



Lumière structurée codée pour une reconstruction 3-D robuste.

Application à la chirurgie mini-invasive.

mémoire présenté pour obtenir le grade de

Docteur de l'Université de Strasbourg
Discipline : Signal, Image, Automatique et Robotique
(Spécialité : Robotique et vision)

par

Xavier Maurice

Directeur de Thèse : Christophe DOIGNON, Maître de Conférences HDR
à l'Université de Strasbourg
Rapporteur Externe : Joaquim SALVI, Professeur à l'Université de Girona, Espagne
Rapporteur Externe : Michel DEVY, Directeur de recherche au LAAS, CNRS, Toulouse
Examineur : David FOFI, Professeur à l'Université de Bourgogne, Dijon
Examineur : Michel DE MATHELIN, Professeur à l'Université de Strasbourg

Remerciements

Toutes ces personnes ont participé, d'une façon ou d'une autre à ce travail de thèse et je tiens à les en remercier vivement.

Merci à Pierre Graebing, parti trop tôt, qui m'a apporté son expérience, sa motivation, m'a amené à mieux expliquer mon travail. Il était d'une compagnie toujours très agréable, amateur de bons mots et cherchait toujours la complicité. Je regarderai, entre autres, un très bon souvenir de notre déplacement à Caen.

Je remercie tout particulièrement Christophe Doignon pour m'avoir proposé ce sujet de thèse, pour avoir été au plus près de mon travail du début à la fin, pour m'avoir poussé à toujours mieux formaliser, mieux communiquer sur celui-ci. Merci pour la flexibilité dont il a fait preuve vis-à-vis du chemin initialement engagé quant à mon sujet de thèse et à la miniaturisation du prototype notamment. Merci de m'avoir aidé à me focaliser et à mettre en évidence les aspects scientifiques de mon travail. Enfin, je me souviendrai des nombreux moments passés ensemble, qui étaient souvent l'occasion de discussions plus personnelles.

Merci à Michel de Mathelin de m'avoir accueilli dans l'équipe AVR, de s'être intéressé à l'approche que je proposais, avoir montré une forte volonté de valorisation et de m'avoir aidé à financer quelques mois supplémentaires pour assurer la pérennité du travail effectué.

Je tiens à remercier les membres du jury, Joaquim Salvi et Michel Devy pour avoir rapporté sur ce mémoire et David Fofi pour avoir accepté d'examiner mon travail. Merci pour vos appréciations et vos remarques constructives, c'était un honneur d'avoir dans mon jury parmi les meilleurs spécialistes en lumière structurée.

Merci au CNRS et à la région Alsace pour le cofinancement de cette thèse.

Merci également à Chadi Albitar pour m'avoir transmis ses compétences sur la problématique de la lumière structurée codée et de m'avoir mis le pied à l'étrier.

Merci à l'IRCAD, à Jacques Marescaux et au personnel de l'établissement pour m'avoir permis de travailler dans ce cadre particulièrement agréable et de m'avoir fait bénéficier de votre expérience dans la chirurgie mini-invasive et des moyens impressionnants qui y sont présents. Sans eux, je n'aurais pu effectuer de validations in vivo.

Je dois également beaucoup à la société Karl Storz pour le prêt du matériel nécessaire à la construction du prototype endoscopique. Merci plus particulièrement à Jean-Mickel Kleiss, à Pierre Chaumeny et à Andrea Nino Ruiz pour leur disponibilité et leurs bons conseils.

Je remercie particulièrement les doctorants et permanents de l'équipe AVR : Laure, Houssem, Lennart, Markus, Bérengère, Julien, Mathieu, Fadi, Ryad, Antonio, Norbert, Wael, Salih, Philippe, Edouard, Bernard, Laurent, Lucile, Florent, Loïc, Pierre et Jacques, ainsi que ceux que j'oublie. Il était très agréable de pouvoir discuter au second, voire au troisième degrés (après le bruit peut causer des instabilités :). Je me souviendrai longtemps de notre week-end mémorable à Lyon qui témoigne de la très bonne ambiance dans cette équipe. Merci aussi à son fondateur, Eric Ostertag.

Merci à Jean-Paul Caspar, Alain Rohrfritsch et Pascale Musard pour l'organisation et la mise en place des séances de validation in vivo. Merci à l'équipe audiovisuelle, en particulier à Guillaume qui m'a aidé à mettre en place la chaîne d'acquisition endoscopique.

Je remercie également Pierre Pfeiffer pour m'avoir permis d'effectuer des tests avec des sources laser au Laboratoire des Systèmes Photoniques, lors de la réalisation du prototype endoscopique.

Merci à ma famille et à mes proches pour m'avoir soutenu et avoir supporté un Xavier trop souvent au travail pendant les vacances et les week-ends.

Enfin, je dédie ce mémoire à mes parents, Arielle et François, qui m'ont soutenu et ont toujours été là dans les moments difficiles, aussi bien scolaires que sentimentaux, qui ont émaillé mon parcours.

Résumé

Lumière structurée codée pour une reconstruction 3-D robuste. Application à la chirurgie mini-invasive.

Durant une opération chirurgicale mini-invasive, la capture de la scène est réalisée par un endoscope qui présente les informations au chirurgien sur un écran, en 2-D. Amélioration de la perception visuelle de la scène, guidage de robots, recalage de modèles d'organes, la capture en temps-réel d'une carte de profondeur pour de telles scènes peut présenter de multiples applications.

A cette fin, de nombreuses contributions menant à la réalisation d'un système endoscopique mono-trocart, stéréo, basé sur une projection de lumière structurée codée, sont présentées. Parmi celles-ci, une utilisation avancée de la géométrie épipolaire dès la conception numérique, puis la réalisation du motif, est motivée, proposée et évaluée. D'autres contributions relatives au codage avec redondance, aux différents traitements implémentés sur GPU, à la reconstruction 3-D, à l'étalonnage partiel du dispositif sont décrites. En particulier, les deux problématiques des algorithmes de codage dits "brute-force" : le comportement de recherche, et les tests d'unicité, sont traitées. D'autre part, une compensation a priori des distorsions liées aux optiques du projecteur est proposée après une rectification inverse, réalisée à travers un alignement du motif sur la géométrie épipolaire. De nombreux résultats sur le codage, les reconstructions ex vivo puis in vivo, qualitatifs et quantitatifs, permettent d'apprécier l'approche mise en oeuvre.

Mots-clés : lumière structurée codée, reconstruction 3-D, vision endoscopique.

Abstract

Robust 3-D reconstruction by means of coded structured light. Application to minimally-invasive surgery

During a minimally-invasive surgical intervention, the intra-abdominal scene, which is acquired by an endoscope, is presented to the surgeon by means of a 2-D display. The real-time depth map computation of such a scene could have huge benefits like a better visual perception, robot visual servoing, registration of pre-existing models of organs.

To this end, in this thesis, many contributions leading to the realisation of an endoscopic stereo, mono-trocar, system, based on a coded structured light pattern projection, are presented. We propose to better exploit the epipolar geometry of the system at the pattern design step, first for the numerical coding, and then, for the pattern layout. Other contributions dealing with "brute-force" coding algorithm with redundancy, with real-time GPU processing, with 3-D reconstruction and with weak-calibration of the system, are described. In particular, the two main issues in "brute-force" coding algorithms : the search behaviour and the unicity test, are addressed. Moreover, an a priori compensation of optical distorsions of the projector is proposed after the pattern alignment on the epipolar geometry. Numerous coding results, ex vivo and in vivo 3-D reconstruction with associated statistics, are reported to evaluate the proposed approach.

Keywords : coded structured light, 3-D reconstruction, endoscopic vision.

Table des matières

Introduction	1
Chapitre 1 Travaux antérieurs et motivations	5
1.1 Reconstructions 3-D en chirurgie mini-invasive	6
1.1.1 Scènes intra-abdominales	6
1.1.2 Optiques endoscopiques	6
1.1.3 Vision monoculaire	8
1.1.4 Vision stéréo passive	8
1.1.5 Vision stéréo active	9
1.1.6 Time-of-Flight	12
1.2 Lumière structurée	13
1.2.1 Stratégies de codage	17
1.2.1.1 Codage par voisinage spatial 1-D	17
1.2.1.2 Codage par voisinage spatial 2-D	20
1.2.1.3 Codage par multiplexage temporel	20
1.2.1.4 Autres types de codage	21
1.2.2 Traitements et motifs imperceptibles	22
1.2.3 Lumière structurée et scènes intra-abdominales	25
1.3 L’alignement épipolaire du motif, fil rouge de la conception proposée	25
Chapitre 2 Conception 1 : Codage du motif	29
2.1 Introduction	29
2.1.1 Codage par voisinage spatial	30
2.1.2 Motivations pour un voisinage 3×3	31
2.1.3 Un critère de robustesse : la distance de Hamming	31
2.1.4 Objectifs du codage	32
2.1.5 Motivations pour l’amélioration des algorithmes	32
2.1.6 Contributions proposées	33
2.2 Travaux antérieurs	33

2.2.1	Un large éventail d'applications	34
2.2.2	Matrices parfaites et variantes	36
2.2.3	Les contraintes posées sur le problème de codage	37
2.2.4	Augmenter les contraintes de codage	37
2.2.5	Diminuer les contraintes de codage	38
2.2.6	Algorithmes directs	39
2.2.7	Algorithmes "brute-force"	40
2.2.7.1	Approche aléatoire (M)	40
2.2.7.2	Approche exhaustive (C)	41
2.3	Éléments théoriques sur le codage	42
2.3.1	La distance de Hamming moyenne \bar{H}	42
2.3.2	Capacités de correction	42
2.3.3	Prédiction de la distance de Hamming moyenne \bar{H}	43
2.3.4	Limites supérieures sur les paramètres de codage	45
2.3.4.1	Limites supérieures de Singleton et de Hamming	45
2.3.5	Dénombrement de mots de codes pour une distance de Hamming donnée	46
2.4	Comportement de recherche par le paradigme exploration / exploitation	46
2.4.1	Espace des solutions	47
2.4.2	Application à la recherche de <i>PSM</i>	48
2.5	Test d'unicité dans l'espace des mots de code	50
2.5.1	Réduction de complexité pour $H_{Min} = 1$	51
2.5.2	Généralisation pour $H_{Min} \geq 1$	52
2.5.3	La stratégie de recherche partagée	53
2.5.4	Résultats de complexité algorithmique	54
2.6	Tests d'unicité exploitant la géométrie épipolaire	54
2.6.1	Unicité par proximité épipolaire	56
2.6.2	Unicité par alignement épipolaire	57
2.7	Codage optimal et alignement épipolaire	58
2.8	Résultats de codage	59
2.8.1	Codage <i>PSM</i>	59
2.8.2	Codage <i>SPSM</i>	60
2.8.3	Résultats temps-réel	62
2.8.4	Résultats de codage optimal	64
2.9	Conclusion	65

Chapitre 3 Conception 2 : Agencement du motif et traitements	67
3.1 Organisation des primitives	68
3.1.1 Utilisation de la géométrie épipolaire pour la conception du motif	68
3.1.2 Configurations de l'épipole	70
3.1.3 Alignement local	71
3.1.4 Alignement global	72
3.1.5 Autres travaux proposant un alignement global	74
3.1.6 Alignement local ou global ?	75
3.2 Etude de la déformation géométrique des voisinages	75
3.2.1 Retrouver les primitives d'un voisinage 2-D	75
3.2.2 Observations	76
3.2.3 Interprétation géométrique	77
3.2.4 Modélisation et perte d'adjacence verticale par proximité	78
3.2.5 Conclusion	81
3.3 Choix des primitives visuelles	81
3.3.1 Types de scène	81
3.3.2 Travaux antérieurs et primitives proposées	84
3.3.3 Primitives pour l'étalonnage	84
3.3.4 Primitives pour scènes inconnues	86
3.3.5 Choix de la couleur	87
3.4 Chaîne de traitements	87
3.4.1 Alignement épipolaire du motif	89
3.4.2 Distorsion du motif	89
3.4.3 Correction de la distorsion de l'image	90
3.4.4 Rectification de l'image	91
3.4.5 Recalage épipolaire (ENDO)	92
3.4.6 Segmentation des franges	93
3.4.7 Estimation des champs de déformation	94
3.4.8 Classification des primitives	96
3.4.9 Construction du graphe d'adjacence	96
3.4.10 Décodage	98
3.5 Conclusion	100
Chapitre 4 Reconstruction 3-D	101
4.1 Reconstruction 3D	102
4.1.1 Modèles de caméra	102
4.1.2 Stéréovision	104

4.1.3	Stratification	105
4.1.4	Matrice Fondamentale et géométrie épipolaire	108
4.1.5	Une reconstruction 3-D projective avec contrainte de quasi-orthogonalité	108
4.2	Estimation de la géométrie épipolaire et des distorsions géométriques	113
4.2.1	Estimation du cercle endoscopique	113
4.2.2	Distorsions géométriques de la caméra et du projecteur	113
4.2.3	Résultats de l'étalonnage partiel	118
4.2.4	Création de la carte de correspondance	119
4.2.5	Estimation de la matrice Fondamentale	122
4.3	Précision des reconstructions	126
4.4	Résultats des reconstructions	130
4.4.1	Reconstructions non endoscopiques	130
4.4.2	Reconstructions endoscopiques Ex Vivo	131
4.4.3	Reconstructions mini-invasives In Vivo	137
4.5	Conclusion	142
Chapitre 5 Développements système		165
5.1	Développement matériel pour la vision endoscopique	165
5.1.1	Un système de projection/acquisition vidéo endoscopique mono-trocart	165
5.1.2	Chaîne d'acquisition du signal vidéo	169
5.1.3	Conclusion	169
5.2	Développement logiciel pour la vision endoscopique	170
5.2.1	Une plate-forme toute intégrée	170
5.2.2	Interface graphique	172
5.2.3	Exploitation du processeur graphique	175
Conclusion générale		177
Annexe A Codes optimaux		183
Annexe B Image intégrale		185
Bibliographie		187
Liste des publications		199

Introduction

Avant l'avènement de la chirurgie mini-invasive, lors d'une intervention chirurgicale, pour disposer d'un accès aux organes de l'abdomen, une large incision de celui-ci était réalisée à l'aide d'un scalpel. Les chirurgiens intervenaient alors directement sur les surfaces organiques pour effectuer leur intervention. Longue période d'hospitalisation, risques importants d'infections, douleurs post-opératoires, les conséquences de telles interventions pouvaient être lourdes pour le patient et pour le personnel hospitalier. Depuis la fin des années 1980, la pratique de la chirurgie mini-invasive (CMI) s'est considérablement développée et a permis de réduire de façon très importante tous ces traumatismes. En chirurgie mini-invasive, le praticien n'a plus une visualisation directe sur les scènes intra-abdominales. Un endoscope de diamètre réduit, en général inférieur à 10 mm, est inséré dans la cavité de l'abdomen et la visualisation de la scène

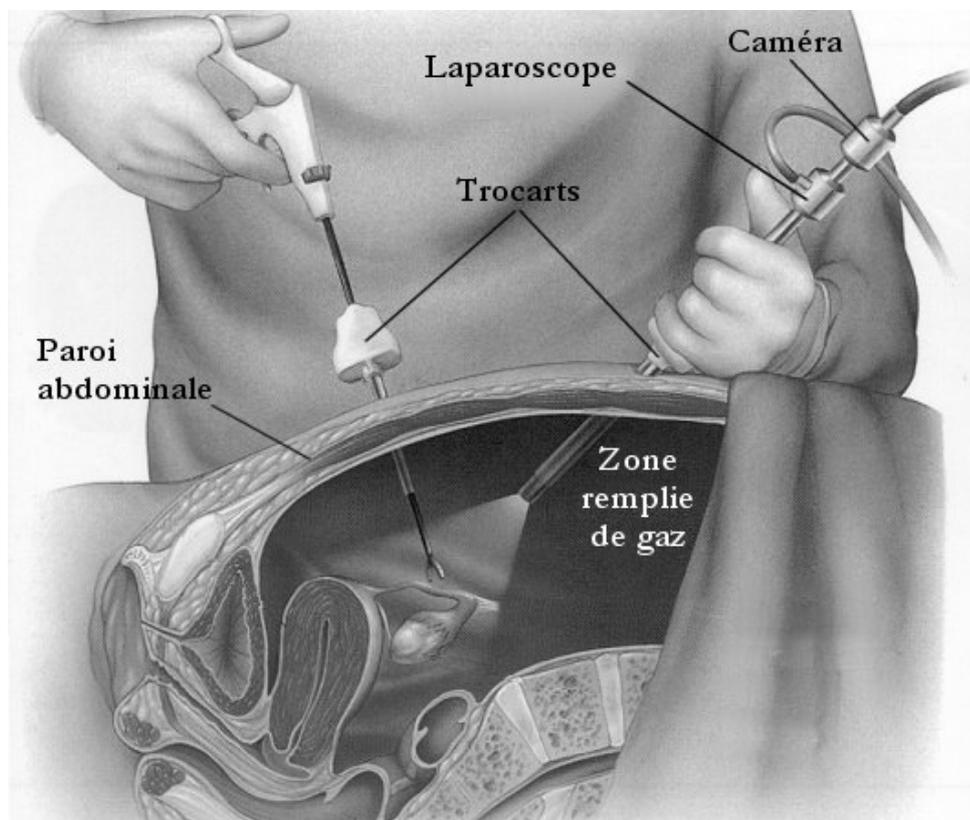


FIGURE 1 – Une intervention chirurgicale mini-invasive.

se fait sur un écran, une tête de caméra étant fixée sur l'embout de l'endoscope de façon à imager l'intérieur de l'abdomen. D'années en années, différents systèmes, de moins en moins invasifs, ont été développés. Trois grandes catégories d'approches coexistent aujourd'hui : l'endoscopie rigide multi-trocarts, l'endoscopie rigide mono-trocart et l'endoscopie flexible. En endoscopie rigide multi-trocarts, un trocart est utilisé pour la visualisation et d'autres servent à l'insertion d'instruments opératoires. Dans le cas mono-trocart, l'ensemble des instruments et optiques sont suffisamment intégrés pour passer dans un trocart unique, l'incision étant généralement réalisée au niveau du nombril et est quasiment invisible après cicatrisation. En endoscopie flexible, les instruments sont insérés par les voies naturelles, c'est la chirurgie transluminale, ou NOTES (*Natural Orifice Transluminal Endoscopic Surgery*), aucune incision extérieure n'est alors réalisée sur le patient.

Quelle que soit la technique employée, l'expérience du chirurgien s'en trouve grandement modifiée par rapport à un accès direct. En particulier, l'affichage sur un écran provoque une perte de la perception de la profondeur. Ainsi il est souvent nécessaire que les instruments soient mis au contact des surfaces organiques pour que la position de ceux-ci dans l'abdomen soit mieux appréciée. Offrir une reconstruction 3-D intra-abdominale permettrait aux chirurgiens de mieux repérer la position de leurs instruments par rapport aux organes lors de l'intervention.

En parallèle du développement des techniques de visualisation et d'opération, la robotique médicale s'est également considérablement développée. Un bras humain tremble davantage qu'un bras robotisé. Ainsi, pour certaines interventions délicates pour lesquelles une grande précision est requise, l'assistance robotisée peut être déterminante. Cette problématique est étudiée par la GMCAO (Gestes Médico-Chirurgicaux Assistés par Ordinateur). Cependant chez le patient endormi, les organes sont en mouvement de façon non rigide par la respiration et les battements cardiaques. Des techniques d'asservissement visuel sont alors étudiées pour maintenir la position relative d'un bras robotisé par rapport à la surface intra-abdominale à opérer. Si des techniques d'asservissement 2-D existent, une reconstruction 3-D de la scène peut permettre d'améliorer ces asservissements, notamment en garantissant la convergence de celui-ci et en augmentant sa rapidité. Cela est d'autant plus vrai que d'importantes discontinuités de courbure sont présentes dans les scènes intra-abdominales. Ainsi, dans ce domaine également, une reconstruction 3-D pourrait être bénéfique.

D'autre part, l'anatomie humaine est aujourd'hui bien connue et on sait créer des modèles 3-D des organes des patients dans une étape pré-opératoire en utilisant l'imagerie IRM, par exemple. Ces modèles peuvent être utilisés pour préparer l'opération réelle à travers des essais virtuels. Des applications de réalité augmentée seraient alors possibles si une reconstruction 3-D était réalisée lors de l'opération réelle. Les modèles virtuels pourraient ainsi être recalés avec les reconstructions per-opératoires.

Enfin, la détection d'excroissances sous-surfaciques anormales, dues à la présence de tumeurs par exemple et invisibles par une simple perception 2-D, pourraient également être détectées par une métrologie tri-dimensionnelle.

Comme nous venons de le voir, les motivations pour une reconstruction 3-D intra-abdominale sont multiples, et peuvent être déterminantes dans plusieurs domaines.

Ce travail de thèse, qui a duré trois années, a été l'occasion d'étudier, d'analyser, de concevoir, de réaliser et de tester un système de vision actif, stéréoscopique, faisant usage d'une projection d'un motif de lumière structurée codée, pour effectuer des reconstructions 3-D en-ligne de scènes intra-abdominales, en endoscopie rigide mono-trocart, dans un contexte mini-invasif. De nombreuses contributions scientifiques sont venues jaloner ce travail. Celles-ci portent aussi bien sur les différentes phases de conception numérique, puis analogique, du motif de lumière

structurée, que sur son exploitation pour mettre en correspondance de façon robuste les indices visuels projetés par une optique du système stéréo avec ceux capturés par la caméra attachée à l'autre optique. Elles portent enfin sur les reconstructions 3-D effectuées à partir de ces appariements. Ce manuscrit relate l'ensemble des contributions proposées et des réalisations effectuées.

Les nombreuses perturbations, tant géométriques que spectrales, présentes dans les scènes intra-abdominales, nous ont mené à nous concentrer sur la robustesse de la mise en correspondance. A ce titre, l'invariance projective de la géométrie épipolaire a été exploitée dès l'étape de conception du motif conjointement avec un codage possédant une forte redondance et/ou utilisant très peu de symboles. Nous avons ainsi introduit les premiers motifs alignés sur la géométrie épipolaire. La redondance du codage par voisinage spatial est assurée par une contrainte d'unicité respectant une distance de Hamming minimale désirée entre les mots de codes associés aux indices visuels.

Ce manuscrit s'articule autour de cinq chapitres et une conclusion générale.

1. Les travaux existants sont abordés au premier chapitre. Dans un premier temps les contributions en vision endoscopique sont présentées. Puis les principales approches utilisées en lumière structurée sont étudiées. Enfin nous motivons l'approche générale : l'alignement épipolaire du motif qui sera détaillé dans la suite.
2. Le chapitre 2 traite de la partie numérique de la conception du motif et plus particulièrement de son codage à l'aide d'un algorithme de type *brute-force*. Après avoir discuté de différents aspects théoriques du codage, quatre contributions sont proposées. Celles-ci portent successivement sur le comportement de la recherche effectuée par l'algorithme de codage, la façon dont est effectué le test d'unicité, puis les deux dernières contributions présentent deux manières d'exploiter la géométrie épipolaire selon la position de l'épipole. Celles-ci permettent de diminuer considérablement les contraintes portant sur le problème de codage et mènent à des solutions inédites. De nombreux résultats sont reportés et comparés à l'état de l'art pour mesurer les progrès effectués. L'utilisation de la distance de Hamming est au coeur de ce chapitre, car elle permet de mesurer la redondance et donc la robustesse des codes générés.
3. Dans le chapitre 3, la représentation physique du motif est présentée. Tout d'abord une étude est menée dans le but de caractériser les déformations des voisinages de primitives dans une image rectifiée en fonction de l'orientation locale des surfaces des objets de la scène sur lesquels ces voisinages se projettent. Celle-ci apporte des clefs pour le choix des primitives présentées par la suite, puis, pour leur organisation sur le plan image du projecteur. L'alignement des primitives le long des lignes épipolaires est également étudié. Finalement la chaîne de traitements exploitant cette conception est présentée.
4. Le chapitre 4 est consacré aux reconstructions 3-D. Après avoir rappelé les principaux modèles utilisés en vision stéréo, une reconstruction projective possédant une contrainte d'orthogonalité est présentée. L'évaluation et la correction des distorsions géométriques de la caméra et du projecteur sont étudiées. En particulier, nous présentons une nouvelle méthode d'estimation des distorsions du projecteur permettant de compenser a priori les distorsions de celui-ci. Enfin, de nombreuses validations qualitatives et quantitatives de reconstructions 3-D sont reportées et comparées à l'état de l'art. Celles-ci sont réalisées avec un système stéréo classique, puis avec le système miniaturisé sur des scènes ex-vivo, et finalement sur des scènes in vivo en situation mini-invasive.
5. Le chapitre 5 présente les différents développements système mis en oeuvre au cours de cette thèse. Dans un premier temps, les développements matériel, portant sur la conception et la réalisation du premier système stéréo endoscopique de projection/acquisition

vidéo mono-trocart sont détaillés. Les développements logiciel qui ont mené à la création d'une plate-forme toute intégrée pour exploiter les solutions proposées dans cette thèse sont ensuite présentés.

6. Le rapport se termine par une conclusion générale dans laquelle nous reprenons l'ensemble des contributions proposées dans cette thèse ainsi que les résultats obtenus dans les différentes étapes de conception du motif et lors de son utilisation pour des reconstructions 3-D en général et en situation *in vivo* en particulier. Finalement, différentes perspectives sont données à court et long terme pour améliorer le système actuel et l'exploiter dans des applications diverses.

Chapitre 1

Travaux antérieurs et motivations

*Hâtez-vous lentement et sans perdre courage
Vingt fois sur le métier remettez votre ouvrage
Polissez-le sans cesse et le repolissez
Ajoutez quelquefois, et souvent effacez.*

(Nicolas Boileau)

Dans ce chapitre nous présentons dans un premier temps les approches utilisées en vision endoscopique. Après une description qualitative des difficultés principales rencontrées dans les scènes intra-abdominales, les approches utilisant des systèmes de vision monoculaires, binoculaires passives puis binoculaires actives et Time-of-Flight sont étudiées. Dans un deuxième temps, nous introduisons l'utilisation d'une projection de lumière structurée et les différentes voies utilisées dans la littérature pour la conception de motifs. Les différentes stratégies de codage, 1-D, 2-D, par multiplexage temporel, sont analysées ainsi que la façon dont ces codages sont exploités : par la projection de franges colorées, de motifs 2-D, la projection successive de plusieurs motifs. Certains développements récents, dont des motifs adaptatifs et le dispositif Kinect, sont également présentés.

Enfin, l'homographie des lignes épipolaires d'un système stéréo est présentée. Cette invariance projective servira de fil rouge à certaines contributions présentées dans la suite de ce rapport, dont l'alignement du motif sur la géométrie épipolaire.

1.1 Reconstructions 3-D en chirurgie mini-invasive

Différents systèmes de vision ont été exploités pour des applications endoscopiques. Ceux-ci peuvent être classés en trois catégories : les systèmes monoculaires, les systèmes stéréo passifs et les systèmes stéréo actifs. Les systèmes proposés sont généralement conçus pour effectuer des reconstructions en temps-réel (> 25 par seconde), sinon en-ligne (quelques secondes par reconstruction maximum). En effet, de par les mouvements du patient, sa respiration et les battements cardiaques, les organes sont en mouvement et se déforment. Ainsi, pour que les reconstructions soient exploitables pendant l'opération (per-opératoire), celles-ci doivent être réalisées rapidement.

Avant d'étudier ces systèmes, nous allons mettre en avant quelques propriétés qualitatives des scènes intra-abdominales et des optiques endoscopiques utilisées pour les imager.

1.1.1 Scènes intra-abdominales

Les scènes intra-abdominales peuvent être complexes, différents organes non rigides et humides y sont présents. De nombreuses perturbations géométriques et spectrales s'y manifestent (Fig. 1.1). Nous notons par exemple, parmi les ***perturbations spectrales*** :

- Le **réseau sanguin** provoquant des discontinuités importantes des propriétés spectrales locales.
- Les **spéularités** nombreuses dues à la moiteur des tissus et surfaces organiques.
- La **diffusion sous surfacique** due à la transparence des tissus.
- Les **réflexions inter-surfaciques** sur les tissus lisses et humides (effet miroir).
- Les **inhomogénéités** importante de la réflectance des surfaces (voir également, les variations de réflectance spectrale entre l'estomac et le côlon [Sambongi00]).
- Le **bruit important** quand l'éclairage est faible, ce qui est le cas quand le motif est projeté (voir Chapitre 5).

Les ***perturbations géométriques*** sont également importantes dans les scènes intra-abdominales :

- Les **discontinuités de courbure** sont nombreuses, en particulier à l'interface entre les organes.
- Les **occultations** des surfaces les unes sur les autres ou dues à l'apparition d'un instrument chirurgical dans le champ de la caméra et/ou du projecteur.

La profondeur de travail se situe dans une plage de 1 à 20 cm, avec 10 cm en moyenne. Chez le patient endormi, la respiration et les battements cardiaques sont les principales sources des mouvements des organes intra-abdominaux.

1.1.2 Optiques endoscopiques

Nous verrons par la suite que plusieurs types d'optiques, mono ou binoculaires, peuvent être employés. D'autre part, contrairement aux systèmes de vision plus classiques, il est courant que ces optiques possèdent une orientation différente de celle de l'axe optique. Cela permet selon les besoins du praticien de pouvoir observer une région d'intérêt différente. Mais cela a également des conséquences en vision, en particulier, cela crée des distorsions tangentielles, sujet sur lequel nous reviendrons au chapitre 4. Une autre propriété importante des optiques utilisées est leur taille, très réduite pour des raisons d'intégration. Cela aura également des conséquences importantes, en particulier sur l'importante quantité de lumière qu'il est nécessaire de projeter

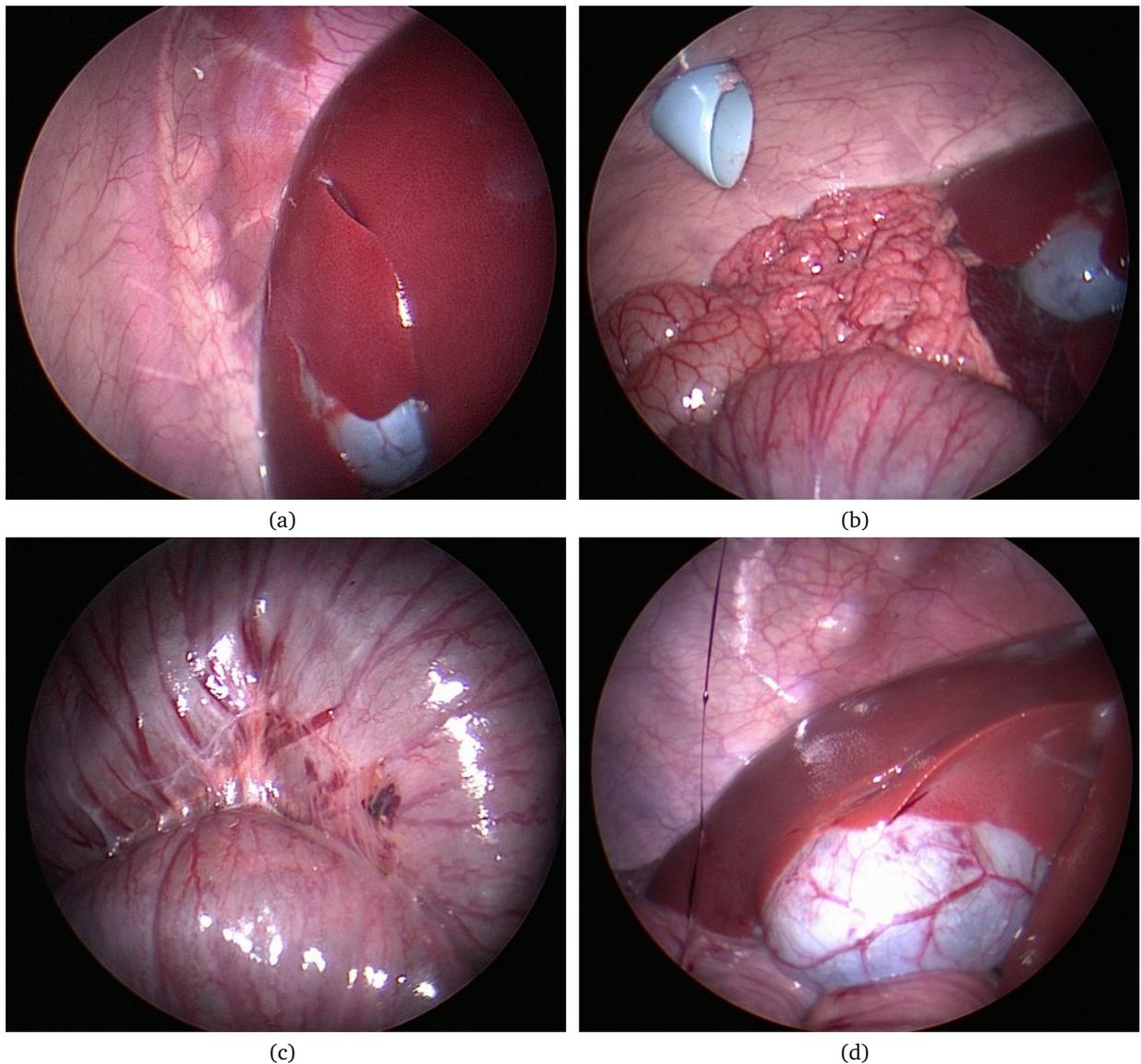


FIGURE 1.1 – (a - d) Exemples de scènes intra-abdominales, de multiples discontinuités géométriques et spectrales sont présentes.

sur la scène pour que l'acquisition puisse se faire avec peu de bruit.

Il existe principalement deux grandes familles d'endoscopes : les endoscopes rigides et les endoscopes flexibles.

- Les **endoscopes rigides** (Fig. 1.2.a) sont formés d'un tube métallique de 3 à 10 millimètres de diamètre et de 15 à 50 cm de longueur. Ils sont insérés dans la cavité abdominale à travers un trocart et leur diamètre peut varier de quelques millimètres à 1 cm. Ils peuvent être associés à des têtes de caméra dotées de capteurs Full-HD pour une très bonne définition de l'image.
- Les **endoscopes flexibles** (Fig. 1.2.b), apparus plus récemment, peuvent être insérés directement par des voies naturelles, et peuvent se déformer pour suivre les courbures internes

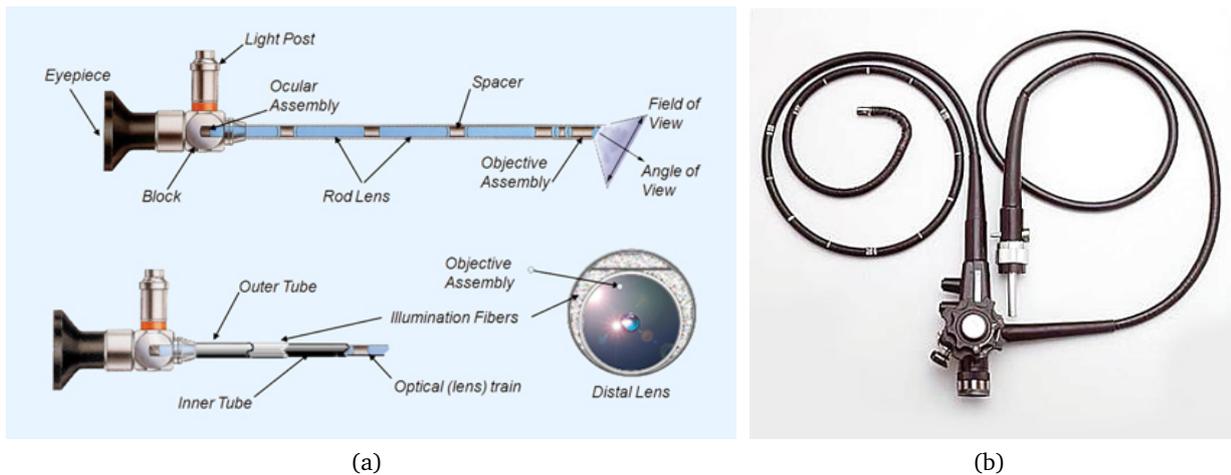


FIGURE 1.2 – (a)Anatomie d'un endoscope rigide.(b) Un endoscope flexible.

de ces organes (trachée, côlon, oesophage), ne laissant ainsi aucune cicatrice apparente au patient. Ils sont utilisés en chirurgie transluminale. L'image est alors formée par un faisceau de fibres optiques (ils sont également appelés fibroscopes), chaque fibre produisant 1 point lumineux. Il y a en général quelques milliers de telles fibres, ainsi la résolution obtenue avec ces dispositifs est bien inférieure à ce qui peut être obtenu avec des endoscopes rigides. Des recherches actives sont menées aujourd'hui en robotique médicale pour la commande de tels dispositifs, bien plus difficiles à modéliser que des endoscopes rigides (voir [Dupont10] et [Kallem09], par exemple).

1.1.3 Vision monoculaire

Dans [Caban04], une méthode est proposée pour retrouver les positions 3-D de l'endoscope et des instruments de chirurgie à partir d'une seule vue. Pour cela des connaissances sur la forme de ces instruments sont utilisées. Notamment la conformation tubulaire des instruments, adaptée à leur insertion dans la cavité intra-abdominale, y est exploitée. Cependant aucune reconstruction de surface abdominale n'est présentée.

1.1.4 Vision stéréo passive

Avec un endoscope monoculaire, la perception de la profondeur se fait difficilement, et il est courant que les praticiens doivent mettre leurs instruments au contact des surfaces intra-abdominales pour mieux apprécier la position de ceux-ci dans la troisième dimension. Cela avait été mis en évidence dans [Durrani95] et des systèmes endoscopiques binoculaires mono-trocart, comme ceux équipant le système robotisé Da-Vinci (Fig. 1.3), ont été adoptés par certaines équipes médicales. De telles optiques ont surtout permis, en vision, d'utiliser la correspondance stéréo pour retrouver la profondeur. La plupart des contributions suivantes en font usage.

Dans [Thormaehlen02], la mise en correspondance des indices visuels segmentés est effectuée par une corrélation croisée normalisée de blocs de dimension 15×15 . La correspondance conservée est alors celle pour laquelle la corrélation la plus élevée est obtenue. Certaines mauvaises correspondances sont détectées lors de l'estimation des mouvements de la caméra avec un algorithme robuste de type *RANSAC* (Random Sample Consensus).

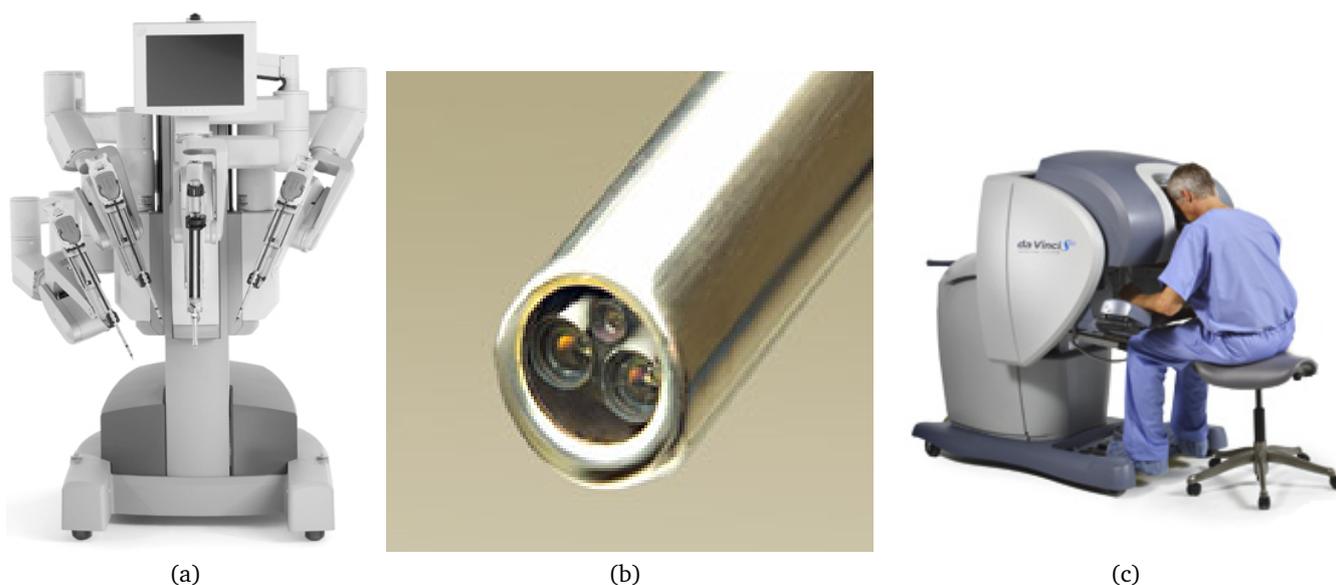


FIGURE 1.3 – Les différentes composantes du système Da-Vinci. (a) Bras articulés manipulant les instruments de chirurgie. (b) Endoscope rigide binoculaire. (c) Système d’observation par stéréovision et de commande à distance.

