidéos coloscopiques. Les résultats obtenus sont similaires concernant la précision des mesures mais AFF-BET s’est avéré moins robuste au bruit et flous de mouvements. Seules 4 séquences ont pu être traitées à l’aide de ce module.

TPS-BET est robuste aux artefacts communément rencontrés lors d’une coloscopie et plus particulièrement au bruit (dégradation de la qualité optique liée à l’humidité, saturations, ...) et aux flous de mouvements. Il est également robuste aux légères déformations de l’environnement coloniques liées à l’action du praticien ou aux mouvements naturels des organes. Le module *BMF* permet d’extraire de façon robuste et précise le point de rupture de netteté malgré le déplacement approximativement constant du coloscope. Le système proposé est adapté à la mesure de néoplasies d’une taille inférieure à 15 mm. Pratiquement, l’erreur de mesure est inférieure au millimètre ce qui permet notamment de définir des intervalles de temps de surveillance précis et fiables.

Différentes perspectives peuvent être envisagées afin d’améliorer ce système ou de s’affranchir des contraintes du protocole de mesure proposé. Selon la position de la tumeur au sein de la cavité colonique, il peut être difficile de réaliser une vidéo dans laquelle cette dernière est entièrement visible à chaque instant (cas d’une lésion située au sein d’un replis colonique). Les occultations partielles ou totales de la néoplasie doivent être considérées afin d’éviter toute erreur de mesure du flou optique impliquant une ambiguïté d’extraction du point de rupture de netteté. Il peut également être difficile d’appliquer un déplacement approximativement constant au coloscope. Il serait ainsi nécessaire d’inclure au modèle de flou considéré dans le module *BMF* une fonction

de déplacement non constante. Cette dernière pourrait être estimée *a posteriori* en fonction du facteur de grossissement de la tumeur au sein des images. Enfin, la détection automatique des flous de mouvements permettait d'améliorer la robustesse des mesures de flou optique notamment lorsque le coloscope est proche de la région d'intérêt.

Les limitations majeures de la méthode proposée sont relatives à la mesure du flou optique basée région ainsi qu'à l'a priori de planarité. Dans le contexte de l'imagerie endoscopique, où l'apparence des lésions est peu texturée, il est difficile de s'affranchir de ces hypothèses. Par ailleurs, les évaluations ont démontrées que l'hypothèse de planarité avait un impact peu significatif sur l'estimation de la taille des néoplasies considérant l'échelle de l'environnement exploré ainsi que les contraintes diagnostiques liées à ce critère. La méthode de mesure par extraction du point de rupture de netteté est cependant difficilement généralisable. Elle n'est pas adaptée à la reconstruction 3D dense et robuste de scènes et ne s'applique qu'à des environnements rigides. Elle est également difficilement adaptable à l'estimation de l'échelle de scènes de taille importante pour lesquelles l'hypothèses d'équifocalité pourrait entraîner un biais de mesure significatif. Le chapitre 6 présente, sous la forme d'une étude théorique, une perspective d'extension de cette méthode à une approche générative. Cette dernière a pour objectif de répondre aux limitations précédemment explicitées en permettant la reconstruction 3D dense et à l'échelle de scènes déformables.

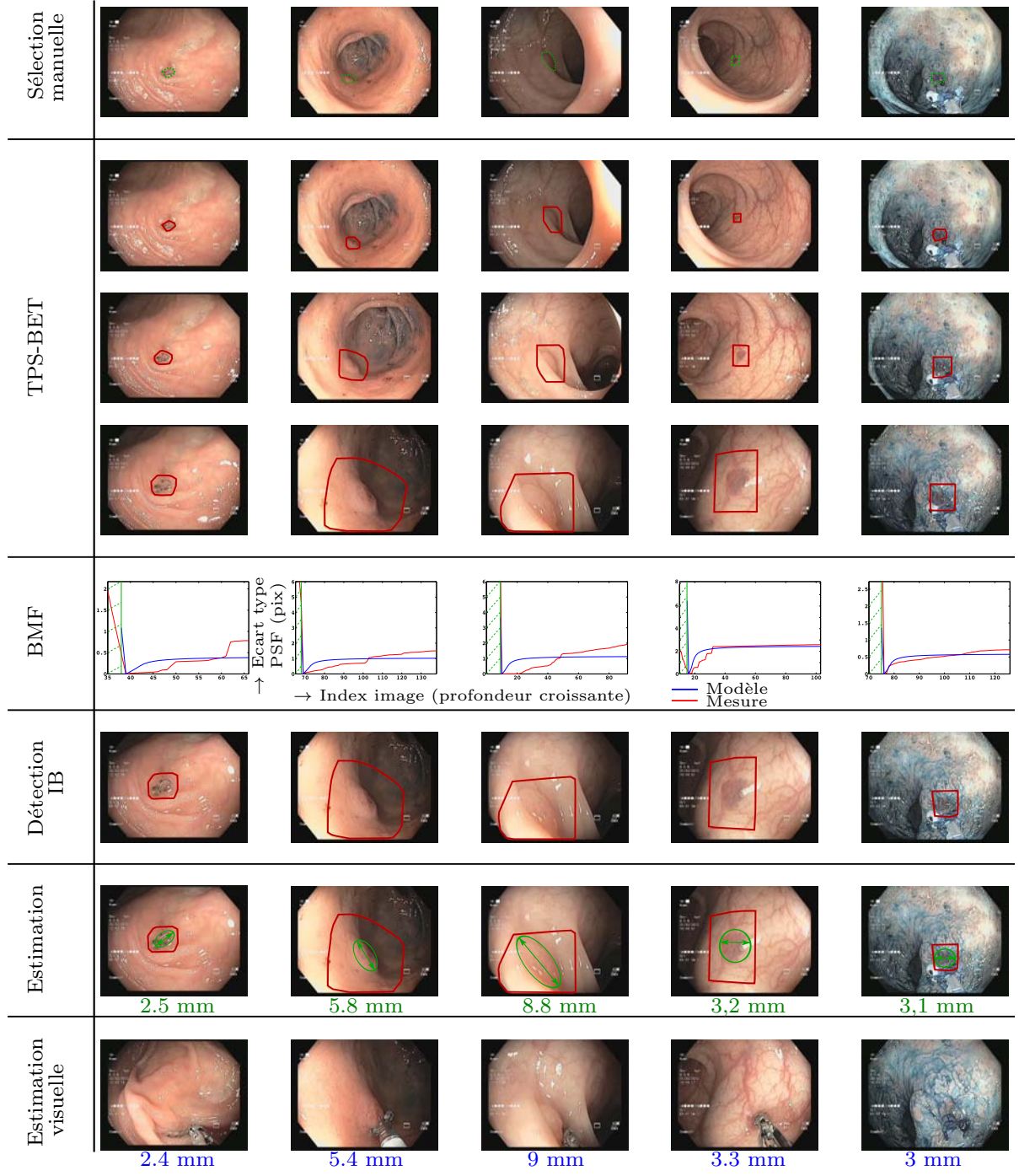


Figure 5.4.6 – Estimation de la taille des néoplasies à l'aide du système de mesure par extraction du point de rupture de netteté : évaluation sur données coloscopiques.

6

Perspective d'extension par un modèle génératif

Nous proposons une perspective d'extension des travaux présentés dans ce manuscrit par une méthode générative de reconstruction 3D à l'échelle de surfaces déformables. Cette approche *NRSfM* (*Non-Rigid Structure-from-Motion*) considère la reconstruction d'une surface (dénommée *template*) subissant un ensemble de déformations isométriques. Nous proposons un modèle de formation d'images original intégrant un modèle de flou optique. L'inférence de ce modèle, basée sur un algorithme d'alternation, permet l'estimation conjointe de la déformation subie par le template ainsi que de sa carte de couleurs. Nous présentons tout d'abord un état de l'art des méthodes de reconstruction 3D de surfaces déformables. Nous y introduisons notamment les travaux [Bartoli et al., 2012, 2015] portant sur les méthodes *SfT* (*Shape-from-Template*) auxquels nous avons contribué. Nous détaillons par la suite le modèle génératif de formation d'images proposé. Nous proposons enfin un algorithme d'alternation permettant son inférence. La méthode proposée fait actuellement l'objet d'une évaluation expérimentale dont les résultats ne sont pas présentés dans ce manuscrit.

La reconstruction 3D de scènes rigides a fait l'objet de nombreuses études. Les méthodes géométriques telles que *SfM* sont aujourd'hui matures bien que l'ambiguïté d'échelle subsiste [Hartley and Zisserman, 2004]. La reconstruction de surfaces déformables est un problème plus complexe car il existe une infinité de solutions générant la même observation [Moreno-Noguer and Fua, 2013; Moreno-Noguer et al., 2010]. La résolution de cette ambiguïté nécessite de considérer des *a priori* sur la nature physique de la surface, le type de déformations qu'elle subit et les propriétés de la caméra opérante. Nous considérons le cas de surfaces déformables non extensibles.

Les approches considérant des *a priori* physiques [Bhat et al., 2003; Cohen and Cohen, 1993;

[McInerney and Terzopoulos, 1993; Tsap et al., 1998] s'appuient généralement sur la minimisation d'une fonction d'énergie liant l'erreur de reprojection de la surface dans l'image aux comportements physiques de cette dernière. Ces méthodes ne sont pas généralisables et reposent sur des modèles physiques dont les paramètres sont difficilement estimables. La reconstruction de surfaces fortement déformées nécessite la complexification des modèles utilisés ce qui rend leurs inférences délicate. Les approches statistiques basées sur l'apprentissage de l'espace de déformation d'une surface [Brand, 2001; Salzmann and Fua, 2009; Salzmann et al., 2008; Xiao and Kanade, 2005] ne sont également pas généralisables. La précision des reconstructions est fortement dépendante de la nature de la surface ainsi que du volume des données d'apprentissage.

Les méthodes s'appuyant sur des *a priori* relatifs à la nature des déformations ont fait l'objet d'un intérêt croissant [Bartoli et al., 2012, 2015; Brunet et al., 2014; Chhatkuli et al., 2014a,b]. Nous considérons des surfaces subissant des déformations isométriques¹. Si cet *a priori* est théoriquement inadapté aux caractéristiques physiques d'un grand nombre de surfaces, il s'est avéré efficace dans de nombreux cas pratiques (tels que la reconstruction d'organes en endoscopie par exemple [Malti and Bartoli, 2014; Malti et al., 2011]).

Deux principales catégories d'approches géométriques peuvent être distinguées : (i) les approches basées modèle appelées *Shape-from-Template (SfT)* et (ii) les approches dérivées des méthodes *SfM* appelées *Non-Rigid Shape-from-Motion (NRSfM)*. Afin d'éviter toute confusion lexicale nous désignerons par l'anglicisme *template* le modèle de la surface aplanie. Nous distinguons le template 3D correspondant à la surface réelle aplanie et le template 2D correspondant à son image lorsqu'elle est fronto parallèle à la caméra. Les modèles considérés par les approches *SfT* et *NRSfM* sont illustrés dans la figure 6.0.1.

L'approche *SfT* considère le template connu. Ce dernier est généralement obtenu en effectuant une reconstruction 3D de la surface non déformée (à l'aide d'une méthode SfM par exemple), puis en réalisant un aplanissement conforme² de cette dernière. Cela est notamment adapté aux examens endoscopiques où l'organe est d'abord inspecté avant qu'il ne subisse diverses déformations liées à une action du praticien (en négligeant les déformations naturelles de l'environnement). Les méthodes *SfT* permettent la reconstruction 3D de la surface déformée à partir d'une observation unique. Parmi celles-ci, les approches basées sur une contrainte d'isométrie permettent une reconstruction à l'échelle. Les méthodes *NRSfM* s'appuient sur l'acquisition d'un ensemble d'images de la surface déformée (au moins 3 observations [Chhatkuli et al., 2014a]). Le template

1. Une déformation isométrique conserve les distances entre les espaces métriques du domaine de définition et du domaine image.

2. Une déformation conforme est une bijection qui préserve localement les angles.

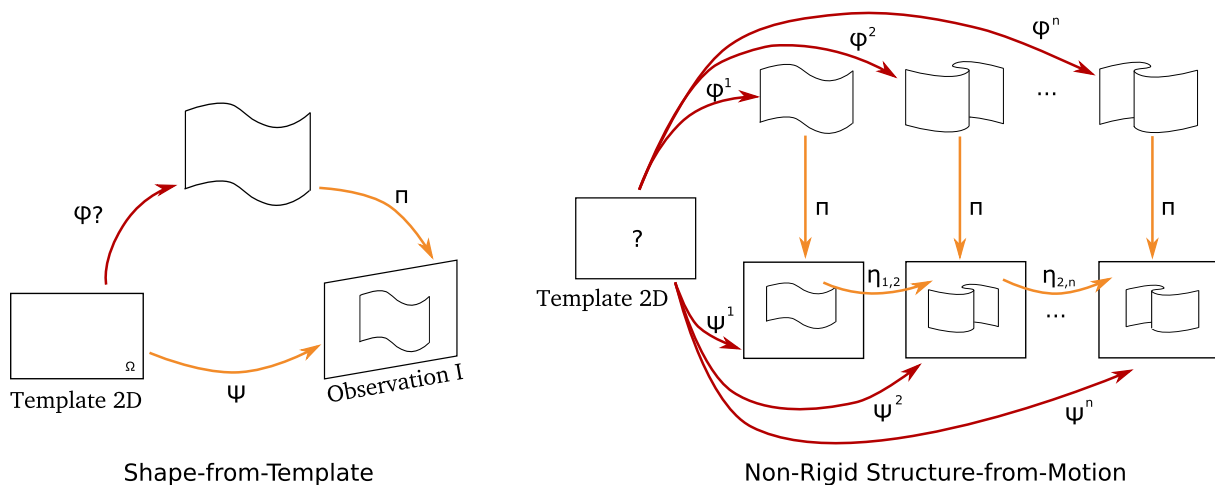


Figure 6.0.1 – Problèmes *Shape-from-Template (SfT)* et *Non-Rigid Shape-from-Motion (NRSfM)*. [Chhatkuli et al., 2014a])

est indéterminé mais l'appariement inter-images est connu. A l'instar des méthodes SfM, NRSfM souffrent de l'ambiguïté d'échelle.

Nous présentons dans les sections suivantes les contributions majeures de l'état de l'art relatif à ces approches. Nous introduirons par la suite une approche NRSfM générative intégrant la modélisation du flou optique afin de résoudre l'ambiguïté d'échelle. Elle s'inscrit dans une étude perspective et théorique préliminaire à une évaluation pratique. Nous soulignerons les contributions majeures apportées par cette approche originale visant la reconstruction dense de surfaces déformables.

6.1 Méthodes géométriques pour la reconstruction 3D de surfaces déformables

6.1.1 *Shape-from-Template (SfT)*

Trois catégories d'approches SfT peuvent être distinguées; (i) les approches basées sur une optimisation convexe [Brunet et al., 2014; Perriollat et al., 2011; Salzmann and Fua, 2011], (ii) celles basées sur une optimisation non-convexe [Brunet et al., 2014], (iii) et enfin les approches analytiques [Bartoli et al., 2012, 2015, 2013; Chhatkuli et al., 2014b].

(i) **Approches basées sur une optimisation convexe** : Ces méthodes s'appuient sur la relaxation de la contrainte d'isométrie à l'aide d'une contrainte d'inextensibilité de la surface. [Perriollat et al., 2011] proposent une approche heuristique (*Maximum-Depth Heuristic*) basée sur la maximisation de la profondeur des points 3D de la surface déformée. La distance séparant chaque paire de points est bornée par sa distance géodésique³ calculée à partir des points du template 3D. [Salzmann and Fua, 2011] propose une méthode basée sur l'apprentissage d'un modèle de déformation de la surface (le modèle est appris en considérant un ensemble de déformations subi par des sous-régions de la surface). La surface déformée est exprimée sur la base vectorielle apprise en considérant l'heuristique de profondeur maximum. [Brunet et al., 2014] combinent cette heuristique et l'*a priori* d'inextensibilité pour formuler un problème d'optimisation convexe SOCP (*Second Order Cone Programming*). Deux formulations sont proposées, l'une basée sur les correspondances de points template-surface (équation (6.1)) et l'autre basée sur une représentation paramétrique de la surface.

$$\begin{aligned}
\max_{\mathbf{Q}} \quad & \mathbf{p}_k^\top \sum_{i=1}^n \mathbf{Q}_i \\
sc. \quad & \left\| \begin{bmatrix} \mathbf{p}_1^\top \\ \mathbf{p}_2^\top \end{bmatrix} \mathbf{Q}_i - \mathbf{q}'_i \mathbf{p}_3^\top \mathbf{Q}_i \right\| \leq \epsilon_{\mathcal{I}} \mathbf{p}_3^\top \mathbf{Q}_i \quad \forall i \in 1, \dots, n \\
& \|\mathbf{Q}_i - \mathbf{Q}_j\| \leq d_{ij} + \epsilon_{\mathcal{T}} \quad \forall (i, j) \in \epsilon \\
& \mathbf{p}_3^\top \mathbf{Q}_i \geq 0 \quad \forall i \in 1, \dots, n
\end{aligned} \tag{6.1}$$

où \mathbf{p}_k^\top correspond à la $k^{\text{ième}}$ ligne de la matrice des paramètres internes de la caméra, \mathbf{Q}_i est un point 3D appartenant à la surface déformée, \mathbf{q}_i est un point 2D appartenant au template et \mathbf{q}'_i est le point 2D correspondant dans l'image. $d_{ij} = \|\mathbf{q}_i - \mathbf{q}_j\|$ correspond à la distance géodésique mesurée à la surface du template 3D. Cette formulation étend [Salzmann and Fua, 2009] en étant robuste au bruit dans le template comme dans l'image (respectivement exprimés par $\epsilon_{\mathcal{T}}$ et $\epsilon_{\mathcal{I}}$). Les problèmes SOCP peuvent être résolus efficacement à l'aide d'un algorithme du point intérieur [Boyd and Vandenberghe, 2004].

Les approches basées sur une relaxation convexe sont particulièrement efficaces au détriment de la précision de reconstruction. Elles sont pratiquement peu robustes aux faibles effets de perspectives [Chhatkuli et al., 2014b].

(ii) **Approches basées sur une optimisation non convexe** : [Brunet et al., 2014] pro-

3. La distance géodésique séparant deux points d'une surface correspond à la distance du plus court chemin sur la surface.

posent une troisième formulation du problème SfT basée sur une optimisation non-convexe. Cette dernière repose sur la contrainte d’isométrie (et non sur sa relaxation) exprimée dans le domaine continu par :

$$\int_{\Omega} \left\| \mathcal{J}(\mathbf{q})^{\top} \mathcal{J}(\mathbf{q}) - Id_{2 \times 2} \right\|^2 d\mathbf{q} = 0 \quad (6.2)$$

Le problème est alors exprimé comme la minimisation de la somme des contraintes d’isométrie, de projection et de lissage de la surface. La contrainte de reprojection est exprimée par :

$$\epsilon(d, \ell) = \sum_{i=1}^n \left\| \varphi_{\ell}(\mathbf{q}_i) - d_i \mathbf{K}^{-1} \bar{\mathbf{q}}'_i \right\|^2 \quad (6.3)$$

où φ_{ℓ} correspond à la fonction de déformation du template de paramètres ℓ (par exemple les coordonnées de points de contrôle agissant sur la déformation de la surface). d_i représente la profondeur du $i^{\text{ième}}$ point de référence de la surface et $\bar{\mathbf{q}}'_i$ représente les coordonnées homogènes du $i^{\text{ième}}$ point observé dans l’image.

Le problème de minimisation est résolu à l’aide de l’algorithme de Levenberg-Marquardt [Levenberg, 1944]. Cette méthode permet d’obtenir une reconstruction 3D précise de la surface déformée. Elle nécessite néanmoins une initialisation proche de la solution finale qui peut être obtenue à l’aide d’une méthode d’optimisation convexe. Dans le cas d’image réelles correspondant à la déformation d’une feuille de papier, l’erreur RMS de reconstruction 3D est de 1,99 mm alors que les méthodes convexes obtiennent dans le meilleur des cas une erreur de 2,26 mm.

(iii) approches analytiques : [Bartoli et al., 2012, 2015] démontrent que le problème de reconstruction SfT est bien posé considérant un opérateur de projection perspective et la déformation d’une surface développable⁴, isométrique ou conforme. Ces *a priori* sont exprimés à l’aide de contraintes sur la matrice Jacobienne de la fonction de déformation du template φ ([Pressley, 2010]) :

$$\begin{cases} \pi \circ \varphi & = & \Psi \\ \mathcal{J}_{\varphi}^{\top} J_{\varphi} & = & \lambda Id \end{cases} \quad (6.4)$$

4. Une surface développable est une surface réglée (par chaque point d’une surface réglée passe une droite contenue dans la surface) possédant une courbure Gaussienne nulle. En d’autres termes, cette surface peut être aplaniée sans distorsion.

avec :

$$\lambda^2 = \begin{cases} 1 & \text{développable} \\ \det(\mathcal{J}_\Delta^\top \mathcal{J}_\Delta) & \text{isométrie} \\ \text{inconnue } \Omega \mapsto \mathbb{R} & \text{conforme} \end{cases} \quad (6.5)$$

avec Id de taille 2×2 et λ la fonction d'échelle locale $\Omega \mapsto \mathcal{R}$. Les cas développable et isométrique possèdent une solution analytique unique (à un facteur d'échelle près) et le cas conforme un sous-ensemble de solutions. Elles sont dues à une ambiguïté de signe définissant les changements de courbures de la surface. [Pizarro et al., 2013] résolvent cette ambiguïté considérant une projection perspective faible (projections quasi affines pour lesquelles la majorité des méthodes précédemment citées sont peu efficaces). Les auteurs proposent une méthode basée sur une optimisation non linéaire. Elle repose sur le calcul de fonctions de déformation φ contraintes par les *a priori* d'isométrie et de conformité.

De récentes méthodes cherchent à s'affranchir des contraintes imposées aux méthodes SfT. [Bartoli and Collins, 2013; Bartoli et al., 2013] considèrent des modèles de reprojection perspectives et quasi-perspectives non calibrés. La première de ces méthodes consiste à retenir, parmi un ensemble de distances focales admissibles, la solution minimisant un critère d'isométrie global (approximation Euclidienne des distances géodésiques entre deux points du template et les deux points correspondants situés sur la surface). La seconde méthode repose sur une solution analytique (contrainte par les propriétés des plans tangents à la surface déformée). Cette dernière est particulièrement efficace (l'erreur de reconstruction est inférieure à 10% de même que l'erreur d'estimation de la distance focale). A l'instar des approches basées sur une optimisation convexe, il a été démontré que les méthodes analytiques sont instables dans de nombreux cas pratiques. [Chhatkuli et al., 2014b] améliorent la stabilité des méthodes *SfT* pour les modèles de projection perspectives ou quasi orthographiques. L'équation aux dérivées partielles (6.4) est reformulée en considérant chacune de ces configurations aboutissant à deux solutions analytiques distinctes. L'ambiguïté de signe est résolue sans adjonction de contraintes relatives à la scène. Cette méthode est particulièrement robuste et peut être combinée à [Brunet et al., 2014] afin d'améliorer la précision des reconstructions (erreur RMS de reconstruction de 3,62 mm pour une erreur de 7,76 mm pour [Salzmann et al., 2007]).

[Malti and Bartoli, 2014; Malti et al., 2011] combinent une approche *SfT* et *SfS* afin d'effectuer la reconstruction 3D d'organes en endoscopie. Dans ce contexte spécifique où les organes observés possèdent des propriétés de réflectance complexes, les méthodes SfT et SfS échouent. La combinaison de ces deux approches (respectivement suivant un modèle de réflectance Lambertien et

de Cook-Torrance [Cook and Torrance, 1982] pour les deux travaux suscités) s’est avérée efficace avec une erreur RMS inférieure à 2 mm. Ces méthodes nécessitent néanmoins l’ajustement de paramètres de poids associés à chacune des contraintes de la fonction de coût proposée.

6.1.2 Non-Rigid Shape-from-Motion (NRSFM)

Nous distinguons les méthodes basées sur un modèle de surface dit de rang faible (*i*) et celles basées sur un modèle de surface supposé localement planaire (*ii*).

(*i*) **Modèle de surface de rang faible** : le modèle de rang faible repose sur la représentation de la surface par une combinaison de formes de bases (exprimées sur une base vectorielle) [Bregler et al., 2000]. Cette dernière peut être obtenue grâce à un apprentissage statistique [Del Bue et al., 2004] ou estimée conjointement avec la pose de la caméra à l’aide d’une méthode de factorisation [Bartoli et al., 2008; Gotardo and Martinez, 2011a,b; Torresani et al., 2008]. Ces approches consistent à factoriser la matrice des points d’intérêts (suivis dans la vidéo) de façon à isoler les composantes relatives à la forme de la surface ainsi que celles relatives à la position de la caméra :

$$\mathbf{S} = \sum_{i=1}^K l^i \mathbf{S}^i \quad \mathbf{S}, \mathbf{S}^i \in \mathbb{R}^{3 \times n} \quad l^i \in \mathbb{R} \quad (6.6)$$

où S correspond à la surface déformée, l^i correspond au poids de déformation d’indice i et S^i est appelé forme de base avec (S^1, S^2, \dots, S^K) des matrices de taille $3 \times n$ contenant les coordonnées des points d’intérêt pour chacune des observations (en d’autres termes la matrice de suivi).

Considérant un modèle de projection orthographique les coordonnées des points 2D observés dans l’image peuvent être exprimés par :

$$\mathbf{W}^k = \begin{bmatrix} x^{k,1} & x^{k,n} \\ y^{k,1} & y^{k,n} \end{bmatrix} = \mathbf{R}^k \left(\sum_{i=1}^K l^{k,i} \mathbf{S}^i \right) + \mathbf{t}^k \quad (6.7)$$

où $(x^{k,i}, y^{k,i})^\top$ représente les coordonnées du point d’intérêt i dans l’image k , \mathbf{R}^k est une matrice orthonormale de taille 2×3 (première et seconde ligne de la matrice de rotation) et \mathbf{t}^k correspond aux deux premières composantes du vecteur de translation. En tirant partie des propriétés de rang faible de la matrice \mathbf{W}^k , cette dernière peut être factorisée et les composantes relatives de la forme ainsi que celles relatives à la pose de la caméra peuvent être isolées. Les méthodes basées sur des modèles de rang faible nécessitent un nombre d’images important. Elles sont adaptées à la reconstruction d’objets subissant un ensemble de déformations limitées (visage, ...).

(ii) **Modèle de surfaces localement planes** : [Varol et al., 2009] considèrent un modèle de caméra perspective. La méthode proposée consiste à calculer un ensemble d’homographies entre différentes régions supposées planaires de paires d’observations. La décomposition des homographies permet d’estimer les normales à la surface et ainsi d’effectuer sa reconstruction. La méthode est rendue robuste en utilisant des régions recouvrantes et en appliquant des contraintes de continuité entre régions voisines. Elle souffre cependant de deux inconvénients majeurs. L’utilisation de la contrainte de continuité (lissage) rend la méthode pratiquement instable [Chhatkuli et al., 2014a]. L’estimation d’homographies entre régions de petites tailles est un problème mal conditionné [Lobay and Forsyth, 2006].

[Collins and Bartoli, 2010] considèrent un modèle de projection localement affine (le modèle de caméra perspective est approximé par un modèle orthographique). Considérant la surface déformée localement plane et rigide, la reconstruction 3D de cette dernière correspond à l’estimation d’un ensemble de poses de la caméra. Les auteurs considèrent un modèle de projection localement affine afin de s’affranchir des problèmes de stabilité et de robustesse liés à ce type d’approche (bruit, estimation de l’orientation des normales aux plans, ...). Une solution analytique au problème d’estimation multi poses est formulée (pour un minimum de trois images de la surface déformée). Cette dernière permet l’estimation d’un template (sous forme d’un graphe) pouvant être employé de façon efficace afin d’estimer la déformation correspondant à une nouvelle observation.

[Vicente and Agapito, 2012] formulent *NRSfM* comme la minimisation d’une fonction d’énergie basée sur la relaxation de la contrainte d’inextensibilité formulée dans [Perriollat et al., 2011]. Cette dernière est exprimée par la consistance des distances entre paires de point de la surface déformée (correspondant à chacune des observations). Le problème est non convexe et la solution proposée ne garantit pas d’obtenir un maximum global.