Dans [Hu07], l’endoscope binoculaire du système Da-Vinci a été utilisé. Des points d’intérêt sont identifiés, mis en correspondance et suivis dans les images droite et gauche. Cette mise en correspondance spatiale a été complétée par une analyse temporelle utilisant la *Structure par le mouvement*. Cependant les résultats reportés montrent que qualitativement, les discontinuités de courbure ne peuvent pas être correctement capturées. En effet, comme souvent avec les méthodes passives, la distribution des points d’intérêt n’est pas suffisamment homogène.

Très récemment, dans [Röhl11], la même optique a été utilisée pour effectuer un recalage non-rigide temps-réel avec une précision millimétrique de modèles pré-opératoires (Fig. 1.4). Les spéularités sont détectées par une méthode de seuillage et une mise en correspondance dense (par pixel) y est effectuée par une méthode appelée *Hybrid Recursive Matching* (HRM), proposée initialement par [Atzpadin04]. Dans cette méthode, une disparité est calculée pour chaque pixel de l’image. Des informations du voisinage temporel et spatial sont utilisées de façon récursive pour générer la carte de disparité. Une différence relative d’intensité est proposée, au lieu d’une différence absolue (SAD) dans la contribution initiale, pour effectuer la mesure de similarité entre les blocs. Cela permet de prendre en compte, jusqu’à un certain degré, les variations d’intensité de la scène.

Ces derniers résultats sont prometteurs mais butteront sur les inconvénients classiques rencontrés en vision passive : si les objets ne sont pas suffisamment texturés, ou si ces textures sont répétitives, comme cela peut être le cas sur foie par exemple (Fig. 1.1.a), ces approches seront mises en défaut soit par le manque d’indices visuels présents, soit par l’apparition de solutions multiples pour leur mise en correspondance.

1.1.5 Vision stéréo active

En vision active, l’un des dispositifs du système stéréo est utilisé comme source de lumière. Différents types de sources sont possibles comme les projecteurs vidéos ou les sources laser. Dans

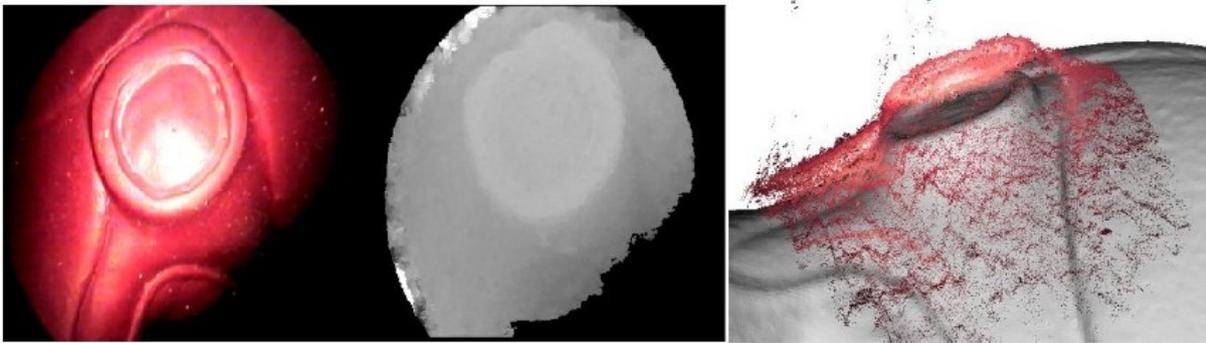


FIGURE 1.4 – Image capturée, disparité puis recalage du modèle 3-D dans [Röhl11].

le cas de la vision endoscopique, les sources laser ont largement été utilisées car elles permettent d'émettre une forte puissance lumineuse, elles possèdent un faible encombrement et une netteté sur une grande plage de profondeur, davantage que les dispositifs utilisant des lentilles optiques. Comme il n'existe pas de tels systèmes endoscopiques dans le commerce, les différentes équipes ont souvent dû confectionner un prototype pour valider leurs approches in-vivo, c'est également notre cas.

Systemes exploitant un projecteur laser

Projection laser directe

Dans le système proposé par Hasegawa [Hasegawa01], figure 1.6, un motif laser binaire encodé spatialement par déphasage est utilisé. Un des deux canaux optiques d'un endoscope binoculaire est utilisé pour la projection. La méthode de projection limite l'acquisition à 0.6 images par seconde, ce qui est peu pour les applications ciblées. De plus, aucune validation In-vivo n'est montrée.

Dans [Krupa03] un guidage d'instruments chirurgicaux est réalisé par asservissement visuel grâce à la projection, sur les surfaces abdominales, de 4 points laser dont les faisceaux sont parallèles (Fig. 1.7.b). Ce guidage permet le déplacement de l'instrument vers une position désirée de manière sécurisée. La localisation de l'instrument dans l'image et dans la scène a pu être réalisée grâce au marquage de celui-ci avec trois LEDs.

Optiques diffractives

Plusieurs contributions exploitant les optiques diffractives ont été proposées. Celles-ci permettent la création d'une image d'un motif à partir d'une source laser incidente grâce à un *élément diffractif*. Il prend le plus souvent la forme d'une petite plaque de verre sur laquelle est gravée la transformée de Fourier du motif par microphotolithographie. L'image du motif se forme alors par diffraction du rayon incident (Fig. 1.5.a). Cette méthode a été utilisée dans un contexte endoscopique par [Chan03] (Fig. 1.6.a) et récemment dans [Albitar09a]. Dans [Chan03], une simple grille de points est projetée à travers l'un des canaux d'un endoscope binoculaire. La mise en correspondance est alors effectuée en exploitant la contrainte épipolaire du système stéréo. De bons résultats ont été reportés In-vivo dans une cavité orale. Mais l'absence de codage du motif rend la mise en correspondance sensible aux perturbations présentes dans des scènes intra-abdominales. La société Storz propose également un système de métrologie basé sur une approche similaire (Fig. 1.5.c).

Un motif plus évolué, reposant également sur une projection laser par diffraction optique a été récemment proposé par [Albitar07]. Il y est fait usage d'un motif monochrome codé par voisinage spatial. Une matrice pseudo-aléatoire composée de 27×29 éléments utilisant 3 symboles

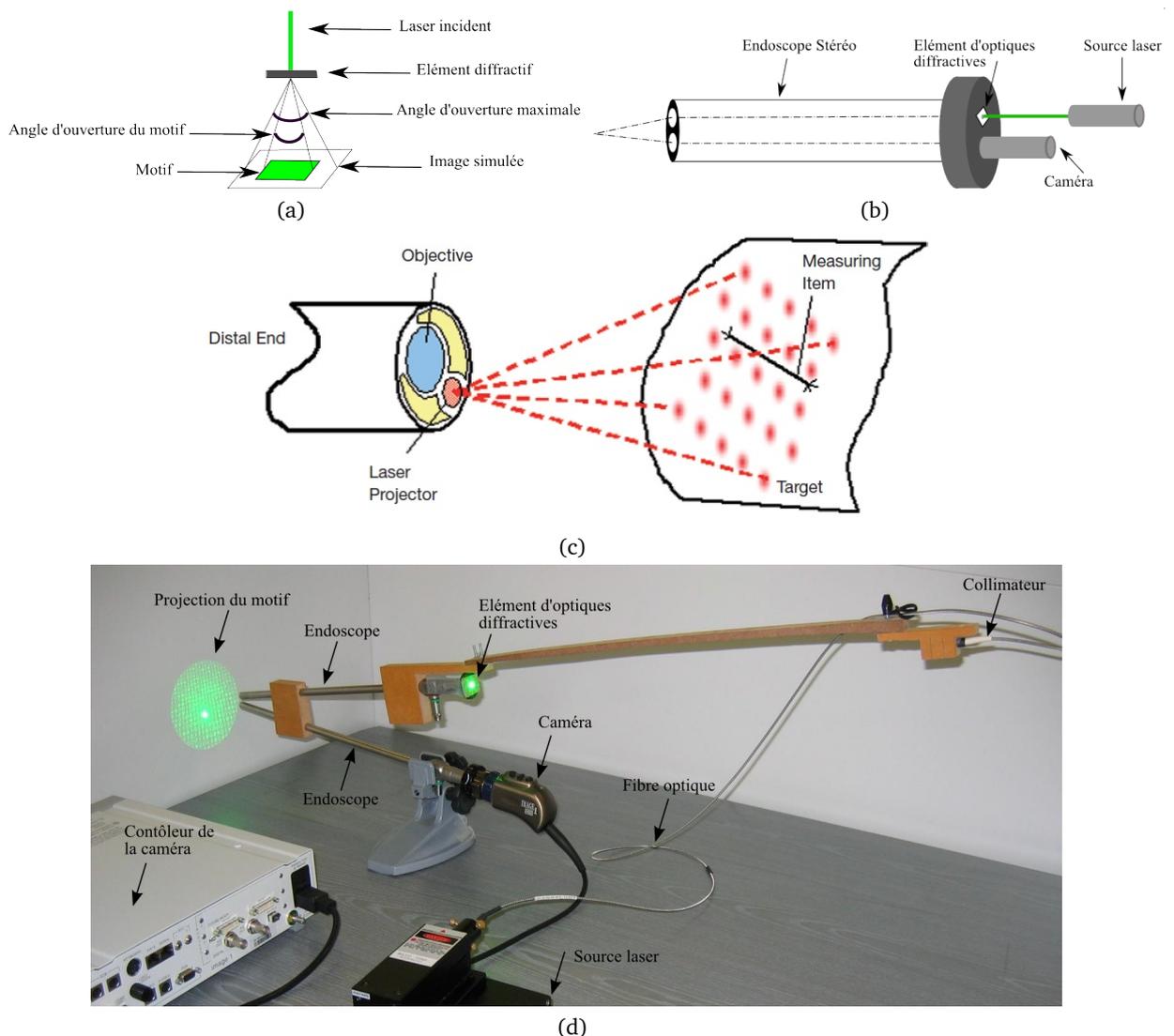


FIGURE 1.5 – Optiques diffractives : a) Principe, b-d) Conception et réalisation dans [Albitar09a], c) Système de métrologie endoscopique proposé par la société Karl Storz exploitant une projection de points

différents est générée en assurant l'unicité de tous les voisinages 3×3 . Cela permet une mise en correspondance par une analyse locale des voisinages et offre donc davantage de robustesse aux occultations. De plus le codage est symétrique central, une contrainte qui a été rendue possible par l'usage d'un algorithme de type *brute-force* (voir section suivante). Cette contrainte de symétrie permet de doubler (de 40% à 80% de l'énergie lumineuse incidente) l'efficacité lumineuse à l'issue de la diffraction du faisceau laser. Enfin les symboles sont représentés par des formes différentes, non par des couleurs différentes comme c'est souvent le cas en lumière structurée. Cela a permis d'une part l'utilisation d'une source laser, donc monochromatique et d'autre part une certaine robustesse aux perturbations spectrales. L'utilisation de la couleur verte a été choisie car le vert se trouve dans la plage de spectre qui est la mieux réfléchiée dans les scènes intra-abdominales. Des validations In-vivo ont été réalisées sur des organes de porc mais en situation non mini-invasive. En effet, le prototype réalisé (Fig. 1.5.d) n'était pas suffisamment intégré

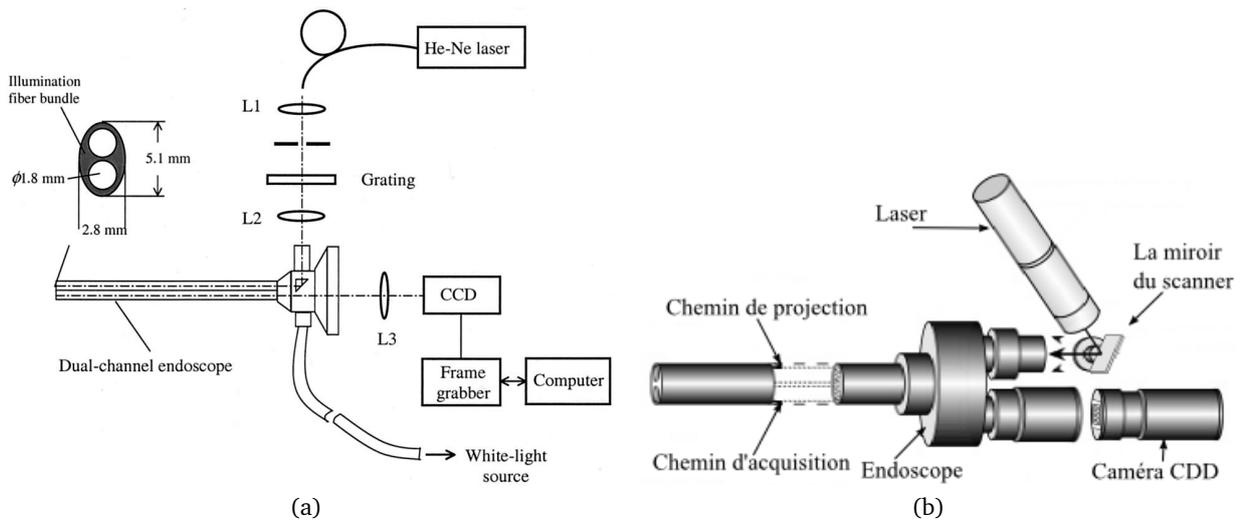


FIGURE 1.6 – Systèmes endoscopiques proposés dans [Chan03] (a), et dans [Hasegawa01] (b).

pour une insertion par un trocart. Les expériences In-vivo ont cependant révélé un manque de robustesse dû, notamment, à une forte présence de bruit dans les images acquises. Une précision de reconstruction de l'ordre du millimètre a été reportée pour d'autres scènes.

Illumination par fibres optiques

Récemment, [Yanjun10] et [Clancy11] (Fig. 1.7.a) ont proposé des illuminations par fibres optiques exploitant un motif de lumière codé spectralement. Si l'illumination par fibres optiques peut laisser entrevoir une utilisation pour les endoscopes flexibles, une telle technique mène inévitablement à une résolution très basse par rapport à ce qui peut être obtenu avec les autres types d'approches.

Systèmes exploitant une projection vidéo

Dans [Keller00] un premier système basé sur la projection de lumière structurée à l'aide d'un projecteur vidéo a été proposé. Le système était composé de deux endoscopes, l'un en émission, l'autre en réception. Six motifs binaires étaient projetés et acquis à haute vitesse (jusqu'à 180 images par seconde) et nécessitaient un système de synchronisation pour le décodage. L'approche utilisée pour la mise en correspondance exploitait donc un multiplexage temporel (voir section suivante). Cependant le mécanisme de synchronisation s'est révélé défectueux et aucune démonstration n'a pu être effectuée sur des objets réels. De plus le système réalisé était composé de deux endoscopes (Fig. 1.8.a) et ne pouvait donc se satisfaire d'un seul trocart. Quelques années plus tard, dans [Keller03], a été proposé un système mono-trocart dans lequel le flux photonique en émission était différencié de celui en réception par une polarisation de la lumière (Fig. 1.8.b), mais aucun prototype n'a été réalisé.

1.1.6 Time-of-Flight

La technique du *Time-of-Flight* a récemment été testée pour la première fois dans un contexte endoscopique par [Penne09]. Dans les techniques *Time-of-Flight* (ToF), la carte de profondeur est créée par analyse du temps de retour de la lumière. C'est une sorte d'échographe pour photons. Cette technique a l'avantage d'offrir une reconstruction par pixel et de n'utiliser qu'une seule optique endoscopique, mais un capteur ToF est nécessaire. Les capteurs ToF, disponibles

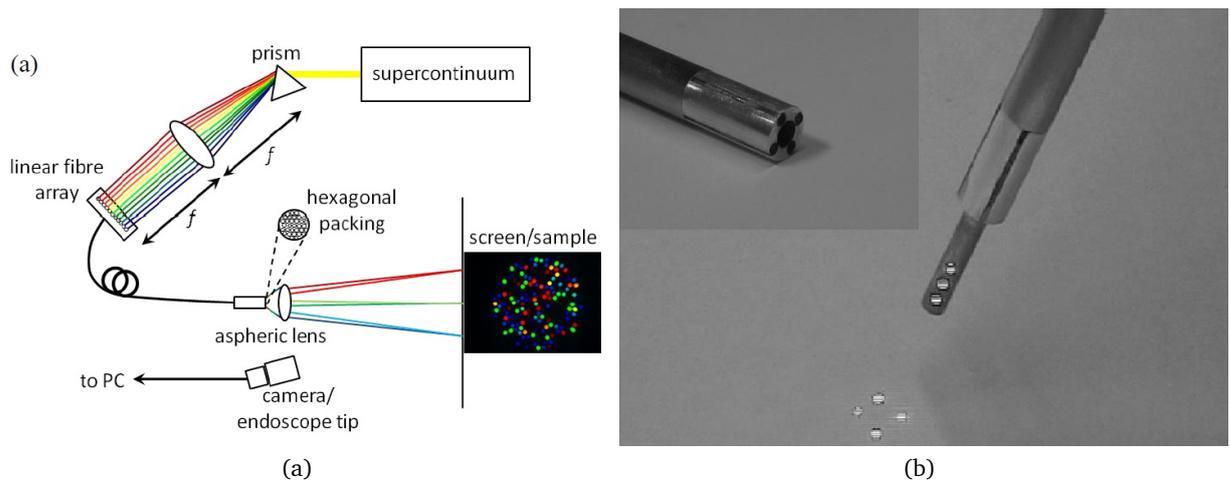


FIGURE 1.7 – (a) Système flexible proposé dans [Clancy11]. (b) Dispositif utilisé dans [Krupa03].

depuis une dizaine d'années à peine, sont constitués d'une matrice de pixels et d'une unité d'illumination externe dans laquelle l'intensité de la lumière est modulée. Chaque pixel est alors synchronisé avec l'unité d'illumination et le délai de phase, dépendant du délai de propagation entre la lumière émise et celle réfléchi, est calculé. Le système testé avait une matrice de 68×48 pixels et opérait à une vitesse de 20 fps. Les reconstructions 3-D ont été effectuées en exploitant les contraintes géométriques d'un cube de dimensions connues placé dans la scène et ont été réalisées avec une précision moyenne mesurée de 0.89 mm pour des scènes se trouvant à une distance moyenne de 3 cm. Rapportée aux distances moyennes constatées en chirurgie intra-abdominale, environ 10 cm, la précision est alors de ~ 3 mm, ce qui est faible comparé à ce que proposent d'autres méthodes de vision plus classiques qui offrent souvent une précision sous le millimètre. De plus, dans [Penne09], il n'est pas dit que cette technique peut être sensible aux nombreuses réflexions inter-organes existantes qui peuvent venir perturber les mesures effectuées par un allongement du trajet des photons.

La vision active offrant plusieurs avantages déjà mentionnés par rapport à la vision passive, c'est vers cette approche que nous nous sommes dirigés. La vision active suppose la conception d'un motif de lumière structurée, que ce motif soit réduit à un point, ou qu'il corresponde à une image, par exemple. Nous allons donc à présent étudier plus en détails les différents travaux existants dans ce domaine.

1.2 Lumière structurée

Contrairement à la vision passive, dans les systèmes de vision active, au moins un dispositif du système est utilisé comme source de lumière. La projection de lumière structurée entre dans cette catégorie. Un motif connu est alors projeté sur la scène à reconstruire et les déformations de ce motif sont capturées par la caméra et mesurées dans le but de retrouver la profondeur (Fig. 1.9). Le motif projeté peut être un simple point, une ligne, ou une image par exemple, les possibilités sont très vastes. Dans un système stéréovision, les indices visuels¹ émis par le projecteur doivent alors être mis en correspondance, ou appariés, avec ceux acquis par la caméra

1. Dans ce chapitre et dans le reste de la thèse, nous emploierons indifféremment les termes d'*indices visuels* et de *primitives visuelles* pour désigner chaque élément du motif projeté. Ceux-ci peuvent correspondre à des points, des lignes, ou d'autres formes diverses.

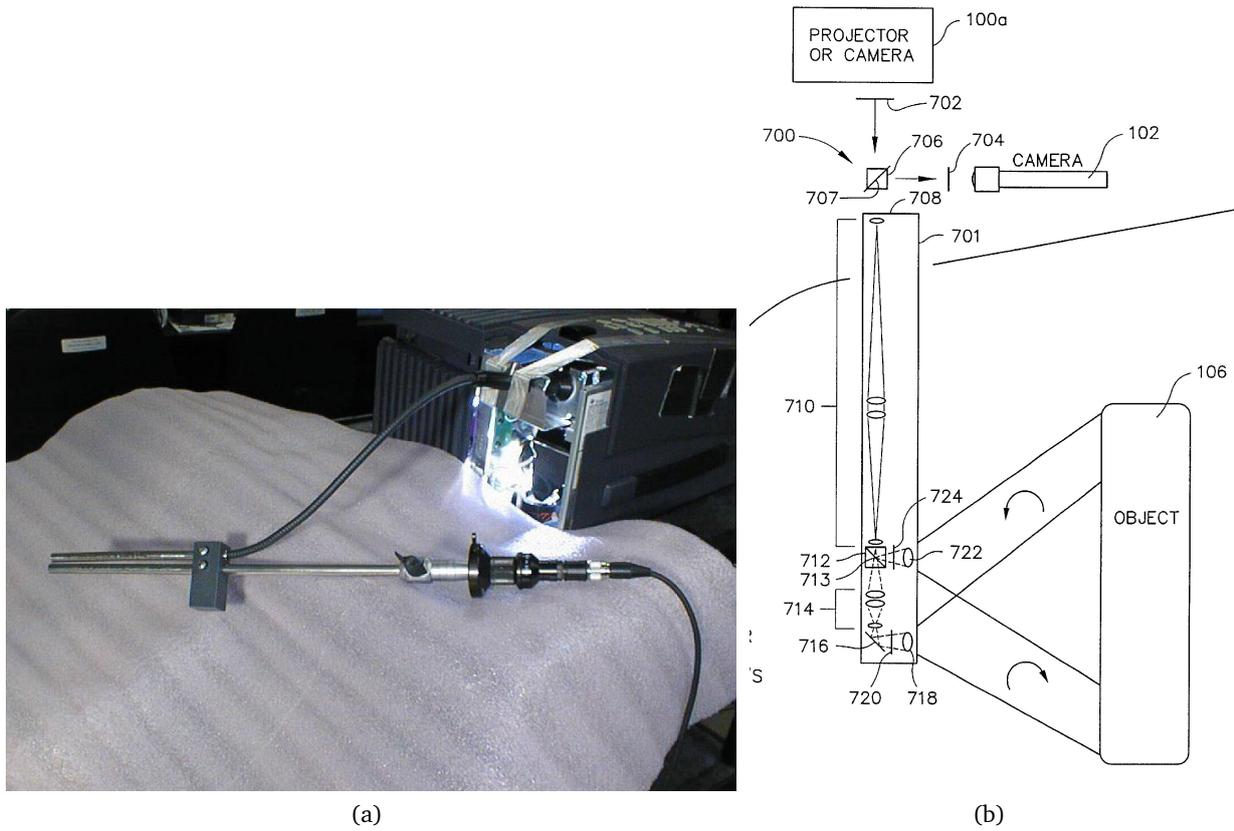


FIGURE 1.8 – a) Système vidéo bi-trocart dans [Keller00]. b) Système vidéo mono-trocart avec lumière polarisée dans [Keller03].

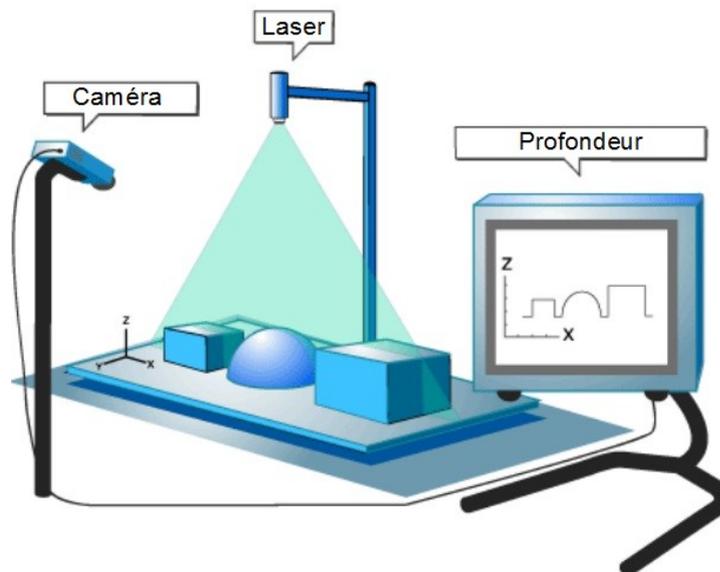


FIGURE 1.9 – Un système stéréo de lumière structurée.

pour que la profondeur puisse être estimée (voir chapitre 4). Il existe également des systèmes utilisant un projecteur et deux ou plus de caméras.

Si en vision passive, la texture des objets de la scène, leurs bords, leurs coins, les discontinuités spectrales, représentent la principale source d'information exploitée pour localiser des points d'intérêt à mettre en correspondance, en vision active, tous ces éléments sont le plus souvent autant de perturbations indésirables qui viennent gêner la segmentation et le décodage du motif projeté dans la scène. Cependant, comme nous le verrons plus en détails dans le chapitre suivant, il est possible, en vision active, d'émettre à une longueur d'onde hors du spectre visible et de diminuer significativement, voire totalement de telles perturbations. De plus les avantages de la projection d'un motif de lumière structurée sont nombreux : la distribution de la reconstruction est bien plus uniforme et souvent plus dense qu'en vision passive et la mise en correspondance plus robuste car le signal projeté est connu et cette connaissance peut être utilisée pour piloter la segmentation et la mise en correspondance des indices visuels projetés. De plus, bien souvent, cette mise en correspondance peut se faire plus rapidement qu'avec les systèmes de vision passifs donnant lieu à des reconstructions calculées en-ligne, voire en temps-réel.

La source de lumière

D'un point de vue de l'information, un signal est alors transmis, par projection de lumière, dans un canal bruité, puis récupéré par une caméra. Mais contrairement à l'échange de documents informatiques, par exemple, on connaît ici le signal initial aussi bien que celui acquis par la caméra. Au lieu de transmettre des données numériques : des "0" et des "1", on transmet des données analogiques : des photons. Différents types de sources de lumière peuvent alors être utilisés. De façon générale, les laser et les projecteurs vidéo sont le plus souvent utilisés. Les lasers offrent l'avantage de produire un faisceau lumineux qui peut être focalisé et dont la projection reste nette sur une grande distance, contrairement aux projecteurs vidéo qui font usage de lentilles qui restreignent cette zone de netteté. Par contre ces derniers possèdent beaucoup plus de flexibilité quant aux types de motifs qui peuvent être projetés. En effet, les laser sont, par définition, monochromatiques. Lors de l'utilisation d'optiques diffractives, évoquée à la section précédente, une gravure du motif est effectuée et celui-ci ne peut donc être modifié, offrant moins de flexibilité qu'avec des projecteurs vidéo. Enfin nous notons que des projecteurs vidéo laser miniaturisés ont fait leur apparition sur le marché depuis 2010 (le *ShowWX+ de Microvision* et le *L1 v2 de AAXA*). Ceux-ci n'utilisent pas d'optiques et sont susceptibles de cumuler les avantages des deux types de projection. Mais pour l'instant ils souffrent d'une puissance lumineuse trop faible (~ 20 lumens contre plusieurs milliers pour des projecteurs vidéos plus classiques) pour la plupart des utilisations en lumière structurée. Mais ces dispositifs sont à surveiller de très près car ils évoluent rapidement. D'autres dispositifs sont envisageables pour l'émission du motif de lumière, par exemple, un simple affichage de motif de lumière structurée sur un écran plat est utilisé dans [Sagawa05] à des fins d'étalonnage.

Applications de la lumière structurée

Les applications exploitant la lumière structurée se développent aujourd'hui dans des domaines très divers. Le contrôle-qualité (Fig. 1.10.a), la reconstruction de sites archéologiques immergés (Fig. 1.10.b), la reconstruction d'empreintes digitales (Fig. 1.10.c) et même les productions artistiques (Fig. 1.10.d), font usage d'une projection de lumière structurée. Il en va de même pour la capture des déformations du visage (Fig. 1.10.e) servant, par exemple, à l'animation de personnages virtuels dans les films d'animation, ou la capture des mouvements du corps entier pour concevoir de nouvelles interactions homme-machine (Fig. 1.10.f).

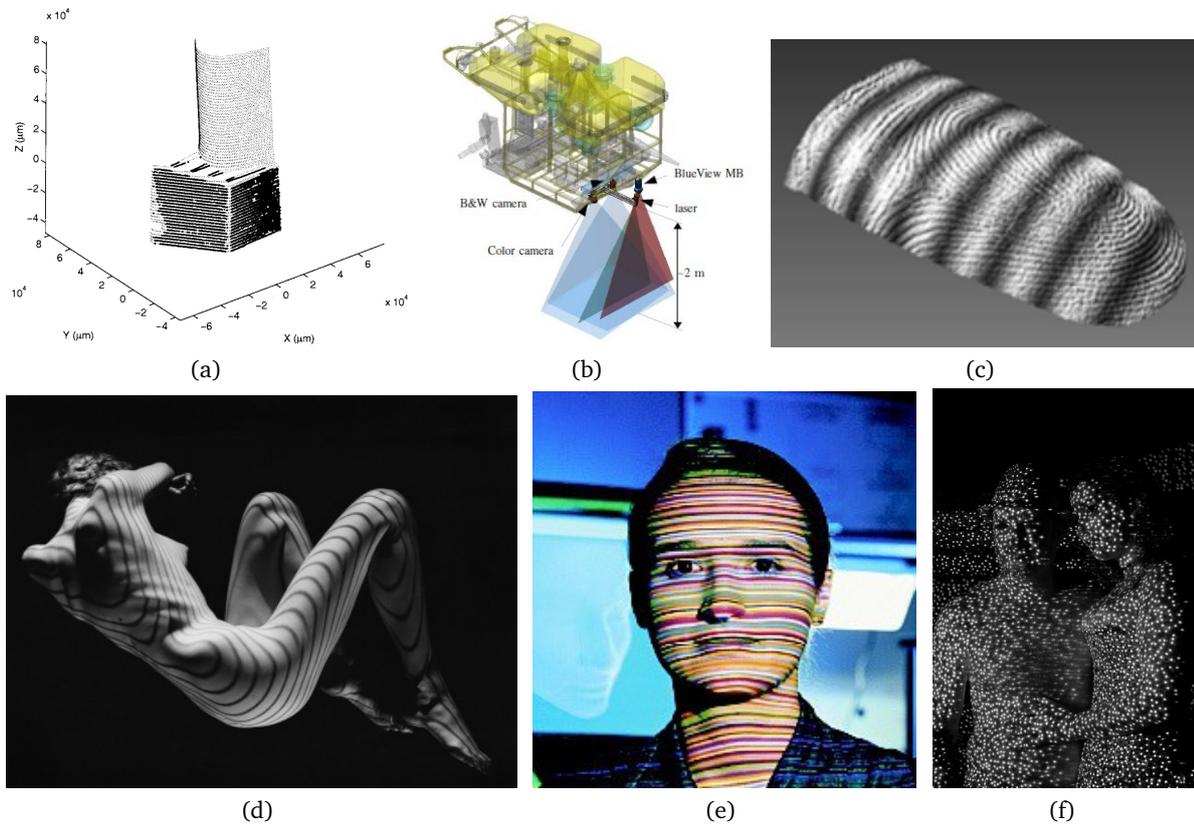


FIGURE 1.10 – Diverses applications de la lumière structurée : (a) Métrologie et contrôle-qualité dans [Graebing02]. (b) Reconstruction de sites archéologiques immergés dans [Roman10]. (c) Reconstruction d'empreintes digitales (système FlashScan3D). (d) Productions artistiques (Fernand Fonssagrives). (e) Reconstruction et reconnaissance faciale. (f) Motif infra-rouge du dispositif Kinect de la société Microsoft.

Codage du motif et mise en correspondance

Si pour des scènes simples, sans occultation ni discontinuité de courbure importante, comme on peut en trouver en vision industrielle, des motifs de lumière structurée non codés peuvent convenir, que ce soit par la projection de bandes (dans [Doignon08], par exemple) ou de points, dans la majorité des cas et pour éviter toute perte de généralité, la possibilité d'occultation doit être envisagée. C'est notamment le cas dans le contexte de la vision endoscopique où les organes possèdent des reliefs importants.

Le codage a pour objectif d'associer à chaque primitive projetée (bande, point, ou toute autre forme géométrique), un code unique qui permettra une mise en correspondance sans ambiguïté si la primitive est correctement décodée lors des traitements faisant suite à l'acquisition. En pratique, de nombreuses perturbations spectrales, géométriques ainsi que le bruit de l'image, sont autant de sources de nuisances pouvant mener à un mauvais décodage. C'est souvent le cas des zones de la scène comportant des in-homogénéités géométriques ou spectrales. Cela peut aussi être le cas pour les objets en dehors de la zone de netteté ou si les primitives ont une taille trop réduite par rapport au bruit du système d'acquisition. Le rapport signal/bruit n'est alors pas suffisant. Pour toutes ces raisons, en pratique, associer un code unique à chaque primitive n'est souvent pas suffisant. Il est possible d'ajouter davantage de redondance au code en lui imposant

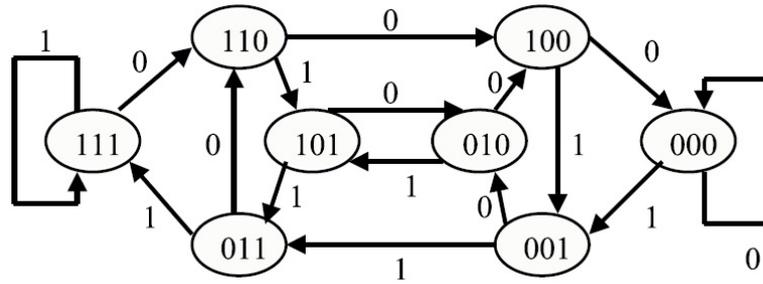


FIGURE 1.11 – Un graphe de De-Bruijn.

une distance de Hamming minimale (voir chapitre 2). Cela permettra ainsi d'envisager des mécanismes de décodage plus robustes.

Travaux antérieurs

De nombreuses études récentes ([Batlle98],[Salvi04],[Ribo05], [Pan10], [Salvi10], [Geng11]) répertorient les nombreux types de motifs proposés dans la littérature. Parmi ceux-ci, [Salvi10] est sans doute le plus complet. Dans cette section nous ne donnons pas une description exhaustive de chacune d'entre elles, le lecteur est invité à consulter les études indiquées. Cependant nous synthétisons ici les grandes approches utilisées, nous mettons à jour les revues citées avec les contributions les plus récentes et nous pointons les avantages et inconvénients de ces approches pour l'application qui nous intéresse : la reconstruction rapide de scènes intra-abdominales.

Il est possible de classer les motifs selon plusieurs critères. Cependant, ceux-ci ayant pour objectif premier d'offrir une bonne mise en correspondance, la stratégie de codage utilisée est souvent considérée pour les différencier.

1.2.1 Stratégies de codage

Il existe de nombreuses stratégies de codage. Celles-ci peuvent exploiter l'espace en une ou deux dimensions et/ou le temps à travers un codage discret ou continu. Certaines approches n'utilisent pas de codage explicite, comme dans [Harvent08] (Fig. 1.12), où un mouchetis est projeté sur la scène à reconstruire et la correspondance est établie par corrélation entre différentes vues issues de capteurs passifs. Il est donc ici nécessaire de disposer d'au moins trois dispositifs optiques, un projecteur et au moins deux caméras, ce qui rend difficile la miniaturisation dans un système endoscopique.

1.2.1.1 Codage par voisinage spatial 1-D

Les séquences de De-Bruijn

Les séquences de De-Bruijn sont au cœur du codage spatial 1-D avec une seule projection de motif. Une séquence de De-Bruijn $B(n, m)$ est une séquence cyclique réalisée dans un alphabet de n symboles différents, dans laquelle tous les mots de code de longueur m sont présents une et une seule fois. Cette unicité des mots de code permet de réaliser une mise en correspondance sans ambiguïté, si la classification des indices visuels se fait sans erreur. Aux séquences de De-Bruijn est associé un graphe (Fig. 1.11) dans lequel tous les mots de codes sont succes-

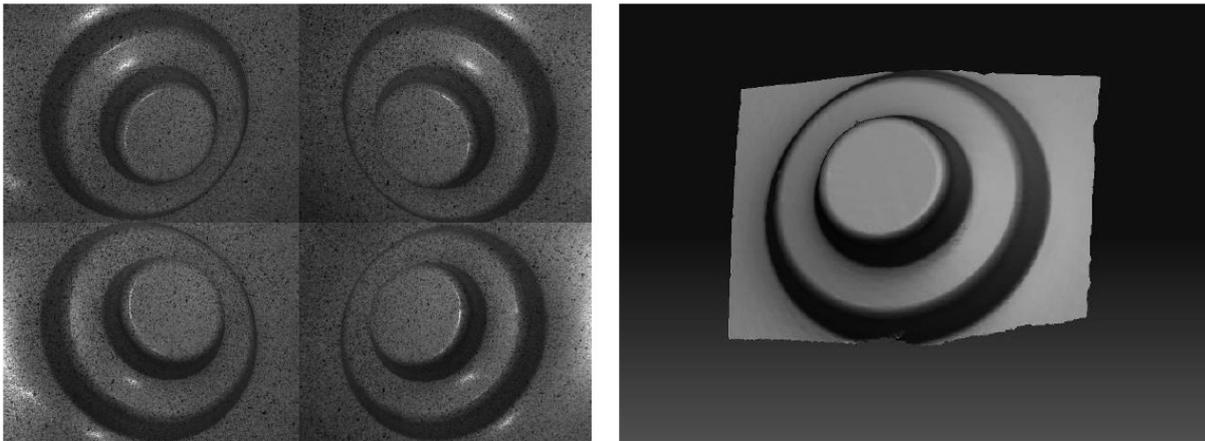


FIGURE 1.12 – Projection d'un mouchetis dans [Harvent08], pour l'inspection de pièces aéronautiques. Embouti vu par 4 caméras (à gauche) et reconstruction 3-D (à droite).

sivement reliés. Par exemple, dans la séquence $B(2, 4) = 0000111101100101$, les $2^4 = 16$ mots de code possibles sont présents, les 3 derniers étant construits par cyclicité de la séquence. Ce type de séquence permet également la constructions de codes de grande taille en introduisant une périodicité dans le codage. Ils sont alors construits par répétition de la séquence. Dans ce cas les voisinages ne sont plus globalement uniques mais seulement périodiquement, toutes les 16 primitives avec la séquence $B(2, 4)$. Cette non-unicité peut être géométriquement justifiée. En effet, chaque primitive se projette dans l'image de la caméra sur la ligne épipolaire associée en fonction de la profondeur des objets de la scène. Dans la plupart des cas, la profondeur minimale et maximale des objets mène à une disparité limitée des primitives dans le plan image. La projection du point dans le plan de la caméra ne parcourt qu'une portion de la ligne épipolaire. Ainsi il est possible de ne pas assurer une unicité tout le long de la ligne épipolaire. Mais ceci se fait au prix d'une perte de généralité car cela dépend de la géométrie du système stéréo utilisé et des propriétés (profondeur maximale/minimale) des surfaces de la scène à reconstruire. A ce codage sont généralement associées des franges de couleurs différentes, mais des niveaux de gris peuvent également être utilisés. Cette technique est, de façon générale, la plus utilisée en lumière structurée codée. De nombreuses contributions ont été proposées et le nombre de nouvelles propositions récentes montrent l'intérêt pour cette approche. C'est le cas par exemple dans [Fechteler08], [Hu09],[Wei10], [Cheng10], [Keeratitivittayanun11] et [Li11].