[Chhatkuli et al., 2014a] proposent une solution analytique au problème *NRSfM* formulé comme un système d’équation aux dérivées partielles fonction de la profondeur de la surface et de ses normales. Les auteurs démontrent que ce système ne peut être résolu localement sans *a priori*. Ils intègrent une contrainte de planarité infinitésimale et proposent une solution basée sur une optimisation linéaire convexe. Elle permet d’extraire les normales à la surface en décomposant un ensemble d’homographies. L’ambiguïté de signe relative à cette formulation, et commune aux méthodes *SfT*, est résolue pour un minimum de trois vues. Cette approche est particulièrement rapide et efficace. L’évaluation sur données synthétiques et réelles démontre le gain de précision et de robustesse de cette méthode. Pour un jeu de données contenant dix observations d’une feuille de papier déformée, l’erreur RMS de reconstruction est d’environ 11 degrés (erreur angulaire sur

les normale) et 8 mm (erreur d'estimation de la profondeur). Cette erreur est supérieure à 12 degrés et 19 mm pour [Varol et al., 2009]. Elle est supérieure à 34 degrés et 87 mm pour [Vicente and Agapito, 2012].

6.1.3 Contributions

Nous présentons une étude théorique préliminaire à la formulation d'une méthode *NRSfM* générative. Nous considérons le cas de surfaces subissant un ensemble de déformations isométriques. Les objectifs visés par cette étude sont énumérés dans les points suivants. Nous formulons une approche générative originale permettant la reconstruction de surfaces déformables (*i*). Cette classe d'approches n'a jamais été appliquée au problème *NRSfM* ou *SfT*. La méthode proposée permet par ailleurs de s'affranchir des contraintes d'occultations dont souffrent généralement les approches génératives [Gargallo et al., 2007]. Nous proposons une méthode permettant l'estimation dense et conjointe d'un ensemble de déformations ainsi que de la carte de couleurs du template (*ii*). Elle ne nécessite pas l'extraction et l'appariement de points d'intérêts. Nous intégrons au modèle génératif proposé le modèle de flou défini par les équations (2.12) et (2.13) (*iii*), voir chapitre 2). Cela permet d'estimer une carte de profondeur dense et à l'échelle de la surface déformée.

Nous considérons un système optique monofocale dont les paramètres internes ainsi que les paramètres optiques de mise au point sont connus. Nous disposons d'un ensemble d'observations d'une surface subissant des déformations isométriques. A des fins de clarté, nous considérons la déformation d'une feuille de papier de format *a priori* connu (rapport longueur/largeur). L'image de la feuille non déformée (plane) et fronto parallèle à la caméra représente le template. Cette dernière correspond à l'image de la scène \hat{I} entièrement mise au point. Elle est *a priori* inconnue. Contrairement aux approches discriminatives de l'état de l'art *NRSfM* nous proposons une méthode générative intégrant un modèle de flou optique. Notre objectif étant de réaliser la reconstruction 3D à l'échelle de la surface déformée, il est nécessaire de connaître le ratio du template. Nous considérons également que la position du template dans les images est connue (sa détection au sein de la scène n'est pas considérée).

Nous présenterons dans une première section le modèle de génération d'images proposé. Nous définirons par la suite les éléments du modèle *a priori* connus ainsi que éléments devant être appris *a posteriori*. Un algorithme d'alternation est proposé afin de réaliser cet apprentissage et inférer le modèle.

6.2 Modélisation

Le problème de reconstruction *NRSfM* considéré est illustré dans la figure 6.2.1.

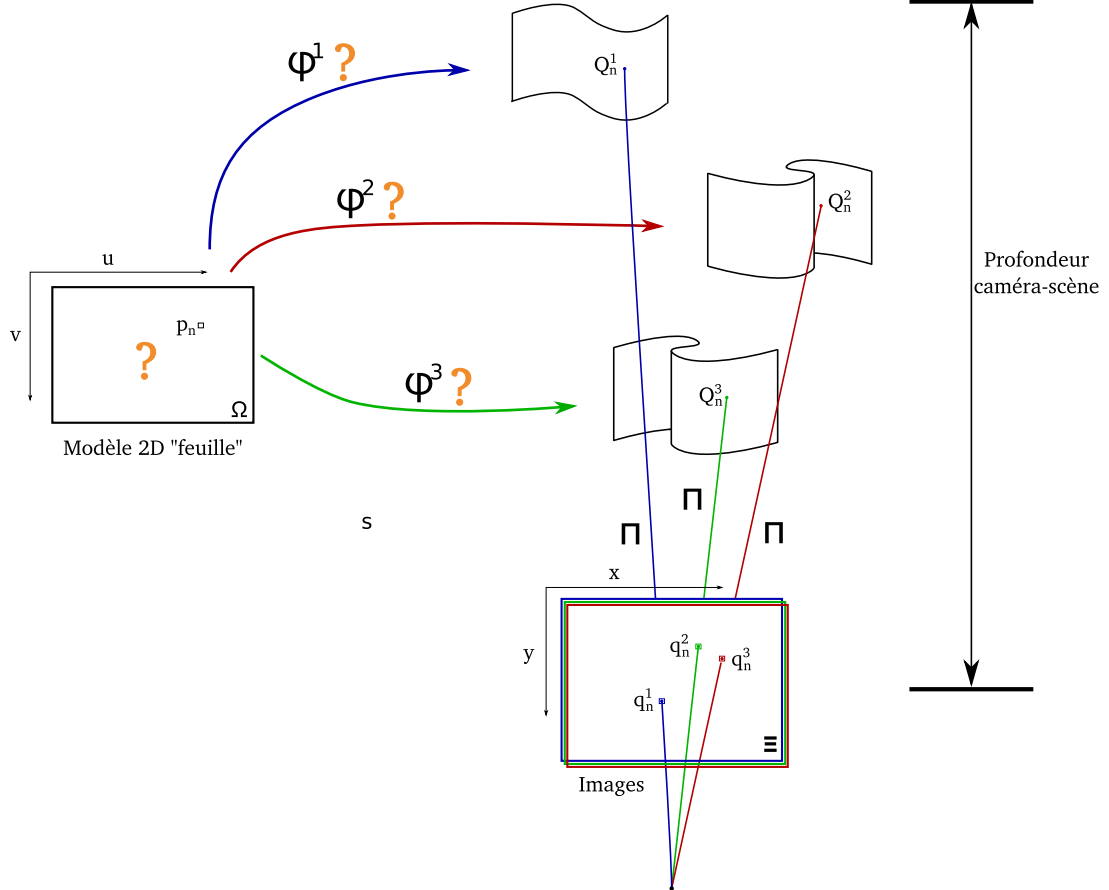


Figure 6.2.1 – Problème de reconstruction *Non-Rigid Structure-from-Motion (NRSfM)*. Un template (modèle de feuille plane) subit un ensemble de déformations isométriques dénotées φ^j . Les surfaces générées sont projetées dans l'image (suivant une projection unique π) formant un ensemble d'observations I^j .

Le template \mathcal{T} est défini dans $\Omega \in \mathbb{R}^2$. Il est modélisé par une grille de taille $N \times M$ avec $\frac{N}{M} = \rho$ fixé. A des fins de clarté nous noterons les pixels du template p_n de coordonnées (u, v) . Le codomaine associé à cette fonction forme la carte de couleurs du template. Cette dernière est *a priori* inconnue. Le template subit un ensemble de déformations isométriques φ^j associant à chaque point p_n le point correspondant $\mathbf{Q}_n^j = \varphi^j(p_n) \in \mathbb{R}^3$. L'observation de la surface déformée est donnée par la projection π dans l'image I^j des points \mathbf{Q}_n^j : $\mathbf{q}_n^j = \pi(\mathbf{Q}_n^j) \in \Xi$ ($\Xi \in \mathbb{R}^2$). Si l'on considère un modèle de caméra sténopé, nous pouvons appliquer à l'ensemble d'images la

transformation K^{-1} (obtenue par calibration des paramètres internes de la caméra). L'opérateur de projection π s'écrit alors simplement : $\pi(\mathbf{Q}_n^j) = \frac{1}{\mathbf{Q}_{n,3}^j}(\mathbf{Q}_{n,1}^j \mathbf{Q}_{n,2}^j)^\top$.

Le processus de génération d'images considéré est exprimé par le réseau Bayésien décrit dans la figure 6.2.2.

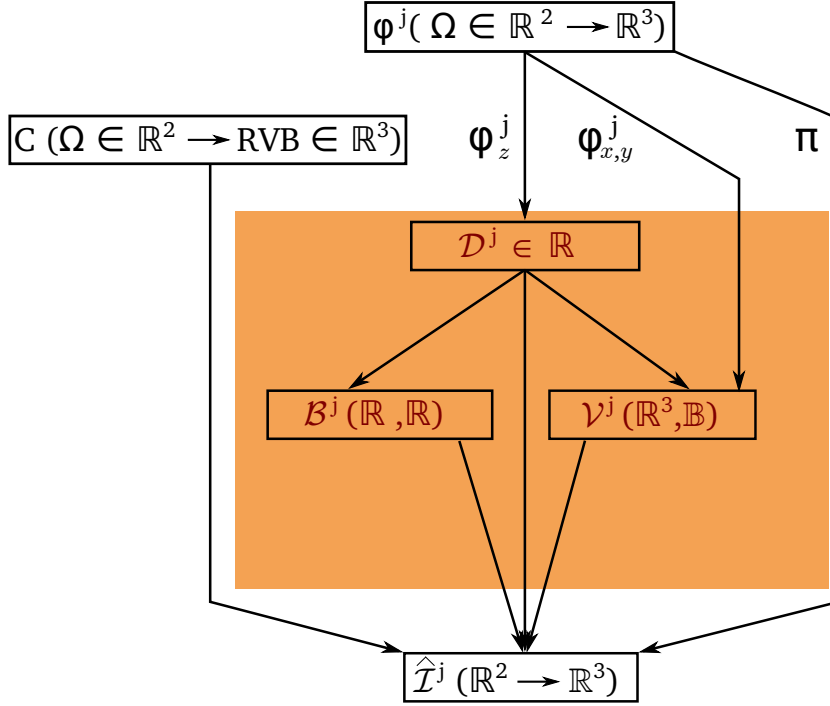


Figure 6.2.2 – Réseau Bayésien modélisant le processus de formation des images. \mathcal{C} et φ représentent respectivement la carte de couleurs du template et la déformation de la surface. Les nœuds situés dans le cadre coloré correspondent aux éléments du réseau définis par calibration ou directement inférés à partir de φ . \mathcal{D} représente la carte de profondeur de la scène, \mathcal{B} et \mathcal{V} représentent respectivement la carte de flou et la carte de visibilité correspondante. L'observation (image) générée par le modèle est dénoté $\hat{\mathcal{I}}$.

Au sein de ce réseau, nous cherchons à estimer la carte de couleurs \mathcal{C} du template ainsi que l'ensemble des déformations isométriques φ^j . Les autres nœuds (encadrés dans la figure) sont directement inférés à l'aide des paramètres de calibration et de la fonction de déformation du template. $\mathcal{D}^j = \varphi_3^j$ représente la carte de profondeur de la surface. Cette dernière permet d'estimer une carte de flou notée \mathcal{B}^j . La carte de flou est spatialement dépendante (dense). Elle définit les paramètres d'étalement de la PSF Gaussienne pour chaque pixel du template. La carte de profondeur permet par ailleurs de définir une carte de visibilité notée \mathcal{V}^j définissant les points de la surface 3D visibles dans l'image générée. Elle est directement obtenue à l'aide d'un algorithme

de Z-Buffer [Catmull, 1974] qui ne sera pas détaillé dans ce manuscrit à des fins de concision. L'image I^j est générée en projetant les points 3D de la surface déformée dans le repère caméra puis en leurs associant les couleurs obtenues par convolution de la carte de couleurs du template et de la carte de flou Gaussien spatialement dépendante (équation 2.13).

Les paramètres du système optique sont nécessaires à l'inférence du modèle. La calibration des paramètres internes K est obtenue à l'aide de [Bouguet, 2008]. Les paramètres de mise au point sont obtenus à l'aide d'une calibration de type DfF . Comme nous l'avons présenté au chapitre 3, elle consiste à établir une table de correspondance liant la profondeur de la prise de vue à l'écart type de la PSF. Dans le cadre de la méthode générative proposée, une calibration fine est requise (à l'aide d'un système mécanique de translation millimétrique par exemple).

La déformation φ est exprimée dans cette étude par un B-Spline bi-cubique. Nous définissons dans ce manuscrit (et par abus de langage) une B-Spline (*Basis-Spline*) comme une combinaison linéaire de fonctions polynomiales par morceaux (formant une base de polynôme). Le degré b de la B-Spline (d'ordre $b + 1$) correspond au degré des polynômes. Les intervalles de la B-Spline sont définis par un vecteur de nœuds $k = (k_0, \dots, k_m)$ avec $k = k_0 < \dots < k_i < \dots < k_m$ correspondant à des échantillons de la fonction interpolée (dans notre cas la fonction de déformation). Chaque polynôme interpole $d + 1$ nœuds. Ils sont définis par l'expression récursive suivante :

$$N_{i,0}(x) = \begin{cases} 1 & \text{si } x \in k_i \leq x < k_{i+1} \\ 0 & \text{otherwise} \end{cases} \quad (6.8)$$

$$N_{i,j}(x) = \frac{x - k_i}{k_{i+b} - k_i} N_{i,b-1}(x) + \frac{k_{i+b+1} - x}{k_{i+b+1} - k_{i+1}} N_{i+1,b-1}(x)$$

La fonction B-Spline, paramétrée par l points de contrôle $w = (x_1; y_1; z_1 \dots x_l; y_l; z_l)$, est définie par :

$$B(x) = \sum_{i=0}^l w_i N_{i,n}(x) \quad (6.9)$$

Le choix de cette fonction de déformation est motivé par la simplicité de son expression et de ses dérivées partielles (en fonction des points de contrôle notamment). Une étude exhaustive des B-Spline et de leurs propriétés est présentée dans [Brunet, 2010].