Certaines contributions permettent de dissocier l'analyse dans le domaine spectral (pour la classification) de l'intensité lumineuse (pour la segmentation) dans [Pages04] et plus récemment dans [Ji09].

Des approches utilisant deux codages 1-D et projetant un motif grillagé ont également été proposées dans [Salvi98] et [Sagawa09]. Dans [Salvi98], l'unicité est assurée sur tout le motif, 8 couleurs ont été nécessaires (4 dans chaque dimension), alors que dans [Sagawa09], seulement 2 couleurs sont utilisées sur des voisinages de 3 franges avec une périodicité toutes les 8 franges. Le décodage est alors effectué par propagation en utilisant conjointement le codage horizontal et vertical. Les structures grillagées permettent une reconstruction dense le long des franges créées dans les deux dimensions. Cependant, notons que la précision de l'intersection entre les franges et les lignes épipolaires ne peut être optimale (si elles sont orthogonales) dans les deux dimensions de ces grilles à la fois. En effet, les 3 intersections de 3 lignes (franges horizontales, franges verticales et lignes épipolaires) dans une image ne peuvent toutes posséder un angle

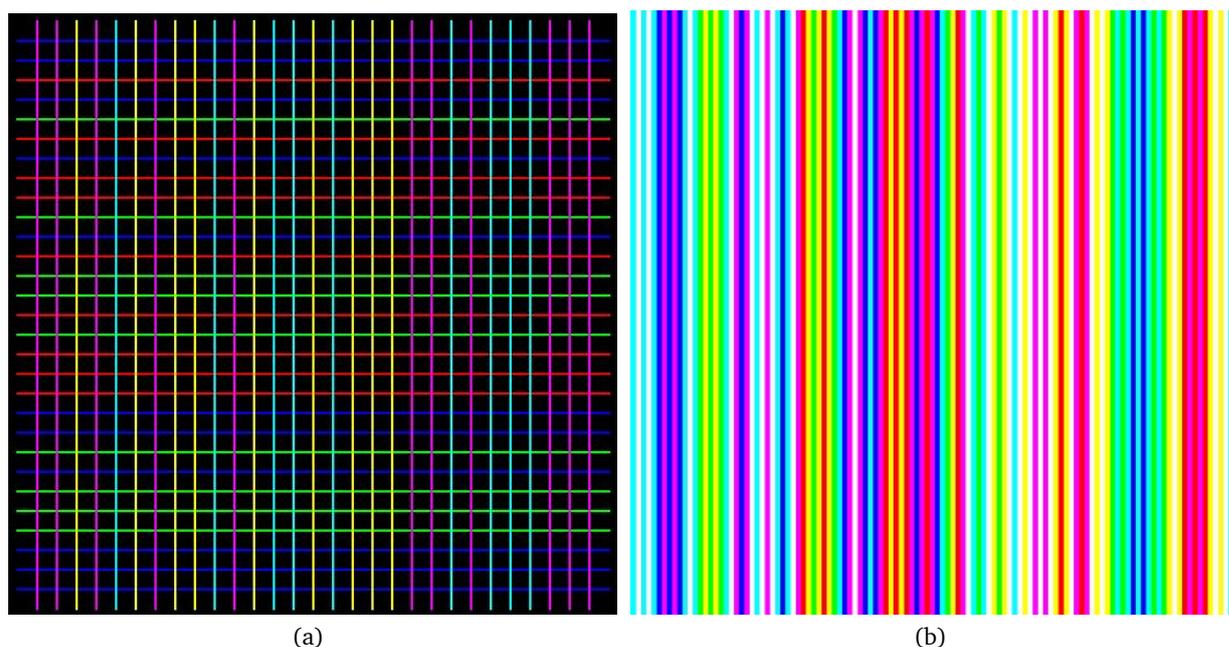


FIGURE 1.13 – Codage 1-D par : (a) une séquence de De-Bruijn dans [Salvi98] (solution analytique), (b) un algorithme de recherche dans Yamazaki (solution "brute-force") [Yamazaki11].

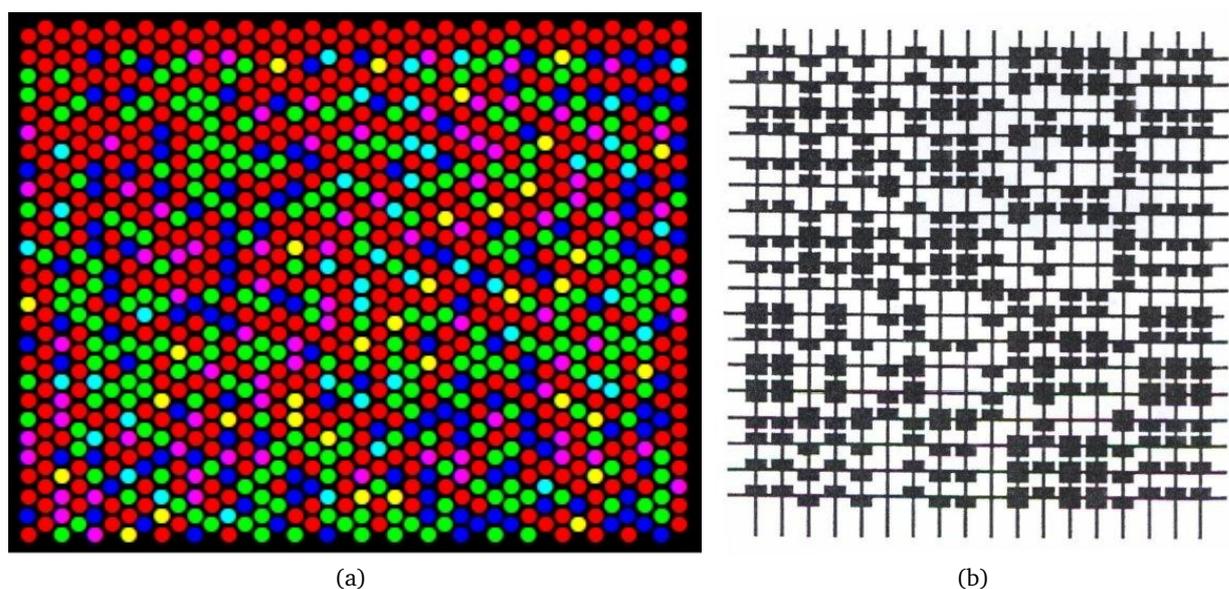


FIGURE 1.14 – Codage 2-D dans (a) [Claes07], (b) [Yee94] .

proche de 90° à la fois (dans un espace euclidien). Soit une des deux orientations de la grille est favorisée, soit une netteté d'intersection moyenne (à 45°) est choisie.

Très récemment [Yamazaki11] (motif de la figure 1.13.b) a proposé d'imposer une contrainte supplémentaire : une distance de Hamming de 1 entre les symboles adjacents. Cela permet d'éviter la présence de bandes noires entre les franges du motif et d'offrir plus de robustesse lorsque l'arrière-plan est éclairé par exemple. Les méthodes de codage analytiques (type séquences de De-Buijn) ne permettent pas d'imposer toutes les contraintes désirées. Notons cependant que

pour cette contrainte en particulier un code de Gray (où chaque code successif ne diffère que de 1 symbole avec le précédent) pourrait être utilisé. Ainsi, le code utilisé dans [Yamazaki11] est généré par une analyse exhaustive, donc grâce à un algorithme de type "brute-force", procédant par essais-erreurs. Nous reviendrons largement sur ce type d'algorithmes dans la suite de ce rapport.

Il a été noté dans plusieurs contributions que ces approches, faisant usage de la couleur, étaient sensibles aux perturbations spectrales. Par exemple dans [Salvi98].

"However, the identification is completely constrained by the kind of scene to be measured and the light conditions. Then, when the light conditions decrease, the projection of colour weakens considerably the use of such kind of pattern".

De plus, les diverses contraintes de réalisation d'un prototype mini-invasif nous ont mené à la création d'un motif monochrome. Pour différencier les indices visuels projetés, il nous reste donc les niveaux de gris et la forme des primitives. Or, comme nous le verrons, le codage 2-D offre plus de liberté pour le choix des formes.

1.2.1.2 Codage par voisinage spatial 2-D

Les images dans lesquelles les motifs sont acquis sont, dans leur grande majorité, en deux dimensions (des images 3-D existent également). L'utilisation de la deuxième dimension pour le codage du motif permet d'utiliser des mots de code de taille beaucoup plus élevée tout en conservant une fenêtre de décodage dans un voisinage restreint. Par exemple, avec un voisinage 3×3 et un codage 2-D, on obtient des mots de code de longueur 9 contre 3 pour le codage 1-D, alors qu'il suffit pour cela de considérer la deuxième dimension de l'image. Cette longueur plus élevée des mots de codes mène à un problème de codage moins contraint qu'en 1-D et donc aboutit à des motifs de plus grande résolution ou utilisant moins de symboles ou permettant davantage de redondance et donc de robustesse. Des mauvaises classifications d'indices visuels peuvent alors être détectées et éventuellement corrigées. En témoigne l'objectif des concepteurs de codes 1-D ou 2-D. Avec le codage 1-D, l'objectif est généralement d'assurer l'unicité des mots de code, alors que pour le codage 2-D on cherche également à obtenir la distance de Hamming minimale (voir chapitre 2) la plus élevée possible dans [Morano98] et [Claes07] par exemple.

Le codage 2-D fera l'objet du chapitre suivant. Nous indiquons simplement ici qu'il existe deux grandes classes d'algorithmes :

- Les algorithmes directs qui sont basés sur une extension en deux dimensions des séquences de De-Bruijn. Ils génèrent des solutions dont les paramètres, notamment le rapport d'aspect de la matrice générée, sont contraints et offrent peu de flexibilité. En outre, il n'est pas possible d'imposer une distance de Hamming minimale arbitraire et donc d'augmenter la redondance du codage. Les premières constructions ont été proposées dans [Etzion88] et ont été exploitées avec certaines variations dans [Griffin92] où toutes les sous-fenêtres ne sont pas utilisées.
- Les algorithmes itératifs, dits "brute-force" procèdent par essais-erreurs. La première contribution a été proposée dans [Morano98] à travers l'utilisation d'un algorithme aléatoire. Ce type d'algorithme permet de définir des contraintes de codage spécifique, comme la symétrie centrale du codage dans [Albitar07], ou leur invariance par rotation dans [Claes08] et dans [Adán09] où des voisinages hexagonaux sont utilisés. Il est également possible d'imposer une distance de Hamming minimale entre tous les couples de mots de code. Nous y reviendrons dans le chapitre suivant.

Le tableau 1.1 résume les propriétés de plusieurs motifs obtenus avec un codage 2-D.

1	2	3	4	5	6	7
Codage spatial						
[Morita88]	PM	NA	24 × 24	2	12	spots noir et blanc
[Griffin92]	PM	NA	18 × 66	3	5	formes
[Morano98]	PSM	1 to 4*	45 × 45	3	9	spots colorés
[Albitar07]	PSM	6 (average)	27 × 29	3	9	formes
[Claes08]	PSM	3*	64 × 85	6	9	spots colorés
[Adán09]	PSM	3 (average)	986	7	7	spots colorés

TABLE 1.1 – Techniques utilisant des motifs pseudo-aléatoires : 1.Référence ; 2.Type de codage (PM : Perfect Maps ; PSM : Perfect Submaps;) ; 3.Distance de Hamming moyenne (* : minimale) ; 4.Taille du motif ; 5.Nombre de symboles utilisés ; 6.Longueur des mots de code ; 7. Type de primitives projetées.

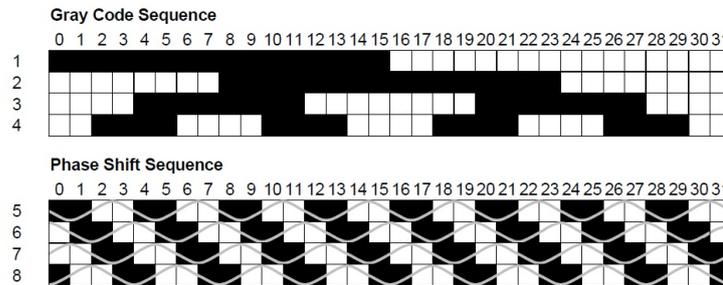


Figure 3. Gray code (top) and Phase Shift (bottom) of a $n = 32$ stripe code sequence.

(a)

FIGURE 1.15 – Code de Gray (en haut) et déphasage (en bas), utilisés pour le codage par multiplexage temporel.

1.2.1.3 Codage par multiplexage temporel

Les approches exploitant le codage par multiplexage temporel procèdent par la projection successive de plusieurs motifs codés le plus souvent avec un codage de Gray et/ou un déphasage des franges projetées (Fig. 1.15) pour affiner les reconstructions dans le temps. Pour n colonnes à projeter, $\log(n)$ motifs sont généralement utilisés, par exemple dans [Posdamer82].

Une telle approche procède par affinements successifs et s'avère très robuste car le contraste du motif est maximisé, mais elle est avant tout dédiée aux objets statiques, plusieurs acquisitions étant nécessaires pour un décodage complet. Des variantes ont été proposées par [Altschuler81] pour capturer des objets en mouvement par un codage en fréquence de chacun des motifs qui sont alors simultanément projetés à des fréquences différentes. Mais le système nécessite alors une caméra par motif, et devient trop encombrant. De plus, le champ commun à toutes les caméras est forcément réduit. Cette idée a récemment été reprise dans [Young07]. Enfin, dans [Benveniste10] un invariant spectral est utilisé pour offrir davantage de robustesse aux spécularités. Et dans [Hu09] un décalage de franges est utilisé pour capturer des scènes statiques en haute résolution.

Comme nous souhaitons une haute réactivité pour des scènes en mouvement, ce type d'approche n'est pas approprié à notre problématique.



FIGURE 1.16 – Motif adaptatif dans [Li11]. a) La scène. c) Motif projeté. b) Les composantes bleue et verte de l'image. d) Troisième composante issue de l'analyse PCA extraite de l'image c).

1.2.1.4 Autres types de codage

Si les codages précédents sont généralement les plus utilisés, d'autres types de codages, appelés codages continus dans [Salvi10], existent. Il sont appelés ainsi car la variation d'intensité lumineuse ou spectrale se fait de façon continue, par opposition aux codages précédents, discrets, car les variations se font par échelon et des contours sont discernables entre les indices visuels projetés. Il existe divers types de codages continus utilisant le multiplexage temporel ([Srinivasan85], [Pribanic10]), le multiplexage fréquentiel ([Takeda83], [Cobelli09]) ou encore le multiplexage spatial ([Carrhill85],[Tajima90]). Parmi ces approches, on retrouve le codage direct, dans lequel chaque primitive est codée par la seule information que sa représentation visuelle véhicule. Bien que des reconstructions denses y sont possibles, ces approches ont été écartées car elles sont de façon générale très sensibles au bruit, qui, comme nous le verrons, sera présent de façon significative dans les images traitées.

D'autres approches récentes utilisent des réseaux de neurones, conjointement à une projection de lumière structurée, pour identifier dans la scène des objets connus, dans [Espinal10] par exemple.

Des travaux récents ont également vu le développement de la lumière structurée adaptative. Le motif est alors modifié pour mieux prendre en compte les propriétés géométriques/spectrales présentes dans la scène. Par exemple dans [Li11], les propriétés spectrales du motif sont modifiées dans une approche à deux étapes pour maximiser le contraste colorimétrique de celui-ci (Fig. 1.16). Pour cela, une analyse PCA (analyse en composantes principales) est effectuée. Les composantes couleur sont transformées de l'espace RGB (Fig. 1.16.b) vers un autre espace (Fig. 1.16.d), dans lequel la direction de la troisième composante principale est utilisée pour la couleur du motif, de façon à maximiser le contraste colorimétrique. De plus, une adaptation géométrique est également réalisée : l'écart entre les franges colorées est également adapté (Fig. 1.16.c) pour diminuer les ambiguïtés de mise en correspondance qui peuvent apparaître avec des franges possédant un écart régulier. D'après [Li11], les performances obtenues sont alors

meilleures qu'avec des motifs non adaptatifs pour des scènes comportant de nombreux objets fortement colorés.

1.2.2 Traitements et motifs imperceptibles

En dehors de la conception même des motifs, les traitements associés sont également importants car ce sont eux qui permettront une bonne exploitation du système réalisé. Dans cette thèse, les traitements font également l'objet d'une attention particulière.

Si les traitements sont souvent abordés en même temps que la conception des motifs, la littérature présente plusieurs contributions concentrées uniquement sur les traitements. En effet certains d'entre eux, comme la segmentation et la localisation précise des points sur les franges, sont communs à un grand nombre de conceptions. Ainsi les problèmes de segmentation de franges colorées sont abordés dans [D. Fofi01], [Koninckx06] et [Pagès03].

Des estimations des contours des franges avec une précision inférieure au pixel sont proposées dans [Ji09], [Chen10] et [Lu10].

La problématique de labellisation des franges colorées par un algorithme adaptatif est proposée dans [Fechteler08] ou des franges binaires noir et blanc dans [Chien09] dans lequel un noyau probabiliste adaptatif est employé. Dans [Xu07], une classification des pixels est effectuée pour détecter si un point de la scène est illuminé ou non par un motif binaire. En effet, la lumière du jour dans la scène peut colorer des objets se trouvant dans une bande noire, donc non illuminé par le motif.

Enfin le décodage de motifs 1-D utilisant une projection de franges colorées est abordé dans [D. Fofi01]. Une approche exploitant un découpage de graphe est proposé dans [Schmalz10b] et a été améliorée dans [Schmalz10a] en utilisant une propagation de croyance.

Dans [Li10] une étude est réalisée sur l'influence de la géométrie du système stéréo sur la précision finale obtenue. Il est mis en évidence que c'est la ligne de base séparant le centre optique du projecteur de celui de la caméra qui a le plus fort impact par rapport à d'autres paramètres, comme l'orientation des axes optiques des dispositifs.

Une étude comparative des différentes approches permettant de projeter un motif invisible (ou imperceptible) est réalisée dans [Fofi04]. Plusieurs approches, de l'utilisation de couleurs hors du spectre visible, à l'alternance entre un motif et son négatif y sont présentées. Une des possibilités souvent employée est d'utiliser l'infra-rouge (IR) avec une caméra sensible à cette plage de longueur d'ondes. D'une part, dans certaines applications, la gêne visuelle que représente la projection des indices dans la scène est ainsi évitée. D'autre part peu d'objets (surtout les matières non vivantes) émettent naturellement dans la plage spectrale de l'infra-rouge, donc les perturbations spectrales y sont très faibles comparé à ce qui est observé dans le spectre visible. Un motif infra-rouge est utilisé, par exemple, dans [Modrow07]. De plus le premier système de lumière structurée codée grand public, le dispositif Kinect de Microsoft, utilise également une projection IR. L'hétérogénéité des conditions dans lesquelles ce dispositif est capable d'effectuer des mise en correspondances montre la bonne robustesse spectrale que permet la projection de lumière structurée dans cette plage de longueur d'ondes.

Cependant, pour des scènes intra-abdominales, l'infra-rouge n'est pas adapté, car les longueurs d'onde proches du rouge diffusent énormément dans les tissus des surfaces organiques. Cet aspect sera davantage détaillé au chapitre 3.

Le cas Kinect

Nous ne pouvons terminer cette étude de l'existant sans évoquer le dispositif Kinect de la société Microsoft.

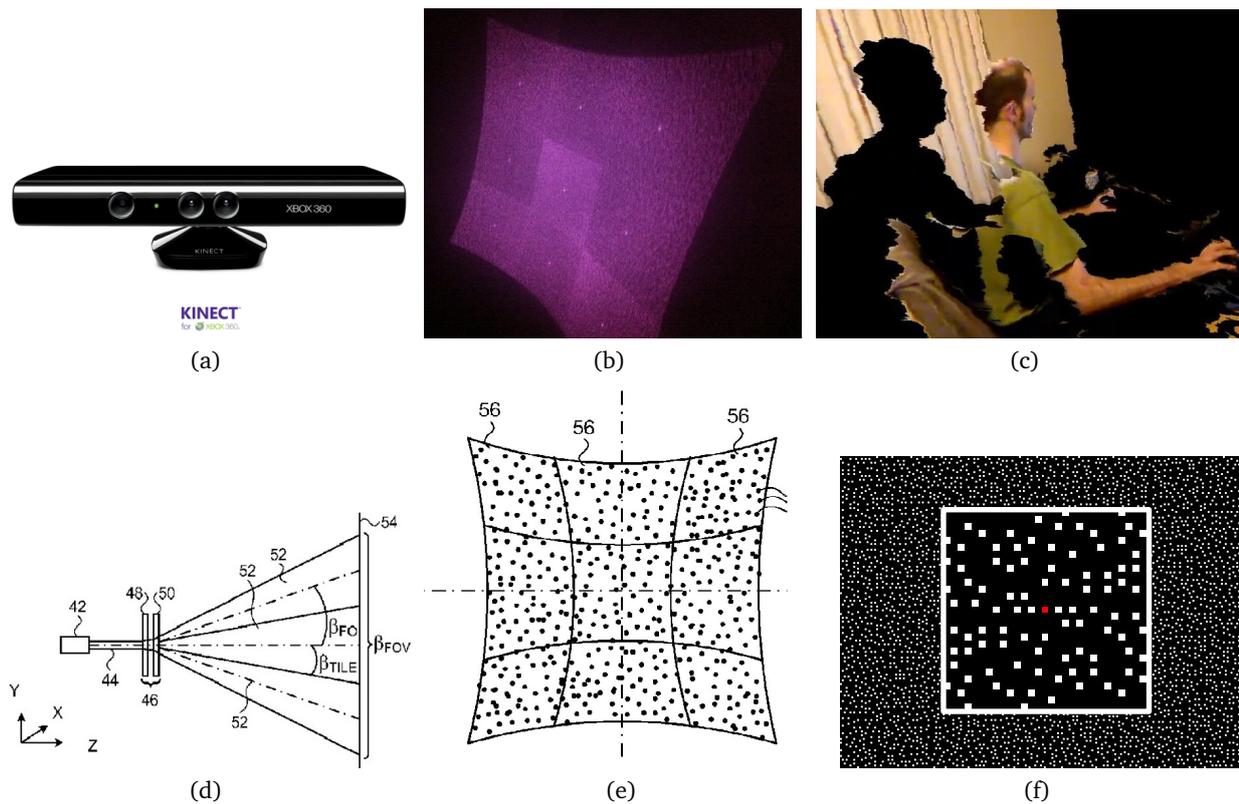


FIGURE 1.17 – Le dispositif Kinect (Microsoft)

A l'origine conçu comme périphérique pour la Xbox de Microsoft lors de son lancement en Novembre 2010, le périphérique Kinect a rapidement été détourné pour d'autres utilisations. Son boîtier (Fig. 1.17.a) intègre un projecteur laser IR (633×495), une caméra VGA sensible aux IR, et une deuxième caméra VGA qui capture la texture de la scène dans le spectre visible. Cette compacité, son prix très abordable (~ 115 euros, rendu possible grâce à une production de masse) et sa capacité à effectuer des reconstructions robustes dans des environnements très divers, possédant de nombreuses perturbations, ont mené de nombreuses équipes scientifiques à l'adopter. Celles-ci notent cependant une précision moyenne de l'ordre du centimètre. Aujourd'hui, on retrouve fréquemment ce périphérique dans les conférences internationales ayant pour thème la Vision.

Ce qui nous intéresse plus particulièrement ici, c'est le motif de lumière structurée codée qui est utilisé. Celui-ci a été conçu et protégé par la société PrimeSense. Pour des raisons compréhensibles d'avantage concurrentiel, les détails des reconstructions opérées par le dispositif Kinect n'ont pas été donnés, mais PrimeSense a indiqué que leur technologie employait une lumière codée quand certains pensaient initialement à l'exploitation du *Time-of-Flight*. Cependant nous disposons de plusieurs éléments permettant de nous faire un avis. D'une part la société PrimeSense a déposé plusieurs brevets ([Shpunt10a] et [Shpunt10b]) pour protéger sa conception, d'autre part, l'observation du motif projeté avec une caméra sensible au spectre IR peut nous apprendre beaucoup de choses, enfin, très récemment (en août 2011), dans [Reichinger11] une analyse poussée du motif a été publiée. Celle-ci met en évidence la structure symétrique centrale (Fig. 1.17.f) des neuf (3×3) sous-motifs (Fig. 1.17.b,e), composés chacun de 211×165 points lumineux binaires (un point est allumé ou non) arrangés sur une structure grillagée. Des éléments

publiés dans le brevet [Shpunt10b] viennent corroborer cette analyse. En effet il y est fait mention de deux éléments d'optique diffractive (*DOE*) utilisés successivement (éléments numéro 48 et 50, Fig. 1.17.d), l'un permettant la création de chaque sous-motif - ce qui explique la symétrie centrale des sous-motifs (pour maximiser l'efficacité lumineuse) et la luminosité supérieure du point correspondant à la composante principale (au centre des sous-motifs sur la figure 1.17.b) - et l'autre *DOE* créant l'agencement (3×3) du motif complet. Les différences de luminosité des sous-motifs visibles sur la figure 1.17.b) correspondraient alors à une in-homogénéité lors de la diffraction par ce deuxième *DOE*. Enfin, si nous continuons l'analyse, en supposant une unicité par voisinage spatial 2-D dans chacun des sous-motifs, il y aurait environ (en comptant les bords des voisinages) $211 \times 165 = 34815$ voisinages. En considérant des mots de codes utilisant 2 symboles, il faudrait une longueur minimale de 16 symboles, ce qui est possible avec des voisinages 4×4 , par exemple.

Le périphérique Kinect, de par la configuration de son système stéréo et la précision reportée, n'est pas adapté, tel quel, à une reconstruction en situation mini-invasive. De plus ce dispositif a été rendu public trop tard par rapport au début de cette thèse pour être pris en considération. Enfin nous avons déjà noté que le spectre proche du rouge peut diffuser beaucoup sur des surfaces intra-abdominales.

1.2.3 Lumière structurée et scènes intra-abdominales

Les scènes reconstruites avec une projection de lumière structurée et exposées dans la littérature comportent généralement peu de perturbations spectrales. Elles sont souvent réalisées sur des objets très peu texturés, voire blanc ou mat et qui réfléchissent très bien la lumière (c'est moins vrai pour les publications récentes). C'est en effet dans ces conditions que de tels systèmes peuvent fonctionner au mieux. Des visages ou des mains sont souvent montrés, mais ceux-ci réfléchissent également bien la lumière. La réflectance de la peau est relativement homogène et ne présente pas de difficulté particulière.

Ce qui fait la particularité de notre problématique et donc un véritable défi, ce sont les très nombreuses perturbations présentes dans les scènes intra-abdominales.

1.3 L'alignement épipolaire du motif, fil rouge de la conception proposée

Nous l'avons vu, nous serons confrontés, dans les scènes intra-abdominales, à d'importantes perturbations géométriques et spectrales lors de la mise en correspondance des primitives projetées. Le bruit sera également très important. Nous en expliquerons les causes par la suite. L'objectif premier de la conception proposée dans cette thèse est ainsi la **robustesse** à ces perturbations et donc la faculté d'effectuer une mise en correspondance malgré celles-ci. De façon générale, les **invariants** et la **redondance** sont autant d'éléments permettant de renforcer la robustesse d'un système. C'est vrai dans beaucoup de domaines.

Nous venons de présenter un grand nombre de conceptions de motifs possibles. Cependant malgré l'hétérogénéité importante des motifs proposés dans la littérature, il existe un point commun remarquable présent dans l'immense majorité de ces contributions. Qu'il s'agisse de la projection de points, de franges, de lignes ou de formes géométriques, l'agencement général de ces indices visuels dans le plan image du projecteur se fait selon un alignement strictement

horizontal et/ou vertical de ces indices. Or dans les systèmes stéréo, il existe un invariant géométrique qui peut être exploité : l'homographie des lignes épipolaires.

Homographie des lignes épipolaires

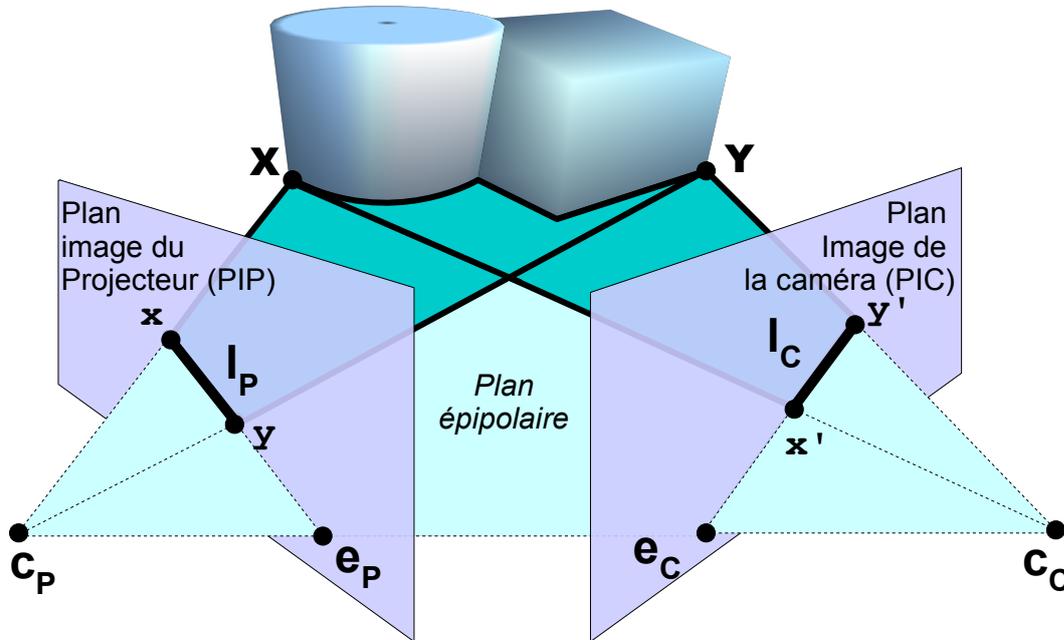


FIGURE 1.18 – Géométrie épipolaire d'un système stéréo utilisant le modèle sténopé. c_p , e_p , l_p , x et y : respectivement, le centre optique, l'épipôle, une ligne épipolaire passant par x, y et e_p sur le PIP. c_c , e_c , l_c , x' et y' : les équivalents pour le PIC. X et Y sont les projections de x et y dans la scène 3D.

Il est depuis longtemps établi que, connaissant les coordonnées d'un point dans un des plans image (projecteur (PIP) ou caméra (PIC)) d'un système stéréo (Fig.1.18), la géométrie épipolaire permet de retrouver la ligne épipolaire correspondante dans l'autre image. Cela permet, lors de l'appariement, de réduire l'espace de recherche à une seule dimension. Si cette contrainte épipolaire de positionnement est couramment utilisée dans les systèmes stéréo pour la mise en correspondance, ce n'est pas le cas pour la conception du motif. En effet, dans le cas général, si la scène à reconstruire est inconnue, il n'est pas possible de savoir comment une ligne projetée sur la scène se comporte dans le PIC (Fig.1.19.a), car cela va dépendre des objets présents dans la scène. Cependant, il existe un cas particulier : si les lignes passent par l'épipôle dans le PIP (Fig.1.18 et Fig.1.19.b), alors elles resteront dans le plan épipolaire associé dans l'espace 3-D et leur projection dans le PIC correspondra à la ligne épipolaire associée (intersection entre le plan épipolaire et le PIC). Ainsi, quelle que soit la scène considérée, cette projection restera invariante. Pour que cette invariance reste valide sur une large profondeur de champ (Fig.1.19.b) et pour disposer d'un volume de travail important, l'estimation de la matrice Fondamentale F (voir chapitre 4) - qui contient toute l'information relative à la géométrie épipolaire - doit être suffisamment précise.

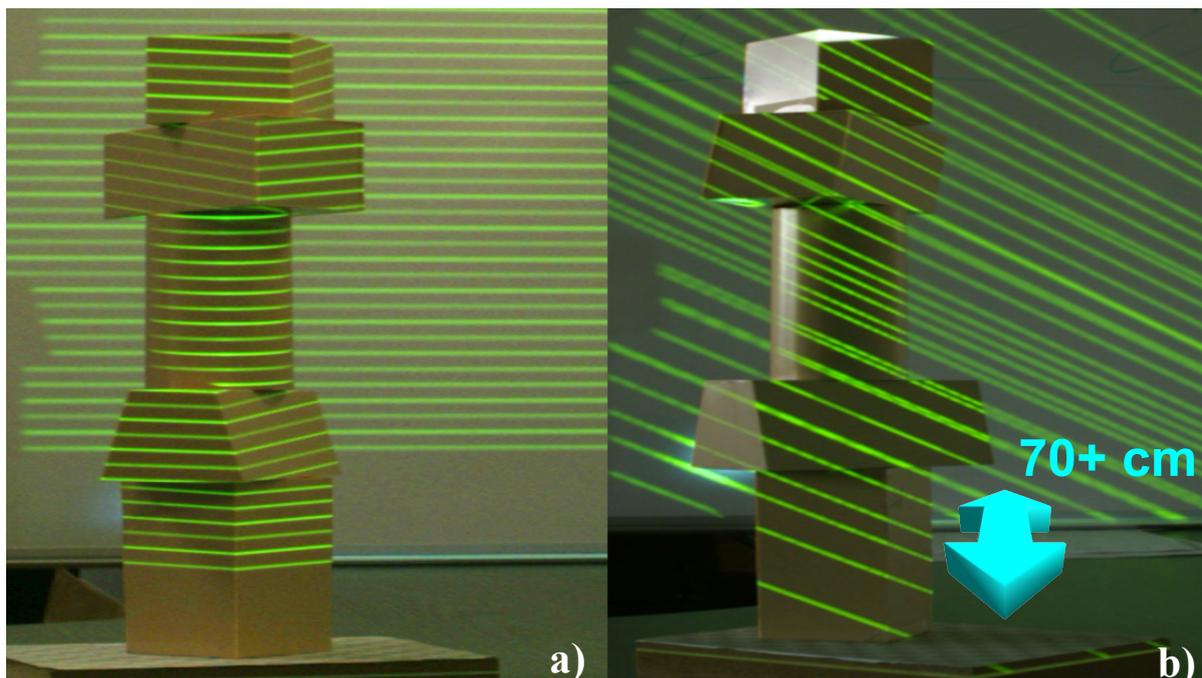


FIGURE 1.19 – Projection de lignes sur le PIC : a) Les lignes sont toutes horizontales dans le PIP, leur projection sur le PIC dépend de la scène. b) Les lignes passent par l'épipoles dans le PIP, leur projection sur le PIC est indépendante de la scène et se fait le long des lignes épipolaires associées. Cette invariance reste valable sur une grande plage de profondeur : plus de 70 cm ici.

Cette propriété est due à l'homographie 1-D qui relie les deux ensembles de lignes épipolaires, dans le PIP et dans le PIC, indépendamment des objets de la scène. Ainsi, si un ensemble de lignes \mathbf{P}_p s'intersectent exactement à l'épipoles e_p dans le PIP, alors les lignes épipolaires correspondantes dans le PIC, \mathbf{P}_c , seront aussi des lignes épipolaires, et vérifieront :

$$\mathbf{P}_p \equiv \mathbf{F} [e_c]_{\times} \mathbf{P}_c . \quad (1.1)$$

Dans cette thèse, nous proposons d'exploiter la géométrie épipolaire pour la conception du motif à travers un alignement géométrique sur celle-ci. Ainsi, au lieu d'être disposées horizontalement/verticalement, les lignes de primitives prendront l'orientation des lignes épipolaires dans le plan image du projecteur. Cette opération peut être mise en regard d'une rectification inverse d'image côté projecteur. En effet, à l'issue d'une rectification d'image, après acquisition par la caméra, les lignes épipolaires se retrouvent orientées horizontalement, ce qui permet d'effectuer la mise en correspondance par simple analyse des lignes de l'image rectifiée. Ici nous proposons l'inverse : à partir d'une grille régulière, les lignes de cette grille se retrouvent, à l'issue de l'alignement du motif, orientées selon la direction des lignes épipolaires. Cette opération a plusieurs conséquences bénéfiques à différentes étapes de la conception et des traitements : sur le codage du motif, sur la segmentation, sur la classification de primitives orientées et enfin sur la netteté de l'intersection entre les colonnes de primitives et les lignes épipolaires.

- Sur le codage, l'unicité des codes associés aux primitives n'a plus besoin d'être assurée globalement sur l'ensemble du motif, mais seulement entre les codes associés à des pri-

mitives alignées sur la même ligne épipolaire. En effet seules celles-ci peuvent mener à des ambiguïtés d'appariement. La conséquence est une **réduction dimensionnelle des contraintes** portant sur le problème de codage et donc la possibilité de créer des codages possédant des propriétés inédites par rapport à ce qui était proposé jusqu'à maintenant dans l'état de l'art. Le chapitre 2 traite de la problématique du codage.

- Sur la segmentation, pour un motif de taille $n \times m$ il n'y aura pas une ligne épipolaire à étudier pour chaque primitive, mais **une ligne épipolaire pour m primitives**, ce qui peut permettre des gains substantiels en rapidité de traitements.
- Pour la classification, il est désormais possible de **connaître a priori l'orientation générale des primitives** comportant une information directionnelle après leur projection sur le plan image de la caméra. En effet si leur orientation est celle de la ligne épipolaire sur laquelle elles se trouvent dans le PIP, alors leur orientation sera celle de la ligne épipolaire correspondante dans le PIC **quels que soient les objets sur lesquels elle se projette dans la scène 3-D**. En effet, les lignes épipolaires sont invariantes par rotation (Fig. 1.19.b). Le chapitre 3 traite de ces aspects géométriques.
- Enfin, lors de la projection de franges, ou de formes possédant une continuité verticale, la précision de la segmentation est optimale lorsque les franges sont orthogonales avec les lignes épipolaires. Si le motif est aligné dans le plan projecteur, alors les colonnes de primitives, ou franges, seront globalement orthogonales aux lignes épipolaires dans le PIP et cette orthogonalité sera également **globalement maximisée** dans le PIC, en particulier si l'orientation des surfaces illuminées par le motif est parallèle aux plans images du système stéréo (voir chapitre 3, déformation des voisinages).

Ainsi, une nouvelle classe de motifs de lumière structurée est proposée : **les motifs alignés sur la géométrie épipolaire**.

Si cet alignement est le fil rouge et la contribution principale de cette thèse, d'autres contributions sont présentées.

Le codage du motif, purement numérique, se fait indépendamment de la façon dont les symboles issus de ce codage sont interprétés pour représenter visuellement les primitives. C'est pour cette raison que la conception de notre motif est séparée en deux chapitres : le premier se concentre sur le codage : propriété numérique du motif, en lien avec la théorie de l'information, alors que le second s'intéresse à la représentation physique du motif.

Ainsi, dans le prochain chapitre, nous aborderons les problématiques de codage. Le choix, l'organisation des primitives et les traitements associés à la conception seront abordés au chapitre 3. Le chapitre 4 traitera des reconstructions 3-D. Et enfin, avant de conclure, nous présenterons, dans le chapitre 5, les développements système tant matériel que logiciel, élaborés au cours de cette thèse.