6.3 *A priori* du modèle

Nous avons défini que les *a priori* du modèle suivent la loi de distribution d'un champ de Markov aléatoire [Gargallo et al., 2007]. Elle facilite notamment l'expression de l'inférence du modèle. Elle est exprimée par :

$$p(X = x) = \exp\left(-\sum_k \omega(k)\right) \quad (6.10)$$

où $\omega(k)$ exprime une fonction de coût relative à un nœud du modèle.

carte de couleurs \mathcal{C} du template : Contrairement à la majorité des approches génératives de reconstruction 3D, nous ne définissons aucun *a priori* relatif au voisinage d'un pixel du template (par exemple deux pixels voisins de profondeur identique ont une probabilité importante de posséder la même couleur). Le modèle que nous proposons dissocie les *a priori* liés à la carte de couleurs \mathcal{C} et ceux liés aux déformations φ^j . Pour chaque pixel du template, toutes les combinaisons de couleurs sont équiprobables :

$$p(C_{color(p_n)=r,g,b}) = \left(\frac{1}{256}\right)^3 \quad (6.11)$$

où r, g, b représentent des valeurs d'intensité pour les trois canaux couleur.

Nous prévoyons que la carte de couleurs estimée correspond à l'image entièrement nette de la surface. La définition d'un *a priori* de netteté est complexe car la majorité des opérateurs de mesure de flou optique [Pertuz et al., 2013a] s'appuient sur l'utilisation de filtres dont la dérivée est nulle (et ne contraint pas l'optimisation). La contrainte de netteté est conjointement imposée par l'intégration du modèle de flou au modèle génératif.

Déformation isométrique φ : L'approche proposée étant générative, il est nécessaire de définir la position de la surface déformée dans les observations. Nous considérons que la position des angles de la feuille est connue. Nous définissons conjointement l'*a priori* (φ_c^j) afin que les angles du template déformé soient projetés sur les positions correspondantes (pour chaque image j de l'ensemble d'observation) :

$$p(\varphi c^j) = \exp - \underbrace{\sum_i^{1,\dots,4} dist \left(\left[\begin{array}{c} \varphi_1^j(\mathcal{T}_c^{j,i}) \\ \varphi_3^j(\mathcal{T}_c^{j,i}) \end{array} \right], I_c^{j,i} \right)}_{\varphi c^j} \quad (6.12)$$

où φ_n^j est la fonction de déformation correspondant à l'image j appliquée à la $n^{\text{ième}}$ coordonnée du template, $\mathcal{T}_c^{j,i}$ définit les coordonnées du coin i du template et $I_c^{j,i}$ définit les coordonnées de ce même coin dans l'observation j . $dist$ est la fonction de distance définie par $dist(\mathbf{q}, \mathbf{q}') = \sqrt{(\mathbf{q}_1 \sim \mathbf{q}'_1)^2 + (\mathbf{q}_2 \sim \mathbf{q}'_2)^2}$.

La contrainte d'isométrique dénotée φ^j est exprimée à l'aide de la matrice Jacobienne de φ [Bartoli et al., 2012] :

$$p(\varphi^j) = \exp - \underbrace{\iint_{(u,v)} \|\mathcal{J}_{\varphi^j}^\top(u,v) \mathcal{J}_{\varphi^j}(u,v) - \lambda Id\|^2}_{\varphi^j} du dv \quad (6.13)$$

avec Id de taille 3×3 . Nous considérons dans cette expression la norme matricielle de Frobenius définie par :

$$\|\mathbf{A}\|^2 = \sqrt{\sum_{i=1}^m \sum_{j=1}^n |a_{ij}|^2} \quad (6.14)$$

où a_{ij} est l'élément d'indice (i, j) de la matrice $\mathbf{A}_m \times n$.

Il est important de noter que la taille réelle du template n'est pas connue. Nous introduisons ainsi un facteur d'échelle λ commun à l'ensemble des observations considérées.

Nous considérons enfin le cas de déformations régulières (le template déformé ne présente pas de discontinuités topologiques). Cet *a priori* de lissage dénoté φ^j est exprimé par l'équation suivante (également basée sur la norme de Frobenius) :

$$p(\varphi^j) = \exp - \underbrace{\iint_{(u,v)} \|\mathcal{H}_{\varphi^j}(u,v) - O\|^2}_{\varphi^j} du dv \quad (6.15)$$

où $\mathcal{H}_{\varphi^j}(u,v)$ correspond à la matrice des dérivées seconde de la fonction de déformation φ^j et $O_{3 \times 3}$ représente la matrice nulle. En combinant les précédentes expressions (équations (6.12), (6.13), (6.15)), nous formons l'*a priori* de déformation suivant :

$$p(\varphi) = \sum_j^{1, \dots, N} p(\varphi^c^j) p(\varphi^i^j) p(\varphi^s^j) \quad (6.16)$$

6.4 Fonction de vraisemblance

La fonction de vraisemblance, pour chaque image de l'ensemble d'observation, correspond à la probabilité de générer l'image I^j considérant les paramètres φ^j et \mathcal{C} courant du modèle. Le processus de génération des images n'étant pas directement réversible, il est difficile d'exprimer la fonction de vraisemblance dans le domaine de l'image. Cette dernière est exprimée dans le domaine du template et pour chaque image de l'ensemble d'observation par l'équation suivante :

$$p(I|\varphi^j, \mathcal{C}) = \exp - \underbrace{\iint_{u,v} \Xi \|(\mathcal{C} * g(\mu, \sigma(\varphi_3^j(u, v))))(u, v) - I((\pi \circ \varphi^j)(u, v))\|^2 du dv}_{\overline{\mathcal{T} \ell^i}} \quad (6.17)$$

où Ξ est un masque binaire correspondant aux pixels du template réellement projetés dans l'image. Il est calculé grâce à un algorithme de Z-buffer. g est une PSF Gaussienne spatialement variante de moyenne μ et d'écart type $\sigma(\varphi_3(u, v))$.

6.5 Inférence du modèle

Nous cherchons à maximiser la probabilité jointe de la déformation φ et de la carte de couleurs \mathcal{C} considérant un ensemble d'observation I^j , $j \in (1, \dots, n)$:

$$\operatorname{argmax}_{\varphi^1, \dots, \varphi^n, \mathcal{C}} \mathcal{L}(\varphi^1, \dots, \varphi^n, \mathcal{C}; I^1, \dots, I^n) = \operatorname{argmax}_{\varphi^1, \dots, \varphi^n, \mathcal{C}} p(\varphi^1, \dots, \varphi^n, \mathcal{C} | I^1, \dots, I^n) \quad (6.18)$$

Afin de simplifier les notations nous désignerons par I l'ensemble (I^1, \dots, I^n) et φ l'ensemble des transformations $(\varphi^1, \dots, \varphi^n)$. En appliquant la loi de Bayes, cette équation peut être reformulée par :

$$\operatorname{argmax}_{\varphi, \mathcal{C}} p(\varphi, \mathcal{C} | I) = \operatorname{argmax}_{\varphi, \mathcal{C}} \frac{p(I|\varphi, \mathcal{C})p(\varphi, \mathcal{C})}{p(I)} \quad (6.19)$$

Les probabilités φ et \mathcal{C} étant disjointes, l'équation 6.19 devient :

$$\operatorname{argmax}_{\varphi, \mathcal{C}} p(\varphi, \mathcal{C} | I) = \operatorname{argmax}_{\varphi, \mathcal{C}} \frac{p(I|\varphi, \mathcal{C})p(\varphi)p(\mathcal{C})}{p(I)} \quad (6.20)$$

Comme nous cherchons à maximiser la fonction de vraisemblance, le dénominateur $p(I)$ peut être supprimé de l'équation (6.20). L'optimisation est réalisée à l'aide d'un algorithme d'alternation

défini dans la section suivante. Ce dernier consiste à réaliser alternativement la maximisation de la fonction de vraisemblance considérant les paramètres de déformation fixes et la maximisation de la fonction de log vraisemblance considérant la carte de couleurs fixe.

6.5.1 Maximisation de la fonction de vraisemblance, estimation de la carte de couleurs

Cette étape consiste à maximiser la fonction de vraisemblance $\mathcal{L}(\varphi, \mathcal{C}; I)$ en considérant l'estimation φ courante fixe.

$$\operatorname{argmax}_{\mathcal{C}} p(\varphi, \mathcal{C}|I) = p(I|\varphi, \mathcal{C}) p(\varphi) p(\mathcal{C}) \quad (6.21)$$

Le problème d'optimisation est résolu en utilisant un algorithme de descente de gradient [Martínez, 1994]. L'expression de la dérivée de la fonction de vraisemblance en fonction de la carte de couleurs est donnée par :

$$\frac{\partial p(\varphi, \mathcal{C}|I)}{\partial \mathcal{C}} = \frac{\partial (\overline{\varphi_c} + \overline{\varphi_s} + \overline{\varphi_i} + \overline{\mathcal{T}\ell})}{\partial \mathcal{C}} \exp(\overline{\varphi_c} + \overline{\varphi_s} + \overline{\varphi_i} + \overline{\mathcal{T}\ell}) \quad (6.22)$$

Pour cette étape de l'alternation, $\overline{\varphi_c}$, $\overline{\varphi_s}$ et $\overline{\varphi_i}$ sont nulles. L'équation 6.22 devient :

$$\frac{\partial p(\varphi, \mathcal{C}|I)}{\partial \mathcal{C}} = \frac{\partial \overline{\mathcal{T}\ell}}{\partial \mathcal{C}} \exp(\overline{\varphi_c} + \overline{\varphi_s} + \overline{\varphi_i} + \overline{\mathcal{T}\ell}) \quad (6.23)$$

Pour chaque pixel du template, le flou Gaussien (écart type de la PSF) dépendant de φ_3 est pré calculé. Lors de la phase d'alternation, la somme des contributions permet d'obtenir efficacement la dérivée partielle de la fonction de vraisemblance :

$$\frac{\partial \overline{\mathcal{T}\ell}}{\partial \mathcal{C}} = -\Xi \iint_{u,v} \frac{\alpha(u, v) \sum_{R,G,B} (\mathcal{C} * g(\mu, \sigma(\varphi_3(u, v))))(u, v) - I((P \circ \varphi)(u, v))}{\sqrt{\sum_{R,G,B} (\mathcal{C} * g(\mu, \sigma(\varphi_3(u, v))))(u, v) - I((P \circ \varphi)(u, v))^2}} du dv \quad (6.24)$$

où $\alpha(u, v)$ correspond à la dérivée de $(\mathcal{C} * g(\mu, \sigma(\varphi_3(u, v))))(u, v)$ qui est constante pour cette étape de l'algorithme d'alternation.

6.5.2 Maximisation de la fonction de log-vraisemblance, estimation de la déformation

Nous supposons la carte de couleurs \mathcal{C} courante fixe. Nous cherchons alors à maximiser la fonction de log vraisemblance dépendante des paramètre $w\ell$ de la déformation φ :

$$\operatorname{argmax}_{\varphi} \mathcal{L}(p(\varphi, \mathcal{C}|I) = \ln(p(I|\varphi, \mathcal{C})) + \ln(p(\varphi)) + \ln(p(\mathcal{C})) \quad (6.25)$$

En développant cette équation, nous obtenons :

$$\operatorname{argmax}_{\varphi} \mathcal{L}(p(\varphi, \mathcal{C}|I) = \overline{\varphi c} + \overline{\varphi s} + \overline{\varphi i} + \overline{\mathcal{T}\ell} \quad (6.26)$$

L'optimisation est réalisée à l'aide d'une descente de gradient. Les dérivées partielles de l'équation (6.26) en fonction des points de contrôle $w\ell = (w\ell^1, \dots, w\ell^m)^\top$ peuvent être calculées à l'aide des expressions suivantes :

$$\frac{\partial \mathcal{L}(p(\varphi, \mathcal{C}|I))}{\partial w\ell^j} = \frac{\partial \overline{\varphi c}}{\partial w\ell^j} + \frac{\partial \overline{\varphi s}}{\partial w\ell^j} + \frac{\partial \overline{\varphi i}}{\partial w\ell^j} + \frac{\partial \overline{\mathcal{T}\ell}}{\partial w\ell^j} \quad (6.27)$$

Les dérivées partielles de la fonction B-spline bi-cubique sont obtenues à l'aide des formules suivantes (Brunet [2010]) :

$$\begin{aligned} \frac{\partial \overline{\mathcal{T}\ell}}{\partial w\ell^j} = & - \iint_{(u,v)} \frac{\frac{\partial((\mathcal{C}*g(\mu, \sigma(\varphi_3(u,v))))(u,v) - I((P \circ \varphi)(u,v))}{\partial w\ell^j}}{\sqrt{((\mathcal{C}*g(\mu, \sigma(\varphi_3(u,v))))(u,v) - I((P \circ \varphi)(u,v)))^2}} du dv \\ & \times (((\mathcal{C}*g(\mu, \sigma(\varphi_3(u,v))))(u,v) - I((P \circ \varphi)(u,v))) \end{aligned} \quad (6.28)$$

Les dérivées partielles correspondant à $w\ell_1^j$ et $w\ell_2^j$ étant nulles $\left(\frac{\partial((\mathcal{C}*g(\mu, \sigma(\varphi_3(u,v))))(u,v)}{\partial w\ell_{1,2}^j} = 0 \right)$, nous obtenons :

$$\frac{\partial \overline{\varphi c}}{\partial w\ell^j} = - \sum_i^{1, \dots, 4} \frac{\frac{\partial \varphi_1^i}{\partial w\ell^j} \varphi_3^i - \varphi_1^i \frac{\partial \varphi_3^i}{\partial w\ell^j}}{(\varphi_3^i)^2} \left(\frac{\varphi_1^i}{\varphi_3^i} - I_{c,1}^i \right) + \frac{\frac{\partial \varphi_2^i}{\partial w\ell^j} \varphi_3^i - \varphi_2^i \frac{\partial \varphi_3^i}{\partial w\ell^j}}{(\varphi_3^i)^2} \left(\frac{\varphi_2^i}{\varphi_3^i} - I_{c,2}^i \right) \sqrt{\left(\frac{\varphi_1^i}{\varphi_3^i} - I_{c,1}^i \right)^2 + \left(\frac{\varphi_2^i}{\varphi_3^i} - I_{c,2}^i \right)^2} \quad (6.29)$$

$$\frac{\partial \overline{\varphi i}}{\partial w\ell^j} = -2 \iint_{(u,v)} A + 2B + D du dv \quad (6.30)$$

où

$$A = \left[\frac{\partial(\frac{\partial\varphi_1}{\partial u})}{\partial w^l j} \frac{\partial\varphi_1}{\partial u} + \frac{\partial(\frac{\partial\varphi_2}{\partial u})}{\partial w^l j} \frac{\partial\varphi_2}{\partial u} \right] \left[\frac{\partial\varphi_1^2}{\partial u} + \frac{\partial\varphi_2^2}{\partial u} - 1 \right] \quad (6.31)$$

$$B = \left[\frac{\partial(\frac{\partial\varphi_1}{\partial u})}{\partial w^l j} \frac{\partial\varphi_1}{\partial v} + \frac{\partial\varphi_1}{\partial u} \frac{\partial(\frac{\partial\varphi_1}{\partial v})}{\partial w^l j} + \frac{\partial(\frac{\partial\varphi_2}{\partial u})}{\partial w^l j} \frac{\partial\varphi_2}{\partial v} + \frac{\partial\varphi_2}{\partial u} \frac{\partial(\frac{\partial\varphi_2}{\partial v})}{\partial w^l j} \right] \left[\frac{\partial\varphi_1}{\partial u} \frac{\partial\varphi_1}{\partial v} + \frac{\partial\varphi_2}{\partial u} \frac{\partial\varphi_2}{\partial v} \right] \quad (6.32)$$

$$D = \left[\frac{\partial(\frac{\partial\varphi_1}{\partial v})}{\partial w^l j} \frac{\partial\varphi_1}{\partial v} + \frac{\partial(\frac{\partial\varphi_2}{\partial v})}{\partial w^l j} \frac{\partial\varphi_2}{\partial v} \right] \left[\frac{\partial\varphi_1^2}{\partial v} + \frac{\partial\varphi_2^2}{\partial v} - 1 \right] \quad (6.33)$$

$$\begin{aligned} \frac{\partial\overline{\varphi_s}}{\partial\varphi_{w_j}} = -2 \iint_{(u,v)} & \frac{\partial(\frac{\partial^2\varphi_1}{\partial u^2})}{\partial w^l j} \frac{\partial^2\varphi_1}{\partial u^2} + \frac{\partial(\frac{\partial^2\varphi_1}{\partial u\partial v})}{\partial w^l j} \frac{\partial^2\varphi_1}{\partial u\partial v} + \frac{\partial(\frac{\partial^2\varphi_1}{\partial v^2})}{\partial w^l j} \frac{\partial^2\varphi_1}{\partial v^2} + \\ & \frac{\partial(\frac{\partial^2\varphi_2}{\partial u^2})}{\partial w^l j} \frac{\partial^2\varphi_2}{\partial u^2} + \frac{\partial(\frac{\partial^2\varphi_2}{\partial u\partial v})}{\partial w^l j} \frac{\partial^2\varphi_2}{\partial u\partial v} + \frac{\partial(\frac{\partial^2\varphi_2}{\partial v^2})}{\partial w^l j} \frac{\partial^2\varphi_2}{\partial v^2} + \\ & \frac{\partial(\frac{\partial^2\varphi_3}{\partial u^2})}{\partial w^l j} \frac{\partial^2\varphi_3}{\partial u^2} + \frac{\partial(\frac{\partial^2\varphi_3}{\partial u\partial v})}{\partial w^l j} \frac{\partial^2\varphi_3}{\partial u\partial v} + \frac{\partial(\frac{\partial^2\varphi_3}{\partial v^2})}{\partial w^l j} \frac{\partial^2\varphi_3}{\partial v^2} du dv \end{aligned} \quad (6.34)$$

6.6 Conclusion

L'approche générative *NRSfM* présentée dans ce manuscrit est destinée à la reconstruction 3D, à l'échelle, d'une surface subissant des déformations isométriques. Cette classe de méthode nécessite une initialisation proche de l'optimum global recherché. Considérant l'état de l'art, elle peut notamment être réalisée à l'aide de la méthode proposée par [Chhatkuli et al., 2014a].

Contrairement au système de mesure par extraction du point de netteté présenté dans la seconde partie de ce manuscrit, nous proposons une approche de reconstruction 3D dense basée sur un modèle génératif. Elle permet d'estimer des mesures géodésiques sur la surface déformée. La carte de couleurs du template est conjointement estimée. Elle correspond à l'image aplanie de la surface lorsque celle-ci est entièrement nette. Ceci permet de réaliser la mise au point de la surface déformée *a posteriori*. L'avantage principal de cette approche est la possibilité de générer efficacement de nouvelles observations de la scène.

Une évaluation statistique est actuellement réalisée. Il est envisagé de compléter cette dernière par une étude sur des données issues d'examen endoscopiques tels que la laparoscopie (examen de l'abdomen) ou l'environnement exploré est soumis à de nombreuses déformations (liées à l'ac-

tion du praticien ou aux déformations naturelles des organes). La validation de cette approche permettrait de proposer un système de mesure et de reconstruction 3D dense de surfaces déformables adapté à la vision monofocale. Ce dernier nécessiterai alors d'être combiné à une approche de segmentation adaptée (permettant de dissocier les différents organes) afin d'être étendu à la reconstruction de scènes.

Conclusion

Les systèmes de métrologie basés vision sont généralement équipés de dispositifs optiques stéréo (ou multi vues) ou de dispositifs de mesure actifs. Ils permettent de résoudre l’ambiguïté d’échelle dont souffrent la majorité des méthodes de reconstruction tridimensionnelles. Cette dernière est inhérente au processus d’acquisition d’image et provient de la relation équivoque entre la profondeur de la scène imagée et la taille des objets la composant. Les approches *Depth-from-Focus (DfF)* et *Depth-from-Defocus (DfD)*, reposant sur l’analyse du flou optique, permettent de résoudre l’ambiguïté d’échelle. Elles s’appuient sur les lois de l’optique géométrique (et ondulatoire) qui mettent en évidence la relation liant le degré de mise au point d’un objet à la profondeur de la prise de vue. Ces approches nécessitent de contrôler les paramètres internes de la camera (distance focale, ouverture, . . .) ou son déplacement.

De nombreux dispositifs de vision sont équipés d’un système optique monofocale pour lesquelles les paramètres internes sont fixes. La modification de ces systèmes ou le contrôle de leurs déplacements est complexe et peut ne pas satisfaire aux contraintes du contexte applicatif visé. Ceci est notamment le cas de la coloscopie. Cet examen endoscopique permet l’exploration du côlon à l’aide d’un endoscope flexible appelé coloscope. La taille des néoplasies (excroissances anormales de tissu) est un critère essentiel à leurs diagnostics et permet de définir des intervalles de surveillance appropriés. [Chaptini et al., 2014] révèlent que l’erreur visuelle d’estimation de la taille des néoplasies est supérieur à 20% dans 52% des cas (2812 cas étudiés). Cela conduit dans 10% des cas à la définition d’intervalles de temps de surveillance inappropriés.

Nous proposons un système de métrologie basé vision adapté à un système optique monofocale. Ce système passif repose sur l’extraction du point de rupture de netteté (nommé *IB (Infocus Breakpoint)*) au sein d’une vidéo correspondant à un mouvement d’approche de vitesse approximativement constante de la camera vis à vis d’un objet d’intérêt supposé planaire. Les paramètres optiques de la camera étant fixes, l’IB correspond à une profondeur d’acquisition singulière pouvant être calibrée. Le système proposé se compose de deux modules nommés *BET (Blur Estimating Tracking)* et *BMF (Blur Model Fitting)*. Le premier module permet d’effectuer le suivi et la mesure conjointe du degré de mise au point de l’objet d’intérêt. Il est robuste aux déformations de l’environnement colonique ainsi qu’aux artefacts communs rencontrés en coloscopie (flous de mouvements ponctuels, présence de liquide sur l’optique, . . .). Le second module permet l’extraction robuste de l’IB grâce à l’ajustement d’un modèle de flou optique aux mesures réalisées par le module BET. L’évaluation statistique (données synthétiques) et pratique (données fantômes et réelles issues de coloscopies) démontre la faisabilité de l’approche proposée. Le protocole clinique de mesure a pu être réalisé sans difficultés par les gastroentérologues impliqués dans notre étude. L’erreur de mesure obtenue est inférieure au millimètre. La méthode de mesure

par extraction du point de rupture de netteté est particulièrement adaptée à la coloscopie. Les lésions coloniques étant faiblement texturées, il est nécessaire de réaliser une mesure de flou base région. Cette méthode n'ai pas adaptée à la reconstruction dense et à l'échelle de surfaces dans ce contexte.

Nous proposons par ailleurs, sous la forme d'une étude préliminaire, une méthode de reconstruction *NRSfM* (*Non-Rigid Structure-from-Motion*) générative. Cette dernière a pour objectif de s'affranchir des limitations imposées par la méthode discriminative de mesure par extraction du point de rupture de netteté. Elle repose sur l'acquisition d'un ensemble d'images de la surface déformée. Par ailleurs, elle permet la reconstruction dense de scènes déformables et ainsi une réelle mesure 3D de celle ci. Elle permet conjointement d'estimer la carte de couleur du template correspondant à l'image de la surface aplanie et entièrement mise au point. Différents points de vues de la scène peuvent être générés ainsi que la mise au point *a posteriori*. Elle nécessite néanmoins de posséder une approximation initiale de la surface proche de la solution finale. Cette dernière peut notamment être obtenue grâce à la méthode proposée par [Chhatkuli et al., 2014a]. Une évaluation statistique et pratique est actuellement réalisée afin de valider cette nouvelle approche *NRSfM*. Il est également envisagé de réaliser cette évaluation sur des données issues d'examens laparoscopiques (examen endoscopique permettant l'exploration de l'abdomen).

Différentes perspectives d'extension du système de mesure par extraction du point de rupture de netteté peuvent être envisagées :

- (i) Afin d'améliorer la robustesse du système de mesure, il est nécessaire de considérer la présence de flou de mouvement. Celui-ci doit être détecté afin d'ignorer ou de traiter les images concernées. De nombreux articles tels que [Ma et al., 2015] s'appuient sur l'analyse des gradients de l'image ou de la déformation des contours afin de détecter l'information de flou directionnelle. L'image est généralement restaurée en s'appuyant sur des *a priori* de netteté. Il est ainsi difficile de dissocier l'information de flou de mouvement de l'information de flou optique qui est dégradée par l'opération de déconvolution. Il peut néanmoins être envisagé de tirer parti de l'information de flou de mouvement afin d'estimer le déplacement de la caméra. Le modèle de flou optique peut également être utilisé afin de restaurer de façon cohérente les images affectées par ces artefacts au sein de la video.
- (ii) Afin de s'affranchir de la contrainte de mouvement approximativement uniforme de la caméra, il est nécessaire d'intégrer au modèle de flou optique une fonction de déplacement non constante. Cette dernière peut être estimée en considérant, par exemple, le facteur de

changement d'échelle des objets au sein de la vidéo.

- (iii) En s'appuyant notamment sur l'information de flou optique, il est possible de réorganiser de façon cohérente les images au sein d'une vidéo afin de prendre en considération un mouvement aléatoire de la caméra. Cette hypothèse suppose néanmoins que le point de rupture de netteté est observable au sein de la séquence.
- (iv) Considérant le cas d'un système optique monofocale, nous nous sommes appuyé sur la redondance de l'information de flou au sein d'une vidéo afin d'extraire de façon robuste le point de rupture de netteté. Cependant, les récents équipements coloscopiques proposent deux modes de visualisation correspondant à deux distances de focalisation différentes (système *dual focus*). Il peut ainsi être envisagé de s'appuyer sur le module BET afin d'estimer, à partir de deux acquisitions réalisées suivant chacun de ces modes, la variation de flou optique et ainsi inférer l'échelle de la scène. Différentes techniques *DfD* peuvent également être envisagées dans ce contexte.

- Adelson, E., Wang, J., 1992. Single Lens Stereo with a Plenoptic Camera. *IEEE Transaction on Pattern Analysis and Machine Intelligence* 14, 99–106. doi :[10.1109/34.121783](https://doi.org/10.1109/34.121783).
- Ahmad, M., Choi, T., 2007. Application of three dimensional shape from image focus in LCD/TFT displays manufacturing. *IEEE Transactions on Consumer Electronics* 53, 1–4. doi :[10.1109/TCE.2007.339492](https://doi.org/10.1109/TCE.2007.339492).
- Ahmad, M., Choi, T., 2008. Optimization technique for three-dimensional shape recovery from image focus. *SPIE Journal of Electronic Imaging* 17, 1–12. doi :[10.1117/1.3013460](https://doi.org/10.1117/1.3013460).
- Alcantarilla, P., Bartoli, A., Chadebecq, F., Tilmant, C., Lepilliez, V., 2013. Enhanced imaging colonoscopy facilitates dense motion-based 3D reconstruction, in : *IEEE Conference of the Engineering in Medicine and Biology Society, IEEE*. pp. 7346–7349. doi :[10.1109/EMBC.2013.6611255](https://doi.org/10.1109/EMBC.2013.6611255).
- An, Y., Kang, G., Kim, I., Chung, H., Park, J., 2008. Shape from focus through laplacian using 3D window, in : *IEEE Conference on Future Generation Communication and Networking, IEEE*. pp. 46–50. doi :[10.1109/FGCN.2008.139](https://doi.org/10.1109/FGCN.2008.139).
- Anish, A., Jebaseeli, T., 2012. A Survey on Multi-Focus Image Fusion Methods. *International Journal of Advanced Research in Computer Engineering & Technology* 1, 319–324.
- Asif, M., Malik, A., C., T.S., 2005. 3D shape recovery from image defocus using wavelet analysis, in : *IEEE International Conference on Image Processing, IEEE*. pp. 1025–1033. doi :[10.1109/ICIP.2005.1529928](https://doi.org/10.1109/ICIP.2005.1529928).
- Bartoli, A., Collins, T., 2013. Template-based isometric deformable 3D reconstruction with sampling-based focal length self-calibration, in : *IEEE Conference on Computer Vision and Pattern Recognition, IEEE*. pp. 1514–1521. doi :[10.1109/CVPR.2013.199](https://doi.org/10.1109/CVPR.2013.199).
- Bartoli, A., Gay-Bellile, V., Castellani, U., Peyras, J., Olsen, S., Sayd, P., 2008. Coarse-to-fine low-rank structure-from-motion, in : *IEEE Conference on Computer Vision and Pattern Recognition, IEEE*. pp. 1–8. doi :[10.1109/CVPR.2008.4587694](https://doi.org/10.1109/CVPR.2008.4587694).
- Bartoli, A., Gérard, Y., Chadebecq, F., Collins, T., 2012. On template-based reconstruction from a single view : Analytical solutions and proofs of well-posedness for developable, isometric and conformal surfaces, in : *IEEE Conference on Computer Vision and Pattern Recognition, IEEE*. pp. 2026–2033. doi :[10.1109/CVPR.2012.6247906](https://doi.org/10.1109/CVPR.2012.6247906).
- Bartoli, A., Gérard, Y., Chadebecq, F., Collins, T., Pizarro, D., 2015. Shape-from-Template. *IEEE Transactions on Pattern Analysis and Machine Intelligence* PP, 1–39. doi :[10.1109/TPAMI.2015.2392759](https://doi.org/10.1109/TPAMI.2015.2392759).
- Bartoli, A., Perriollat, M., Chambon, S., 2010. Generalized Thin-Plate Spline Warps. *IEEE Conference on Computer Vision and Pattern Recognition* , 1–8doi :[10.1109/CVPR.2007.382998](https://doi.org/10.1109/CVPR.2007.382998).