Chapitre 2

Conception 1 : Codage du motif

Overall, if the technical contribution and novelty is strongly confirmed by other reviewers, weakly accept, otherwise weakly reject.

(CVPR 2010, reviewer n° 2)

La redondance n'a de sens que si une indépendance existe entre les sources. Si une des sources est biaisée ou perturbée, alors les sources dépendant de celle-ci seront également affectées. Elles ne rempliront pas leur rôle de redondance, mettant à mal la robustesse générale du système. Il en va de même pour le codage des motifs.

Dans ce chapitre, nous étudions les algorithmes de codage par voisinage spatial, exploitables pour la conception de motifs de lumière structurée robustes. Par robustesse, nous entendons que les codes générés possèdent des capacités de correction à même de prendre en compte certaines perturbations de l'environnement provoquant des erreurs de décodage. Après un état de l'art sur les principales approches algorithmiques existantes et un rappel des bases théoriques pour un tel codage, nous proposons plusieurs contributions, tant théoriques que pratiques, permettant de générer de nouveaux codes plus robustes ou permettant la création de motifs de résolution supérieure avec moins de symboles que ce qui est proposé dans la littérature. En particulier, nous proposons d'exploiter la géométrie épipolaire dès cette étape de conception du motif pour diminuer significativement la complexité portant sur les contraintes du problème de codage et autoriser de nouvelles solutions aux propriétés très intéressantes. Outre les gains en terme de robustesse et/ou de taille des motifs générés, le temps d'exécution du codage permis par les algorithmes proposés ouvrent la voie à des motifs adaptatifs recodés en temps-réel. De plus les contributions proposées sont valables quelles que soient la forme, la couleur, l'intensité lumineuse, ou les autres modalités de discrimination utilisées pour différencier les primitives projetées, tant que celles-ci s'inscrivent dans une logique de codage par voisinage spatial.

2.1 Introduction

Le codage a une place très importante dans la conception du motif et donc dans ce travail de recherche pour les raisons suivantes :

- Le **problème de mise en correspondance** est le principal problème pointé dans la littérature lors de reconstructions 3-D basées sur plusieurs vues. Le codage a pour objectif, entre autres, de rendre cette mise en correspondance la plus fiable possible.

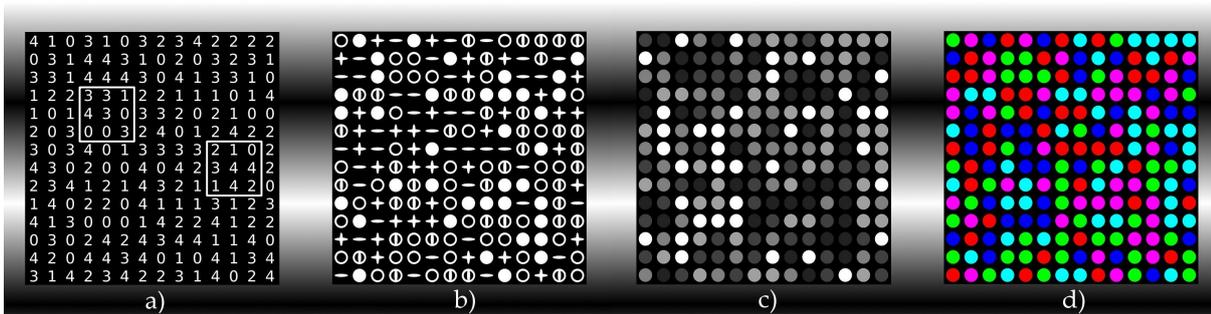


FIGURE 2.1 – Une matrice 5-aire de (14×14) éléments codée, dans laquelle tous les couples de voisinages (3×3) possèdent au minimum 5 symboles différents position à position. Les symboles sont représentés avec a) des chiffres, b) des formes, c) des niveaux de gris, d) des couleurs différentes.

- De la qualité du codage découle la **qualité du signal transmis** sur la scène, et sa capacité à résister aux perturbations présentes dans le canal de transmission, entre la projection des primitives par le projecteur et leur acquisition par la caméra. Il est donc important, en amont, d'obtenir la meilleure qualité de signal possible.
- **L'indépendance de cette étape** : si ce travail de recherche a pour finalité des reconstructions dans un contexte médical mini-invasif, l'étape de codage, première de la séquence des traitements à effectuer, peut-être employée pour d'autres types de scènes et d'autres motifs que ceux qui vont être présentés ici, également pour des motifs qui exploitent les codes générés de façon différente. Certaines contributions (sections 2.4 et 2.5) permettent également d'améliorer le codage 1-D de motifs qui utilisent, le plus souvent, une projection de franges.
- Du codage va dépendre la robustesse, le nombre de symboles utilisés et la résolution du motif projeté, autant d'**éléments qui sont essentiels** pour les reconstructions 3-D visées.

2.1.1 Codage par voisinage spatial

Comme indiqué dans le chapitre précédent, nous nous focalisons ici sur le codage par voisinage spatial 2-D (exemple figure 2.1), qui permet de reconstruire la scène à partir d'une seule projection et est donc plus adapté aux scènes comportant des objets en mouvement que l'utilisation du multiplexage temporel, par exemple. Le codage direct, présenté dans le chapitre précédent peut être considéré comme un codage par voisinage spatial avec un voisinage de 1×1 , réduit à la primitive centrale; et le codage par franges (souvent colorées), utilisant le plus souvent des séquences de De-Bruijn, utilise un voisinage 1-D, donc de type $1 \times X$. Le codage par voisinage spatial 2-D emploie un voisinage de type $X \times Y$, mais d'autres organisations spatiales, par exemple hexagonales ([Adán09]), sont aussi possibles. Dans cette étude, bien que nous nous focalisons sur un voisinage de taille 3×3 , les contributions proposées pourront être adaptées pour toute taille de voisinage 2-D et certaines contributions seront également utilisables pour le codage 1-D.

2.1.2 Motivations pour un voisinage 3×3

Comme rappelé dans le chapitre précédent, les approches utilisant le codage 2-D emploient principalement un voisinage 3×3 , avec quelques variations comme 2×3 ou 2×4 par exemple. Ainsi la dynamique des tailles de voisinage est proche de 3×3 . La raison est qu'un tel voisinage permet de n'analyser qu'une primitive autour de la primitive centrale du voisinage dans chaque direction. Un tel voisinage permet ainsi de n'utiliser que les informations les plus proches spatialement (le plus proche possible en dehors du codage direct) pour le décodage et donc d'être capable de décoder une primitive sans nécessiter qu'une zone importante de l'image soit examinée (ce serait le cas avec un voisinage 5×5 par exemple), ce qui est souhaitable en cas de forte discontinuité de courbure et/ou d'occultations, d'ombres, etc... Comme nous le verrons dans ce chapitre, avec un voisinage 3×3 des mots de code de longueur 9 sont créés, ce qui est suffisant pour générer, avec peu de symboles, des motifs de résolution suffisamment importante pour la majorité des applications. De plus cette taille de mots de code permet également d'ajouter de la redondance au codage. Des mécanismes de décodage robustes en cas de mauvaise classification des primitives projetées peuvent alors être mis en place.

2.1.3 Un critère de robustesse : la distance de Hamming

Définition de la distance de Hamming

Etant donnés deux mots de code A et B de longueur W , et s , la position du symbole étudié. La distance de Hamming est alors le nombre de positions en lesquelles les symboles de A et B diffèrent, tel que :

$$\text{Ham}(A, B) = \sum_{s=1}^W \delta_s \quad ; \quad \delta_s = \begin{cases} 0 & \text{si } A_s = B_s \\ 1 & \text{sinon,} \end{cases} \quad (2.1)$$

où $A = \{A_1, \dots, A_W\}$ et $B = \{B_1, \dots, B_W\}$.

Définition de la distance de Hamming minimale

La distance de Hamming minimale, H_{Min} , correspond à la plus petite distance de Hamming entre tous les couples de mots de code possibles. Soit Ω , l'ensemble des mots de code entre lesquels la distance de Hamming minimale, H_{Min} , doit être imposée, alors $\forall (A, B) \in \Omega, A \neq B, \text{Ham}(A, B) \geq H_{Min}$.

Bien souvent, pour ne pas alourdir le propos, quand nous parlerons simplement d'*unicité* des codes, nous entendrons *unicité assurant la distance de Hamming minimale désirée*.

La distance de Hamming est une mesure de robustesse bien connue en *Théorie de l'Information*. Elle est très souvent utilisée comme paramètre pour assurer des capacités de correction lors du codage de documents numériques. C'est le cas notamment pour les *codes de Reed-Solomon*, largement utilisés dans l'industrie pour le stockage des données sur CD, DVD, disques Blue-Ray

par exemple, et les *codes de Hamming* ([Reed60], [Hamming50]).

Comme nous le verrons dans la section suivante (section 2.3), pour le même nombre de symboles utilisés, plus la distance de Hamming minimale d'un code est élevée, plus de mauvaises classifications des indices visuels peuvent être tolérées lors du décodage.

2.1.4 Objectifs du codage

Le codage prend la forme d'une matrice pseudo-aléatoire k -aire, si k symboles sont utilisés pour la coder. Chaque élément de la matrice sera associé à un type différent d'indice visuel du motif projeté.

Les algorithmes de codage présentés ont pour objectif de :

- Maximiser le nombre d'éléments de la matrice. Celle-ci doit être de taille maximale si l'objectif est lié à l'obtention d'une grande résolution et par conséquent à une meilleure densité dans la reconstruction et la localisation 3-D. Il y a tout de même une limitation physique car des images de primitives trop petites peuvent poser des problèmes durant la segmentation et la classification.
- Minimiser la taille de l'alphabet. Il doit être de taille minimale, car utiliser le moins de symboles possible permettra de les représenter par des primitives plus dissemblables les unes des autres (en termes d'intensité lumineuse, de chromaticité ou de géométrie) et donc, de réduire les erreurs de classification.
- Maximiser la distance de Hamming minimale. Pour augmenter le pouvoir discriminant du motif, celle-ci doit être maximale. En effet, si des primitives n'ont pu être correctement identifiées sur les images acquises, une valeur élevée de la distance de Hamming minimale offre plusieurs possibilités de décodage sans ambiguïté.

Bien sûr, ces paramètres sont interdépendants, par exemple moins on utilise de symboles, plus le problème de codage sera contraint, moins la taille du motif et la distance de Hamming minimale pourront être élevées. Il s'agira donc de trouver le meilleur équilibre en fonction de l'application visée.

2.1.5 Motivations pour l'amélioration des algorithmes

Des algorithmes de codage par voisinage spatial sont déjà disponibles. Notamment, Morano [Morano98] est souvent utilisé dans la littérature. Comme nous le verrons dans l'état de l'art, en pratique, de tels algorithmes utilisent une approche "brute-force", procédant par "essais-erreurs". Ainsi plus le temps alloué à la recherche d'une solution est important, meilleurs seront les codes trouvés par ces algorithmes. En pratique nous disposons d'un temps limité pour générer ces motifs. Pouvoir en générer rapidement est très important pour sélectionner celui ou ceux qui conviennent le mieux à une application donnée. En fonction des résolutions des périphériques utilisés, du type de scène, on voudra par exemple une résolution de motif, un nombre de symboles ou encore une distance de Hamming minimale différents.

De plus, le codage rapide peut ouvrir la porte à l'utilisation de codage pour des motifs adaptatifs, ce qui n'est pas le cas de ceux utilisés aujourd'hui qui nécessitent souvent plusieurs minutes à plusieurs jours de calculs. Dans notre cas, les améliorations algorithmiques sont employées à

la fois pour pouvoir générer rapidement et tester de nouveaux motifs ; et pour trouver des motifs avec des propriétés beaucoup plus intéressantes (en termes de taille, nombre de symboles utilisés et/ou de la distance de Hamming minimale) que ce que les algorithmes actuels peuvent produire.

2.1.6 Contributions proposées

Dans ce chapitre, après une revue des principales voies algorithmiques présentes dans la littérature, nous présentons plusieurs contributions. Les contributions portent à la fois sur les aspects théoriques du codage et sur la présentation de nouveaux algorithmes de génération de codes.

Contributions analytiques

- Prédiction de la distance de Hamming moyenne d'un motif (section 2.3.3).
- Dénombrement de codes en fonction de leur distance de Hamming (section 2.3.5).

Nouveaux algorithmes de codage

A travers de nouveaux algorithmes de génération de codes, nous proposons de :

- Trouver un meilleur équilibre entre un comportement aléatoire et exhaustif de la recherche (section 2.4).
- Diminuer la complexité des tests d'unicité :
 - Dans l'espace des mots de code (section 2.5),
 - En exploitant la géométrie par proximité épipolaire (section 2.5),
 - En exploitant la géométrie par alignement épipolaire (section 2.6).

La plupart de ces contributions sont complémentaires. Il est par exemple possible d'effectuer les tests d'unicité dans l'espace des mots de code en n'assurant l'unicité qu'entre les primitives alignées sur une même ligne épipolaire et en effectuant une recherche avec une fenêtre de correction qui optimise le comportement aléatoire/exhaustif de recherche dans l'espace des solutions. Ces contributions répondent aux deux composantes principales qu'un algorithme de type "brute-force" se doit de maîtriser :

- Le test d'unicité permettant d'assurer que la contrainte de distance de Hamming minimale est respectée.
- Le comportement de recherche qui détermine, comme pour un processus d'optimisation, la façon dont l'algorithme va se comporter pour trouver de nouvelles solutions plus intéressantes.

2.2 Travaux antérieurs

Après avoir montré quelques exemples pratiques de l'utilisation, par l'industrie, de codages spatiaux au sens large, nous verrons les approches s'inscrivant dans le cadre de la théorie des matrices parfaites et de leurs variantes à travers les algorithmes directs et ceux dits "brute-force". Finalement, nous montrerons les limites des algorithmes existants et proposerons, dans les sections suivantes, diverses contributions à même de dépasser ces limites.

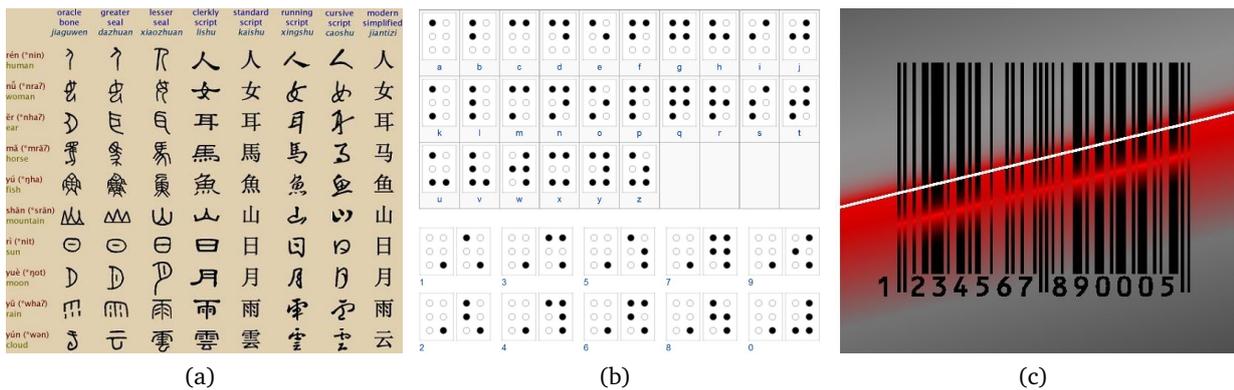


FIGURE 2.2 – (a) Evolution des synogrammes utilisés dans le Mandarin. (b) Le braille. (c) Un code barre.

2.2.1 Un large éventail d'applications

Le codage spatial, au sens large, consiste à utiliser l'espace, en 1, 2 ou N dimensions, pour former des mots de code constitués d'une suite de symboles appartenant à un alphabet de dimension finie.

Voici quelques exemples utilisés au quotidien qui exploitent un codage spatial :

- Les systèmes d'écriture (1-D),
- Le braille (2-D),
- Le codage de documents numériques (texte, images, vidéo) avec / sans pertes,
- Les codes barre (1-D) , et code DataMatrix (2-D).

Les systèmes d'écriture

Des mots, comme ceux composant ce document, de taille variable, sont formés à partir de caractères appartenant à un alphabet servant de base au langage utilisé. Les caractères d'écriture, destinés avant tout à être lus par un humain sont distingués par des formes différentes. L'origine de ces formes montre que, bien souvent, il existe une relation sémantique entre la forme et ce qu'elle représente pour l'homme. Un des exemples les plus flagrants provient des idéogrammes utilisés en Mandarin (Fig. 2.2.a).

Le braille

Dans le système d'écriture Braille, du nom de son inventeur , chaque signe est représenté en relief par une matrice 2 × 3, il y a donc un maximum de 63 signes codables plus l'espace (matrice sans points) (Fig. 2.2.b). Le codage est différent selon la langue utilisée, mais les langues



FIGURE 2.3 – Différents formats de code : Code QR, Datamatrix, Shotcode, Colorzip, Maxicode.

utilisant des caractères similaires (langues latines par exemple) tendent à utiliser un codage semblable pour les caractères communs.

Le codage de documents numériques

Le codage de documents numériques est omni-présent dans l'industrie aujourd'hui, que ce soit avec perte (images, vidéos) ou sans (texte). L'objectif le plus fréquent est de compresser au maximum le document avec un minimum de perte qualitative. Le codage de documents 1-D (texte), 2-D (images), 3-D (modèles CAO, vidéos), N-D est une part importante des problématiques soulevées par l'ère numérique. Souvent, plusieurs techniques sont combinées. Par exemple pour les fichiers images *jpeg*, un codage avec perte sélectionne dans le domaine fréquentiel (transformée en cosinus discret ou ondelettes) des coefficients qui seront eux-même codés avec un codage arithmétique, sans pertes. Les techniques de codage sans perte (arithmétique, Huffman), peuvent être complétées par des bits de redondance (codes de Reed-Solomon, Hamming) permettant la détection et/ou la correction d'erreurs en assurant, pour certains d'entre eux, une distance de Hamming minimale. Ces techniques représentent une part importante des mathématiques appliquées.

Les code barre

De nombreuses variantes de codes barre existent, la plus utilisée est celle de Woodland [Woodland57]. Ce codage 1-D est basé sur la différence d'épaisseur entre les bandes noires et les bandes blanches successives composant le code (Fig. 2.2.c).

Les codes DataMatrix

Les codes DataMatrix sont des codes 2-D qui permettent une densité d'informations bien plus importante que les codes barre et sont composés de carrés blancs et noirs juxtaposés. Selon la norme de codage utilisée, ils peuvent posséder en outre une redondance permettant la détection et la correction d'erreur. Jusqu'à 2335 caractères alphanumériques ou 3116 caractères numériques peuvent être codés sur $1cm^2$. Ici des mots clefs de 8 bits (voisinage 3×3 moins le coin supérieur-droit) servent à coder les données. Dans une étape supplémentaire, il est possible d'ajouter de la redondance à l'ensemble (Fig. 2.3).

Limites et remarque générale

Parmi les approches présentées ici, l'objectif est principalement de réaliser le codage de caractères, numériques ou alphanumériques, ou d'images, que ce soit en utilisant les codes barre, dataMatrix, etc... Dans notre cas la finalité sera de rendre possible une mise en correspondance robuste en assurant l'unicité, à H_{Min} près, des codes proposés, ce qui n'est pas le cas ici car plusieurs caractères identiques ou plusieurs pixels de même valeur peuvent apparaître dans les documents à coder. Ces types de codages spatiaux ne sont donc pas adaptés à notre problématique.

Cependant, comme ces codes sont utilisés dans l'industrie, ils doivent être très robustes à tout ce que la vie au quotidien peut provoquer de perturbations. Nous remarquons en particulier que le braille, les code barre, aussi bien que la majorité des codes 2-D type Datamatrix, sont des codes binaires (utilisant seulement 2 symboles de base) : soit le contraste blanc/noir, soit l'existence/l'absence de relief (braille). Ceci montre qu'en pratique, utiliser le nombre minimum de symboles peut être souhaitable en cas de fortes perturbations ou quand les conditions d'utilisation ne sont pas restreintes a priori.

1	0	1	1	1	1	1	1	1	1	1	2	0	1	2
1	0	0	0	1	1	2	2	1	0	1	2	0	1	2
0	0	0	1	0	1	2	0	2	1	1	2	0	1	2
1	1	0	1	1	0	1	2	2	1	1	2	0	1	2
1	0	1	1	1	1	1	1	1	1	1	2	0	1	2

(a)
(b)
(c)

FIGURE 2.4 – Variantes de matrices parfaites : (a) Une matrice parfaite 2 – aire (5,5 ; 2,2). (b) Une sous-matrice parfaite 3 – aire (5,5 ; 3,3) avec $H_{Min} = 5$. (c) Une sous-matrice sous-parfaite 3 – aire (5,5 ; 3,3) avec $H_{Min} = 9$, l'unicité n'est assurée qu'entre les sous-matrices se trouvant sur la même ligne.

2.2.2 Matrices parfaites et variantes

Dans cette partie, nous allons présenter les matrices parfaites (*PM*), les sous-matrices parfaites (*PSM*) et les sous-matrices sous-parfaites (*SPSM*), que nous introduisons ici. Grâce à ces matrices codées, le problème de mise en correspondance peut être résolu en assurant une unicité entre les sous-matrices qui les composent.

En outre des *PM* et *PSM*, déjà connues de la communauté, nous avons jugé bon d'ajouter un nouveau groupe de matrices : les *SPSM*, car, comme nous allons le voir, de telles matrices ne répondent plus à la définition des *PM* et *PSM*.

Les différences entre *PM*, *PSM* et *SPSM* concernent, d'une part, les propriétés d'unicité qui sont différentes et, d'autre part, les types d'algorithmes qui les produisent. Nous donnons dans cette section les définitions et spécificités de ces différents objets.

Matrices parfaites (*PM*)

Une matrice A , C -aire est une matrice parfaite (**Perfect map**) ($m,n ; u,v$), s'il s'agit d'une matrice de taille ($m \times n$) dans laquelle toutes les sous-matrices C -aire ($u \times v$) possibles apparaissent exactement une fois. Les *PM* fournissent une solution directe à la mise en correspondance. Elles sont construites par une extension en 2-D des séquences de De-Bruijn [deBruijn46] et ont été largement étudiées (voir [Etzion88, Mitchell95, Patterson94]). La figure 2.4-a montre un exemple de *PM* 2-aire (5,5 ; 2,2). Sur cette figure, nous pouvons remarquer que toutes les 16 sous-matrices de taille 2×2 , sont présentées une et une seule fois.

Sous-matrices parfaites (*PSM*)

Les sous-matrices parfaites (**Perfect Submap**) ont les mêmes propriétés que les *PM* excepté que seulement un sous-ensemble des sous-matrices possibles est utilisé. Ces matrices sont généralement générées par des algorithmes de type *brute-force* et sont adaptées pour imposer davantage de redondance ($H_{Min} > 1$). Cela n'est pas possible avec les *PM* car toutes les sous-matrices doivent être utilisées. Un exemple de sous-matrice parfaite est illustrée par la figure 2.4-b.

Sous-matrices sous-parfaites (*SPSM*)

Les sous-matrices sous-parfaites (**Subperfect Submap**) ont les mêmes propriétés que les *PSM* excepté que l'unicité n'est assurée qu'entre des sous-ensembles de sous-matrices. Ainsi, le pro-

Niveau de contraintes	Principe	Références
Plus contraint + + +	Voisinages non ordonnés	[Adan04]
Plus contraint + +	Invariance par rotation des voisinages	[Claes07]
Plus contraint +	Symétrie centrale du motif	[Albitar07]
PSM classique	Voir définition d'une PSM	[Morano98]
Moins contraint - (SPSM)	Disparité limitée des primitives dans l'image, périodicité du code	[Sagawa09] (1-D)
Moins contraint - - (SPSM)	Exploitation de la géométrie épipolaire, codage 1,5D	[Maurice10], [Maurice11]
Moins contraint - - - (SPSM)	Disparité limitée et géométrie épipolaire	Non utilisé

TABLE 2.1 – Tableau comparatif des contraintes posées sur le codage à partir d'une PSM

blème de codage est moins contraint qu'avec les *PM* et les *PSM*. Une H_{Min} plus élevée, ou une résolution supérieure, ou moins de symboles peuvent être utilisés. Deux des algorithmes proposés dans ce chapitre permettent d'en générer. Généralement, des hypothèses portant sur la géométrie du système stéréo sont utilisées pour guider le choix des primitives entre lesquelles une unicité n'est pas nécessaire. Un exemple de sous-matrice sous-parfaite est peut être observé sur la figure 2.4-c.

2.2.3 Les contraintes posées sur le problème de codage

Hormis la distance de Hamming minimale et le nombre de symboles utilisés avec les *PSM*, d'autres types d'informations peuvent être utilisés pour rendre le codage plus ou moins contraint. Le plus souvent ce sont les contraintes géométriques du système stéréo qui sont utilisées.

Partant de la définition d'une *PSM*, le tableau 2.1 résume les différentes variations qui sont utilisées.

2.2.4 Augmenter les contraintes de codage

Symétrie centrale du motif

Dans la thèse d'Albitar [Albitar09a], une symétrie centrale avait été imposée au motif et donc à son codage. La raison était liée à l'utilisation d'optiques diffractives pour la projection du motif. De telles optiques projettent la transformée de Fourier du motif. Ainsi le ratio d'énergie lumineuse projetée passait de 40% à 80% de l'énergie transmise aux optiques si le motif était à symétrie centrale. Désirant reprendre, au début de ce travail de thèse, ce principe de projection, nous avons proposé un algorithme exploitant une des contributions discutées plus loin dans ce chapitre (voir section 2.4) [Maurice09]. Dans le cas de la symétrie centrale, l'unicité doit être vérifiée pour chaque voisinage ainsi que pour son homologue par symétrique centrale, ce qui augmente les contraintes du codage. Mais seulement la moitié du motif doit être codée.

Invariance par rotation $\pi/2$ et $3\pi/2$

Dans [Claes07], le système stéréo utilisé était de type *hand-eye*. Le projecteur, fixé sur une table, venait illuminer la scène et la caméra était fixée sur l'organe terminal d'un bras robotisé à des fins d'asservissement visuel. Ainsi la configuration projecteur-caméra n'était pas fixe. Pour

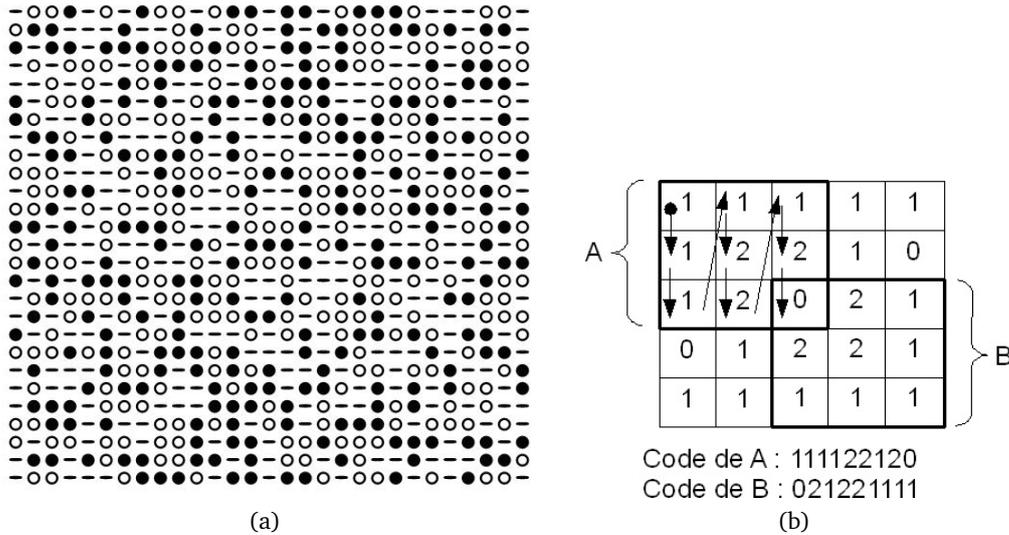


FIGURE 2.5 – (a) Motif à symétrie centrale dans [Albitar07]. (b) Remplissage symétrique de la matrice.

$C_{0,2}$	$C_{1,2}$	$C_{2,2}$	$C_{0,0}$	$C_{0,1}$	$C_{0,2}$	$C_{2,0}$	$C_{1,0}$	$C_{0,0}$
$C_{0,1}$	$C_{1,1}$	$C_{2,1}$	$C_{1,0}$	$C_{1,1}$	$C_{1,2}$	$C_{2,1}$	$C_{1,1}$	$C_{0,1}$
$C_{0,0}$	$C_{1,0}$	$C_{2,0}$	$C_{2,0}$	$C_{2,1}$	$C_{2,2}$	$C_{2,2}$	$C_{1,2}$	$C_{0,2}$

FIGURE 2.6 – Invariance par rotation ($\pm\pi/2$) des voisinages. Au centre, le voisinage initial (v_i). A gauche et a droite, les voisinages obtenus par rotation de v_i de $\pm\pi/2$.

prendre en compte les variations de la géométrie du système stéréo pendant l’asservissement, un codage assurant une unicité invariante par rotation a été proposé. Dans celui-ci, l’unicité de chaque code est alors assuré à une rotation $\pi/2$ et $3\pi/2$ près. Ce qui augmente les contraintes de codage, car il y a alors trois fois moins de possibilités.

Voisinages non ordonnés

Dans [Adan04] et [Adán09], les symboles des codes sont considérés indépendamment de leur position dans les mots de code (Fig. 2.7), donnant lieu à des voisinages non ordonnés. L’invariance par rotation dans [Claes07] est cas particulier de cette approche. Le problème de codage est ici encore beaucoup plus contraint que dans [Claes07].

2.2.5 Diminuer les contraintes de codage

Disparité limitée des primitives

Comme nous l’avons évoqué au chapitre 1, les primitives projetées dans le plan caméra peuvent, en fonction de la profondeur des objets sur lesquels elles se projettent, ne pas se déplacer sur toute la section de leur ligne épipolaire dans la partie visible de l’image. Dans certaines conditions du système stéréo, la plage de déplacement des primitives peut être assez réduite. C’est le cas par exemple si le projecteur et la caméra ont axe optique quasiment parallèle et si la ligne de base est faible. La disparité générale sera alors moins importante. Certaines approches exploitent cette donnée géométrique à travers un codage périodique en exploitant la propriété

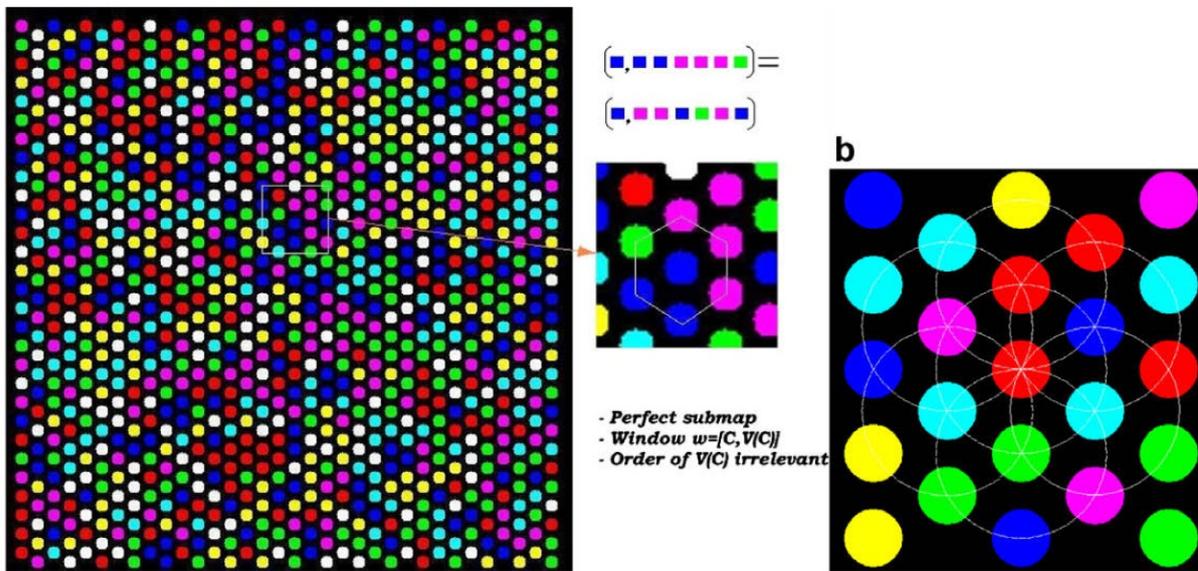


FIGURE 2.7 – Voisinages hexagonaux non ordonnés proposés dans [Adán09].

cyclique des séquences de De-Bruijn. L'unicité n'est plus assurée tout le long des lignes épipolaires mais sur des portions de celle-ci. Cela permet d'utiliser des motifs de plus grande taille, ou codés avec moins de symboles. C'est le cas dans [Sagawa09], par exemple. Une sous-matrice sous-parfaite (*SPSM*) est alors créée.

Exploitation de la géométrie épipolaire

Deux des contributions présentées dans ce chapitre permettent de diminuer significativement les contraintes reposant sur le codage en exploitant la géométrie épipolaire (voir sections 2.6.2 et 2.6.1). La conséquence est une réduction dimensionnelle (1-D vs 2-D) du nombre de voisinages entre lesquels l'unicité doit être assurée.

Disparité limitée et géométrie épipolaire

Les deux approches précédentes contribuent à la réduction des contraintes peuvent être combinées pour aboutir à une réduction accrue des contraintes de codage. Dans un souci de généralité et pour ne pas restreindre la méthodologie proposée à des applications trop spécifiques, nous nous sommes contentés, dans cette thèse, de l'utilisation de la géométrie épipolaire.

Comme les approches qui diminuent les contraintes du problème de codage ne correspondent plus à la définition d'une *PSM* (l'unicité n'étant pas assurée entre toutes les sous-matrices présentes), nous les avons nommées sous-matrices sous-parfaites (subperfect submaps, *SPSM*).

Il existe deux grandes catégories d'approches permettant de générer des codes par voisinage spatial : le codage direct, qui est issu de l'analyse et qui peut donner une solution directe à certains problèmes de codage, ne nécessitant aucune recherche ; et les algorithmes de codage dits "brute-force".

2.2.6 Algorithmes directs

Pour réaliser des codages par voisinage spatial, il existe des solutions directes, basées sur une extension bidimensionnelle des séquences de De-Bruijn, présentées au chapitre 1, qui permettent de générer des matrices parfaites a périodiques, périodiques ou semi-périodiques ([Etzion88],

[Mitchell95],[Patterson94]). Le principe de construction le plus souvent utilisé consiste à disposer des séquences de De-Buijn dans une des dimensions, et les répéter dans l'autre avec des décalages, de façon à assurer l'unicité dans les deux dimensions.

Ces solutions ne permettent cependant pas d'assurer une distance de Hamming minimale supérieure à 1 et donc d'assurer une robustesse à d'éventuelles mauvaises classification des indices visuels. De plus, le choix des paramètres n'est pas libre et le rapport des dimensions des codes obtenus est souvent disproportionné et ne correspond pas, le plus souvent, aux rapports d'aspect les plus utilisés pour des projecteurs/caméras (4 :3, 16 :9, etc). L'avantage principal réside dans la génération de codes optimaux, codes pour lesquels le nombre maximal de voisinages est utilisé, car tous les voisinages possibles sont présents.

Généralement, cette approche est moins flexible que celle utilisant les algorithmes "brute-force" présentés dans la section suivante. Par exemple, il n'est pas ici possible d'ajouter des contraintes d'invariance par rotation ou par symétrie aux voisinages.

Les recherches sont cependant toujours actives dans ce domaine. Très récemment, [Jia11] a proposé une nouvelle construction pour des matrices de taille importante. Un décalage de sous-matrices construites à partir de séquences de De-Buijn y est utilisé avec un algorithme nommé *piece growing*. La matrice produite est de dimensions 79×59 et utilise des voisinages 2×2 avec 10 symboles. Ce nombre de symboles est bien trop élevé pour notre application étant donné le niveau de bruit et mènerait inévitablement à de nombreuses mauvaises classifications des indices visuels. De plus, les reconstructions montrées dans [Jia11] étaient réalisées sur des plans simples sans difficultés géométriques ni spectrales.

2.2.7 Algorithmes "brute-force"

Les algorithmes permettant d'imposer une H_{Min} arbitrairement choisie et d'imposer un rapport d'aspect à la matrice à coder sont basés sur une approche "brute-force", procédant par essais-erreurs successifs.

Principe général

A partir d'un motif vide, des symboles sont ajoutés un par un et la contrainte sur la valeur de H_{Min} est vérifiée pour chaque nouveau symbole ajouté. Pour cela, le voisinage $W_A \times H_A$ créé par l'ajout du nouveau symbole est comparé à tous les voisinages $W_i \times H_i$ déjà inclus dans la matrice. A chaque comparaison, la distance de Hamming H est alors calculée entre le code associé à $W_A \times H_A$ et celui associé à $W_i \times H_i$. Si H est supérieure ou égale à H_{Min} alors ce nouveau symbole est conservé et le symbole suivant est traité, sinon différentes stratégies sont mises en place :

- Stratégie de recherche aléatoire dans [Morano98] (M).
- Stratégie de recherche exhaustive dans [Claes08] (C).

2.2.7.1 Approche aléatoire (M)

Dans [Morano98] (Fig. 2.8), le comportement est aléatoire : si le test de contrainte sur H_{Min} échoue, les autres symboles sont testés pour cette position. Si aucun d'entre eux ne permettent d'obtenir un test d'unicité positif, alors une nouvelle matrice est initialisée aléatoirement. Cela est répété jusqu'à ce qu'une matrice de la taille désirée soit complètement remplie.

Limite de cette approche :

Cette approche ne prend pas suffisamment avantage de la matrice incomplète déjà trouvée pour tenter d'améliorer le résultat. Cela est d'autant plus vrai lorsque les contraintes de codage (faible

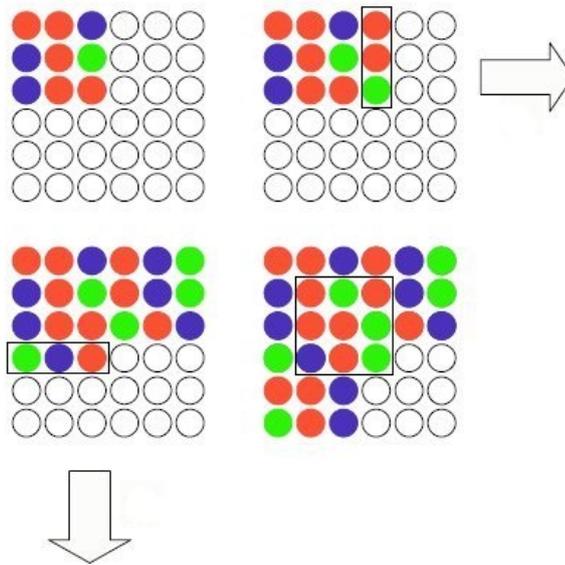


FIGURE 2.8 – Dans [Morano98], depuis un voisinage 3×3 , les 3 premières lignes sont d'abord remplies, puis les 3 premières colonnes, cela permet de gérer les bords des voisinages. Puis chaque nouveau symbole crée un nouveau voisinage.

nombre de symboles, H_{Min} élevée) sont plus importantes, car les tests d'unicité seront négatifs plus tôt dans la recherche de solutions.