- Bartoli, A., Pizarro, D., Collins, T., 2013. A Robust Analytical Solution to Isometric Shape-from-Template with Focal Length Calibration, in : IEEE International Conference on Computer Vision, IEEE. pp. 961–968. doi :[10.1109/ICCV.2013.123](https://doi.org/10.1109/ICCV.2013.123).
- Ben-Ari, R., 2014. A Unified Approach for Registration and Depth in Depth-from-Defocus. IEEE Transactions on Pattern Analysis and Machine Intelligence , 1041–1055doi :[10.1109/TPAMI.2014.14](https://doi.org/10.1109/TPAMI.2014.14).
- Bhasin, S., Chaudhuri, S., 2001. Depth-from-Defocus in presence of partial self occlusion, in : IEEE Conference on Computer Vision, IEEE. pp. 488–493. doi :[10.1109/ICCV.2001.937556](https://doi.org/10.1109/ICCV.2001.937556).
- Bhat, K., Twigg, C., Hodgins, J., Khosla, P., Popović, Z., Seitz, S., 2003. Estimating Cloth Simulation Parameters from Video, in : ACM Eurographics SIGGRAPH Symposium on Computer Animation, Eurographics Association. pp. 37–51.
- Bonnans, J., Gilbert, J., Lemaréchal, C., Sagastizábal, C., 2006. Numerical optimization : theoretical and practical aspects. Springer Science & Business Media.
- Bookstein, F.L., 1989. Principal warps : thin-plate splines and the decomposition of deformations. IEEE Transactions on Pattern Analysis and Machine Intelligence 11, 567–585. doi :[10.1109/34.24792](https://doi.org/10.1109/34.24792).
- Born, M., Wolf, E., 1999. Principles of optics : electromagnetic theory of propagation, interference and diffraction of light 7th Edition. Cambridge University Press.
- Bouguet, J.Y., 2008. Camera calibration toolbox for matlab. - URL : http://www.vision.caltech.edu/bouguetj/calib_doc/.
- Boyd, S., Vandenberghe, L., 2004. Convex optimization. Cambridge University Press .
- Brand, M., 2001. Morphable 3D models from video, in : IEEE Conference on Computer Vision and Pattern Recognition, IEEE. pp. 456–463. doi :[10.1109/CVPR.2001.990997](https://doi.org/10.1109/CVPR.2001.990997).
- Brauer-Burchardt, C., Heinze, M., Munkelt, C., Kühmstedt, P., Notni, G., 2006. Distance Dependent Lens Distortion Variation in 3D Measuring Systems Using Fringe Projection, in : British Machine Vision Conference, BMVA Press. pp. 34–44. doi :[10.5244/C.20.34](https://doi.org/10.5244/C.20.34).
- Bregler, C., Hertzmann, A., Biermann, H., 2000. Recovering non-rigid 3D shape from image streams, in : IEEE Conference on Computer Vision and Pattern Recognition, IEEE. pp. 690–696. doi :[10.1109/CVPR.2000.854941](https://doi.org/10.1109/CVPR.2000.854941).
- Britton, D., 2006. Generalized Gaussian Decompositions for Image Analysis and Synthesis. Ph.D. thesis. Georgia Institute of Technology.
- Brunet, F., 2010. Contributions to Parametric Image Registration and 3D Surface Reconstruction. Ph.D. thesis. Université d’Auvergne, Technische Universität Munchen.

- Brunet, F., Bartoli, A., Hartley, R., 2014. Monocular template-based 3D surface reconstruction : Convex inextensible and nonconvex isometric methods. *Computer Vision and Image Understanding* 125, 138–154. doi :[2014.04.003](#).
- Catmull, E., 1974. A Subdivision Algorithm for Computer Display of Curved Surfaces. Ph.D. thesis. The University of Utah.
- Chadebecq, F., Tilmant, C., Bartoli, A., 2012a. Measuring the size of neoplasia in colonoscopy using depth-from-defocus, in : *IEEE Conference on Engineering in Medicine and Biology Society*, IEEE. pp. 1478–1481. doi :[10.1109/EMBC.2012.6346220](#).
- Chadebecq, F., Tilmant, C., Bartoli, A., 2013a. Estimation de l'échelle des néoplasies en colonoscopie par détection de la profondeur de défocalisation, in : *ORASIS-Journées francophones des jeunes chercheurs en vision par ordinateur*, AFRIF. pp. 1–8. doi :[hal-00829442](#).
- Chadebecq, F., Tilmant, C., Bartoli, A., 2013b. Using the infocus-breakpoint to estimate the scale of neoplasia in colonoscopy, in : *IEEE International Symposium on Biomedical Imaging*, IEEE. pp. 354–357. doi :[10.1109/ISBI.2013.6556485](#).
- Chadebecq, F., Tilmant, C., Bartoli, A., 2015. How big is this neoplasia ? Live colonoscopic size measurement using the Infocus-Breakpoint. *Medical Image Analysis* 19, 58–74. doi :[10.1016.2014.09.002](#).
- Chadebecq, F., Tilmant, C., Peyras, J., Collins, T., Bartoli, A., 2012b. Estimation de l'échelle en coloscopie monoculaire par quantification du flou optique : étude de faisabilité, in : *Reconnaissance des Formes et Intelligence Artificielle*, AFIA. pp. 1–8. doi :[hal-00656530](#).
- Chakrabarti, A., Zickler, T., Freeman, W., 2010. Analyzing spatially-varying blur, in : *IEEE Conference on Computer Vision and Pattern Recognition*, IEEE. pp. 12–19. doi :[10.1109.2010.5539954](#).
- Chang, C., Chiu, H., Wang, H., Lee, C., Tai, J., Tu, C., Tai, C., Chiang, T., Huang, J., Chang, D., Lin, J., 2010. An endoscopic training model to improve accuracy of colonic polyp size measurement. *Colorectal Disease* 25, 655–660. doi :[10.1007/s00384-010-0878-9](#).
- Chaptini, L., Chaaya, A., Depalma, F., Hunter, K., Peikin, S., Laine, L., 2014. Variation in polyp size estimation among endoscopists and impact on surveillance intervals. *Gastrointestinal Endoscopy* 80, 652–659. doi :[10.1016/j.gie.2014.01.053](#).
- Chhatkuli, A., Pizarro, D., Bartoli, A., 2014a. Non-Rigid Shape-from-Motion for Isometric Surfaces using Infinitesimal Planarity, in : *British Machine Vision Conference*, BMVA Press. pp. 1–5.
- Chhatkuli, A., Pizarro, D., Bartoli, A., 2014b. Stable Template-Based Isometric 3D Reconstruction in All Imaging Conditions by Linear Least-Squares, in : *IEEE Conference on Computer Vision and Pattern Recognition*, IEEE. pp. 708–715. doi :[10.1109/CVPR.2014.96](#).

- Choi, T., Yun, J., 2000. Three-dimensional shape recovery from the focused-image surface. *Optical Engineering* 39, 1321–1326. doi :[10.1117/1.602498](https://doi.org/10.1117/1.602498).
- Claxton, C., 2007. Colour depth-from-defocus incorporating experimental point spread function measurements. Ph.D. thesis. University of Warwick.
- Claxton, C., Staunton, R., 2008. Measurement of the point-spread function of a noisy imaging system. *Optical Society of America A* 25, 159–170. doi :[10.1364/JOSAA.25.000159](https://doi.org/10.1364/JOSAA.25.000159).
- Cohen, L., Cohen, I., 1993. Finite-element methods for active contour models and balloons for 2-D and 3-D images. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 15, 1131–1147. doi :[10.1109/34.244675](https://doi.org/10.1109/34.244675).
- Collins, T., Bartoli, A., 2010. Locally affine and planar deformable surface reconstruction from video, in : *Workshop on Vision, Modeling and Visualization*, Eurographics Association. pp. 339–346. doi :[10.2312/PE/VMV/VMV10/339-346](https://doi.org/10.2312/PE/VMV/VMV10/339-346).
- Cook, R.L., Torrance, K.E., 1982. A Reflectance Model for Computer Graphics. *ACM Transactions on Graphics* 1, 7–24. doi :[10.1145/357290.357293](https://doi.org/10.1145/357290.357293).
- Crofts, W., 2007. The Generation of Depth Maps via Depth-from-Defocus. Ph.D. thesis. University of Warwick.
- Daubechies, I., 1992. Ten lectures on wavelets. doi :[10.1137/1.9781611970104](https://doi.org/10.1137/1.9781611970104).
- Del Bue, A., Smeraldi, F., Agapito, L., 2004. Non-Rigid Structure-from-Motion using non-Parametric Tracking and Non-Linear Optimization, in : *IEEE Conference on Computer Vision and Pattern Recognition*, IEEE. pp. 1–8. doi :[10.1109/CVPR.2004.132](https://doi.org/10.1109/CVPR.2004.132).
- Deschênes, F., Ziou, D., Fuchs, P., 2004. An unified approach for a simultaneous and cooperative estimation of defocus blur and spatial shifts. *Image and Vision Computing* 22, 35–57. doi :[10.1016/j.imavis.2003.08.003](https://doi.org/10.1016/j.imavis.2003.08.003).
- Deschênes, F., Ziou, D., Fuchs, P., 2008. A homotopy-based approach for computing defocus blur and affine transform simultaneously. *Pattern Recognition* 41, 63–82. doi :[10.1016/j.patcog.2007.12.005](https://doi.org/10.1016/j.patcog.2007.12.005).
- D’Orazio, L., Bartoli, A., Baetz, A., Beorchia, S., Calvary, G., Chabane, Y., Chadebecq, F., Collins, T., Laurillau, Y., Martins-Baltar, L., Mohamad, B., Ponchon, T., Rey, C., Tilmant, C., Torti, S., 2014. Multimodal and multimedia image analysis and collaborative networking for digestive endoscopy. *Innovation and Research in BioMedical engineering* 35, 88–94. doi :[10.1016/j.irbm.2014.02.006](https://doi.org/10.1016/j.irbm.2014.02.006).
- Durou, J.D., Falcone, M., Sagona, M., 2008. Numerical methods for shape-from-shading : A new survey with benchmarks. *Computer Vision and Image Understanding* 109, 22–43. doi :[10.1016/j.cviu.2007.09.003](https://doi.org/10.1016/j.cviu.2007.09.003).

- Ens, J., Lawrence, P., 1993. An investigation of methods for determining depth from focus. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 15, 97–108. doi :[10.1109/34.192482](https://doi.org/10.1109/34.192482).
- Favaro, P., 2010. Recovering thin structures via nonlocal-means regularization with application to depth from defocus, in : *IEEE Conference on Computer Vision and Pattern Recognition*, IEEE. pp. 33–40. doi :[10.1109/CVPR.2010.5540089](https://doi.org/10.1109/CVPR.2010.5540089).
- Favaro, P., Duci, A., 2008. A theory of defocus via Fourier analysis, in : *IEEE Conference on Computer Vision and Pattern Recognition*, IEEE. pp. 1–8. doi :[10.1109/CVPR.2008.4587412](https://doi.org/10.1109/CVPR.2008.4587412).
- Favaro, P., Soatto, S., 2005. A geometric approach to shape from defocus. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 27, 406–417. doi :[10.1109/TPAMI.2005.43](https://doi.org/10.1109/TPAMI.2005.43).
- Favaro, P., Soatto, S., Burger, M., Osher, S.J., 2008. Shape from Defocus via Diffusion. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 30, 518–531. doi :[10.1109/TPAMI.2007.1175](https://doi.org/10.1109/TPAMI.2007.1175).
- Florczak, J., Petko, M., 2014. Usage of Shape From Focus Method For 3D Shape Recovery And Identification of 3D Object Position. *International Journal of Image Processing* 8, 116–124.
- Fusaglia, M., Wallach, D., Peterhans, M., Beldi, G., Weber, S., 2012. Accuracy evaluation of an image overlay in an instrument guidance system for laparoscopic liver surgery. *Computer Assisted Radiology and Surgery* , 1–6.
- Gaganov, V., Ignatenko, A., 2009. Robust Shape from Focus via Markov Random Fields, in : *GraphiCon*, pp. 74–80.
- Gargallo, P., Sturm, P., Pujades, S., 2007. An Occupancy–Depth Generative Model of Multi-view Images, in : *Asian Conference on Computer Vision*, Springer. pp. 373–383. doi :[10.1007/978-3-540-76390-1_37](https://doi.org/10.1007/978-3-540-76390-1_37).
- Goodman, J., 2004. *Introduction to Fourier optics*, 3rd Edition. Roberts and Company Publishers.
- Gotardo, P., Martinez, A., 2011a. Computing smooth time trajectories for camera and deformable shape in structure from motion with occlusion. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 33, 51–65. doi :[10.1109/TPAMI.2011.50](https://doi.org/10.1109/TPAMI.2011.50).
- Gotardo, P., Martinez, A., 2011b. Non-rigid structure from motion with complementary rank-3 spaces, in : *IEEE Conference on Computer Vision and Pattern Recognition*, IEEE. pp. 65–72. doi :[10.1109/CVPR.2011.5995560](https://doi.org/10.1109/CVPR.2011.5995560).
- Granlund, G., Knutsson, H., 1995. *Signal Processing for Computer Vision*. Springer.