2.2.7.2 Approche exhaustive (C)

Dans [Claes07], l'approche est purement exhaustive : à chaque étape, le symbole 0 est ajouté. Si le code associé au voisinage créé par ce nouveau symbole ne respecte pas H_{Min} , tous les symboles précédemment trouvés sont concaténés selon leur ordre d'insertion pour former un grand nombre N , k -aire (s'il y a k symboles). N est alors incrémenté en base k et à chaque incrémentation, les voisinages correspondant aux chiffres ayant subi des modifications sont testés pour vérifier H_{Min} . Ainsi cet algorithme explore toutes les combinaisons possibles, du nombre composé uniquement de 0 à celui composé uniquement de $k - 1$, en n'explorant cependant que les branches prometteuses de l'arbre des possibilités. En effet, le nombre de chiffres de N dépendra des éléments précédents déjà trouvés et n'est donc, jamais, sauf à la fin du remplissage, égal au nombre d'éléments de la matrice à remplir. De plus, contrairement à [Morano98], [Claes07] propose d'étendre la matrice au fur et à mesure de son remplissage, en ajoutant, de façon alternée, des lignes et des colonnes en fonction du rapport d'aspect choisi (voir Fig. 2.9). Cela permet d'obtenir des nouvelles matrices complètes et exploitables au fur et à mesure de l'exploration des solutions.

Cette approche exhaustive montre qu'il est possible d'obtenir des **codes optimaux** avec des algorithmes "brute-force". Dans les codes optimaux, aucun paramètre de la matrice (taille, nombre de symboles, H_{Min}), ne peut être amélioré sans modifier les autres. Si une solution existe, elle sera forcément trouvée, car tous les cas sont testés. Mais le temps de recherche demandé progresse exponentiellement avec la taille de la matrice et dépasse en pratique très rapidement l'espérance de vie d'un être humain.

Dès que la taille de la matrice devient supérieure à une vingtaine d'éléments et que le nombre

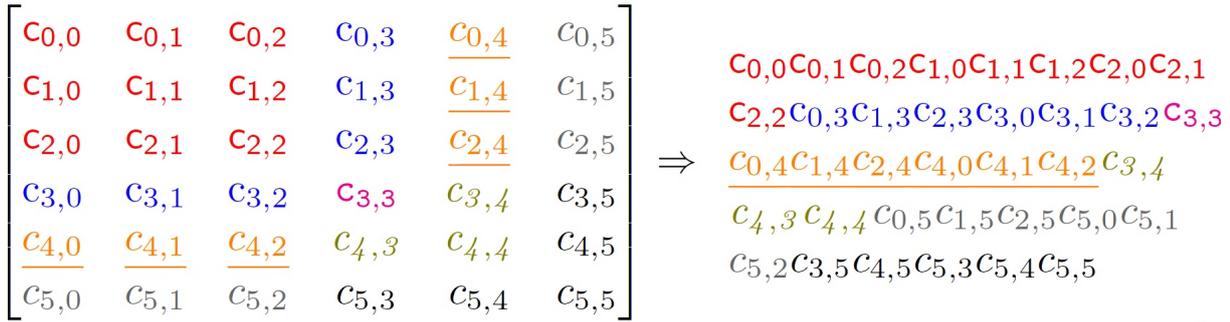


FIGURE 2.9 – Dans [Claes07], depuis un voisinage 3×3 , k -aire, des lignes et des colonnes sont ajoutées alternativement. A droite, la conversion en un grand nombre k -aire qui est toujours incrémenté pour une recherche exhaustive.

de symboles augmente, l'explosion combinatoire fait que très peu de possibilités peuvent être explorées en un temps raisonnable. En effet en pratique seules des matrices de très petite taille et avec peu de symboles ont pu être explorées exhaustivement en plusieurs jours de calculs. Dans [Claes07] seuls les cas ($k = 2$, $H_{Min} = 3$, $taille = 4 \times 5$), ($k = 3$, $H_{Min} = 5$, $taille = 3 \times 4$), ($k = 3$, $H_{Min} = 5$, $taille = 4 \times 5$) ont pu être explorés exhaustivement. Ceux ci sont inutilisables en pratique de part leur résolution trop faible.

2.3 Éléments théoriques sur le codage

Dans cette section, nous présentons quelques éléments théoriques, dont certains sont nouveaux, pouvant aider à la conception des motifs utilisant un codage par voisinage spatial. En particulier nous proposons d'estimer a priori la distance de Hamming moyenne d'un motif. De même, nous proposons des éléments analytiques de dénombrement qui seront exploités dans une approche portant sur l'utilisation de l'espace des mots de codes pour effectuer le test d'unicité.

2.3.1 La distance de Hamming moyenne \bar{H}

Soit S , un ensemble de N mots de code de longueur W . La distance de Hamming moyenne, \bar{H} , de S , est donnée par :

$$\bar{H} = \frac{\sum_{i=1}^N \sum_{j=i+1}^N \text{Ham}(C_i, C_j)}{N(N-1)/2}. \quad (2.2)$$

2.3.2 Capacités de correction

Il est bien connu, en théorie de l'information, que pour détecter M erreurs dans un message, il est nécessaire de disposer d'un code assurant une distance de Hamming minimum $H_{Min} \geq M + 1$, et que pour les corriger sans ambiguïté, il est nécessaire de vérifier $H_{Min} \geq 2M + 1$.

Cependant ces valeurs sont reliées à des erreurs correspondant le plus souvent à des inversions de bits. En lumière structurée cependant, deux éléments viennent modifier cette approche :

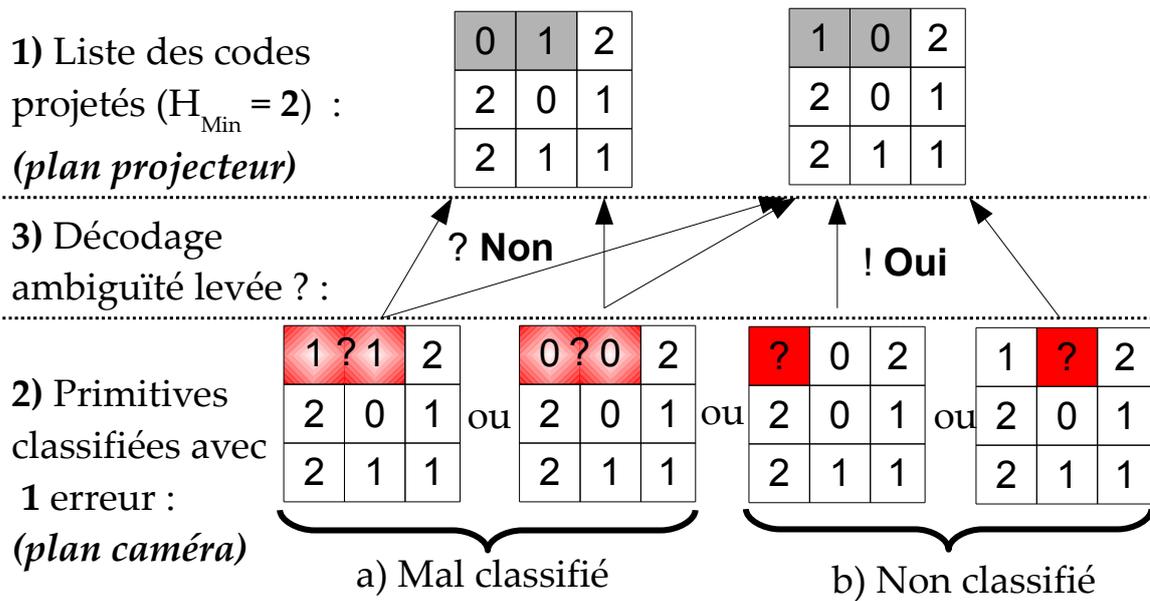


FIGURE 2.10 – Exemple d’une correction. Avec $H_{Min} = 2$, correction d’une erreur impossible si la primitive est **MAL** classifiée, possible si la primitive est **NON** classifiée

- La liste des codes transmis (projetés) est **connue**, il est donc plus facile de retrouver les codes initiaux par comparaison.
- Il est possible de ne pas classer une primitive en cas de trop forte incertitude.

Les erreurs peuvent apparaître pendant l’étape de classification des primitives qui relient les primitives visuelles segmentées au symbole du code qui leur est associé. Or, il est possible de décider de ne pas classer une primitive si on estime que la confiance dans cette classification n’est pas assez bonne. A la charge du concepteur de mettre en place l’évaluation de ce critère de confiance. Comme l’illustre la figure 2.10, des primitives non classifiées peuvent être corrigées avec une H_{Min} inférieure à celle requise pour corriger des erreurs de classification. En effet, dans le cas de primitives mal classifiées, l’erreur a un poids plus important car une mauvaise information supplémentaire est fournie (la mauvaise classification).

Dans le cas de N primitives non classifiées, la contrainte $H_{Min} \geq N + 1$ est suffisante pour effectuer une correction. En effet, il suffit alors qu’une seule primitive supplémentaire au nombre d’erreurs de NON classification permette de différencier les mots de code pour que l’association avec les mots de codes projetés puisse se faire sans ambiguïté (voir exemple Fig. 2.10).

Généralisation

Pour corriger M erreurs dues à des mauvaises classifications et N erreurs dues à des absences de classification, la distance minimale de Hamming requise est alors : $H_{Min} \geq 2M + N + 1$.

2.3.3 Prédiction de la distance de Hamming moyenne \bar{H}

Alors que la distance de Hamming moyenne permet d’avoir une idée statistique sur la robustesse du codage, la distance de Hamming minimale donne des garanties sur la capacité du décodage à détecter et/ou corriger un certain nombre d’erreurs. Une distance de Hamming

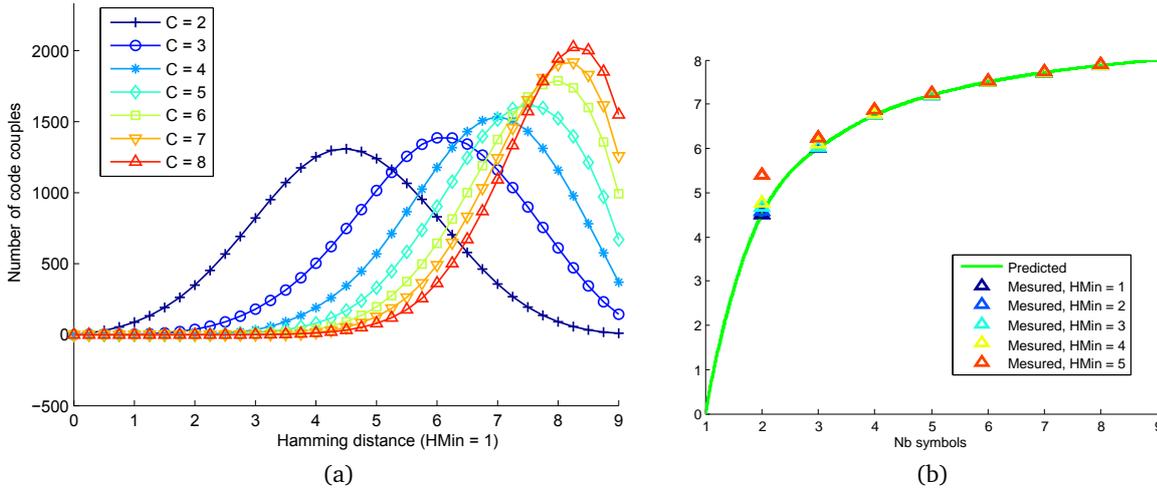


FIGURE 2.11 – (a) Distribution lissée de la distance de Hamming pour des motifs codés avec C symboles, $H_{Min} = 1$ et $W = 9$ (moyenne sur 100 codes PSM), pour de nombreuses valeurs de C . (b) \bar{H} estimée (en vert) pour $W = 9$ en fonction du nombre de symboles et \bar{H} calculée pour de nombreux codes (moyenne sur 100 codes) avec différentes valeurs de H_{Min} .

moyenne ne garantit pas qu'il sera possible de détecter un certain nombre d'erreurs, même si ceci sera vrai dans la plupart des cas. En effet, une distance de Hamming moyenne plus élevée indique une meilleure robustesse statistique aux erreurs.

Il est alors intéressant de pouvoir la prédire avant même de générer le codage. Cela permet, par exemple, de savoir à l'avance quel est le nombre minimum de symboles requis pour assurer une robustesse statistique suffisante du motif. En effet, selon les applications, les perturbations peuvent être plus ou moins importantes et disposer d'une prédiction sur les propriétés du motif qui va être généré peut éviter de nombreux essais infructueux.

Pendant le codage de PM , PSM ou $SPSM$, aucun symbole ne doit, a priori, être favorisé. Donc il est raisonnable de poser l'hypothèse d'une distribution uniforme des symboles pour l'estimation de \bar{H} .

Proposition : En supposant une distribution uniforme des symboles choisis dans un alphabet C -aire et formant des mots de codes de longueur W , la distance de Hamming moyenne \bar{H} entre tous les couples de mots codes est donnée par :

$$\bar{H} = W - W/C. \quad (2.3)$$

Preuve : Soient $A = \{A_1, \dots, A_W\}$ et $B = \{B_1, \dots, B_W\}$ deux mots de code de longueur W , avec $A_i, B_i \in \{0, 1, \dots, C-1\}$. A_i et B_i peuvent prendre chacune des C valeurs avec une équiprobabilité de $1/C$. En conséquence, la probabilité que le même symbole apparaisse dans A et B à la i^{eme} position est donnée par :

$$\begin{aligned} p(A_i = B_i) &= p(A_i = 0) \times p(B_i = 0) \\ &+ \dots \\ &+ p(A_i = C-1) \times p(B_i = C-1) \\ &= \sum_{j=0}^{C-1} p(A_i = C_j) \times p(B_i = C_j) \\ &= \sum_{j=0}^{C-1} (1/C) \times (1/C) = 1/C. \end{aligned} \quad (2.4)$$

Donc l'espérance E du nombre de symboles identiques, position à position, dans A et B est :

$$\begin{aligned} E &= p(A_1 = B_1) + \dots + p(A_W = B_W) \\ &= \sum_{i=1}^W p(A_i = B_i) = W/C. \end{aligned} \quad (2.5)$$

Finalement, par complémentarité, on obtient $\bar{H} = (W - E)$.

Comme le montre la figure 2.11.b, la valeur \bar{H} estimée (equ. 2.3) est très proche des valeurs mesurées a posteriori avec un grand nombre de motifs. Seul le cas $(C, H_{Min}) = (2, 5)$, très peu utilisé en pratique car produisant des matrices de taille trop petite, n'est pas bien estimé. Ces résultats montrent que l'évolution de \bar{H} par rapport au nombre de symboles C utilisés est très similaire à ce qui peut être observé dans la distribution des distances de Hamming de tous les couples de mots de code générés (Fig. 2.11.a). C'est à dire que le gain est plus important quand les valeurs successives de C sont faibles.

2.3.4 Limites supérieures sur les paramètres de codage

Il n'est pas ici question d'étudier de manière exhaustive les limites existant sur les paramètres de codage et leurs calculs mais de donner un aperçu nous permettant de nous situer par rapport à ce qui se fait et à l'efficacité des algorithmes qui vont être présentés dans les sections suivantes. A contrario de la prédiction de la distance de Hamming moyenne qui donne une appréciation statistique, ces limites donnent une idée a priori des paramètres avec lesquels il est certain qu'une solution au problème de codage ne pourra pas être trouvée, car celui-ci est trop contraint. Ce sont des limites supérieures.

2.3.4.1 Limites supérieures de Singleton et de Hamming

Des limites théoriques sont connues en théorie de l'information, notamment concernant l'existence de codes correcteurs qui utilisent également la distance de Hamming minimale et un alphabet restreint de symboles.

Soit $A_k(w, H_{Min})$, la taille maximale possible d'une séquence de codes C , k -aire de longueur w et de distance de Hamming minimale H_{Min} , alors la **limite de Singleton** est définie par [Singleton64] :

$$A_k(w, H_{Min}) \leq k^{w-(H_{Min}-1)}. \quad (2.6)$$

Cette limite est assez grossière, la **limite de Hamming** [Hamming50] est en effet plus fine. Elle est donnée par :

$$A_k(w, H_{Min}) \leq \frac{k^w}{\sum_{q=0}^t \binom{w}{q} (k-1)^q}, \quad \text{avec } t = \lfloor \frac{H_{Min}-1}{2} \rfloor, \quad (2.7)$$

où $\lfloor X \rfloor$, correspond à l'entier inférieur le plus proche de X et $\binom{w}{q}$, est le nombre de façons de choisir q éléments parmi w sans tenir compte de l'ordre.

Un code qui atteint la limite de Hamming est appelé un code parfait.

En pratique, ces limites ne sont atteintes que par les méthodes de codage directes.

2.3.5 Dénombrement de mots de codes pour une distance de Hamming donnée

Soit un mot de code C , k -aire de longueur w et de distance de Hamming minimale H_{Min} , le nombre de mots de code ayant une distance de Hamming $h < H_{Min}$ avec C est donné par ([Maurice11]) :

$$N_f(H_{Min,w}) = 1 + \sum_{h=1}^{H_{Min}-1} \frac{w!}{h!(w-h)!} (k-1)^h. \quad (2.8)$$

Preuve : Pour tout $h \in [0, H_{Min}[$, le nombre de mots de codes (N_C) ayant une distance de Hamming de h avec le mot de code A est

$$N_C = C_w^h \times (k-1)^h, \quad (2.9)$$

où

$$C_w^h = \frac{w!}{h!(w-h)!}. \quad (2.10)$$

C_w^h est le nombre de façons de choisir h positions parmi w , sans tenir compte de l'ordre. Ce sont tous les ensembles possibles de h positions auxquelles les symboles ne sont pas les mêmes que ceux de A entre les w possibles. Le terme $(k-1)^h$ correspond, pour chaque ensemble de h positions, au nombre de combinaisons de symboles différents de ceux de A aux mêmes positions. En effet, pour chaque position, il reste $(k-1)$ symboles différents possibles et donc $(k-1)^h$ combinaisons possibles pour les h positions. Le premier terme (1) prend en compte le cas $h = 0$ qui mènerait C_w^h à une singularité.

Ce dénombrement, bien que présenté ici, sera utilisé dans la section 2.5 dans laquelle est décrite notre contribution sur les tests d'unicité dans l'espace des mots de code.

2.4 Comportement de recherche par le paradigme exploration / exploitation

Comme nous venons de le voir dans la section précédente, la littérature des codages dits "brute-force" propose deux algorithmes ayant des comportements opposés, l'un purement exhaustif (figure 2.12.a), l'autre purement aléatoire (figure 2.12.b). Il existe un autre domaine qui présente une telle analogie, l'apprentissage par renforcement (*reinforcement learning*) très utilisé en intelligence artificielle, notamment pour la problématique de recherche du plus court chemin. Certains de ces algorithmes (par exemple le *SARSA* (State Action Reward State Action) ou le *Q-Learning*, voir [Sutton98]) utilisent dans leur politique de mise à jour, un terme qui permet de régler l'algorithme vers un comportement plus aléatoire ou plus exhaustif. Il n'existe pas à ce jour de bases théoriques permettant de savoir à l'avance comment fixer ce paramètre pour obtenir le meilleur résultat possible. Mais l'expérience montre que les meilleurs résultats sont obtenus avec un réglage qui cherche un juste milieu et qui peut varier en fonction des ressources (de calcul, de transferts en mémoire, etc...) à disposition. Cette recherche d'équilibre est également valable dans la plupart des fonctions d'optimisation susceptibles d'être confrontées à plusieurs minima locaux (voir Fig. 2.12).

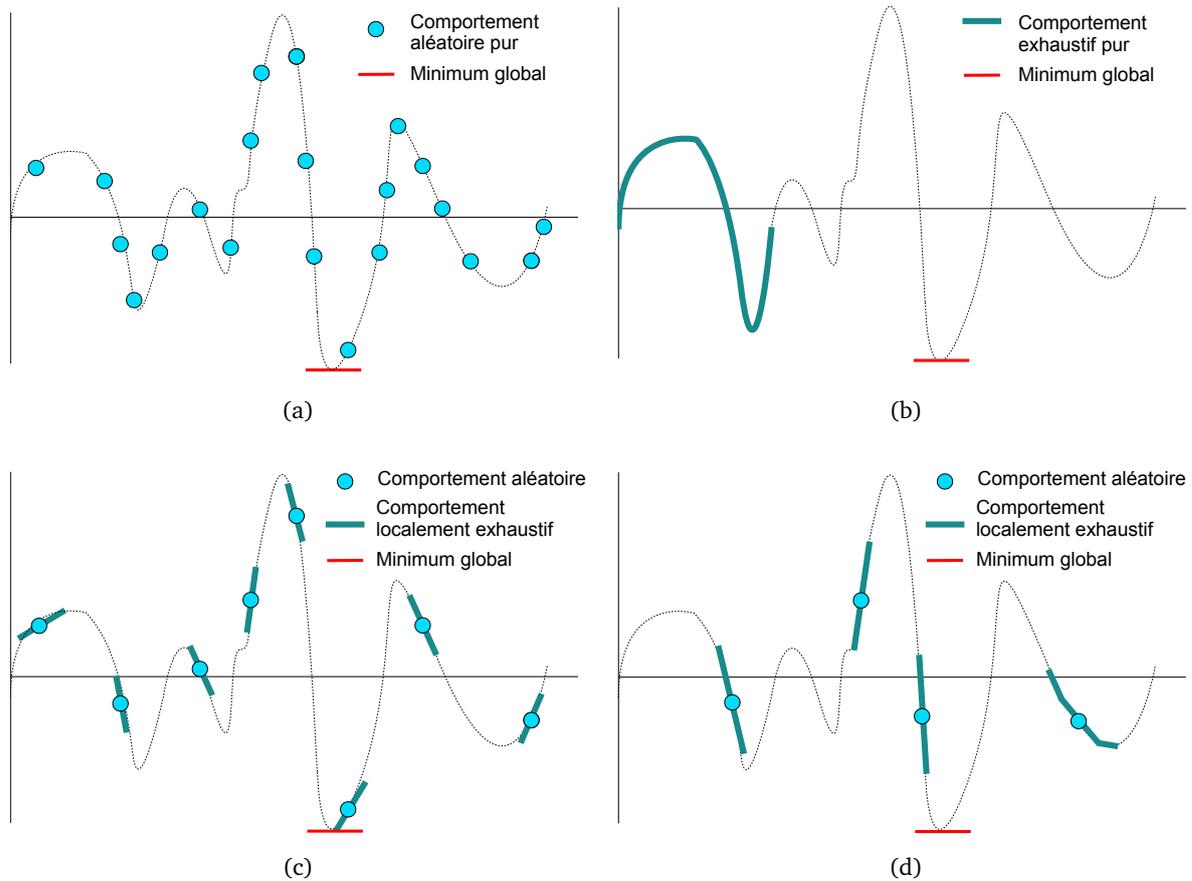


FIGURE 2.12 – Illustration de différents types de comportements de recherche dans un espace de solutions fictif (courbe noire). En abscisse, les solutions possibles, en ordonnée, le résultat correspondant à chaque solution. (a) Comportement purement aléatoire. (b) Comportement purement exhaustif. (c,d) Comportements intermédiaires privilégiant (c) le hasard, (d) l'exhaustivité.

2.4.1 Espace des solutions

Comme dans cette section nous faisons plusieurs fois référence à l'espace des solutions, nous précisons ici ce que nous entendons par ce terme en ce qui concerne les PM , PSM et a fortiori les $SPSM$, aussi bien que les séquences 1-D. Dans la suite, le terme *séquence* désignera aussi bien une séquence 1-D de symboles qu'une séquence 2-D issue d'une matrice.

Définition

Soient k , le nombre de symboles utilisés, w la longueur des mots de code et H_{Min} , la distance de Hamming minimale désirée. L'objectif du codage est de trouver la plus grande séquence de mots de codes respectant H_{Min} . Or comme évoqué précédemment, il est possible de convertir cette séquence de symboles en un grand nombre k -aire qui est incrémenté en permanence de manière à effectuer une recherche exhaustive. Pour une séquence de taille M à remplir, il existe en tout k^M combinaisons possibles de symboles. L'espace des solutions correspond alors l'application qui à chaque combinaison de symboles, associe le nombre de mots de codes de longueur w successifs qui respectent H_{Min} , le premier symbole du premier mot de code correspondant

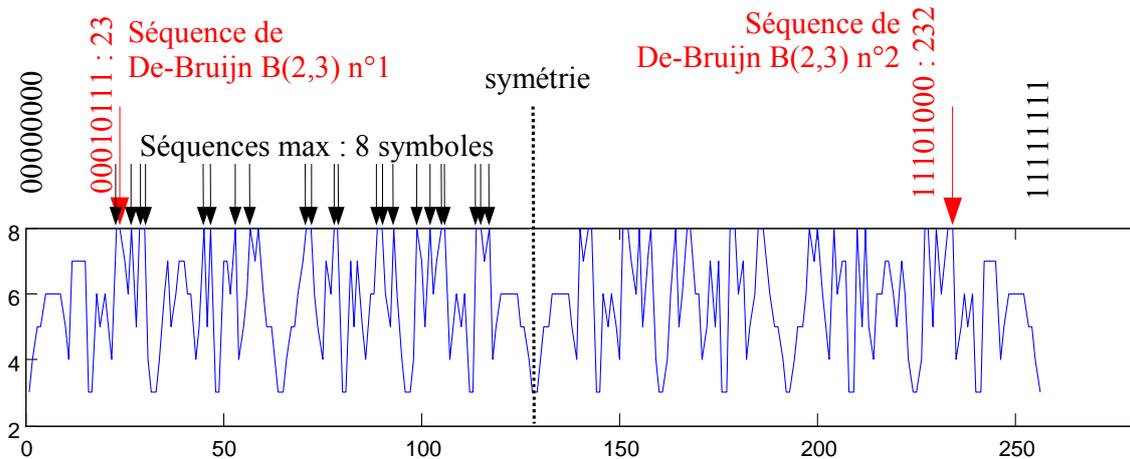


FIGURE 2.13 – Espace des solutions pour des mots de code de longueur 3 utilisant 2 symboles et $H_{Min} = 1$. En abscisse, la séquence de symboles étudiée, en ordonnée, le nombre de mots de codes successifs qui vérifient H_{Min} dans la séquence.

au premier symbole de la séquence. Dans l'exemple donné sur la figure 2.13, $k = 2$, $w = 3$ et $H_{Min} = 1$. Il y a donc en tout $2^8 = 256$ combinaisons de symboles et les séquences ont une longueur maximale de $2^3 = 8$. On observe, dans cet exemple, qu'il existe plusieurs solutions optimales, dont les séquences de De-Bruijn. De plus les minima locaux sont nombreux, l'espace des solutions est donc très loin d'être convexe, d'où l'intérêt d'une contribution aléatoire dans le comportement de recherche.

Dans cet exemple très simple, un comportement exhaustif pur serait suffisant car la taille de l'espace des solutions est très petite. En pratique les paramètres utilisés sont : voisinage 3×3 , nombre de symboles de 2 à 8, pour des matrices de taille $> 50 \times 50$, l'espace des solutions est bien plus important : au minimum $2^{2500} = 3.75 \times 10^{752}$, pour seulement 2 symboles.

Dans cette section, nous proposons de généraliser le comportement de recherche de l'algorithme entre exhaustivité et hasard, grâce à l'introduction du concept de "fenêtre de correction".

2.4.2 Application à la recherche de PSM

L'algorithme proposé exploite le paradigme exploration/exploitation bien connu en intelligence artificielle. Il peut être vu comme une généralisation des approches de Morano [Morano98] - purement aléatoire, qui explore l'espace des solutions au hasard - et de Claes [Claes07] - purement exhaustif, qui exploite et affine des solutions localement trouvées - en paramétrant le comportement de recherche qui sera alors plus ou moins aléatoire. Cela permet de tirer au mieux parti des ressources disponibles en puissance de calcul, pour en extraire les meilleures solutions.

Pour cela nous introduisons le concept de *fenêtre de correction*. Celle-ci permet de définir les éléments de la matrice qui contribueront localement à une recherche exhaustive.

Algorithme : Soient $H_P \times W_P$, les dimensions du motif désiré et $W_A \times H_A$ celles des voisinages (Fig. 2.14.a). Les éléments sont traités ligne par ligne, puis colonne par colonne.

1. Chaque nouvel élément $S_{i,j}$ de la matrice est choisi aléatoirement, comme dans [Morano98].

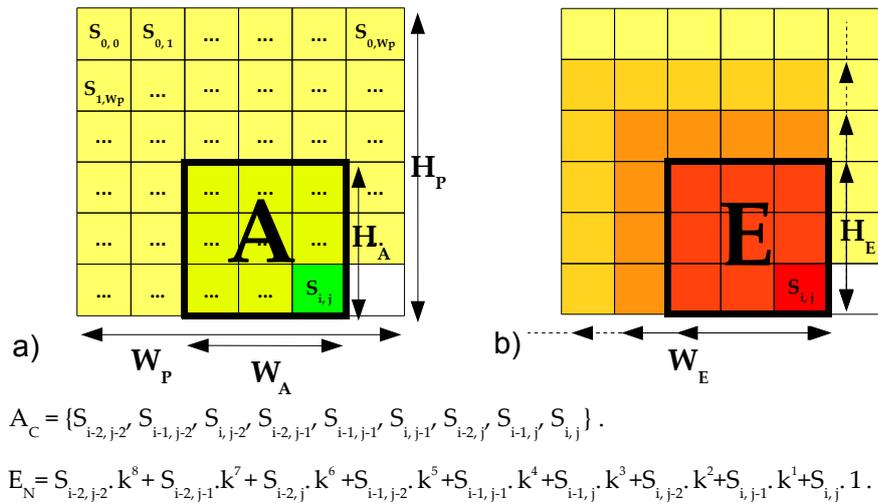


FIGURE 2.14 – a) Ajout d'un nouveau symbole dans la matrice créant un nouveau voisinage. b) Création de la fenêtre de correction suite à un test négatif sur H_{Min} .

Cela permet, contrairement à [Claes08] de maximiser la variance initiale des mots de code et donc la distance de Hamming moyenne.

2. Chaque nouvel élément $S_{i,j}$ ($i \geq H_A$, $j \geq W_A$) crée un nouveau mot de code A_C de taille $W_A \times H_A$ correspondant au voisinage A . La distance de Hamming de A_C , est calculée avec tous les autres mots codes précédemment introduits pour vérifier H_{Min} .
3. Si le test est positif, le symbole est associé à cet élément de la matrice et l'élément suivant est traité.
4. Sinon, une fenêtre de correction E de taille $H_E \times W_E$ est créée et les symboles s'y trouvant sont convertis en un nombre k -aire de correction E_N (Fig. 2.14.b). Les chiffres de E_N sont incrémentés en évoluant de leur valeur initiale $I_{i,j}$ à $(I_{i,j} + C - 1)$ modulo C . De cette manière les k valeurs possibles sont essayées avant que le chiffre de poids supérieur ne soit à son tour incrémenté. Cela permet d'effectuer une recherche localement exhaustive (seulement dans E). Le nombre d'incrémentations de chaque élément est stocké dans un compteur propre.
5. E_N est incrémenté tant que les modifications induites ne permettent pas d'obtenir un test H_{Min} positif ou que tous les chiffres de E_N ont été incrémentés au maximum de leur valeur.
6. Si E_N ne peut plus être incrémenté (toutes les valeurs ont été testées), la matrice est effacée complètement et la recherche recommence, comme dans [Morano98].

Propriétés du comportement de recherche :

- *Comportement exhaustif* : plus la taille de la fenêtre de correction E est grande, plus le comportement de l'algorithme sera exhaustif, une plus grande zone contiguë de l'espace des solutions sera exploitée pour corriger les éléments conflictuels.
- *Comportement aléatoire* : plus la taille de la fenêtre de correction E est petite, plus le comportement sera aléatoire. La recherche commencera par des états initiaux aléatoires mis à jour plus fréquemment, permettant de sortir plus facilement des minimums locaux rencontrés.

Taille de la fenêtre de correction :

Il est très difficile de prédire quel est l'équilibre aléatoire/exhaustif qui donnera le meilleur résultat. Sur la figure 2.12, qui présente un espace de solution fictif, la solution (c), qui privilégie l'aspect aléatoire par rapport à la solution (d), semble mieux approcher le minimum local. En pratique, différentes valeurs devront être essayées pour trouver celle qui permet d'obtenir le meilleur résultat.

Cependant, il est possible de donner **une règle générale** : plus les ressources disponibles sont importantes (forte puissance de calculs disponibles et/ou faibles contraintes sur le problème de codage) plus le comportement pourra être exhaustif. En effet, plus de temps pourra alors être alloué à affiner la solution courante.

Généralisation à N dimensions :

La fenêtre de correction peut aussi bien être adaptée à la génération de tableaux en N dimensions, la fenêtre de correction aura alors N dimensions. Par exemple, des séquences de nombres peuvent être générées pour le codage 1-D de franges à projeter, et proposer des séquences tout comme celle de De-Bruijn avec la possibilité en plus, dans le cas de cet algorithme brute-force, d'imposer une valeur de H_{Min} supérieure à 1.

2.5 Test d'unicité dans l'espace des mots de code

Une tâche commune à tous les algorithmes de génération de PSM présentés précédemment est le test d'unicité relatif à la distance de Hamming minimum que l'on veut imposer. Durant le codage de la matrice, les approches brute-force testent, les uns après les autres, les codes que produisent les nouveaux symboles introduits dans la matrice pour vérifier que la contrainte sur la valeur de H_{Min} désirée est bien satisfaite. Pour cela la distance de Hamming h avec tous les codes déjà validés dans la matrice est calculée. Si pour l'un de ces codes $h < H_{Min}$, alors il n'est pas nécessaire d'effectuer le calcul avec les codes précédents, le test d'unicité est négatif. Ce test est donc effectué **dans l'espace de la matrice**.

Si N codes sont déjà présents dans la matrice, le prochain code créé avec le nouvel élément introduit devra être comparé, dans le pire des cas, à ces N codes. Ainsi, la complexité algorithmique de ce test d'unicité est de $\mathcal{O}(N)$. Pour une matrice de N éléments à remplir, la matrice sera remplie, durant la recherche, de plus de $N/2$ éléments en moyenne. En effet, plus il y aura d'éléments dans la matrice, moins la contrainte H_{Min} sera satisfaite en moyenne, plus l'algorithme effectuera des corrections. Dans le cas où le codage est très peu contraint (grand nombre de symboles et faible valeur de H_{Min} par rapport à la taille de la PSM recherchée), le test d'unicité sera très souvent positif et la matrice sera remplie à moitié en moyenne. Donc **une limite inférieure de la complexité algorithmique du test d'unicité dans l'espace de la matrice** est de $\mathcal{O}(N/2)$.

Il est très intéressant de pouvoir améliorer la complexité algorithmique d'un tel test, car, comme vu précédemment, une telle amélioration pourra être employée quelle que soit la stratégie de recherche (aléatoire, exhaustive, localement exhaustive) de l'algorithme et le type de codage généré (codage 1-D, PSM, SPSM), pour tous les algorithmes de codage utilisant l'approche *essais-erreurs*.

Pour diminuer cette complexité, nous proposons d'effectuer le test d'unicité non pas dans *l'espace de la matrice*, mais dans *l'espace des mots de codes*, ce qui constitue une nouveauté.

L'espace des mots de code. L'espace des mots de code est constitué de l'ensemble, $\Omega_{C,W}$, de tous les mots de code C -aire possibles de longueur W . Ainsi, il y a en tout $\text{card}(\Omega_{C,W}) = C^W$ mots de codes possibles.

Test d'unicité dans l'espace des mots de code. Comme nous l'avons vu précédemment, pour une valeur de H_{Min} donnée, effectuer le test d'unicité dans l'espace de la matrice revient à comparer le code A à tester avec tous les codes déjà utilisés pour vérifier qu'aucun n'a une distance de Hamming avec A qui est inférieure à H_{Min} . Pour tester l'unicité dans l'espace des mots de code, le nouvel algorithme devra considérer tous les mots de codes ayant une distance de Hamming inférieure à H_{Min} avec A et vérifier qu'aucun n'est associé à un voisinage déjà validé dans la matrice.

Nous proposons tout d'abord une méthode exploitant l'espace des mots de code pour réduire la complexité du test d'unicité à un **accès direct** dans le cas où $H_{Min} = 1$, puis nous généraliserons pour $H_{Min} \geq 1$.

2.5.1 Réduction de complexité pour $H_{Min} = 1$

Dans le cas où $H_{Min} = 1$, chaque code ne peut être utilisé qu'une seule fois, mais tous les C^W codes possibles peuvent être utilisés, comme c'est le cas pour les matrices parfaites. L'unicité pour tous les codes possibles peut alors être représentée par un tableau Tab_{MDC} à 1 dimension contenant C^W drapeaux booléens, un drapeau pour chaque mot de code. L'espace des mots de code, E_{MDC} , est alors exploité de la façon suivante : dans Tab_{MDC} , chaque drapeau est uniquement adressé par un mot de code A de $\Omega_{C,W}$ de telle sorte que les W symboles du mot de code C -aire soient les coefficients en base C (par exemple, avec $C = 3$, $W = 9$, le mot de code 000000012 correspond à l'adresse décimale $2 \times 3^0 + 1 \times 3^1 + \dots + 0 \times 3^8 = 5$ du tableau Tab_{MDC}). La formule générale étant :

$$\text{Adresse}(A) = \sum_{i=1}^W A_i C^{W-i}. \quad (2.11)$$

Tous les drapeaux de Tab_{MDC} sont initialement positionnés à *Faux*. Si un drapeau est à *Faux* alors le mot de code correspondant n'est pas encore utilisé, sinon, le code n'est plus disponible. Ainsi, pour chaque nouveau code A à tester, tous les drapeaux correspondant à $H < H_{Min}$ avec A sont vérifiés. Si tous ces drapeaux sont à *Faux*, le test d'unicité est positif, le symbole ayant créé A peut alors être ajouté au motif et le drapeau correspondant prend alors la valeur *Vrai* dans Tab_{MDC} . Sinon un autre code devra être essayé. Dans le cas trivial où $H_{Min} = 1$, il n'y a que le drapeau correspondant à A lui-même qui devra être testé car c'est le seul code ayant une distance de Hamming égale à 0 ($H = (H_{Min} - 1) = 0$) dans E_{MDC} . De cette façon, la complexité du test d'unicité est réduite à $\mathcal{O}(1)$, c'est à dire un accès direct, contre plus de $(m \times n) \div 2 = 5000$ tests pour un motif de taille 100×100 en utilisant les tests dans l'espace de la matrice. Dans ce cas, comme nous le verrons dans les résultats, le gain en réduction de

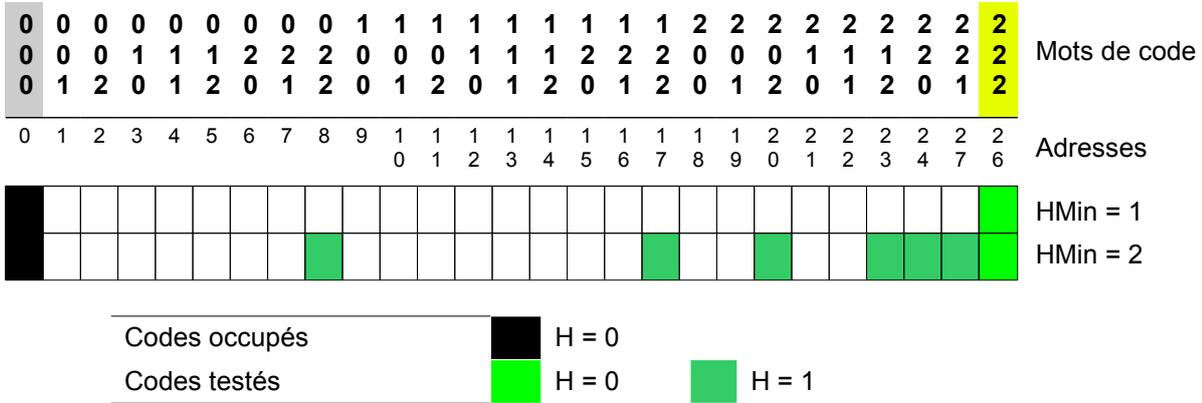


FIGURE 2.15 – Exemple de test d’unicité dans l’espace des mots de code pour un cas simple.

complexité est énorme.

Dans le cas $H_{Min} = 1$, cela fait clairement, du test dans l’espace des mots de code, la technique la plus indiquée par rapport au test dans l’espace de la matrice, quels que soient la taille de la matrice, le nombre de symboles utilisés, la distance de Hamming minimum, ou le type de matrice PSM, SPSM, à générer.