- Grossmann, P., 1987. Depth from focus. *Pattern Recognition Letters* 5, 63–69. doi :[10.1016/0167-8655\(87\)90026-2](https://doi.org/10.1016/0167-8655(87)90026-2).
- Group, E.C.R., 2005. Update on the Paris classification of superficial neoplastic lesions in the digestive tract. *Endoscopy* 37, 570–578. doi :[10.1055/s-2005-861352](https://doi.org/10.1055/s-2005-861352).
- Hansard, M., Lee, S., Choi, O., Horaud, R., 2012. *Time-of-flight cameras : principles, methods and applications*. Springer.
- Hartley, R., Zisserman, A., 2004. *Multiple view geometry in computer vision*, 2nd Edition. Cambridge University Press.
- Helmlí, F., Scherer, S., 2001. Adaptive shape from focus with an error estimation in light microscopy, in : *IEEE International Symposium on Image and Signal Processing and Analysis*, IEEE. pp. 188–193. doi :[10.1109/ISPA.2001.938626](https://doi.org/10.1109/ISPA.2001.938626).
- Hien, K., Tongyan, P., Zhaoyang, W., Minh, L., N., H., Minh, V., 2014. Accurate 3D shape measurement of multiple separate objects with stereo vision. *Measurement Science and Technology* 25, 1–7. doi :[10.1088/0957-0233/25/3/035401](https://doi.org/10.1088/0957-0233/25/3/035401).
- Horn, B., 1968. Focusing. *Artificial Intelligence Memo* 160.
- Huang, W., Jing, Z., 2007. Evaluation of focus measures in multi-focus image fusion. *Pattern Recognition Letters* 28, 493–500. doi :[10.1016/j.patrec.2006.09.005](https://doi.org/10.1016/j.patrec.2006.09.005).
- Huber, P., 1964. Robust Estimation of a Location Parameter. *The Annals of Mathematical Statistics* 35, 73–101.
- Hyun, Y., Han, D., Bae, J., Park, H., Eun, C., 2011. Graduated injection needles and snares for polypectomy are useful for measuring colorectal polyp size. *Digestive and Liver Disease* 43, 391–394. doi :[10.1016/j.dld.2010.12.015](https://doi.org/10.1016/j.dld.2010.12.015).
- Institut National du Cancer, 2014. *Les cancers en France en 2013*. Institut National du Cancer. URL : <http://www.e-cancer.fr/publications/69-epidemiologie/758-les-cancers-en-france-edition-2013>.
- Jarvis, R., 1983. A Perspective on Range Finding Techniques for Computer Vision. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 5, 122–139. doi :[10.1109/TPAMI.1983.4767365](https://doi.org/10.1109/TPAMI.1983.4767365).
- Kaufman, A., Wang, J., 2008. *Visualization in Medicine and Life Sciences*. Springer. doi :[10.1007/978-3-540-72630-2](https://doi.org/10.1007/978-3-540-72630-2).
- Kim, D., Pickhardt, P., Taylor, A., Leung, W., Winter, T., Hinshaw, J., Gopal, D., Reichelderfer, M., Richard, H., Pfau, P., 2007. CT colonography versus colonoscopy for the detection of advanced neoplasia. *The New England Journal of Medicine* , 1403–1412doi :[10.1056/NEJMoa070543](https://doi.org/10.1056/NEJMoa070543).

- King, P., Comley, P., Sansom, C., 2014. Parabolic Trough Surface form Mapping Using Photogrammetry and its Validation with a Large Coordinate Measuring Machine. *Energy Procedia* 49, 118–125. doi :[10.1016/j.egypro.2014.03.013](https://doi.org/10.1016/j.egypro.2014.03.013).
- Krotkov, E., 1987. Exploratory visual sensing with an agile camera. Ph.D. thesis. University of Pennsylvania.
- Kudo, S., Hirota, S., Nakajima, T., Hosobe, S., Kusaka, H., Kobayashi, T., Himori, M., Yagyuu, M., 1994. Colorectal tumours and pit pattern. *Clinical Pathology* 47, 880–885.
- Leberl, F., Irschara, A., Pock, T., Meixner, P., Gruber, M., Scholz, S., Wiechert, A., 2010. Point Clouds : Lidar versus 3D Vision. *Photogrammetric Engineering And Remote Sensing* 76, 1123–1134. doi :[0099-1112/10/7610a\\$1123](https://doi.org/0099-1112/10/7610a$1123).
- Levenberg, K., 1944. A method for the solution of certain non-linear problems in least squares. *Applied Mathematics* 2, 164–168.
- Levin, A., Fergus, R., Durand, F., Freeman, W., 2007. Image and Depth from a Conventional Camera with a Coded Aperture, in : *ACM Eurographics SIGGRAPH Symposium on Computer Animation*, Eurographics Association. pp. 1–9. doi :[10.1145/1275808.1276464](https://doi.org/10.1145/1275808.1276464).
- Levin, A., Lischinski, D., Weiss, Y., 2008. A Closed-Form Solution to Natural Image Matting. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 30, 228–242. doi :[10.1109/TPAMI.2007.1177](https://doi.org/10.1109/TPAMI.2007.1177).
- Li, A., Staunton, R., Tjahjadi, T., 2013a. Rational-operator-based depth-from-defocus approach to scene reconstruction. *Optical Society of America A* 30, 87–95. doi :[10.1364/JOSAA.30.001787](https://doi.org/10.1364/JOSAA.30.001787).
- Li, C., Su, S., Matsushita, Y., Zhou, K., Lin, S., 2013b. Bayesian Depth-from-Defocus with Shading Constraints, in : *IEEE Conference on Computer Vision and Pattern Recognition*, IEEE. pp. 217–224. doi :[10.1109/CVPR.2013.35](https://doi.org/10.1109/CVPR.2013.35).
- Li, S., Yang, B., 2008. Multifocus image fusion by combining curvelet and wavelet transform. *Pattern Recognition Letters* 29, 1295–1301. doi :[10.1016/j.patrec.2008.02.002](https://doi.org/10.1016/j.patrec.2008.02.002).
- Lighthart, G., Groen, F., 1982. A comparison of different autofocus algorithms. *IEEE Pattern Recognition* , 597–600.
- Lin, J., Ji, X., Xu, W., Dai, Q., 2013. Absolute Depth Estimation From a Single Defocused Image. *IEEE Transactions on Image Processing* 22, 4545–4550. doi :[10.1109/TIP.2013.2274389](https://doi.org/10.1109/TIP.2013.2274389).
- Lin, J., Zhang, C., Shi, Q., 2004. Estimating the amount of defocus through a wavelet transform approach. *Pattern Recognition Letters* 25, 407–411. doi :[10.1016/j.patrec.2003.11.003](https://doi.org/10.1016/j.patrec.2003.11.003).

- Lobay, A., Forsyth, D., 2006. Shape from Texture Without Boundaries. *Computer Vision* 67, 71–91. doi :[10.1007/s11263-006-4068-8](https://doi.org/10.1007/s11263-006-4068-8).
- Lowe, D., 2004. Distinctive Image Features from Scale-Invariant Keypoints. *Computer Vision* 60, 91–110. doi :[10.1023/B:VISI.0000029664.99615.94](https://doi.org/10.1023/B:VISI.0000029664.99615.94).
- Ma, Z., Liao, R., Tao, X., Xu, L., Jia, J., Wu, E., 2015. Handling motion blur in multi-frame super-resolution, in : *IEEE Conference on Computer Vision and Pattern Recognition*, IEEE. pp. 5224–5232.
- Malik, A., Choi, T., 2008. A Novel Algorithm for Estimation of Depth Map Using Image Focus for 3D Shape Recovery in the Presence of Noise. *Pattern Recognition* 41, 1–25. doi :[10.1016/j.patcog.2007.12.014](https://doi.org/10.1016/j.patcog.2007.12.014).
- Mallat, S., 2009. *A Wavelet Tour of Signal Processing*, 3rd Edition. Academic Press. doi :[10.1016/B978-0-12-374370-1.00001-X](https://doi.org/10.1016/B978-0-12-374370-1.00001-X).
- Malti, A., Bartoli, A., 2014. Combining Conformal Deformation and Cook Torrance Shading for 3-D Reconstruction in Laparoscopy. *IEEE Transactions on Biomedical Engineering* 61, 1684–1692. doi :[10.1109/TBME.2014.2300237](https://doi.org/10.1109/TBME.2014.2300237).
- Malti, A., Bartoli, A., Collins, T., 2011. Template-based conformal shape-from-motion from registered laparoscopic images, in : *Conference on Medical Image Understanding and Analysis*, pp. 1–7.
- Margulies, C., Krevsky, B., Catalano, M., 1994. How accurate are endoscopic estimates of size? *Gastrointestinal Endoscopy* 40, 174–176. doi :[10.1016/S0016-5107\(94\)70162-8](https://doi.org/10.1016/S0016-5107(94)70162-8).
- Martínez, J., 1994. Algorithms for Solving Nonlinear Systems of Equations, in : *Algorithms for Continuous Optimization*. Springer. volume 434 of *NATO ASI Series*, pp. 81–108. doi :[10.1007/978-94-009-0369-2_4](https://doi.org/10.1007/978-94-009-0369-2_4).
- McInerney, T., Terzopoulos, D., 1993. A finite element model for 3D shape reconstruction and nonrigid motion tracking, in : *IEEE Conference on Computer Vision*, IEEE. pp. 518–523. doi :[10.1109/ICCV.1993.378169](https://doi.org/10.1109/ICCV.1993.378169).
- Meer, P., Weiss, I., 1990. Smoothed differentiation filters for images, in : *IEEE Conference on Pattern Recognition*, IEEE. pp. 121–126. doi :[10.1109/ICPR.1990.119341](https://doi.org/10.1109/ICPR.1990.119341).
- Minhas, R., Mohammed, A., Wu, Q., Sid-Ahmed, M., 2009. 3D Shape from Focus and Depth Map Computation Using Steerable Filters. *Lecture Notes in Computer Science* 5627, 573–583. doi :[10.1007/978-3-642-02611-9_57](https://doi.org/10.1007/978-3-642-02611-9_57).
- Minhas, R., Mohammed, A., A., Q., J.W., 2011. Shape from focus using fast discrete curvelet transform. *Pattern Recognition* 44, 839–853. doi :[10.1016/j.patcog.2010.10.015](https://doi.org/10.1016/j.patcog.2010.10.015).

- Mirota, D., Ishii, M., Hager, G., 2011. Vision-Based Navigation in Image-Guided Interventions. *Annual Review of Biomedical Engineering* 13, 297–319. doi :[10.1146/annurev-bioeng-071910-124757](https://doi.org/10.1146/annurev-bioeng-071910-124757).
- Moreno-Noguer, F., Fua, P., 2013. Stochastic Exploration of Ambiguities for Nonrigid Shape Recovery. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 35, 463–475. doi :[10.1109/TPAMI.2012.102](https://doi.org/10.1109/TPAMI.2012.102).
- Moreno-Noguer, F., Porta, J.M., Fua, P., 2010. Exploring Ambiguities for Monocular Non-rigid Shape Estimation, in : *Lecture Notes in Computer Science*. Springer. volume 6313, pp. 370–383. doi :[10.1007/978-3-642-15558-1_27](https://doi.org/10.1007/978-3-642-15558-1_27).
- Moug, S., Vernall, N., Saldanha, J., McGregor, J., Balsitis, M., Diamant, R., 2010. Endoscopists' estimation of size should not determine surveillance of colonic polyps. *Colorectal Disease* 12, 646–650. doi :[10.1111/j.1463-1318.2009.01870.x](https://doi.org/10.1111/j.1463-1318.2009.01870.x).
- Muhammad, M., Choi, T., 2012. Sampling for Shape from Focus in Optical Microscopy. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 34, 564–573. doi :[10.1109/TPAMI.2011.144](https://doi.org/10.1109/TPAMI.2011.144).
- Myles, Z., da Vitoria Lobo, V., 1998. Recovering affine motion and defocus blur simultaneously. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 20, 652–658. doi :[10.1109/34.683782](https://doi.org/10.1109/34.683782).
- Namboodiri, V., Chaudhuri, S., 2007. On defocus, diffusion and depth estimation. *Pattern Recognition Letters* 28, 311–319. doi :[10.1016/j.patrec.2006.04.011](https://doi.org/10.1016/j.patrec.2006.04.011).
- Namboodiri, V., Chaudhuri, S., 2008. Recovery of relative depth from a single observation using an uncalibrated (real-aperture) camera, in : *IEEE Conference on Computer Vision and Pattern Recognition*, IEEE. pp. 1–6. doi :[10.1109/CVPR.2008.4587779](https://doi.org/10.1109/CVPR.2008.4587779).
- Nayar, S., Nakagawa, Y., 1994. Shape from focus. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 16, 824–831. doi :[10.1109/34.308479](https://doi.org/10.1109/34.308479).
- Ogle, K., 1968. *Optics*. Charles C. Thomas.
- Oren, M., Nayar, S., 1995. Generalization of the Lambertian model and implications for machine vision. *Computer Vision* 14, 227–251. doi :[10.1007/BF01679684](https://doi.org/10.1007/BF01679684).
- Park, S., Choi, E., Lee, S., Byeon, J., Jo, J., Kim, Y., Lee, K., Ha, H., Han, J., 2007. Polyp Measurement Reliability, Accuracy, and Discrepancy : Optical Colonoscopy versus CT Colonography with Pig Colonic Specimens. *Radiology* 244, 157–164. doi :[10.1148/radiol.2441060794](https://doi.org/10.1148/radiol.2441060794).
- Pentland, A., 1982. Depth of Scene from Depth of Field, in : *DARPA Image Understanding Workshop*, DARPA. pp. 253–259.