Ressources mémoire nécessaires : comme un drapeau doit être stocké pour chaque mot de code, C^W valeurs booléennes et donc, C^W bits sont nécessaires. Pour les valeurs utilisées en pratique et dans notre étude, cela équivaut au maximum à 8^9 bits $\simeq 16$ Moctets, ce qui est largement acceptable avec les ordinateurs actuels.

2.5.2 Généralisation pour $H_{Min} \geq 1$

Pour tout C, W et H_{Min} , le nombre de drapeaux, N_f à tester dans l’espace des mots de code E_{MDC} est donné par :

$$N_f(H_{Min}) = 1 + \sum_{H=1}^{H_{Min}-1} \frac{W!}{H!(W-H)!} (C-1)^H . \tag{2.12}$$

Nous avons présenté et démontré cette formule dans la section 2.3 de ce chapitre.

La figure Fig.2.15 illustre cela pour $C = W = 3$, pour des raisons de simplification d’illustration, et pour $H_{Min} = 2$ (dernière ligne). Dans cet exemple, le nouveau mot de code à tester est $A = \{2, 2, 2\}$ et son adresse est 26. Le nombre de tests à effectuer dans l’espace des mots de code est donc $N_f = 1 + \frac{3!}{(3-1)!} \cdot (3-1)^1 = 1 + 6 = 7$ pour vérifier $H_{Min} = 2$. Les 7 cellules correspondantes sont en vert ($H = 0$ en vert clair (le mot de code lui-même !) et $H = 1$ en vert foncé). Il en résulte que toutes les autres adresses correspondent à des mots de code ayant $H \geq 2$ avec $\{2, 2, 2\}$ et sont dessinées en blanc et noir (le code $\{0, 0, 0\}$ est supposé être déjà présent dans le motif).

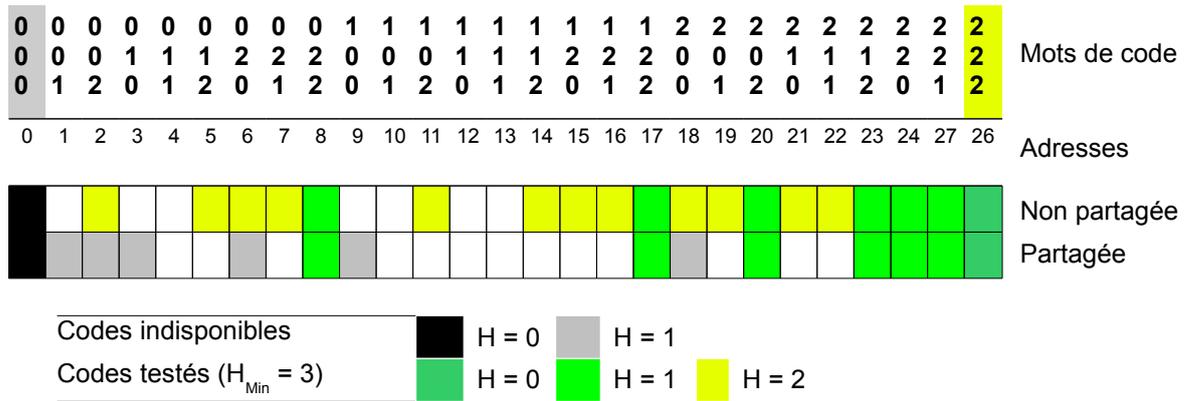


FIGURE 2.16 – Illustration de la différence du nombre de tests à effectuer dans l'espace des mots de codes entre la recherche non partagée (19 tests) et la recherche partagée (14 tests) dans un cas simplifié.

2.5.3 La stratégie de recherche partagée

La méthode précédente est efficace pour des valeurs H_{Min} faibles, mais comme le montre l'équation 2.12, le nombre de tests évolue exponentiellement avec la distance de Hamming H qui évolue entre 0 et $H_{Min} - 1$. Sachant que $exp(H_1 + H_2) \gg exp(H_1) + exp(H_2)$, pour les valeurs de H_{Min} plus importantes, nous proposons de partager l'effort de recherche entre le code à tester et ceux déjà introduits grâce à un code intermédiaire.

En effet, si on considère deux mots de code A_1 et A_2 ayant une distance de Hamming de $H_{1,2}$, alors il existe au moins un mot de code A_3 ayant des distances de Hamming de H_1 et H_2 respectivement avec A_1 et A_2 de telle façon que $H_{1,2} = H_1 + H_2$ avec $H_1 \leq H_{1,2}$ et $H_2 \leq H_{1,2}$.

Preuve

Il suffit pour cela que A_3 possède H_1 symboles de A_2 différents de A_1 et les H_2 autres symboles de A_1 différents de A_2 , les symboles communs à A_1 et A_2 se retrouvant dans A_3 .

Exemple

$A_1 = (000001234)$, $A_2 = (000004321)$, $H_{1,2} = 4$. On peut alors trouver des codes A_3 avec $H_1 = 1$ et $H_2 = 3$ (par exemple : $A_3 = (000001231)$) ou $H_1 = 2$ et $H_2 = 2$ ($A_3 = (000001221)$).

Algorithme

L'algorithme effectuant le test avec une recherche partagée procède alors comme suit :

On suppose que $H_1 + H_2 = (H_{Min} - 1)$, où

- H_1 est la distance de Hamming maximale entre les codes validés $C_{validé}$ et les codes aux adresses desquels un drapeau doit être placé pour chaque code validé.
- H_2 est la distance de Hamming maximale entre les nouveaux codes C_{Test} dont l'unicité doit être testée et les codes aux adresses desquels la présence d'un drapeau doit être testée.

Initialement, tous les drapeaux de l'espace des mots de code sont à *Faux*.

Pour chaque code $C_{validé}$, les drapeaux aux adresses de tous les codes ayant une distance de Hamming comprise entre 0 et H_1 inclu avec $C_{validé}$ sont placés à *vrai*.

Pour chaque code à tester, C_{Test} , tous les drapeaux se trouvant aux adresses associées aux codes ayant une distance de Hamming comprise entre 0 et H_2 inclu avec C_{Test} sont testés jusqu'à ce que l'un d'entre eux ait la valeur *vrai*.

Si l'un d'entre eux retourne la valeur *vrai*, cela signifie qu'il existe au moins un code $C_{validé}$ dans la matrice qui a une distance de Hamming comprise entre 0 et $(H_{Min} - 1)$ avec C_{Test} , le test d'unicité a donc échoué. Sinon, le test d'unicité a réussi.

Le nombre de tests à effectuer est alors :

$$N_{f_{Separee}}(H_{Min}) = N_f(H_1 + 1) + N_f(H_2 + 1), \text{ avec } H_1 + H_2 = H_{Min} - 1. \quad (2.13)$$

Complexité algorithmique

Dans la figure 2.16, nous considérons l'exemple précédent avec $H_{Min} = 3$ cette fois-ci. Le nombre de tests à effectuer est alors $N_f(3) = 1 + 7 + \frac{3!}{(2!(3-2)!)} \cdot (3-1)^2 = 19$ (cellules jaunes et vertes) pour la recherche non partagée. Avec le partage de la recherche, à partir des mots de code $\{0, 0, 0\}$ (déjà utilisé dans la matrice) et $\{2, 2, 2\}$ (à tester), seuls les mots de codes ayant $h = (0, 1)$ avec eux doivent être cherchés, au lieu que les codes ayant $h = (0, 1, 2)$ avec le code $\{2, 2, 2\}$ ne soient cherchés (cas de la recherche non partagée). Comme le nombre de tests augmente exponentiellement avec h (équation (2.12)), cette stratégie requiert alors moins de tests. Le nombre de tests requis est alors simplement l'addition de $N_f(2) = 7$, cellules noire et grises ayant $h = (0, 1)$ avec le mot de code $\{2, 2, 2\}$, et de $N_f(2) = 7$, cellules vertes claires et foncée (deuxième ligne). Ce qui totalise 14 tests pour l'approche partagée au lieu de 19 pour l'approche non partagée. Ce gain est beaucoup plus significatif avec les paramètres typiques des motifs utilisés (notamment $W = 9$ pour les voisinages 3×3 considérés dans cette étude).

2.5.4 Résultats de complexité algorithmique

Nous reportons dans la figure 2.17, les résultats théoriques de complexité (en terme de nombre de tests à effectuer) avec des valeurs typiques pour le nombre de symboles ($C \in [2, 8]$) et la distance de Hamming minimale ($H_{Min} \in [1, 5]$). Nous montrons le nombre de tests $\mathcal{O}(N/2)$ pour des matrices de taille 100^2 (en orange) et 200^2 (en rouge), la recherche non partagée dans l'espace des mots de code (voir équ. 2.12) (en jaune) et la recherche partagée (voir équ. 2.13) (en vert). Pour la recherche partagée, nous avons choisi H_1 et H_2 de telle sorte qu'ils soient le plus équilibrés possible.

Malgré la complexité exponentielle par rapport à H_{Min} , la recherche dans l'espace des mots de code avec l'approche partagée (Fig.2.17, vert) nécessite en moyenne beaucoup moins de tests que l'approche classique dans l'espace de la matrice. Cela est dû à la très faible dynamique de h dans la recherche partagée. En effet, au maximum, pour $H_{Min} = 5$, il faut vérifier que tous les codes ayant $H \in [1, 4]$ avec le code à tester, sont libres. Ce qui mène, avec la recherche partagée, à une valeur de H maximum de 2 pour les codes utilisés et 2 pour les codes à tester.

2.6 Tests d'unicité exploitant la géométrie épipolaire

Jusqu'ici, tous les algorithmes de génération de matrices pseudo-aléatoires exploitant le voisinage spatial, proposés dans la littérature, assurent l'unicité entre toutes les primitives sur l'ensemble du motif. Comme nous l'avons vu, en pratique, certains algorithmes vont plus loin en assurant une unicité invariante aux rotations, ce qui mène à des contraintes supplémentaires

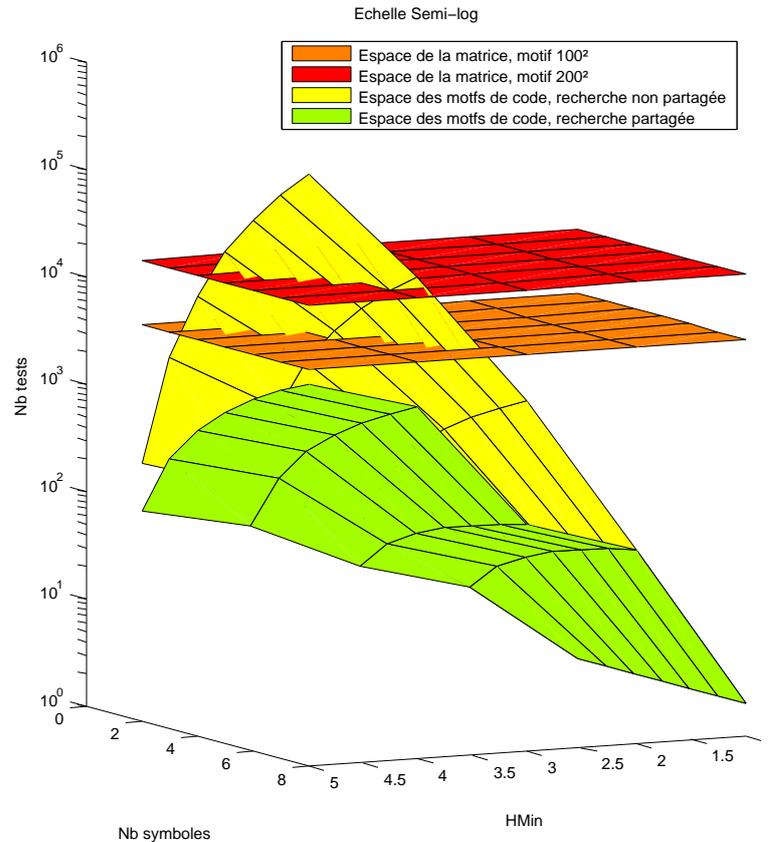


FIGURE 2.17 – Estimation du nombre de tests d'unicité dans l'espace de la matrice (rouge) et dans l'espace des mots de code sans (jaune) et avec (vert) le partage de la recherche.

sur le codage et à des codes aux propriétés moins intéressantes, comme l'ont reporté [Claes08] et [Adán09] par exemple. Comme il est plus intéressant, dans la majorité des cas, d'avoir soit moins de symboles, soit une plus grande taille de motif, soit une distance de Hamming minimale plus importante, il est souhaitable, au contraire, de diminuer les contraintes de codage autant que possible.

Dans le cas où la configuration caméra - projecteur reste fixe durant l'acquisition, ce qui représente la grande majorité des cas d'utilisation d'un tel système stéréo, nous proposons d'exploiter la géométrie épipolaire dès le codage du motif pour diminuer significativement les contraintes de codage, notamment les contraintes qui reposent sur l'unicité des codes.

A la fin du chapitre 1, nous avons introduit l'homographie épipolaire. Nous présentons deux nouvelles approches exploitant la géométrie épipolaire pour diminuer significativement les contraintes de codage. La première, exploitant la proximité épipolaire des primitives à travers un alignement local des primitives est plus indiquée si les épipoles sont relativement proches du centre des images du projecteur et de la caméra. La seconde exploitant l'alignement épipolaire sera préférée si l'épipole est loin du centre de l'image et que les lignes épipolaires apparaissent quasiment parallèles.

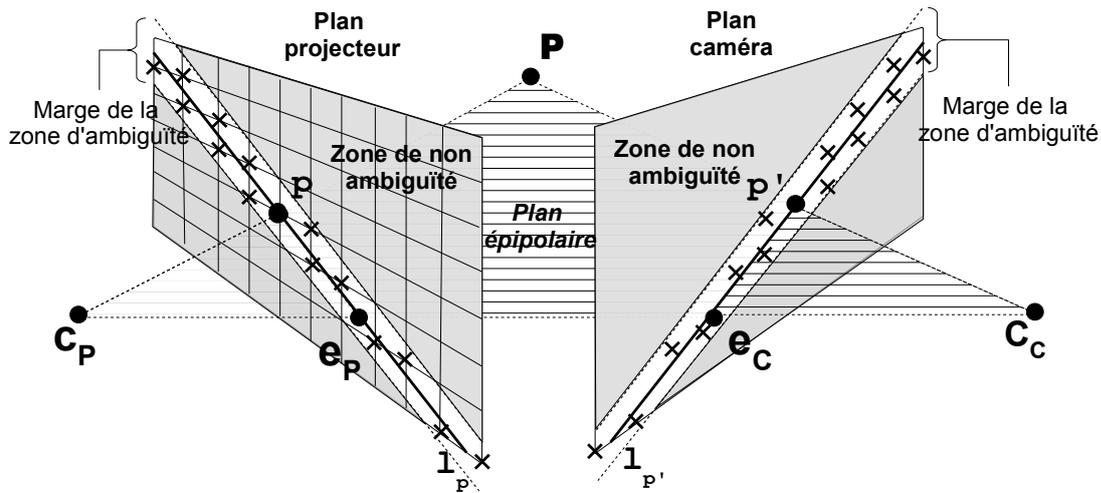


FIGURE 2.18 – Modélisation d'un système stéréo projecteur-caméra avec le modèle sténopé. c_p , e_p , p , et l_p : respectivement, le centre de projection, l'épipôle, le centre d'une primitive du motif et sa ligne épipolaire dans le plan image du projecteur. c_c , e_c , p' et $l_{p'}$: les entités homologues dans le plan image de la caméra. P : projection du point p sur la scène 3-D. *Zone d'ambiguïté* (blanc) : zone qui contient des ambiguïtés potentielles de mise en correspondance du point p .

2.6.1 Unicité par proximité épipolaire

Comme la géométrie épipolaire est connue à une erreur d'estimation près, correspondant généralement à la distance moyenne entre les points utilisés pour l'estimation et leur ligne épipolaire respective, il est nécessaire de considérer une marge autour de chaque ligne épipolaire. Ainsi, le test d'unicité ne doit être effectué qu'entre les primitives se trouvant dans la marge d'une même ligne épipolaire. A cette fin, le test d'unicité dans l'espace de la matrice présenté dans la section précédente, doit être modifié de la façon suivante :

Avant le codage :

- Nous considérons que la matrice Fondamentale est disponible et que la configuration projecteur-caméra est fixe ou que la géométrie peut-être recalculée à chaque image.
- Pour chaque ligne épipolaire, considérer une marge dans le plan projecteur pour la *zone d'ambiguïté* (voir Fig. 2.18).
- Pour chaque voisinage du motif dont l'unicité doit être vérifiée, identifier les coordonnées de la primitive P au centre du voisinage dans le plan projecteur, puis considérer la ligne épipolaire passant par ce point.
- Estimer la zone d'ambiguïté autour de cette ligne épipolaire en prenant en compte la précision de l'estimation de la matrice fondamentale.
- Identifier toutes les primitives se trouvant dans cette zone d'ambiguïté et les ajouter à la *liste de tests* des primitives dont l'unicité avec la primitive P devra être vérifiée.

Durant le codage :

- Pour chaque test d'unicité, au lieu de vérifier la contrainte H_{Min} entre la primitive P et tous les voisinages précédemment trouvés, identifier toutes les primitives dans la *liste de tests* de P .

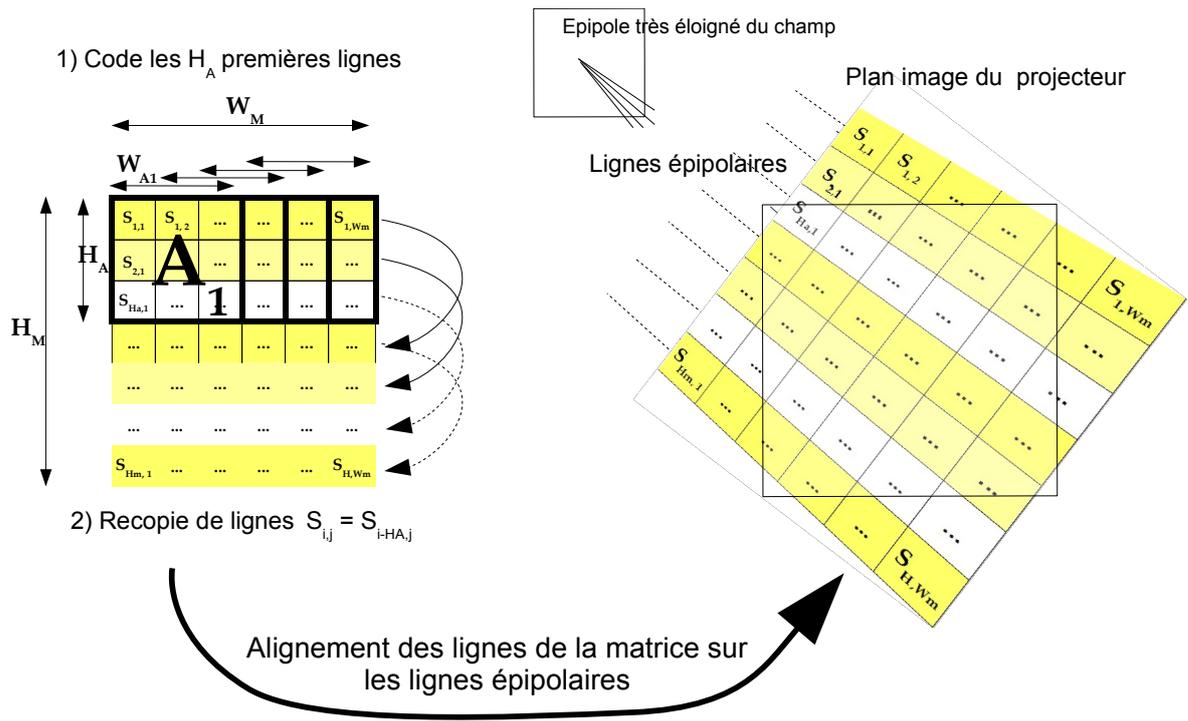


FIGURE 2.19 – Principe du codage pour l’alignement épipolaire.

- Vérifier que la contrainte sur H_{Min} est satisfaite avec tous les voisinages déjà codés, dont la primitive centrale se trouve dans cette *liste de tests*.

De cette façon, seulement un sous-ensemble de mots de codes va contribuer à chaque test d’unicité, générant une SPSM, au lieu de tous les mots de codes comme avec des PSM. Comme le montre la section résultats de ce chapitre, cette approche permet une réduction de complexité du test de codage de \sim 1-D au lieu de 2-D. De plus, peuvent alors être codés en temps-réel, des motifs avec des propriétés bien plus intéressantes que ce qu’il est possible de coder avec des PSM.

Pour exploiter conjointement la proximité épipolaire avec la recherche dans l’espace des mots de code, il faut également stocker un drapeau de plus dans l’espace des mots de code qui indique pour quelle ligne épipolaire le code est utilisé. Ainsi le test d’unicité ne pourra être négatif que pour les codes associés à une même ligne épipolaire. Ceci fonctionne également avec la recherche séparée.

2.6.2 Unicité par alignement épipolaire

Dans le cas où l’épipole est suffisamment loin du centre de l’image, les lignes épipolaires sont pratiquement parallèles et une disposition des primitives avec un alignement direct des lignes de code sur les lignes épipolaires est possible tout en conservant une distribution homogène des primitives dans l’image.

Principe

Dans le cas de l’alignement épipolaire (voir Fig. 2.19), il y a une correspondance directe entre les lignes de la matrice codée et les lignes épipolaires. Ainsi l’unicité ne doit être assurée

qu'entre les primitives qui sont sur une même ligne de la matrice. Mais comme la distance de Hamming se calcule position à position, il est possible, contrairement à l'approche par proximité épipolaire, de ne coder que les H_A premières lignes pour des voisinages de taille $W_A \times H_A$. En effet, ces H_A premières lignes peuvent être recopiées directement sur les lignes suivantes, chacune des H_A premières lignes étant recopiées toutes les H_A lignes (voir Fig. 2.19.a) et ceci indéfiniment, pouvant mener à des motifs possédant un nombre de lignes virtuellement infini, la contrainte portant alors seulement sur le nombre de colonnes qu'il est possible de coder.

Si les H_A premières lignes doivent être codées, les codes des voisinages $W_A \times H_A$, ne sont cependant créés qu'à la H_A ème ligne. Et donc les tests d'unicité ne sont effectués qu'entre les voisinages créés à partir de cette H_A ème ligne, c'est-à-dire selon une dimension. Cela correspond donc à une réduction dimensionnelle de la complexité des tests d'unicité.

Par rapport à l'unicité par proximité épipolaire, cette approche offre plusieurs avantages :

- Il n'y a pas besoin de choisir de marge d'ambiguïté.
- Il n'y a pas besoin de stocker pour chaque primitive, les lignes épipolaires sur lesquelles sont alignées ses primitives voisines.
- Une seule ligne de voisinages doit être codée.
- L'adjacence verticale des voisinages ne contraint pas le codage ici car les lignes de code sont directement recopiées selon la direction verticale pour remplir la matrice.

Ces deux nouvelles approches du codage par voisinage spatial produisent ainsi une nouvelle classe de codes nommés *SPSM* pour subperfect submap ou sous-matrices sous-parfaites à travers un **codage 1,5D** (2-D pour les voisinages 3×3 utilisés, mais 1-D pour l'unicité des codes).

2.7 Codage optimal et alignement épipolaire

Définition

Un codage d'une matrice de taille $W_p \times H_p$, avec k symboles, et imposant une distance de Hamming minimale H_{Min} entre des voisinages de taille $W_A \times H_A$ est dit optimal si aucun des paramètres (k (à minimiser), H_{Min} (à maximiser), $W_p \times H_p$ (à maximiser), $W_A \times H_A$ (à minimiser)) ne peut être amélioré sans modifier les autres.

Les codes optimaux sont très intéressants car il ont une **propriété absolue** : on ne pourra pas faire mieux.

Les approches directes évoquées (section 2.2) permettent l'obtention d'un codage optimal, mais avec des limitations importantes concernant la valeur de H_{Min} que l'on peut imposer et le rapport d'aspect des matrices obtenues.

Nous avons vu (section 2.2) dans [Claes07], qu'il est possible de générer des codes optimaux avec les algorithmes brute-force en explorant l'espace des solutions de façon exhaustive. Cependant, dans l'algorithme proposé par [Claes07], seules des matrices de taille très réduite, $\in (3 \times 4, 4 \times 5)$ ont pu être générées de façon optimale à cause de l'explosion combinatoire des possibilités avec des matrices plus grandes. De telles tailles de matrice sont inexploitable en lumière structurée, car elles mèneraient à la conception de motifs ayant une résolution bien trop limitée. Comme nous venons de proposer une méthode permettant une réduction dimensionnelle de la complexité des tests d'unicité à effectuer et également du nombre d'éléments de la matrice à remplir, il sera intéressant de voir, dans la section résultats, quelles sont les propriétés des motifs qui peuvent être obtenues avec un codage optimal exploitant l'alignement épipolaire.

k/H_{Min}	1				2			3				4			5		
Note					Claes : N.A.							Claes : N.A.			Morano : N.A.		
Algo	Nous <1s	Nous Max	C.	M.	Nous <1s	Nous Max	M.	Nous <1s	Nous Max	C.	M.	Nous <1s	Nous Max	M.	Nous <1s	Nous Max	C.
2	20 ²	22 ²	19x25	NA	10 ²	11 ²	NA	6 ²	7 ²	6x8	NA	5 ²	5 ²	NA	NA	NA	4x5
3	109 ²	111 ²	106x114	>= 45 ²	34 ²	37 ²	< 45 ²	15 ²	16 ²	14x18	< 45 ²	9 ²	9 ²	< 45 ²	6 ²	7 ²	6x8
4	397 ²	418 ²	316x421	>= 45 ²	103 ²	104 ²	< 45 ²	34 ²	37 ²	33x44	< 45 ²	16 ²	17 ²	< 45 ²	10 ²	10 ²	9x11
5	815 ²	1155 ²	429x572	>= 45 ²	242 ²	262 ²	>= 45 ²	67 ²	73 ²	68x90	< 45 ²	27 ²	27 ²	< 45 ²	12 ²	14 ²	12x16
6	855 ²	2580 ²	496x661	>= 45 ²	429 ²	548 ²	>= 45 ²	123 ²	133 ²	124x165	< 45 ²	41 ²	43 ²	< 45 ²	17 ²	20 ²	18x23
7	905 ²	5205 ²	676x901	>= 45 ²	445 ²	1023 ²	>= 45 ²	196 ²	228 ²	204x272	< 45 ²	44 ²	64 ²	< 45 ²	19 ²	27 ²	24x32
8	905 ²	9620 ²	696x925	>= 45 ²	365 ²	1645 ²	>= 45 ²	223 ²	344 ²	292x389	>= 45 ²	22 ²	92 ²	< 45 ²	14 ²	35 ²	32x42

Plus grande matrice, gain en taille :

 < 100% or N.A. 100% - 500% > 500%Plus grande matrice générée en moins de 1 sec : 

FIGURE 2.20 – Comparative results of perfect submaps generation. N.A. : Not Avail., M. : [Morano98] C. : [Claes08]. Espace des mots de code, fenetre 3x3

2.8 Résultats de codage

Nous présentons ici différents résultats de codage obtenus en mettant en pratique les différentes approches présentées dans ce chapitre. Comme le nombre de variantes possibles est très important, nous n'avons pu tester toutes les combinaisons possibles. Cependant les résultats présentés ici permettent de mettre en avant les avantages et les limites des contributions proposées. Nous comparons les résultats obtenus avec ceux de Morano [Morano98], Claes ([Claes07], [Claes08]), Albitar [Albitar07] et Adán[Adán09].

2.8.1 Codage PSM

Le tableau 2.20 reporte les résultats obtenus pour $H_A = W_A = 3$, $H_{Min} \in [1, 5]$ et $k \in [2, 8]$. Ces paramètres nous permettent de les comparer aux résultats de Claes [Claes08] et de Morano [Morano98]. Avec les ressources disponibles (Pentium 4 3Ghz, 4 Go RAM DDR2), les meilleurs résultats sont obtenus en quelques minutes avec une fenêtre de correction de taille $H_E = W_E = 5$. Les tests d'unicité sont effectués, dans notre cas, dans l'espace des mots de code. Dans tous les cas, nous obtenons des matrices plus grandes que Morano [Morano98], qui essayait d'en générer pour une taille de 45². Nous parvenons en moyenne à générer des matrices de taille équivalente avec 2 symboles de moins. Pour $H_{Min} > 2$ les résultats obtenus sont équivalents à ceux de [Claes08] en termes de taille (il est quasiment impossible de faire mieux qu'une approche purement exhaustive, si celle-ci dispose de beaucoup de temps), mais pas en termes de temps de calculs où ces résultats sont obtenus au bout de plusieurs heures voire plusieurs jours dans [Claes08], alors qu'en moins d'une seconde nous pouvons générer des solutions de taille entre 80% et 95% des maximums reportés (en orange dans le tableau). Pour $H_{Min} \leq 2$, notre approche fait également la différence. En effet, ici un comportement trop exhaustif amène à une explosion combinatoire alors que notre algorithme continue d'obtenir des matrices de tailles qui augmentent quasi linéairement avec k , le nombre de symboles utilisés.

Taille	Symboles	k = 2	k = 4	k = 6	k = 8
	H_{min}				
50 ² (15)		2 (M : < 45 ²)	5 (C : 9x11) (Cri : 6x7)	6	6
75 ² (10)		1	4	5	6
100 ² (8)		1*	4*	5	6
150 ² (6)		1	4	5	5
200 ² (4)		1 (C : 19x25) (Cri : 11x14)	3 (C : 33x44) (Cri : 18x24) (M : < 45 ²)	5 (C : 18x23) (Cri : 10x13) (M : < 45 ²)	5 (C : 32x42) (Cri : 14x18) (M : < 45 ²)

TABLE 2.2 – Résultat de codage (SPSM) par proximité épipolaire . Pour chaque taille de matrice et nombre de symboles utilisés (k), la distance de Hamming minimale est reporté. (x) : épaisseur en pixels de la marge de la zone d’ambiguïté. C, Cri, M : tailles obtenues dans [Claes08] ((C) sans et (Cri) avec invariance des codes par rotation) et dans (M) [Morano98] avec des valeurs de H_{Min} et k identiques.

2.8.2 Codage SPSM

Nous reportons ici les différents résultats obtenus pour la génération de codages de type SPSM par proximité épipolaire puis par alignement épipolaire.

Par proximité épipolaire

Pour tester les bénéfices de la contrainte d’unicité par proximité épipolaire, un ensemble de motifs de taille et nombre de symboles variables a été généré. Les résultats sont reportés dans le tableau (Fig. 2.21.a). Comme nous pouvons le voir, comparés aux PSM générées par [Morano98] (M) et [Claes08] (Cri,C : avec ou sans invariance par rotation), des motifs de taille largement plus importante ont pu être obtenus avec une H_{Min} et un nombre de symboles k équivalents. De plus, des motifs de taille plus importante ($\geq 200^2$) que ceux généralement utilisés en pratique dans la littérature (voir chapitre 1), ont pu être générés avec **deux symboles seulement**.

Par alignement épipolaire

Pour les résultats reportés par alignement épipolaire (voir Tab. 2.3), nous nous sommes arrêtés à 5000 colonnes codées. Au-delà ce n’est pas réellement intéressant car difficilement exploitable par les résolutions des périphériques de projection et de capture actuels. A voir, cependant, car la résolution de ceux-ci continue d’augmenter rapidement. Rappelons que le nombre de lignes, pour ces codes, peut être infini (voir section 2.6.2). Nous voyons que les tailles de matrices pouvant être générées sont bien supérieures, dans tous les cas, à ce qui peut être produit aussi bien avec un codage PSM qu’avec un codage SPSM par proximité épipolaire. De plus, nous atteignons le maximum absolu pour $A = 2$ et $H_{Min} = 1$ à 1 voisinage près, 512 voisinages 3×3 maximum (2^9), nous en avons obtenu 511 ($513 - 2$). Cela a été possible ici grâce à la réduction de complexité du problème de codage des tests. Les tests ont été effectués dans l’espace de la

k/H_{Min}	1		2		3		4		5	
Algo	<40ms	SPSM aligné Max								
2	503	513	103	113	30	33	17	18	7	7
3	2079	>5000	1583	2588	327	352	83	90	28	32
4	1910	>5000	1825	>5000	1285	2473	376	399	90	96
5	1781	>5000	1751	>5000	1565	>5000	883	1429	240	260
6	1702	>5000	1694	>5000	1606	>5000	1120	4233	486	597
7	1658	>5000	1648	>5000	1602	>5000	1246	>5000	659	1246
8	1629	>5000	1615	>5000	1596	>5000	1322	>5000	744	2425

TABLE 2.3 – Résultat de codage (SPSM) par alignement épipolaire. Avec une fenêtre de correction de taille 25×3 . Pour chaque cas, colonne de gauche : résultats temps-réel (< 40 ms), colonne de droite : maximum obtenu. Le nombre de colonnes est reporté, le nombre de lignes pouvant être infini.

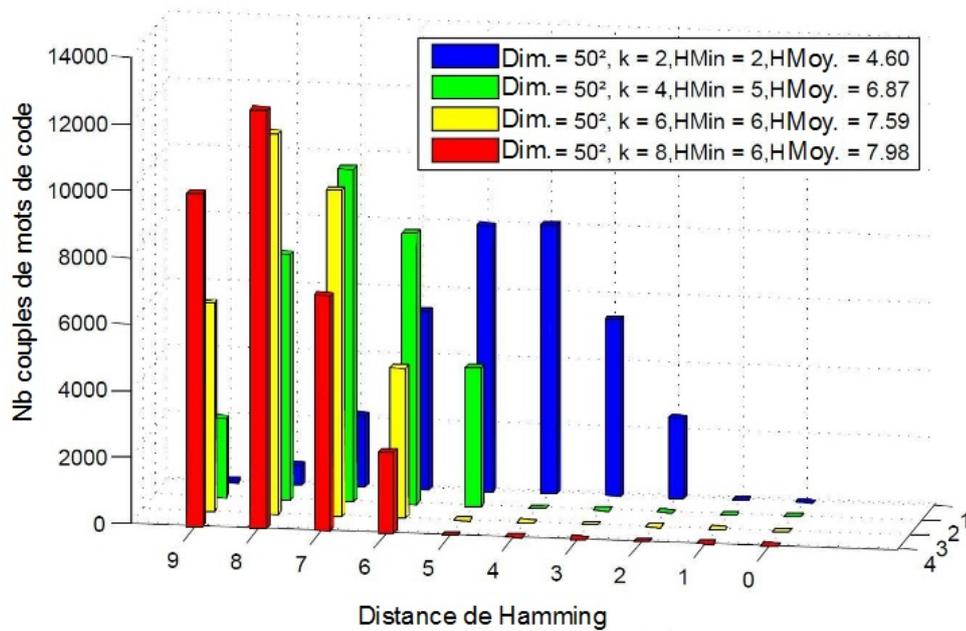


FIGURE 2.21 – Distribution de la distance de Hamming pour des motifs de taille 50×50 codés avec 2 à 8 symboles (k), H_{Min} : distance de Hamming minimum, H_{Mean} : distance de Hamming moyenne (\bar{H}).

matrice, car comme ils y a peu de voisianages (contrainte 1-D) entre lesquels l'unicité doit être assurée, les tests dans l'espace des mots de codes est moins intéressant (voir section 2.5).

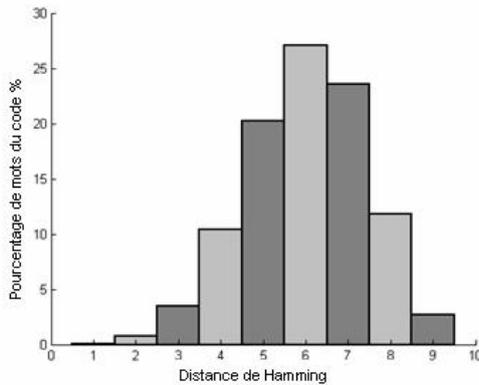


FIGURE 2.22 – Albitar $k = 3$, $w = 9$, \bar{H} de 6.173

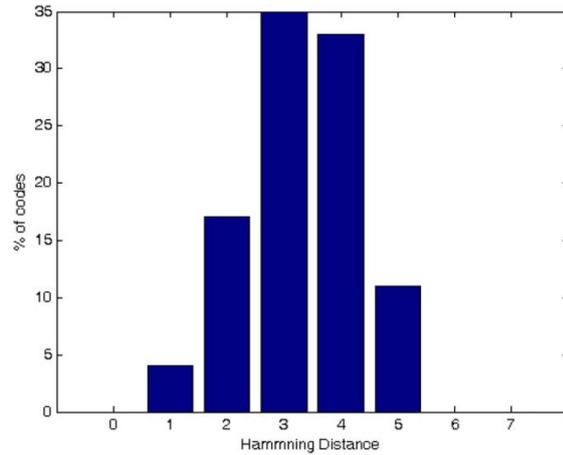


FIGURE 2.23 – Distribution de la distance de Hamming entre codes voisins avec des voisinages non ordonnés dans [?], avec $k = 7$ symboles, $w = 7$, \bar{H} proche de 3.16

Distribution de la distance de Hamming

La figure 2.21 montre la distribution des distances de Hamming entre tous les couples de mots de code pour 4 motifs de taille 50^2 avec 2, 4, 6 et 8 symboles. Tandis que les distances de Hamming moyenne et minimale augmentent avec l'ajout de nouveaux symboles, nous remarquons que le gain, lorsque l'on passe de 2 à 4 symboles, est plus important que de 4 à 6 ou 6 à 8 symboles. Cela correspond à la remarque faite dans la section 2.3, concernant la distance de Hamming moyenne qui évolue plus rapidement quand le nombre de symboles est plus faible. Pour limiter les problèmes de classification des symboles tout en maintenant des distances de Hamming moyenne et minimales élevées, il ne semble pas très efficace d'utiliser plus de 4 symboles. A titre de comparaison (voir figure 2.23), dans [Albitar07] et $\bar{H} = 6.17$, ce qui correspond à l'estimation a priori (voir section 2.3 pour \bar{H} pour $k = 3$ symboles et des mots de code de longueur $w = 9$ ($\bar{H} = w - w/k = 6$). A contrario, dans [Adán09], la distance de Hamming moyenne donnée est de 3.16 pour $k = 7$ et $w = 7$, alors qu'elle devrait être de $\bar{H} = 7 - 7/7 = 6$. Cela est dû au fait que dans [Adán09], seules les distances de Hamming entre voisinages adjacents sont pris en compte dans le calcul de \bar{H} , donc il y a toujours 2 éléments en commun, la distance de Hamming maximale est donc de 5. Nous notons que aussi bien dans [Albitar07] que dans [Adán09], il existe des couples de codes ayant une distance de Hamming de 1. La robustesse est donc entièrement statistique et des détections et/ou corrections d'erreurs ne peuvent être toujours garanties.

2.8.3 Résultats temps-réel

Si dans la majorité des cas, un motif codé une fois pour toute, en prenant tout le temps de codage nécessaire, est envisageable, en pratique, le codage rapide présente plusieurs intérêts :

- Essayer rapidement de nouveaux motifs avec des nouveaux paramètres qui sont mieux adaptés à la scène courante.
- Permettre le recodage en temps réel de motifs adaptatifs.

Dans le cadre de cette thèse, le codage rapide a été très précieux, en particulier lors des tests *in vivo*, alors que la présence d'un chirurgien était nécessaire et que la durée des expériences dans le bloc opératoire était très limitée. Pouvoir générer en moins d'une seconde des codes différents a permis d'étudier différentes projections et de voir quelles propriétés du motif permettaient d'obtenir des reconstructions.

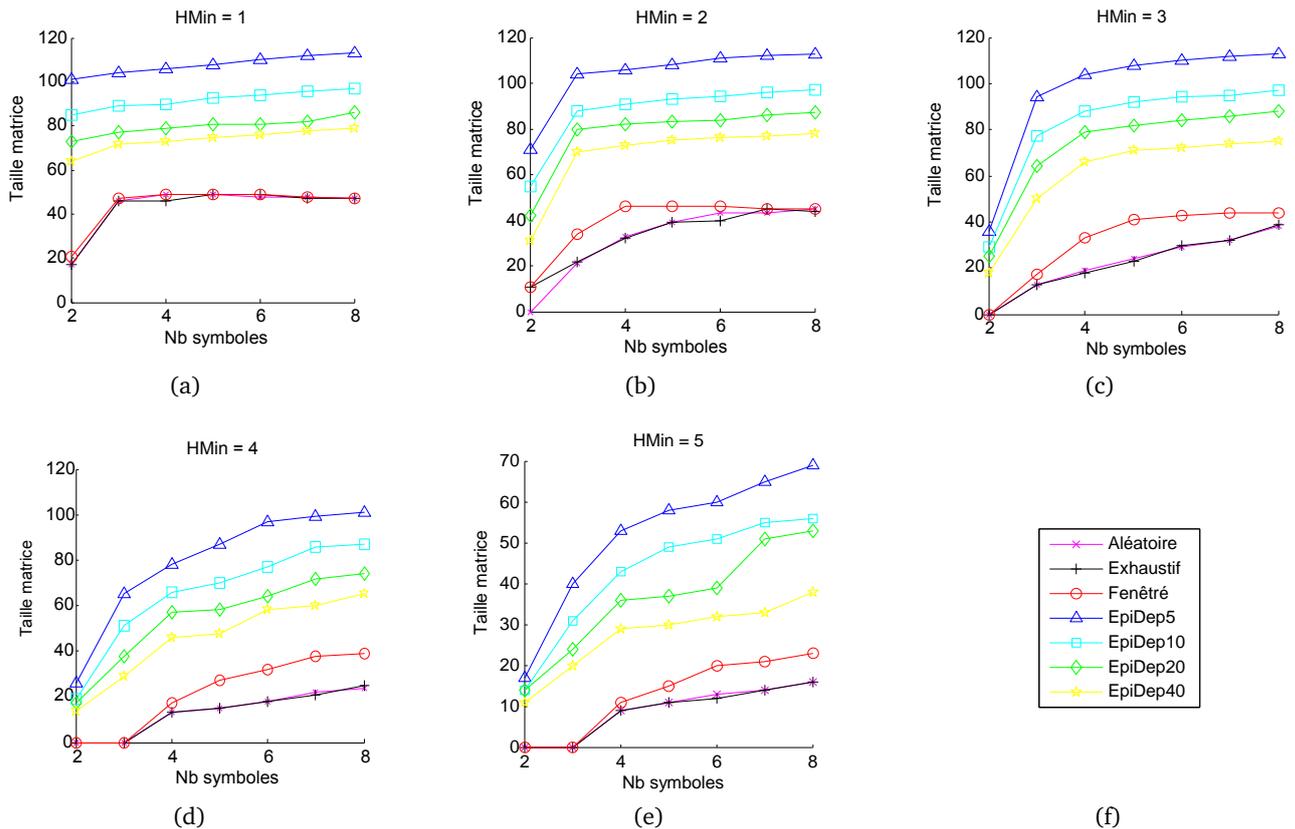


FIGURE 2.24 – Résultats de codage pour 25 codes générés par seconde. La racine carré de la plus grande taille d'une matrice est reportée par rapport au nombre de symboles ($C \in [2 - 8]$), à la distance H_{Min} (de 1 à 5 pour, respectivement (a), (b), (c), (d), (e)) ainsi que l'algorithme utilisé. Pour chaque cas, la moyenne sur 100 codages est reportée.

Méthodologie.

Les résultats de codages temps-réel sont reportés sur la figure 2.24. Pour simuler le comportement aléatoire de [Morano98] (appelé "Aléatoire" dans la Fig. 2.24) nous avons utilisé une fenêtre de correction de taille 1×1 . Pour simuler le comportement exhaustif de [Claes08] (appelé "Exhaustif" dans la Fig. 2.24), nous avons utilisé une fenêtre de correction de la même taille que celle du motif à générer et avons mis 0 comme valeur initiale des symboles, au lieu d'une valeur aléatoire comme dans [Morano98], tel que Claes l'a effectué. Nous avons fixé une taille de fenêtre de 5×5 pour le comportement cherchant un équilibre entre aléatoire et exhaustivité (appelé "Fenêtré" dans la Fig. 2.24). Cette taille produit en moyenne les meilleurs résultats avec un Pentium 4, 3Ghz, en utilisant un seul coeur. Finalement en utilisant à la fois une fenêtre de

correction de taille 5×5 et l'unicité par proximité épipolaire, nous avons généré des SPSM avec une marge de zone d'ambiguïté de respectivement 5, 10, 20, et 40 pixels ("EpiDep5", "EpiDep10", "EpiDep20" et "EpiDep40"), pour une résolution d'image de 1000×1000 pixels.

Pour chaque cas, (H_{Min} et le nombre de symboles varient), nous avons reporté la taille du côté de la matrice carrée la plus grande qui a pu être générée en moins de 1/25 secondes. A cause de la nature pseudo-aléatoire des algorithmes, nous avons considéré la moyenne sur 100 générations pour chaque cas.

Résultats PSM. Comme nous pouvons le voir, les comportements "Aléatoire" et "Exhaustif" montrent des résultats très similaires dans chaque cas. En effet, ils sont tous deux immédiatement pénalisés quand le problème de codage devient plus contraint (H_{Min} plus élevée, moins de symboles) et que la stratégie de correction en cas de tests d'unicité négatifs devient importante. Le comportement "Exhaustive", qui teste toutes les possibilités, n'a pas suffisamment de temps pour trouver des solutions et le "Aléatoire" s'arrête de corriger trop rapidement. De façon générale, ces approches génèrent alors des motifs de résolution trop modeste : $\leq 20^2$ pour $H_{Min} > 2$, $C \leq 5$ et un maximum de taille de $\sim 40^2$ quand le problème est moins contraint. Alors que le comportement "Fenêtré" peut générer des motifs de taille comparable aux autres algorithmes PSM (avec une taille maximale d'environ 40^2), il résiste mieux aux situations où le problème de codage est plus contraint et génère toujours des solutions ayant des résolutions supérieures ou égales aux comportements "Aléatoire" et "Exhaustif", jusqu'à ~ 4 fois plus pour $H_{Min} = 4$ et $C = 6$.

Résultats SPSM. De façon générale, les algorithmes SPSM font bien mieux que les algorithmes PSM. En effet, chaque test d'unicité est effectué beaucoup plus rapidement comme il y a moins de codes à tester. Par exemple pour un motif de taille 60^2 , les algorithmes de type PSM ont au pire 58^2 tests à effectuer (pour des voisinages 3×3), alors que les SPSM par proximité épipolaire ont, en moyenne, 19.21 pour "EpiDep5", 34.88 pour "EpiDep10", 61.52 pour "EpiDep20" et 120.31 pour "EpiDep40" codes dans leur zone d'ambiguïté. Ce qui est de l'ordre d'une réduction dimensionnelle de 2-D à 1-D. Nous notons qu'en pratique, ces marges (jusqu'à 40 pixels) sont très larges par rapport à la précision typique obtenue lors de l'estimation de la matrice Fondamentale. En effet, dans [Armangué03], où une grande variété de scènes (synthétiques et réelles) sont considérées, la précision d'estimation de la matrice Fondamentale, exprimée comme étant la distance moyenne entre les points et les lignes épipolaires estimées, est le plus souvent inférieure à 5 pixels et inférieure à 1 pixel dans la moitié des cas envisagés. Comme règle générale, plus la marge de la zone d'ambiguïté est importante, plus il y aura de codes qui vont contribuer aux tests d'unicité, plus petites seront les résolutions des codes générés.

Finalement, le codage par proximité épipolaire permet de générer des motifs de taille significative, 100^2 , (plus que la majorité des motifs utilisés dans la littérature, voir chapitre 1) tout en offrant des capacités de correction ($H_{Min} \geq 3$) avec un nombre limité de symboles (pas plus de 6). Le codage par alignement épipolaire permet de générer, en temps-réel, des résultats encore bien meilleurs (voir Fig. ??, colonne < 40 ms). De tels motifs sont utilisables en pratique pour effectuer des reconstructions de qualité sur des scènes comportant des perturbations à même de provoquer des erreurs de classification.

2.8.4 Résultats de codage optimal

Des codes SPSM optimaux de différentes tailles, possédant jusqu'à 26 colonnes ($k = 3$ et $H_{Min} = 6$) et 39 colonnes ($k = 2$ et $H_{Min} = 3$) et potentiellement une infinité de lignes ont

pu être obtenus grâce à un alignement global et une recherche exhaustive (voir Annexe 1). Ces tailles sont supérieures à certaines tailles de motifs reportées dans la littérature, comme :

- Albitar [Albitar07] : taille 27×29 .
- Salvi [Salvi98] : taille 29×29 .

Contrairement aux résultats exhaustifs reportés dans [Claes08], grâce à la réduction des contraintes portant sur le problème de codage par l'alignement épipolaire du motif, les codes optimaux générés ici ont une taille suffisante en pratique pour des applications demandant beaucoup de robustesse mais une faible densité, car ils possèdent une distance de Hamming minimale très élevée.

2.9 Conclusion

Dans ce chapitre, après avoir introduit les matrices parfaites et mentionné quelques éléments théoriques (limites supérieures, capacités de correction) sur le codage, nous avons proposé une formule permettant de connaître, a priori, la distance de Hamming moyenne des couples de mots de code d'une matrice pseudo-aléatoire. Celle-ci donne une indication statistique sur la robustesse d'un code.

Ensuite, plusieurs contributions ont été proposées pour améliorer le codage de matrices pseudo-aléatoires grâce à des algorithmes de type "brute-force" :

- Nous avons montré que le codage $1,5D$, dans lequel l'unicité des codes n'est assurée que dans une dimension, grâce à la contrainte épipolaire (codage par proximité et par alignement épipolaire), mène à des nouveaux codes possédant des propriétés (taille, nombre de symboles utilisés, distance de Hamming minimale) plus intéressantes que ce qui avait été proposé dans l'état de l'art jusqu'à maintenant.
- D'autre part, une formule de dénombrement de mots de codes a permis d'étudier la pertinence d'effectuer le test d'unicité dans l'espace des mots de codes par rapport à l'espace de la matrice. Le test d'unicité dans l'espace des mots de codes a surtout un intérêt par rapport au test dans l'espace de la matrice pour le codage de PSM , donc ne prenant pas en compte la géométrie épipolaire. Car pour les $SPSM$, le nombre de tests dans l'espace de la matrice est très faible (réduction du nombre de tests dimensionnelle, de 2-D à 1-D). Cependant pour $HMin = 1$, il n'y a qu'un test à effectuer dans l'espace des mots de code quelle que soit la taille de la matrice. Donc dans ce cas, le test peut être effectué inconditionnellement dans l'espace des mots de code.
- De plus, une généralisation du comportement de recherche d'algorithmes existants, entre exploration et exploitation, a été permise grâce au concept de fenêtre de correction dans laquelle la recherche est localement exhaustive.

En outre, les résultats obtenus ont montré qu'il est possible de coder, en temps-réel, de nouveaux motifs possédant des propriétés permettant une utilisation en pratique, ce qui ouvre la voie à des motifs adaptatifs recodés en temps-réel.

Finalement, dans cette étude, bien que nous nous focalisions sur un voisinage de taille 3×3 , les contributions proposées pourront être adaptées pour toute taille de voisinage 2-D, et certaines contributions (tests dans l'espace des mots de code, fenêtre de correction) sont également utilisables pour le codage 1-D.

Chapitre 3

Conception 2 : Agencement du motif et traitements

Un pour tous, tous pour un

(Les Trois mousquetaires, Alexandre Dumas)

Le type de motif, le choix des primitives, leur nombre, leur codage, la résolution du motif, mènent à un nombre très important de possibilités de conceptions. La littérature, de par la diversité des motifs proposés (Chapitre 1), souvent adaptés à des applications très différentes, en témoigne. Ainsi il serait bien présomptueux de prétendre proposer une conception "absolue", la conception qui serait meilleure que les autres et qui serait adaptée à toutes les applications en faisant mieux que d'autres motifs plus optimaux pour l'application visée. Par exemple, certains motifs, de résolution élevée avec peu de redondance et non codés, mais permettant une localisation très précise des indices visuels projetés, peuvent être adaptés à des scènes possédant peu de perturbations spectrales et géométriques pour faire de la métrologie. A l'inverse, d'autres applications nécessitent davantage de robustesse pour pouvoir effectuer une mise en correspondance correcte malgré un nombre important de perturbations dans des applications où la capture générale de formes est recherchée. Il est cependant intéressant de pouvoir proposer une méthodologie souple pouvant mener à des conceptions de motifs ayant des propriétés adaptées aux scènes visées.

En ce qui concerne ce travail de thèse, les perturbations sont nombreuses et de natures différentes. Les images endoscopiques présentées au chapitre 1 en témoignent. Le codage du motif (Chapitre 2) est une première étape importante qui permet déjà d'effectuer une **conception numérique** du motif. Celui-ci étant projeté sur la scène par le biais de photons, puis acquis par une caméra, nécessite une **représentation physique**, visuelle, analogique, en deux dimensions. C'est le sujet de ce chapitre.

En réalité, la partie numérique, (codage) de la conception du motif et sa représentation physique, bien que présentées séparément, sont, dans notre approche, intimement liées par la géométrie épipolaire, dont ces deux étapes de conception dépendent de par l'**alignement épipolaire** du motif, évoqué précédemment.

Dans ce chapitre, nous étudions la représentation visuelle qui permet de répondre au mieux à la problématique de mise en correspondance dont nous avons parlé dans le chapitre précédent. Pour pouvoir décoder les primitives projetées puis acquises, il est nécessaire d'effectuer

une segmentation, permettant de localiser précisément les primitives dans l'image, puis une classification permettant de faire correspondre les types de primitives aux symboles numériques utilisés lors du codage. Ainsi, il y a toujours autant de primitives différentes que de symboles. Ces étapes traitement sont ainsi adaptées au choix de la représentation visuelle des primitives. De façon générale, il est admis dans la littérature, qu'il est souhaitable d'utiliser le moins de symboles possible pour faire le moins d'erreurs de classification possible. La méthode de codage proposée nous a permis de coder des motifs de taille suffisante ($> 200 \times 200$) pour l'application visée avec deux symboles seulement.

Dans ce chapitre nous présentons tout d'abord l'organisation générale (positionnement et orientation) des primitives dans le cadre d'un motif aligné sur la géométrie épipolaire (section 3.1). Nous étudions par la suite les déformations potentielles des voisinages dans l'image (section 3.2). Ceci nous fournit des indications importantes pour choisir les représentations des primitives (section 3.3). Enfin les traitements associés à ces choix de conception et adaptés aux scènes présentant de fortes perturbations sont présentés, allant des déformations géométriques du motif jusqu'à la mise en correspondance (décodage) des primitives projetées (section 3.3).

3.1 Organisation des primitives

L'étape de **rectification d'image côté caméra**, dans l'objectif d'obtenir des lignes épipolaires horizontales et de segmenter plus facilement les indices visuelles, est bien connue de la communauté vision, en particulier pour les systèmes multi-vues caméra/caméra ou caméra/projecteur.

L'alignement épipolaire du motif correspond alors, à la **rectification inverse côté projecteur**. Les lignes du motif, initialement horizontales prennent alors l'orientation des lignes épipolaires (voir Fig. 3.1).

Dans cette section, nous présentons deux approches, *l'alignement local*, adapté à un épipole proche du centre de l'image et *l'alignement global* adapté lorsque les lignes épipolaires sont quasi-parallèles. Cela permet de prendre en considération des positions différentes de l'épipole.

3.1.1 Utilisation de la géométrie épipolaire pour la conception du motif

Si, dans la littérature, les motifs proposés offrent une organisation le plus souvent purement verticale/horizontale dans l'image du projecteur, il serait erroné de dire que la géométrie épipolaire, n'a pas été prise en compte avant nos travaux et ceux de ([Xu10] et [Dao09]) qui paraissent au même moment (voir note ultérieure). Cependant le niveau de prise en compte varie beaucoup (voir Tab. 3.1). Les approches les plus courantes utilisent une projection de franges. Pour améliorer la netteté de l'intersection (et donc la précision de la reconstruction) entre les franges et les lignes épipolaires, si les lignes épipolaires ont une direction moyenne plutôt verticale, alors les franges seront orientées horizontalement, et vice versa.

Précision d'intersection franges, lignes épipolaires.

Pour une même résolution des périphériques de projection et de capture, il est souhaitable d'obtenir la meilleure précision possible de localisation des indices visuels dans l'image. Des approches aboutissant à une précision subpixel ont ainsi été proposées ([Lu10, Chen10]). Comme nous l'avons déjà noté, maximiser l'orthogonalité entre les franges et les lignes épipolaires, mène à une intersection la plus nette possible (Fig. 3.2). Ainsi, a minima, si les lignes épipolaires sont orientées verticalement, les franges seront orientées horizontalement et réciproquement. Dans notre cas, cette orthogonalité est recherchée dans l'image du projecteur. Dans l'image acquise

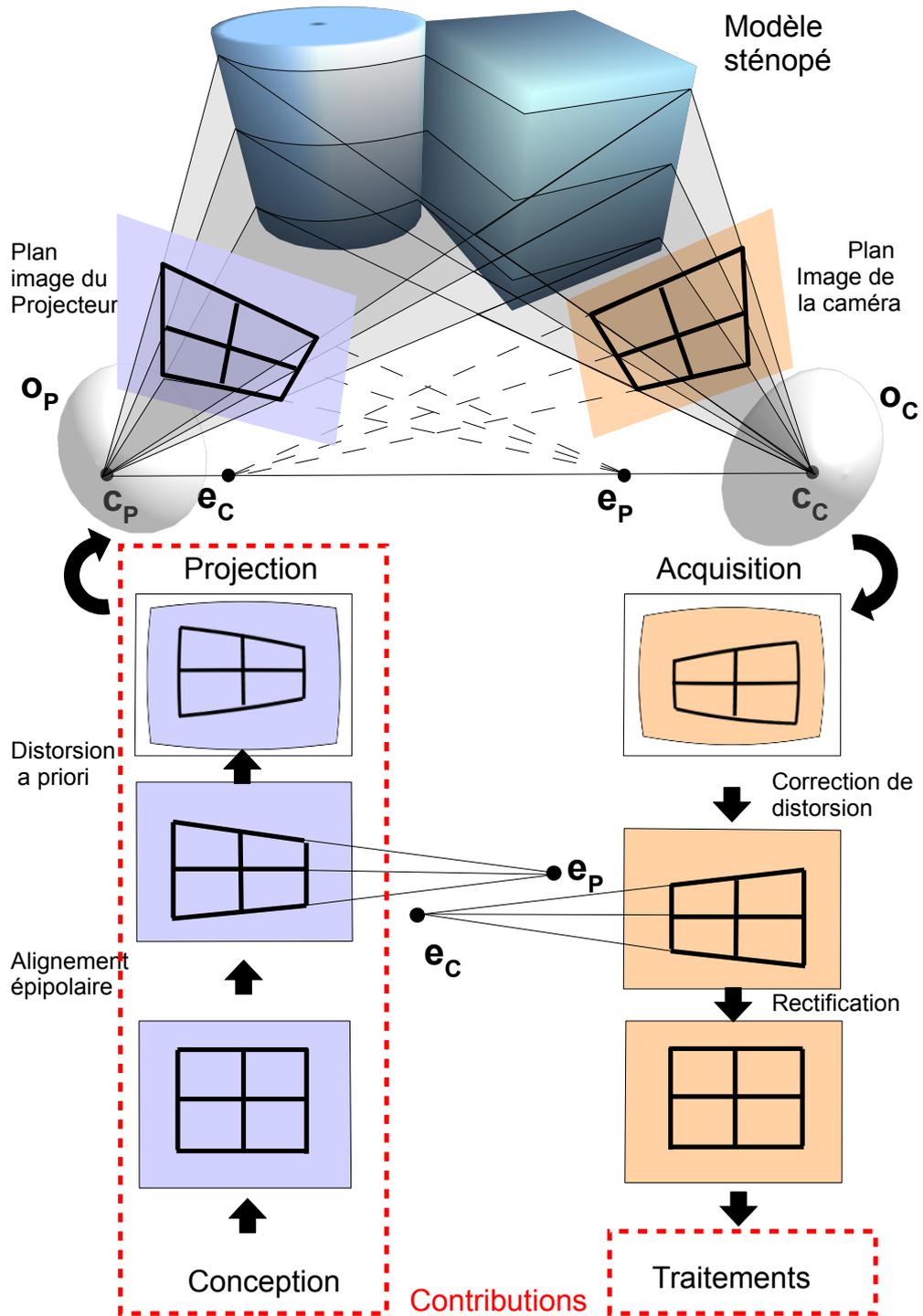


FIGURE 3.1 – Schémas de principe des déformations géométriques du flux d'images. Avec e_c et e_p , les épipoles, o_c, o_p , les optiques, c_c, c_p , les centres optiques.



FIGURE 3.2 – Plus les franges sont orthogonales aux lignes épipolaires, plus la segmentation de leur intersection sera précise.

Type d'utilisation	Références
Aucune	[Majorité des approches utilisant des spots]
Orientation codage 1-D H/V	[Majorité des approches utilisant des franges]
Adaptation codage	[Koninckx 06]
Alignement épipolaire	[Dao09, Xu10]
Alignement + codage	[Maurice10]

TABLE 3.1 – Tableau comparatif de l'exploitation de la géométrie épipolaire pour la conception du motif.

par la caméra, le résultat dépendra de l'orientation des surfaces sur lesquelles le motif est projeté (voir section 3.2).

3.1.2 Configurations de l'épipole

Comme évoqué dans le chapitre précédent, il est souhaitable d'avoir une homogénéité dans la distribution des indices visuels projetés. Or, en fonction de la position de l'épipole, intersection entre les lignes épipolaires, une angularité importante peut apparaître entre les lignes.

Nous distinguons trois cas (Fig. 3.3) :

1. L'épipole est dans la partie visible de l'image (Fig. 3.3.a). Dans ce cas, toutes les lignes épipolaires se croisent et apparaissent dans une même zone d'ambiguïté (voir chapitre 2, section 2.6.1). Le codage est alors le même que pour une PSM classique. Un alignement local est alors utilisé.
2. L'épipole est proche de la partie visible de l'image (Fig. 3.3.b). L'angularité est non négligeable et pour une distribution homogène des primitives, il n'est pas envisageable d'associer directement les lignes du motif aux lignes épipolaires. Le codage est alors de type SPSM par proximité épipolaire associé à un alignement local des primitives du motif.
3. L'épipole est loin de la partie visible de l'image (Fig. 3.3.c). Dans ce cas, les lignes épipolaires sont pratiquement parallèles et un codage de type SPSM par alignement épipolaire peut être associé à un *alignement global*.

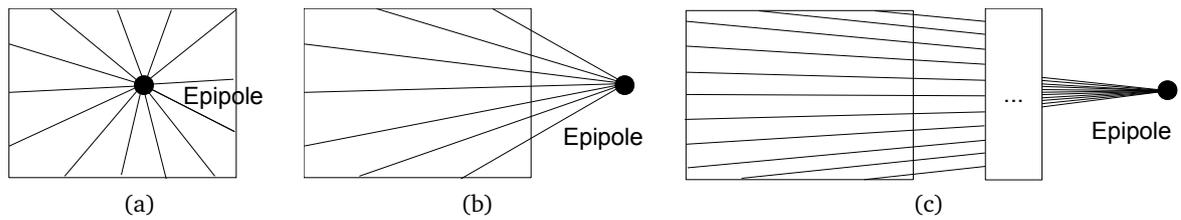


FIGURE 3.3 – Configurations de l'épipôle : a) Epipôle dans la partie visible de l'image : codage PSM, motif aligné localement ; b) Epipôle proche de la partie visible de l'image : angles significatifs entre lignes épipolaires : codage SPSM par proximité, motif aligné localement ; c) Epipôle loin de la partie visible de l'image, lignes épipolaires pratiquement parallèles : codage SPSM par alignement, motif aligné globalement.

Nous laissons à la discrétion du concepteur, le choix de décider, en fonction de l'application à quel niveau de parallélisme des lignes épipolaires, il est souhaitable d'utiliser un alignement local ou un alignement global.

3.1.3 Alignement local

Placement des primitives.

Généralement, les primitives sont régulièrement réparties sur le plan image du projecteur, (*PIP*), sur une grille par exemple, puis elles sont recherchées dans le plan image de la caméra, (*PIC*) en fonction de la proximité de leur ligne épipolaire. Pour tenir compte des imprécisions géométriques résiduelles, en pratique, la recherche est plutôt menée au sein d'une zone d'une certaine largeur dans laquelle d'autres primitives, a priori non connues, mais ayant une ligne épipolaire proche, peuvent apparaître. Cependant il est possible de savoir exactement quelles primitives

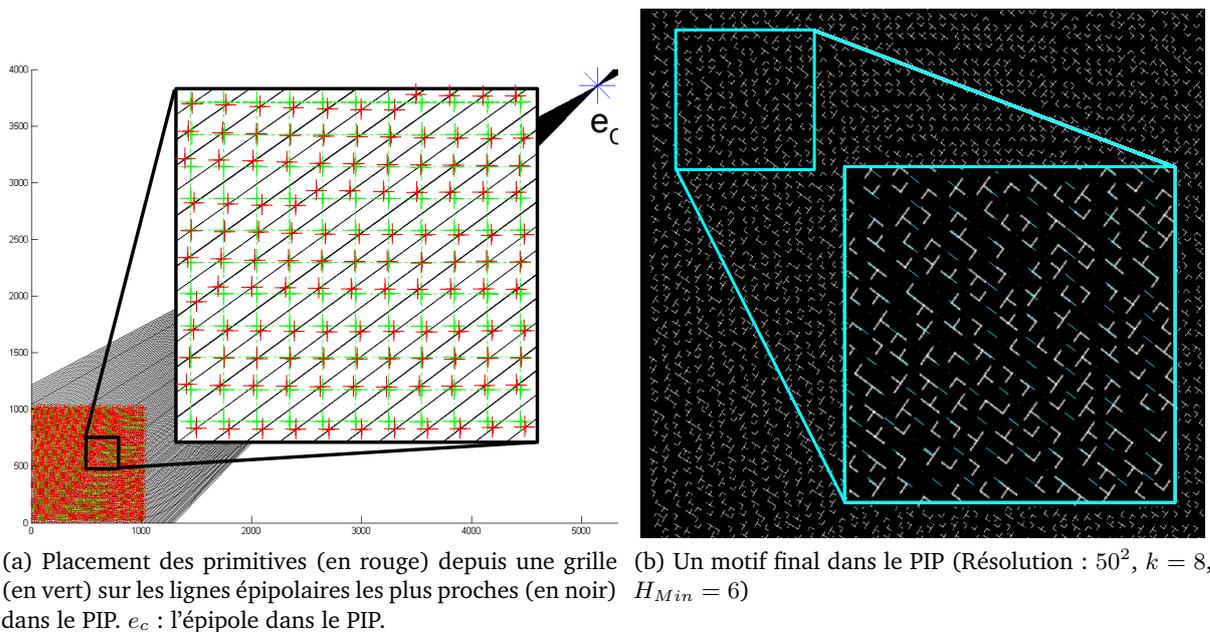


FIGURE 3.4 – Conception d'un motif 50×50 aligné localement sur la géométrie épipolaire. (a) placement des primitives, (b) orientation des primitives.

vont être ou non en concurrence lors de l'appariement en les disposant sur des lignes épipolaires suffisamment distinctes dans le PIP. Cela a un double intérêt. D'une part, il est alors possible de savoir exactement, lors du codage, entre quelles primitives la contrainte H_{Min} doit porter. D'autre part, cela simplifie la segmentation, les primitives étant regroupées par ligne épipolaire dans le PIP et dans le PIC, il sera nécessaire de parcourir moins de lignes épipolaires qu'en considérant celles associées à chaque primitive individuellement.

Pour disposer les primitives par ligne épipolaire tout en gardant une distribution homogène de celles-ci dans le PIP et le PIC, nous proposons l'approche suivante, illustrée sur la figure 3.4a :

1. Choisir dans le PIC un certain nombre de lignes épipolaires, régulièrement réparties, le long desquelles on souhaite que les primitives se projettent. Leur nombre et leur espacement dépendront de la résolution du motif et de la marge que l'on souhaite garder pour limiter les risques d'ambiguïté lors de la segmentation.
2. Identifier les lignes épipolaires correspondantes dans le PIP grâce à la matrice Fondamentale, celles-ci doivent couvrir la surface du motif (lignes en noir, motif en rouge et en vert).
3. Partant d'une grille régulière ayant la résolution souhaitée (en vert) pour le motif final, projeter orthogonalement chaque intersection de la grille (croix vertes) sur la ligne épipolaire la plus proche, identifiée à l'étape 2. Les positions obtenues (croix rouges) correspondent alors aux positions des centres des primitives dans le PIP.
4. A l'issue de cette étape, les sous-ensembles de primitives entre lesquels l'unicité doit être respectée sont les primitives dont les centres se trouvent sur la même ligne épipolaire.

L'alignement est ici **local**, car on recherche, à proximité de chaque intersection de la grille, la ligne épipolaire sur laquelle la primitive étudiée va être alignée. Il n'y a pas de relation globale entre les lignes de la grille initiale (en vert) et les lignes épipolaires (en noir).

Orientation des primitives

Enfin, une fois positionnée dans le PIP, chaque primitive subit une rotation autour de son centre, de telle sorte que son orientation, initialement horizontale, corresponde à la pente de la ligne épipolaire sur laquelle elle est alignée (Exemple Fig. 3.4b). Cette nouvelle orientation, restera alors invariante dans le PIC, indépendamment de la scène, grâce à l'invariance projective des lignes épipolaires.

3.1.4 Alignement global

En pratique, les lignes épipolaires ne sont pas parallèles, elles s'intersectent à l'épipole. Malgré cela, la méthode d'alignement choisie visera à respecter les propriétés suivantes durant la construction du motif :

- Conserver un **espacement vertical et horizontal** des primitives proches pour obtenir une distribution la plus homogène possible des primitives.
- Respecter au mieux l'**orthogonalité** locale des voisinages, pour optimiser la précision de l'intersection en chaque point et éviter des déformations trop importantes qui pourront perturber la recherche des primitives voisines pour le décodage (voir également la section suivante sur la déformation des voisinages).
- Assurer une **couverture spatiale maximale** de l'image du projecteur, par les primitives codées.

Nous différencions ici deux cas selon le degré de non parallélisme des lignes épipolaires. Si l'angularité entre les lignes épipolaires est localement (dans le plan image) négligeable (*Cas Parallèle*) (Fig. 3.5.c), et le cas inverse (*Cas Non Parallèle*) (Fig. 3.5.a).

Voici les différentes étapes :

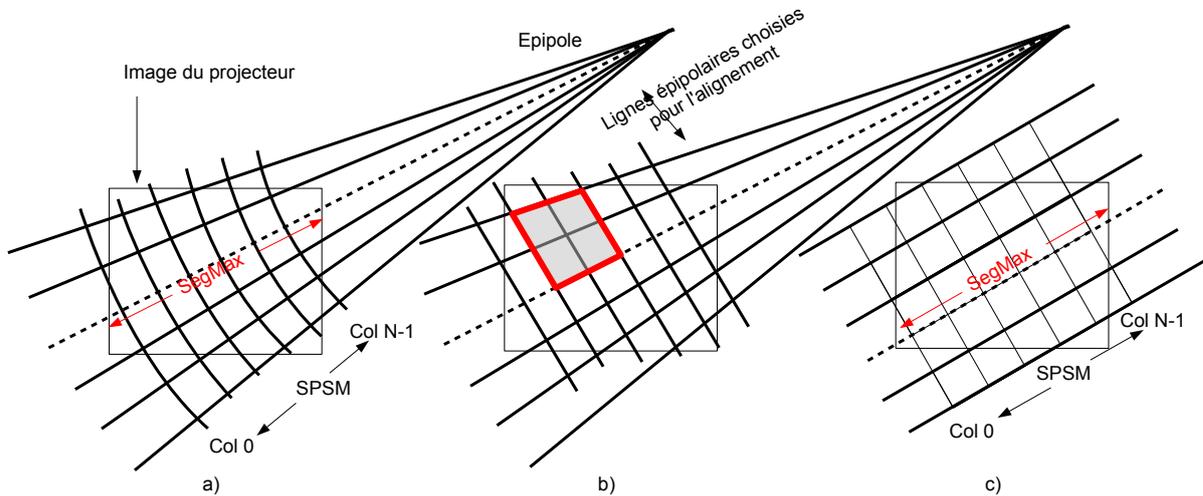


FIGURE 3.5 – Aligement global a) et b) Angularité non négligeable, déformation locale des voisinages dans b) car méthode d'aligement non adaptée. c) Angularité négligeable

A ce stade, on suppose que le motif est codé (SPSM par aligement épipolaire) et que la géométrie épipolaire du système est connue. Les différentes étapes sont les suivantes :

1. **Espacement moyen entre les primitives.** La contrainte du codage par aligement provient du nombre de colonnes de la matrice, le nombre de lignes pouvant être infini (voir Chapitre 2, codage type SPSM par aligement). Ainsi, il est important de s'assurer, dans un premier temps, que toutes les colonnes codées seront bien présentes dans le plan image du projecteur, car le nombre de lignes peut être indéfini. Pour cela, il est nécessaire d'identifier les lignes épipolaires dont le segment *SegMax* d'intersection avec l'image a une longueur maximale (en pointillés Fig. 3.5). L'espacement entre les colonnes sera alors égal à $EspHoriz = SegMax \div (NbCol)$.
2. **Discretisation des lignes épipolaires et des colonnes.** Pour obtenir une distribution la plus homogène possible, l'espacement moyen entre les lignes épipolaires successives devra être égal à $EspHoriz$. Pour cela deux approches sont possibles en fonction du cas considéré (*Cas Parallèle* ou *Cas Non Parallèle*).
 - Dans le *Cas Parallèle*, il suffit d'utiliser la direction moyenne des lignes épipolaires dans l'image, puis de discrétiser l'espace selon la direction orthogonale, tous les $EspHoriz$ pixels. Les lignes épipolaires servant à l'aligement seront alors celles passant par ces points de discrétisation. Puis, les colonnes seront discrétisées de la même façon, régulièrement le long des lignes épipolaires choisies. L'ensemble ressemblera globalement à une grille ayant subi une rotation (Fig. 3.5.c), même si dans les faits, les lignes ne seront pas parallèles mais auront une très légère angularité.
 - Dans le *Cas Non Parallèle*, la méthode précédente peut mener à des artefacts géométriques et à des voisinages distordus (Fig. 3.5.b). Nous proposons dans ce cas de déter-

	Dao 2009	Maurice (RFIA 2010)	Maurice (SPIE 2010)	Xu 2010
Date soumission	?	27 sep. 2009	18 sep. 2009	?
Date publication	déc. 2009	jan. 2010	avr. 2010	mai 2010

TABLE 3.2 – Dates de publication et soumission des travaux proposant un alignement global.

miner l'angle $\alpha_{InterEpi}$ correspondant à un espacement moyen entre les lignes épipolaires de $EspHoriz$ pixels dans l'image. Les lignes épipolaires seront alors choisies tous les $\alpha_{InterEpi}$ degrés.

Les colonnes (alors courbées), seront dessinées de façon similaire à un compas pointé sur l'épipole et dont l'ouverture augmente de $EspHoriz$ pixels pour chaque nouvelle colonne. Comparé à l'exemple Fig 3.5.b, cette façon de procéder permet d'obtenir des voisinages plus réguliers tout en maximisant l'orthogonalité de l'intersection entre les lignes épipolaire et les franges.

3.1.5 Autres travaux proposant un alignement global

Nous sommes en réalité plusieurs équipes à avoir présenté la même idée générale d'alignement épipolaire au même moment, mais avec de multiples différences significatives. Par "au même moment", j'entends que la première soumission de l'une des équipes n'est pas intervenue ultérieurement à la publication sur le sujet d'une des deux autres (voir tableau 3.2).

Je discute donc ici des similitudes et des différences avec deux autres contributions qui sont [Dao09] et [Xu10]. Dans [Dao09], aucun codage n'est effectué et un motif similaire à une mire calibration, très différent de notre approche, est projeté. Les intersections de la mire sont alignées sur des lignes épipolaires (Fig. 3.6.a). Par plusieurs aspects nous notons la proximité entre notre approche et celle de [Xu10]. Notamment, l'utilisation de codage par voisinage spatial, l'utilisation de motifs binaires dont les primitives sont différenciées par la géométrie (voir section suivante). Cependant nous nous en distinguons sur plusieurs points. En particulier dans [Xu10], un codage 1-D réalisé avec une séquence de De-Bruijn est utilisé. Ainsi ils obtiennent un motif de 63 colonnes utilisant 4 symboles sur un voisinage 1×3 avec une distance de Hamming mi-

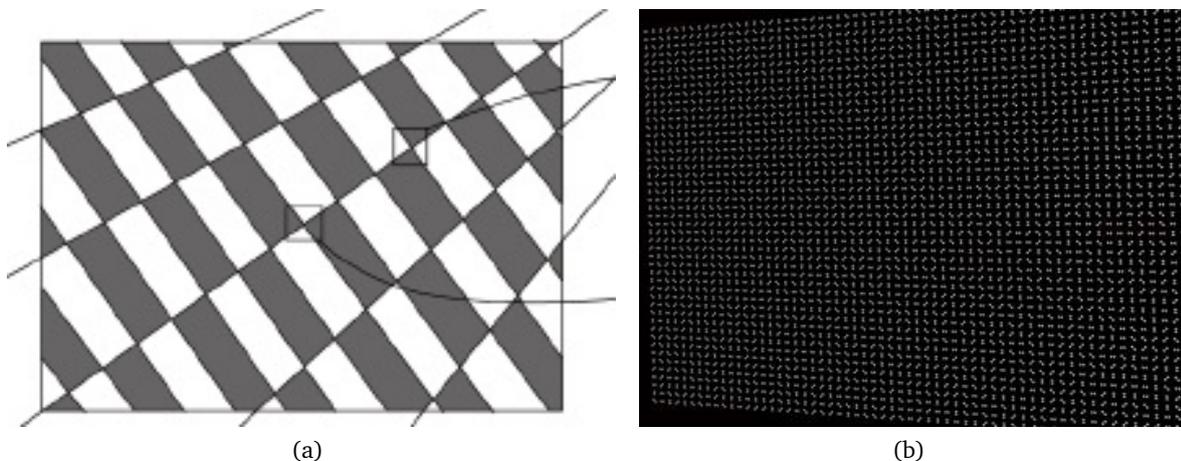


FIGURE 3.6 – Motif aligné dans (a) [Dao09] et dans (b) [Xu10].

nimale de 1. En considérant les primitives également dans l'autre dimensions (voisinage 3×3), nous avons obtenu, avec 2 symboles seulement une distance de Hamming minimale de 2 pour autant de colonnes (voir chapitre 1). De plus les formes de primitives choisies dans [Xu10] ne possèdent pas de continuité verticale et n'offrent donc pas la possibilité d'une reconstruction dense le long de franges. Nous reviendrons sur ce dernier point dans la suite de ce chapitre. Le fait que plusieurs équipes aient choisi cette approche d'alignement épipolaire du motif, est, pour nous, un facteur qui renforce la pertinence de cette approche.

3.1.6 Alignement local ou global ?

En cas d'hésitation sur le type d'alignement, local ou global, à choisir, les points suivants font nettement pencher la balance vers une conception utilisant un alignement global.