- Pentland, A., 1987. A New Sense for Depth of Field. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 9, 523–531. doi :[10.1109/TPAMI.1987.4767940](https://doi.org/10.1109/TPAMI.1987.4767940).
- Perriollat, M., Hartley, R., Bartoli, A., 2011. Monocular Template-based Reconstruction of Inextensible Surfaces. *Computer Vision* 95, 124–137. doi :[10.1007/s11263-010-0352-8](https://doi.org/10.1007/s11263-010-0352-8).
- Pertuz, S., Puig, D., Garcia, M., 2013a. Analysis of focus measure operators for shape-from-focus. *Pattern Recognition* 46, 1415–1432. doi :[10.1016/j.patcog.2012.11.011](https://doi.org/10.1016/j.patcog.2012.11.011).
- Pertuz, S., Puig, D., Garcia, M., 2013b. Reliability measure for shape-from-focus. *Image and Vision Computing* 31, 725–734. doi :[10.1016/j.imavis.2013.07.005](https://doi.org/10.1016/j.imavis.2013.07.005).
- Pizarro, D., Bartoli, A., Collins, T., 2013. Isowarp and Conwarp : Warps that Exactly Comply with Weak-Perspective Projection of Deforming Objects, in : *British Machine Vision Conference*, BMVA Press. pp. 1–11. doi :[10.5244/C.27.104](https://doi.org/10.5244/C.27.104).
- Pressley, A.N., 2010. *Elementary Differential Geometry* 2nd Edition. Springer. doi :[10.1007/978-1-84882-891-9](https://doi.org/10.1007/978-1-84882-891-9).
- Rahmat, R., Malik, A., Kamel, N., Nisar, H., 2013. 3D shape from focus using {LULU} operators and discrete pulse transform in the presence of noise. *Visual Communication and Image Representation* 24, 303–317. doi :[10.1016/j.jvcir.2013.01.005](https://doi.org/10.1016/j.jvcir.2013.01.005).
- Raj, A., 2009. Accurate depth from defocus estimation with video-rate implementation. Ph.D. thesis. University of Warwick.
- Rajagopalan, A., Chaudhuri, S., 1999. An MRF model-based approach to simultaneous recovery of depth and restoration from defocused images. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 21, 577–589. doi :[10.1109/34.777369](https://doi.org/10.1109/34.777369).
- Reichl, T., Luo, X., Menzel, M., Hautmann, H., Mori, K., Navab, N., 2013. Hybrid electromagnetic and image-based tracking of endoscopes with guaranteed smooth output. *Computer Assisted Radiology and Surgery* 8, 955–965. doi :[10.1007/s11548-013-0835-5](https://doi.org/10.1007/s11548-013-0835-5).
- Ren, N., 2006. *Digital Light Field Photography*. Ph.D. thesis. Stanford University.
- Rooms, F., 2002. Estimating image blur in the wavelet domain, in : *IEEE Conference on Acoustics, Speech, and Signal Processing*, IEEE. pp. 1–4. doi :[10.1109/ICASSP.2002.5745673](https://doi.org/10.1109/ICASSP.2002.5745673).
- Rubio, C., Jónasson, J., Nesi, G., Mazur, J., Olafsdóttir, E., 2010. The Size of Colon Polyps Revisited : Intra- and Inter-observer Variations. *Cancer Research and Treatment* 30, 1–2.
- Saitoh, N., Kuroki, K., Kurosawa, K., 1997. Estimation of out-of-focus blur and its effectiveness for deblurring, in : *Investigative Image Processing*, SPIE. pp. 2–9. doi :[10.1117/12.267165](https://doi.org/10.1117/12.267165).

- Salzmann, M., Fua, P., 2009. Reconstructing sharply folding surfaces : A convex formulation, in : IEEE Conference on Computer Vision and Pattern Recognition, IEEE. pp. 1054–1061. doi :[10.1109/CVPR.2009.5206759](https://doi.org/10.1109/CVPR.2009.5206759).
- Salzmann, M., Fua, P., 2011. Linear Local Models for Monocular Reconstruction of Deformable Surfaces. IEEE Transactions on Pattern Analysis and Machine Intelligence 33, 931–944. doi :[10.1109/TPAMI.2010.158](https://doi.org/10.1109/TPAMI.2010.158).
- Salzmann, M., Hartley, R., Fua, P., 2007. Convex Optimization for Deformable Surface 3-D Tracking, in : IEEE Conference on Computer Vision, IEEE. pp. 1–8. doi :[10.1109/ICCV.2007.4409031](https://doi.org/10.1109/ICCV.2007.4409031).
- Salzmann, M., Urtasun, R., Fua, P., 2008. Local deformation models for monocular 3D shape recovery, in : IEEE Conference on Computer Vision and Pattern Recognition, IEEE. pp. 1–8. doi :[10.1109/CVPR.2008.4587499](https://doi.org/10.1109/CVPR.2008.4587499).
- Sang-Yong, L., Kumar, Y., Ji-Man, C., Sang-Won, L., Soo-Won, K., 2008. Enhanced Autofocus Algorithm Using Robust Focus Measure and Fuzzy Reasoning. IEEE Transactions on Circuits and Systems for Video Technology 18, 1237–1246. doi :[10.1109/TCSVT.2008.924105](https://doi.org/10.1109/TCSVT.2008.924105).
- Saxena, A., Chung, S., Ng, A., 2008. 3-D Depth Reconstruction from a Single Still Image. Computer Vision 76, 53–69. doi :[10.1007/s11263-007-0071-y](https://doi.org/10.1007/s11263-007-0071-y).
- Schlag, J., Sanderson, A., C.P., N., Wimberly, F., 1983. Implementation of Automatic Focusing Algorithms for a Computer Vision System with Camera Control. Technical Report CMU-RI-TR-83-14. Robotics Institute.
- Schmalz, C., Forster, F., Schick, A., Angelopoulou, E., 2012. An endoscopic 3D scanner based on structured light. Medical Image Analysis 16, 1063–1072. doi :[10.1016/j.media.2012.04.001](https://doi.org/10.1016/j.media.2012.04.001).
- Schoen, R., 2002. The case for population-based screening for colorectal cancer. Nature Reviews Cancer 2, 65–70. doi :[10.1038/nrc705](https://doi.org/10.1038/nrc705).
- Schoen, R., Gerber, L., Margulies, C., 1997. The pathologic measurement of polyp size is preferable to the endoscopic estimate. Gastrointestinal Endoscopy 46, 492–496. doi :[10.1016/S0016-5107\(97\)70002-6](https://doi.org/10.1016/S0016-5107(97)70002-6).
- Shim, S.O., Choi, T.S., 2010. A novel iterative shape from focus algorithm based on combinatorial optimization. Pattern Recognition 43, 3338–3347. doi :[10.1016/j.patcog.2010.05.029](https://doi.org/10.1016/j.patcog.2010.05.029).
- Sochen, N., Kimmel, R., Bruckstein, A., 2001. Diffusions and confusions in signal and image processing. Journal of Mathematical Imaging and Vision 14, 195–209. doi :[10.1023/A:1011277827470](https://doi.org/10.1023/A:1011277827470).
- Subbarao, M., 1987. Direct recovery of depth-map I : differential methods. IEEE workshop on computer vision , 58–65.

- Subbarao, M., 1988. Parallel depth recovery by changing camera parameters. *IEEE Conference on Computer Vision* , 149–155doi :[10.1109/CCV.1988.589986](https://doi.org/10.1109/CCV.1988.589986).
- Subbarao, M., 1989. Efficient depth recovery through inverse optics. *Machine Vision for Inspection and Measurement* , 101–126.
- Subbarao, M., Choi, T., 1995. Accurate recovery of three-dimensional shape from image focus. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 17, 266–274. doi :[10.1109/34.368191](https://doi.org/10.1109/34.368191).
- Subbarao, M., Gurumoorthy, N., 1988. Depth recovery from blurred edges. *IEEE Conference on Computer Vision and Pattern Recognition* , 498–503doi :[10.1109/CVPR.1988.196281](https://doi.org/10.1109/CVPR.1988.196281).
- Subbarao, M., Surya, G., 1994. Depth from defocus : A spatial domain approach. *Computer Vision* 13, 271–294. doi :[10.1007/BF02028349](https://doi.org/10.1007/BF02028349).
- Talbot, I., 1996. Redetection and growth of colorectal polyps. *Gastroenterology and Hepatology* 39, 1. doi :[10.1136/gut.39.3.492](https://doi.org/10.1136/gut.39.3.492).
- Tao, M., Hadap, S., Malik, J., Ramamoorthi, R., 2013. Depth from Combining Defocus and Correspondence Using Light-Field Cameras, in : *IEEE Conference on Computer Vision, IEEE*. pp. 673–680. doi :[10.1109/ICCV.2013.89](https://doi.org/10.1109/ICCV.2013.89).
- Thelen, A., Frey, S., Hirsch, S., Hering, P., 2009. Improvements in Shape-from-Focus for Holographic Reconstructions With Regard to Focus Operators, Neighborhood-Size, and Height Value Interpolation. *IEEE Transactions on Image Processing* 18, 151–157. doi :[10.1109/TIP.2008.2007049](https://doi.org/10.1109/TIP.2008.2007049).
- Torresani, L., Hertzmann, A., Bregler, C., 2008. Nonrigid Structure-from-Motion : Estimating Shape and Motion with Hierarchical Priors. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 30, 878–892. doi :[10.1109/TPAMI.2007.70752](https://doi.org/10.1109/TPAMI.2007.70752).
- Tsap, L., Goldof, D., Sarkar, S., 1998. Nonrigid motion analysis based on dynamic refinement of finite element models. *IEEE Conference on Computer Vision and Pattern Recognition* , 728–734doi :[10.1109/CVPR.1998.698684](https://doi.org/10.1109/CVPR.1998.698684).
- Varol, A., Salzmann, M., Tola, E., Fua, P., 2009. Template-free monocular reconstruction of deformable surfaces, in : *IEEE Conference on Computer Vision, IEEE*. pp. 1811–1818. doi :[10.1109/ICCV.2009.5459403](https://doi.org/10.1109/ICCV.2009.5459403).
- Vicente, S., Agapito, L., 2012. Soft Inextensibility Constraints for Template-Free Non-rigid Reconstruction, in : *European Conference on Computer Vision*. Springer. volume 7574, pp. 426–440. doi :[10.1007/978-3-642-33712-3_31](https://doi.org/10.1007/978-3-642-33712-3_31).
- Watanabe, M., Nayar, S., 1998. Rational Filters for Passive Depth from Defocus. *Computer Vision* 27, 203–225. doi :[10.1023/A:1007905828438](https://doi.org/10.1023/A:1007905828438).

- Wee, C., Paramesran, R., 2007. Measure of image sharpness using eigenvalues. *Information Sciences* 177, 2533–2552. doi :[10.1016/j.ins.2006.12.023](https://doi.org/10.1016/j.ins.2006.12.023).
- Wei, Y., Dong, Z., Wu, C., 2009. Global depth from defocus with fixed camera parameters, in : *IEEE Conference on Mechatronics and Automation*, IEEE. pp. 1887–1892. doi :[10.1109/ICMA.2009.5246652](https://doi.org/10.1109/ICMA.2009.5246652).
- Wei, Y., Dong, Z., Wu, C., 2012. Depth measurement using single camera with fixed camera parameters. *IET Computer Vision* 6, 29–39. doi :[10.1049/iet-cvi.2010.0017](https://doi.org/10.1049/iet-cvi.2010.0017).
- Wenguo, L., Suping, F., Shaojun, D., 2013. 3D shape measurement based on structured light projection applying polynomial interpolation technique. *Optik - Light and Electron Optics* 124, 20–27. doi :[10.1016/j.ijleo.2011.11.084](https://doi.org/10.1016/j.ijleo.2011.11.084).
- World Health Organization, 2008. *World Cancer Report*. International Agency for Research on Cancer. URL : http://www.who.int/cancer/publications/world_cancer_report2008.
- Xiao, J., Kanade, T., 2005. Uncalibrated perspective reconstruction of deformable structures, in : *IEEE Conference on Computer Vision*, IEEE. pp. 1075–1082. doi :[10.1109/ICCV.2005.241](https://doi.org/10.1109/ICCV.2005.241).
- Xie, H., Rong, W., Sun, L., 2006. Wavelet-Based Focus Measure and 3-D Surface Reconstruction Method for Microscopy Images, in : *IEEE Conference on Intelligent Robots and Systems*, IEEE. pp. 229–234. doi :[10.1109/IRoS.2006.282641](https://doi.org/10.1109/IRoS.2006.282641).
- Xiong, Y., Shafer, S., 1997. Moment and Hypergeometric Filters for High Precision Computation of Focus, Stereo and Optical Flow. *Computer Vision* 22, 25–59. doi :[10.1023/A:1007927810205](https://doi.org/10.1023/A:1007927810205).
- Yang, G., Nelson, B., 2003. Wavelet-based autofocusing and unsupervised segmentation of microscopic images, in : *IEEE Conference on Intelligent Robots and Systems*, IEEE. pp. 2143–2148. doi :[10.1109/IRoS.2003.1249188](https://doi.org/10.1109/IRoS.2003.1249188).
- Yang, J., Schonfeld, D., 2010. Virtual Focus and Depth Estimation From Defocused Video Sequences. *IEEE Transactions on Image Processing* 19, 668–679. doi :[10.1109/TIP.2009.2036708](https://doi.org/10.1109/TIP.2009.2036708).
- Zhang, X., Liu, Z., Jiang, M., Chang, M., 2014. Fast and accurate auto-focusing algorithm based on the combination of depth from focus and improved depth from defocus. *Optics Express* 22, 31237–31247. doi :[10.1364/OE.22.031237](https://doi.org/10.1364/OE.22.031237).
- Zhou, C., Lin, S., Nayar, S., 2009. Coded Aperture Pairs for Depth-from-Defocus, in : *IEEE Conference on Computer Vision*, IEEE. pp. 325–332. doi :[10.1109/ICCV.2009.5459268](https://doi.org/10.1109/ICCV.2009.5459268).
- Zhu, X., Cohen, S., Schiller, S., Milanfar, P., 2013. Estimating Spatially Varying Defocus Blur From A Single Image. *IEEE Transactions on Image Processing* 22, 1–13. doi :[10.1109/TIP.2013.2279316](https://doi.org/10.1109/TIP.2013.2279316).

- Zhuo, S., Sim, T., 2009. On the Recovery of Depth from a Single Defocused Image. *Computer Analysis of Images and Patterns* 5702, 889–897. doi :[10.1007/978-3-642-03767-2_108](https://doi.org/10.1007/978-3-642-03767-2_108).
- Zhuo, S., Sim, T., 2011. Defocus map estimation from a single image. *Pattern Recognition* 44, 1852–1858. doi :[10.1016/j.patcog.2011.03.009](https://doi.org/10.1016/j.patcog.2011.03.009).
- Ziou, D., 1998. Passive depth from defocus using a spatial domain approach, in : *IEEE Conference on Computer Vision*, IEEE. pp. 799–804. doi :[10.1109/ICCV.1998.710809](https://doi.org/10.1109/ICCV.1998.710809).
- Ziou, D., Deschênes, F., 2001. Depth from Defocus Estimation in Spatial Domain. *Computer Vision and Image Understanding* 81, 143–165. doi :[10.1006/cviu.2000.0899](https://doi.org/10.1006/cviu.2000.0899).