- L'alignement global permet l'utilisation d'un codage type SPSM aligné, qui comme nous l'avons vu au chapitre 1, donne des meilleurs résultats que le codage par proximité épipolaire.
- L'alignement local demande de stocker, pour chaque primitive et chaque primitive voisine, l'information sur la ligne épipolaire sur laquelle elle est alignée, ce qui ajoute des lourdeurs notables lors de la recherche des voisinages pour le décodage.
- Il résulte de l'alignement global, des voisinages ayant une forme beaucoup plus régulière qu'avec un alignement local, où les primitives sont translatées dans des directions différentes vers la ligne épipolaire la plus proche.
- Une continuité verticale des primitives le long de franges (et donc des reconstructions denses le long des ces franges) n'est possible qu'avec l'alignement global.

Après avoir testé l'alignement local dans un souci de généralité par rapport à la position de l'épipole, celui-ci se trouvant dans notre cas précis loin de la partie visible de l'image, nous avons finalement opté pour l'alignement global, pour tous les avantages qu'offre cette approche.

3.2 Etude de la déformation géométrique des voisinages

Avant de choisir les primitives à projeter, il est important de savoir quel type d'information elles doivent véhiculer, et donc de connaître au mieux les perturbations qu'elles (et leur voisinage) peuvent rencontrer. Nous proposons ici d'étudier les causes et les effets de la déformation locale des voisinages du motif en fonction des propriétés géométriques des surfaces sur lesquelles le motif de lumière est projeté. Ce thème est en effet peu abordé dans littérature sur la lumière structurée et pourtant cette analyse fournit des éléments déterminants pour mieux choisir les primitives et les traitements à utiliser par la suite.

3.2.1 Retrouver les primitives d'un voisinage 2-D

Le codage étant basé sur un voisinage 2-D (3×3 dans notre cas), une fois chaque primitive segmentée et classifiée, il sera nécessaire de retrouver les 8 primitives voisines pour être en mesure de reconstituer le mot de code complet convenablement et d'effectuer le décodage. En procédant de proche en proche, identifier les 4 primitives respectivement dessus, dessous, à gauche et à droite de la primitive centrale dans l'image rectifiée est suffisant.

Bien que dans cette étude, nous posions l'hypothèse d'une organisation générale sous forme

de grille, une analyse similaire pourrait être faite pour les motifs utilisant des voisinages hexagonaux, comme dans [Adán09].

Notre motif étant basé sur la géométrie épipolaire, trouver les voisins de gauche et de droite se réduit à identifier la primitive segmentée selon un ordonnancement séquentiel sur la même ligne épipolaire. Le problème se pose plutôt de *retrouver les primitives dessus et dessous* la primitive au centre du voisinage. L'approche la plus simple, et largement employée dans littérature, est d'utiliser un **critère de proximité** en considérant que la primitive du dessus serait la plus proche entre celles qui se trouvent sur la ligne épipolaire du dessus, ou, pour les motifs non alignés, la plus proche parmi les primitives se trouvant sur la ligne épipolaire de la primitive du dessus dans le plan projecteur (idem pour celle du dessous).

Nous allons voir dans cette section que cette approche, qui paraît évidente de prime abord, présente des inconvénients majeurs et, dans certains cas, est simplement insuffisante. Pour cela nous proposons d'étudier la déformation des voisinages tels qu'ils apparaissent dans l'image après rectification. Pour cette étude, nous entendons que les primitives d'un voisinage sont suffisamment proches pour considérer que la surface sur laquelle elles se projettent dans la scène 3-D, est localement plane et possède une orientation principale, suivant le principe de linéarité par morceaux. Nous ne nous intéressons pas ici aux discontinuités de courbure qui peuvent intervenir au sein d'un même voisinage.

Dans cette étude, nous nous attachons tout d'abord à **observer** les déformations des voisinages en fonction de l'orientation locale des surfaces sur lesquelles le motif est projeté et à identifier les composantes principales de ces déformations. Puis nous donnons une **interprétation géométrique** expliquant les causes de ces déformations ainsi que leur relation avec les composantes identifiées. Enfin, nous proposons une **modélisation** permettant d'identifier les valeurs critiques de ces composantes qui *mettent en échec l'approche classique* de recherche des primitives voisines par le critère de proximité. Enfin une **discussion** met en avant l'intérêt de cette analyse pour la suite de la conception et des traitements.

3.2.2 Observations

L'observation de la déformation d'une grille alignée sur la géométrie épipolaire et projetée sur des surfaces possédant différentes orientations permet, après rectification de l'image, de mieux appréhender les déformations géométriques des voisinages. Notre prototype de système stéréo de projection/acquisition miniaturisé (Chapitre 5), nous a permis de projeter une telle grille à l'intérieur d'un simple cylindre de papier (Figure 3.7.a). Nous observons alors, après correction de la distorsion et après rectification, l'image de la Figure 3.7.b.

Les composantes de déformation : l'élongation et l'inclinaison

On remarque tout d'abord que les franges horizontales, qui sont des lignes épipolaires, ne subissent aucune déformation, ce qui est normal car l'homographie des lignes épipolaires est invariante par projection perspective. Puis se distinguent deux composantes principales venant déformer les voisinages : un **facteur d'élongation**, S (Figure 3.8.a), et un **angle d'inclinaison**, α (Figure 3.8.b), venant déformer l'espacement et l'orientation des franges verticales. Ces transformations correspondent à une transformation affine (dans l'image rectifiée) de la famille des *transvections*, du ici à la conservation du parallélisme vertical causé par le glissement des points le long des lignes épipolaires.

Les zones de perte d'adjacence verticale

De plus, nous observons deux *zones de perte d'adjacence verticale* (en rouge) qui correspondent

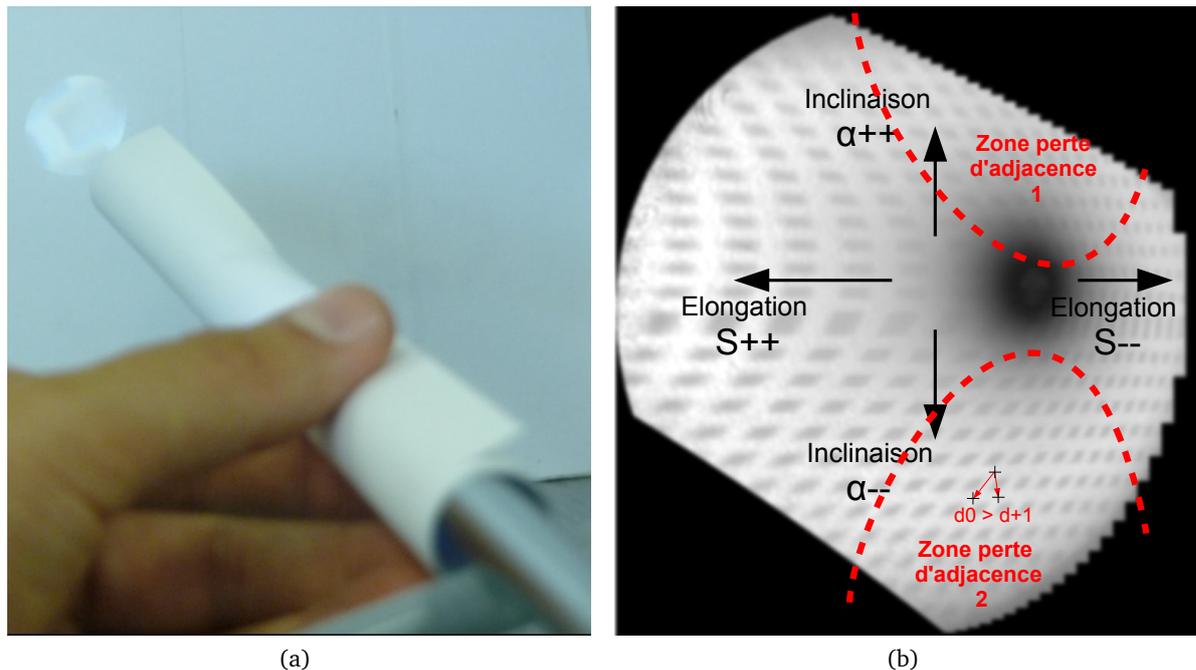


FIGURE 3.7 – a) Projection d’une grille alignée sur la géométrie épipolaire à l’intérieur d’un cylindre de papier, grâce au système endoscopique miniaturisé. b) Observation de la grille projetée après correction des distorsions géométriques et après rectification.

à des régions dans lesquelles le critère de proximité verticale, n’est plus valable pour définir le voisinage. En effet ces zones correspondent à un cas critique pour lequel une primitive à gauche ou à droite de la primitive de dessus (resp. dessous) dans le plan projecteur, devient, plus proche de la primitive centrale, que la primitive de dessus (resp. dessous) dans le plan projecteur et qui fausse donc l’adjacence initiale des primitives voisines (l’ordre établi selon la distance euclidienne n’est pas conservé). Ces zones apparaissent logiquement lorsque *l’élongation* S est faible et que *l’inclinaison* α est forte.

Si la projection est réalisée sur un plan ayant une orientation quasiment parallèle à ces deux plans images (Figure 3.8.c à droite), alors les voisinages subissent peu de déformation et ont une configuration proche de celle de la grille orthogonale initialement projetée.

3.2.3 Interprétation géométrique

Les déformations observées sur l’exemple précédent sont significatives car l’orientation de la surface interne du cylindrique est fortement non parallèle avec les plans image du projecteur et de la caméra.

Découplage de l’élongation et de l’inclinaison Nous observons de plus que les variations d’*angle d’inclinaison* interviennent majoritairement lorsque l’orientation locale se fait autour d’un axe parallèle aux lignes épipolaires, donc horizontal, comme l’illustre la Figure 3.9.a, tandis que l’**élongation** varie lorsque cette orientation se fait autour d’un axe perpendiculaire, donc vertical, aux lignes épipolaires dans l’image rectifiée (Figure 3.9.b et 3.9.c).

Elongation Quand l’axe de rotation est vertical, la variation de profondeur entre les primitives

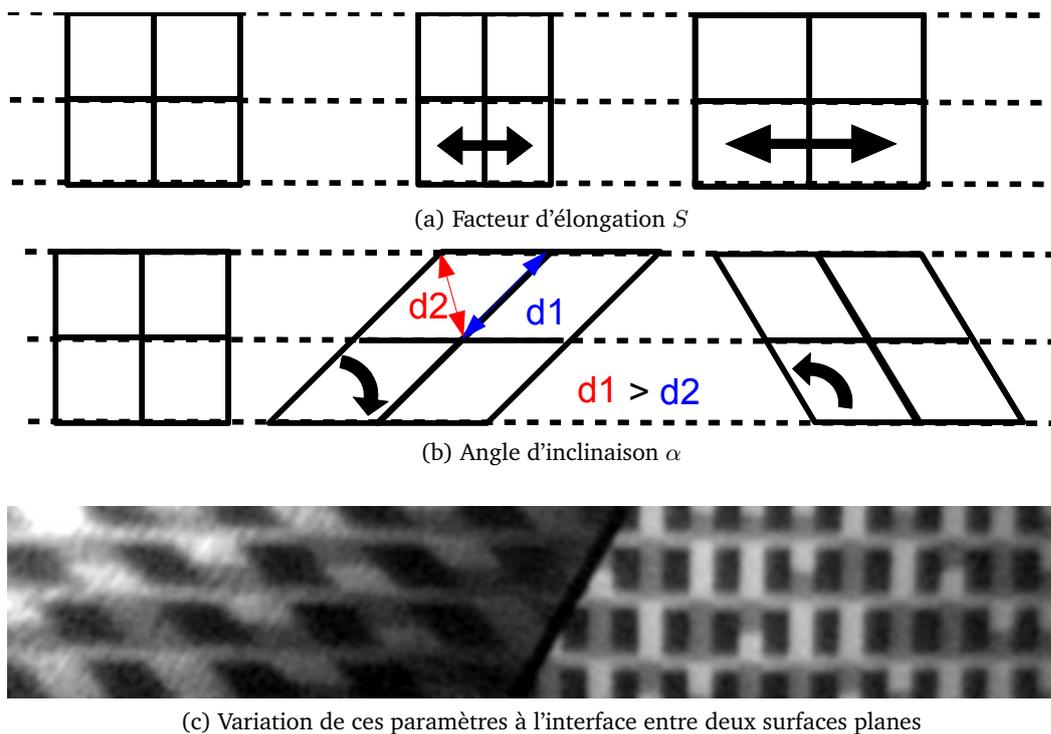


FIGURE 3.8 – Les composantes principales de la déformation des voisinages. a) L'élongation S , b) l'inclinaison α , c) Déformation d'un voisinage par a) et b) sur une scène.

du voisinage est *intra-ligne épipolaire*. Ainsi le déplacement des primitives s'opère entre chaque colonne des voisinages. Et comme les primitives alignées verticalement restent à la même profondeur relative, l'effet est une élongation horizontale du voisinage, l'inclinaison de celui-ci restant inchangé.

Inclinaison A l'inverse, quand l'axe de rotation est horizontal, la variation de profondeur entre les primitives du voisinage est *inter-lignes épipolaires*. Ainsi, pour un voisinage 3×3 , les primitives sur la ligne épipolaire du dessus glissent horizontalement par rapport à celles alignées sur la ligne épipolaire centrale, qui glissent de la même manière par rapport à la ligne épipolaire du dessous, provoquant une non orthogonalité (une transvection), et augmentant l'*angle d'inclinaison* du voisinage alors que l'élongation reste inchangée.

3.2.4 Modélisation et perte d'adjacence verticale par proximité

Pour mieux étudier les conditions de perte d'adjacence, nous proposons de modéliser l'influence que peuvent avoir ces déformations, à travers le facteur d'élongation (noté S) et l'angle d'inclinaison (noté α) sur la perte d'adjacence. Pour cela, nous définissons un *indice d'adjacence*, noté IA . Nous posons l'hypothèse de primitives initialement positionnées à l'intersection de lignes et de colonnes d'une grille orthonormée, ce qui correspond à la disposition la plus fréquente en lumière structurée. Dans ce cas IA représente le plus petit rapport de distance entre $d_{-1} \div d_0$ et $d_{+1} \div d_0$, (Equ. 3.1 et Fig. 3.10).

Où :

- d_0 est la distance entre la primitive P_0 au centre du voisinage et sa voisine de dessous, P_1 .

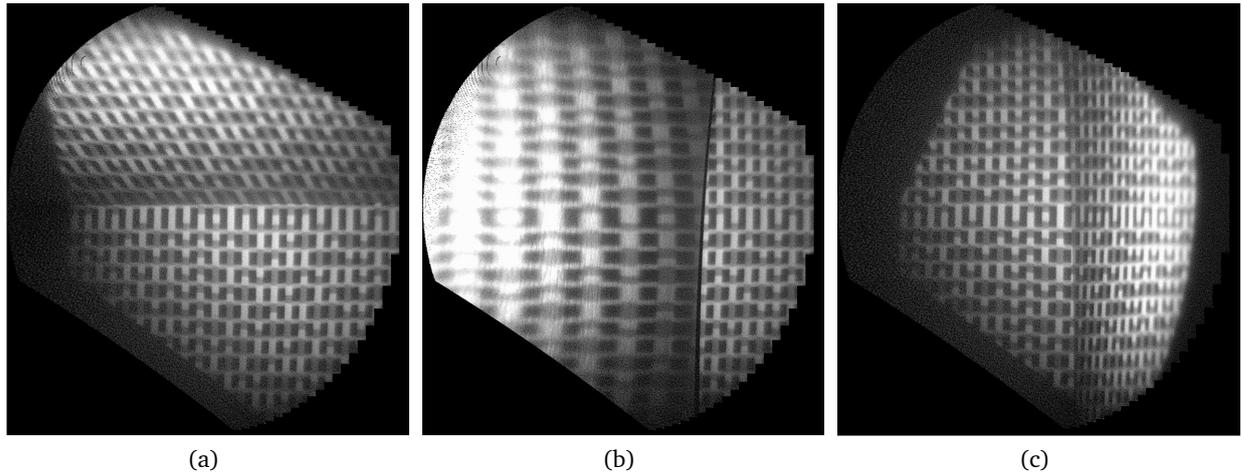


FIGURE 3.9 – Mise en évidence de la relation entre l'orientation des surfaces par rapport aux lignes épipolaires et le type de déformation. a) L'orientation de la surface selon un axe parallèle aux lignes épipolaires entraîne une variation de l'angle d'inclinaison α . b) et c) L'orientation de la surface selon un axe orthogonal aux lignes épipolaires entraîne une variation de l'élongation S

- d_{-1} est la distance entre P_0 et la voisine de gauche de P_1 , soit P_{-1} .
- d_{+1} est la distance entre P_0 et la voisine de droite de P_1 , soit P_{+1} .

Ainsi, lorsque $IA < 1$ alors d_{-1} ou d_{+1} est inférieur à d_0 . Ceci correspond à une situation de *perte d'adjacence verticale*, le cas limite étant obtenu pour $IA = 1$.

Les expressions de d_0 , d_{-1} , d_{+1} , en fonction de S et α , se trouvent par trigonométrie (Equ. 3.2, 3.3, 3.4) :

$$IA(S, \alpha) = \min\left(\frac{d_{-1}}{d_0}, \frac{d_{+1}}{d_0}\right), \quad (3.1)$$

avec

$$d_0 = \frac{1}{\sin \alpha}, \quad (3.2)$$

$$d_{-1} = \sqrt{1 + \left(S + \frac{1}{\tan \alpha}\right)^2}, \quad (3.3)$$

$$d_{+1} = \sqrt{1 + \left(S - \frac{1}{\tan \alpha}\right)^2}. \quad (3.4)$$

Une représentation graphique de cette fonction est visible figure 3.11. Nous avons considéré des dynamiques de valeurs pour $S \in [1/2, 2]$ et $\alpha \in [45^\circ, 135^\circ]$, proches de celles que l'on peut observer en pratique et notamment dans les figures précédemment présentées (3.8, 3.7 et 3.9). Nous avons également représenté en bleu marine le plan ($IA = 1$) qui correspond à la limite des zones de *perte d'adjacence*.

Éléments d'analyse

- On remarque tout d'abord la symétrie de la figure par rapport à la valeur $\alpha = 90^\circ$, ce qui correspond à la symétrie du modèle utilisé.

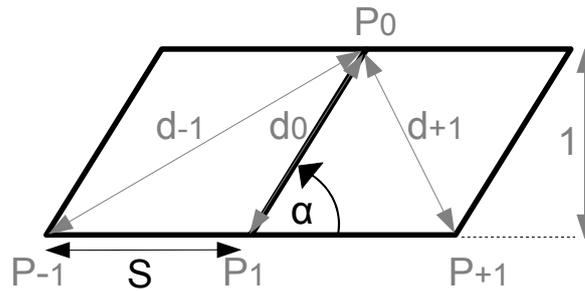


FIGURE 3.10 – Modèle utilisé

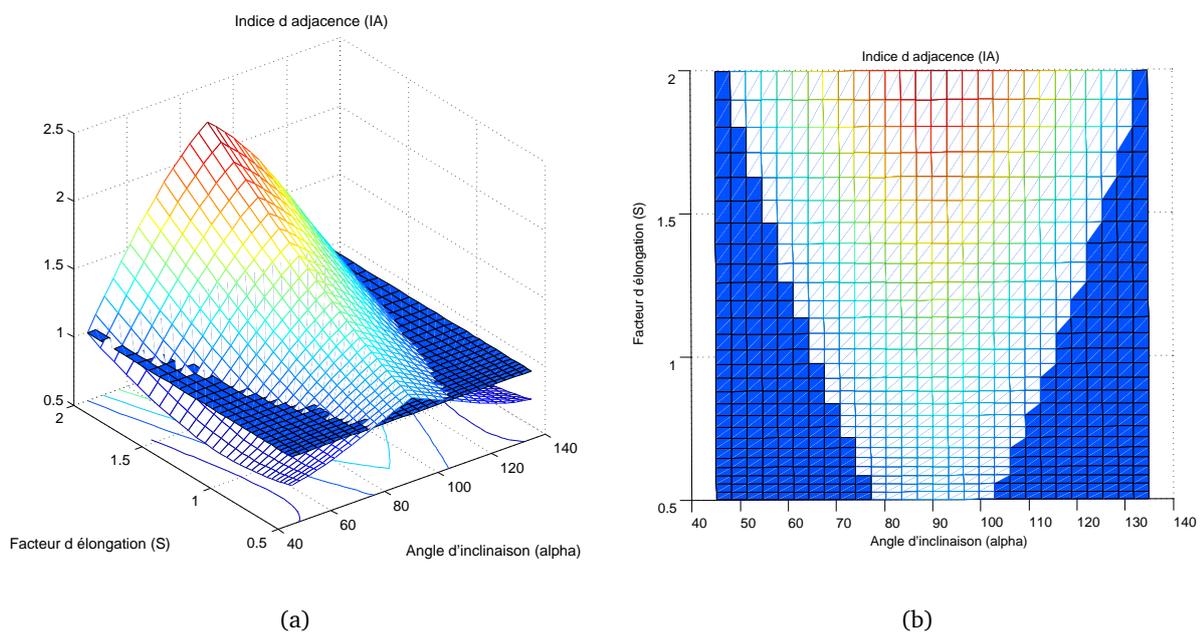


FIGURE 3.11 – Représentations graphique de l'indice d'adjacence en fonction de l'élongation et de l'inclinaison des voisinages, en bleu marine : zones de perte d'adjacence.

- Les deux zones de perte d'adjacence ($IA < 1$) observées figure 3.7 se retrouvent également ici (Fig. 3.11) et représentent une région importante des configurations possibles. En effet, pour les dynamiques de valeurs choisies, que l'on rencontre en pratique, ces zones représentent $\simeq 40.4\%$ des combinaisons (S, α) envisagées. Cette valeur n'était pas attendue à un niveau aussi élevé.

Note sur la discrétisation du domaine. Nous avons utilisé une discrétisation de l'espace régulier pour α et normalisé pour l'élongation, S , de telle sorte qu'il y ait autant d'échantillons entre $S \in [1/2, 1]$ et $S \in [1, 2]$, (voir Figure 3.11.a).

On note les valeurs remarquables de IA suivantes :

- Minimum absolu $IA = 1/\sqrt{2}$ pour $S = 1$ et $\alpha = 90 \pm 45^\circ$.
- Cas limites ($IA = 1$) pour :
 - $S = 2$ et $\alpha = \pi/4$ rads
 - $S = 1$ et $\alpha = \arctan(2)$ rads
 - $S = 1/2$ et $\alpha = \arctan(4)$ rads

3.2.5 Conclusion

Le choix de négliger ou non ces déformations et le risque d'une perte d'adjacence verticale dépendra, comme dans bien des cas, de l'application. Dans un souci de généralité, il est souhaitable de les prendre en compte, car comme nous l'avons vu, plus de 40% des combinaisons d'élongation et d'inclinaisons susceptibles d'être rencontrées en pratique peuvent mener à une situation d'échec de l'approche de recherche de voisinage par proximité, ce qui est très significatif. Dans notre cas précis, pour la reconstruction de surfaces intra-abdominales, les surfaces internes de la paroi se trouvent sur les côtés et donc leur orientation sera fortement non parallèle aux plans image du système stéréo. De plus, des scènes tubulaires (trachée, oesophage, côlon) sont fréquentes dans l'utilisation d'endoscopes flexibles, dans un souci de généralité et de portabilité future, il est évident qu'il ne faut pas négliger ces déformations.

Pour toutes ces raisons, il est souhaitable de prendre en compte ces paramètres dans la suite de la conception ainsi que lors des traitements associés.

3.3 Choix des primitives visuelles

Dans le chapitre précédent, nous avons étudié le codage du motif. Cette étape aboutit à un code utilisant un certain nombre de symboles. Ce code étant numérique, chaque symbole doit maintenant être physiquement représenté par une primitive visuelle propre. Nous avons également vu dans les sections précédentes que ces primitives sont régulièrement disposées sur un grille alignée sur la géométrie épipolaire et qu'il est nécessaire qu'elles contiennent une information directionnelle verticale pour pouvoir reconstituer les voisinages pour le décodage.

Dans cette section, après avoir étudié les différentes propriétés que doivent posséder les primitives, en fonction du type de scène à reconstruire, nous proposons celles qui nous semblent les plus adaptées.

De façon générale, le choix des primitives utilisées doit pouvoir répondre aux deux problématiques suivantes :

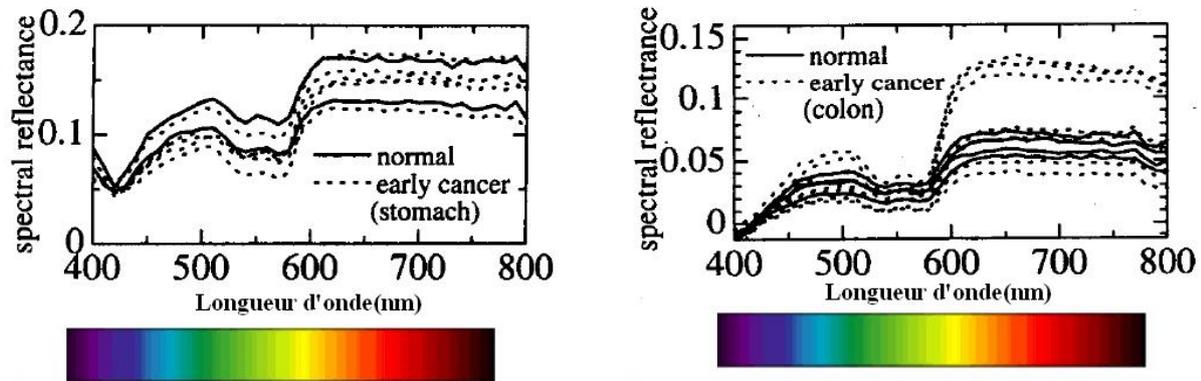
1. La **localisation** précise du centre des primitives, qui est une des composantes déterminant la précision de la reconstruction finale.
2. La **classification** la plus robuste possible aux perturbations géométriques et spectrales, permettant d'établir la correspondance entre la primitive segmentée et le symbole qui a été utilisé pour la coder.

A cela, nous rajoutons les deux problématiques suivantes :

1. L'**identification des primitives voisines** quelle que soit la déformation du voisinage (voir section 3.2).
2. La **continuité verticale** le long de franges, pour permettre la reconstruction d'une infinité de points le long de ces franges et donc être capable de capturer davantage de discontinuités de courbures à travers une reconstruction dense.

3.3.1 Types de scène

La modalité de discrimination des primitives projetées est un choix important. De celle-ci va dépendre la capacité du système à segmenter, puis à classifier les primitives par traitement



(a)

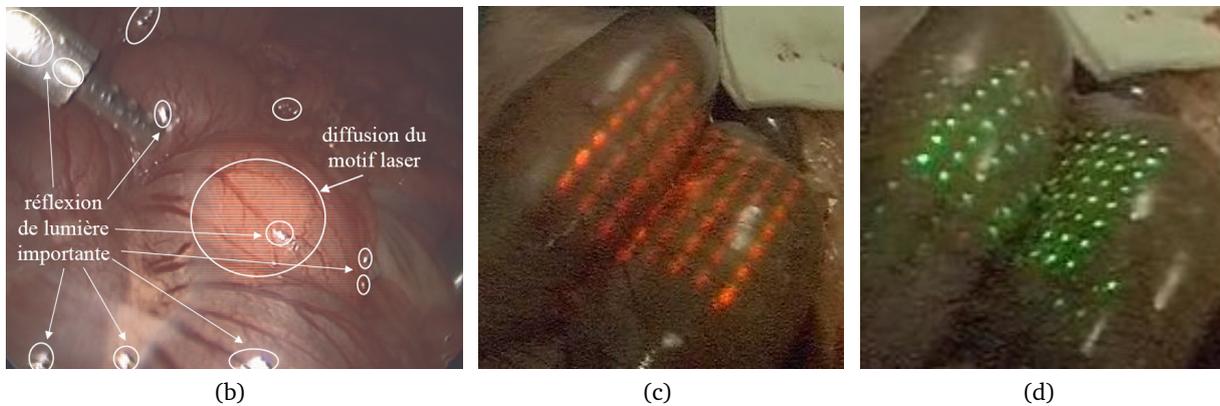


FIGURE 3.12 – a) In-homogénéité de la réflectance en fonction du type de tissus (estomac à gauche, colon à droite) [Sambongi00], b) Diffusion sous-surfacique des longueurs d'onde proche du rouge [Krupa03], c) et d) In-homogénéité de la réflectance en fonction de la longueur d'onde de la lumière projetée [Albitar07].

d'image pour retrouver les symboles associés, puis les mots de codes correspondants (voir Chapitre 2).

Comme évoqué dans le chapitre 1 (Travaux antérieurs), les primitives peuvent être différenciées spectralement, par niveau d'intensité lumineuse ou géométriquement, mais aussi par analyse dans le domaine fréquentiel, ou par une combinaison de de ces modalités. Nous nous intéressons ici à des motifs codés sur une seule image, adaptés à des objets en mouvement. Il n'est donc pas question d'étudier les différenciations par multiplexage temporel.

L'utilisation d'un motif monochrome, dont les primitives sont différenciées par la géométrie, permet d'être plus robuste aux perturbations spectrales présentes dans la scène que l'utilisation de la couleur.

Comme nous l'avons noté au chapitre 1, les perturbations géométriques et spectrales sont en effet très importantes dans des scènes intra-abdominales

Comme l'alignement épipolaire du motif présente l'avantage de rendre l'orientation le long des lignes épipolaires des primitives invariante quelle que soit la discontinuité de courbure locale, ce gain en robustesse géométrique nous a amené à choisir la géométrie comme modalité de discrimination des primitives. Une autre raison évoquée dans [Albitar07] est que cela permet

d'utiliser des sources de lumière laser, donc monochromatiques, comme source de lumière. Cette possibilité est toujours valable avec le système actuel.

Dans le reste de cette section, après avoir étudié les choix proposés dans la littérature concernant les primitives différenciées par la géométrie, nous proposons trois ensembles de primitives, chacun adapté à un type de scène ou à une situation spécifique : une scène plane pour la calibration, des scènes classiques en situation non endoscopiques et pour lesquelles la géométrie épipolaire est stable et enfin des scènes obtenues par projection et vision endoscopiques pour lesquelles le système stéréo est manipulé, provoquant de petites instabilités de la géométrie épipolaire. En effet, comme précisé ci-après, ces scènes possèdent des spécificités et donc des traitements adaptés :

Spécificités des scènes planes pour l'étalonnage

Conception	Motif indépendant de la géométrie épipolaire. Nécessite plus de primitives que pour un codage SPSM (au moins 3,4). Nous expliquons dans le Chapitre 4, pourquoi il est intéressant d'utiliser un motif codé pour la calibration.
Géométrie épipolaire	A pour objectif de l'estimer.
Perturbations spectrales	Quasi, voire totalement absentes.
Perturbations géométriques	Possibles hors zone de netteté (effet de flou). Voir plus de détails dans le Chapitre 4 (Calibration, reconstructions 3-D).
Rapport signal/bruit	Très bon, car environnement lumineux contrôlé.

Spécificités des scènes non endoscopiques

Conception	Motif aligné sur la géométrie épipolaire. Codage de type SPSM, nécessite peu de primitives : 2 suffisent.
Géométrie épipolaire	Généralement stable, si dispositifs de projection / capture sont fixés.
Perturbations spectrales	Potentiellement très importantes.
Perturbations géométriques	Potentiellement très importantes.
Rapport signal/bruit	Généralement bon, en fonction des dispositifs de projection et de capture utilisés.

Spécificités des scènes endoscopiques

Conception	Motif aligné sur la géométrie épipolaire. Codage de type <i>SPSM</i> , nécessite peu de primitives : 2 suffisent.
Géométrie épipolaire	Légèrement instable, car dispositif de projection / capture manipulé. Nécessite d'effectuer un léger <i>recalage épipolaire</i>
Perturbations spectrales	Potentiellement très importantes.
Perturbations géométriques	Potentiellement très importantes.
Rapport signal/bruit	Mauvais, dû à l'utilisation d'optiques endoscopiques qui capturent moins de lumière que des optiques plus grandes et à la nécessité d'ajouter des optiques d'interface - provoquant une perte d'intensité lumineuse - pour la réalisation du prototype de projection / acquisition miniaturisé (voir Chapitre 5, partie Développement Matériel).

3.3.2 Travaux antérieurs et primitives proposées

La figure 3.13, présente un tableau récapitulatif des différents ensembles de primitives discriminées par la forme proposés dans littérature. Les formes que nous proposons (Fig. 3.13.j, 3.13.k, 3.13.l, 3.13.m) sont inspirées de l'écriture cunéiforme Fig. 3.13.a, le premier système d'écriture créé par l'humain (3000 av JC, Mésopotamie) qui est sans doute mieux adapté à la reconnaissance par ordinateur que par l'humain.

Parmi les ensembles proposés, les formes de la figure 3.13.g se démarquent par la méthode de classification utilisée qui consiste en une analyse radiale de chaque primitive dans le domaine fréquentiel, alors que la classification des autres se fait par une analyse dans le domaine spatial. Elles n'ont finalement pas été utilisées et cette approche semble très sensible aux problèmes de recouvrement spectral et de perte de netteté.

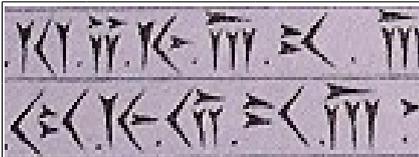
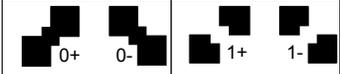
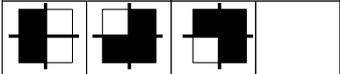
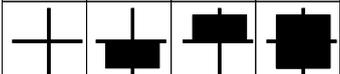
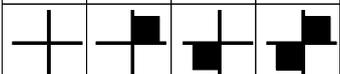
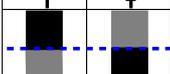
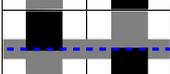
Plus récemment, [Albitar07] a proposé des formes de natures différentes : anneau, disque et trait (Fig.3.13.c). L'avantage de l'information directionnelle du trait, permettant de trouver les primitives voisines, était déjà mis en évidence. Nous réutiliserons ce principe pour toutes les primitives. Cependant ni le trait, ni le disque, ni l'anneau, ne contiennent d'orientation verticale permettant de trouver le voisinage complet en cas de *perte d'adjacence* (voir section précédente). C'est également le cas pour les primitives des figures 3.13.g et 3.13.i.

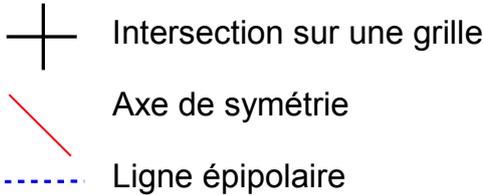
Certaines formes de primitives comportent des inconvénients. Notamment des primitives asymétriques dans la direction des lignes épipolaires peuvent provoquer une perte de précision par un déplacement du maximum d'énergie créé par intégration, lorsque la primitive sort de la zone de netteté ou que la surface provoque une diffusion de la lumière sur la surface, comme l'illustre la figure 3.14, où l'effet d'intégration est simulé par l'application d'un filtre passe-bas gaussien.

Les ensembles des figures (3.13.b,c,e,f,i,j), correspondent à ce cas et ne sont pas adaptés à la reconstruction de scènes, sauf éventuellement, pour un étalonnage, comme nous allons le voir.

3.3.3 Primitives pour l'étalonnage

Pour les **scènes d'étalonnage**, beaucoup de ces ensembles pourraient convenir, mais nous avons choisi de nous inspirer des primitives cunéiformes qui, avec 8 primitives possibles (Fig. 3.13.j) permettent de générer des motifs possédant beaucoup de redondance, avec une distance

a)							[Mésopotamie 3000 av. JC]
b)							[Vuylsteke et al. 90]
c)							[Griffin et al. 92]
d)							[Yee et al. 93]
e)							[Yee et al. 93]
f)							[Yee et al. 93]
g)							[Claes 08]
h)							[Albitar 07]
i)							[Xu 10]
j)							[Nous 10] Etalonnage
k)							[Nous 11]
l)							[Nous 11]
m)							[Nous 11]
n)							[Nous 11]
o)							[Nous 11]



 + Intersection sur une grille
 / Axe de symétrie
 - - - Ligne épipolaire

FIGURE 3.13 – a) Le système d'écriture cunéiforme, (b - i) Différenciation de symboles par les formes proposées dans des travaux antérieurs, (j - o) formes que nous proposons.

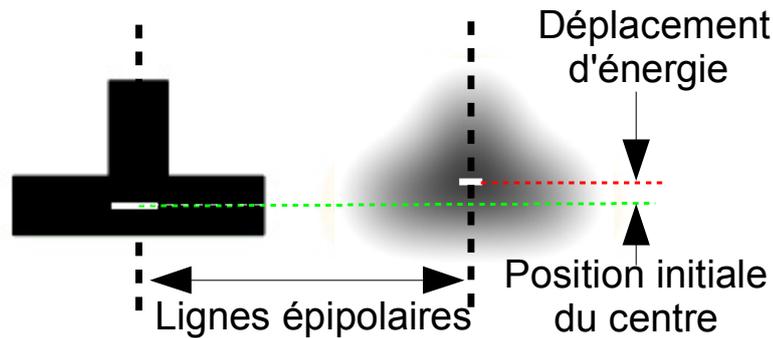


FIGURE 3.14 – Déplacement d'énergie dû à l'intégration sur des primitives non symétriques.

de Hamming minimale très élevée (jusqu'à 5). Ainsi, il est possible de les décoder même dans des zones très distantes en profondeur, ce qui est souhaitable pour capturer au mieux la géométrie de l'espace de travail. Nous avons ainsi sélectionné 8 formes de coins, 4 possédant deux bras (ou segments) et 4 autres en possédant trois.

Ainsi les deux dimensions peuvent être exploitées au mieux. Les avantages de l'utilisation de primitives cunéiformes sont nombreux. Parmi ceux-ci, nous relevons que :

- le centre de la primitive est localement bien défini par l'intersection de deux segments, mieux qu'avec des primitives circulaires,
- les informations directionnelles des bras offrent la possibilité d'exploiter la géométrie épipolaire,
- elles permettent d'utiliser un grand nombre de symboles (8 au maximum),
- elles possèdent des axes de symétrie différents (par pair) permettant de les classer par minimisation de la différence symétrique (voir chapitre 5, calibration).

En ce qui concerne les primitives destinées à l'étalonnage, le déplacement d'énergie évoqué plus tôt pour les formes asymétriques n'est pas problématique avec la méthode de segmentation que nous employons (voir Chapitre 4), car ces légers déplacements d'énergie se compensent statistiquement le long des lignes et des colonnes d'une grille. Cela est d'autant plus vrai que le nombre de primitives du motif d'étalonnage est important.

3.3.4 Primitives pour scènes inconnues

Dans le cas général, l'utilisation des primitives cunéiformes n'est plus possible à cause du déplacement d'énergie mentionné précédemment, du à la non symétrie horizontale (ou plus généralement dans la direction des lignes épipolaires).

Grâce au codage efficace proposé dans le chapitre précédent, deux primitives suffisent. En effet, de façon générale, moins on utilise de primitives moins il y aura d'erreurs de classification. cela est d'autant plus vrai lorsque l'acquisition (et dans certains cas la projection) est bruitée et que les perturbations sont nombreuses, comme dans notre cas.

Les discussions précédentes sur les déformations du voisinage, sur la possibilité de faire apparaître des franges verticales pour une reconstruction plus dense et sur la nécessité de respecter une symétrie horizontale, nous ont mené aux formes 3.13.(k-o), du tableau 3.13.

Dans le cas où le système stéréo est manipulé, la présence de deux bras horizontaux (primitives 3.13.l,n,o) le long des lignes épipolaires peut mener à un recalage robuste. La discrimination des primitives 3.13.k,l,m,n se fait en étudiant la quantité d'énergie lumineuse au-dessus et en dessous de la ligne épipolaire. L'amplitude du gradient vertical, extrême à cet endroit,

peut également servir à rendre le recalage épipolaire plus robuste). Pour les primitives 3.13.o, le rapport de l'amplitude des gradients vertical et horizontal est exploité.

Avec ces primitives, une intensité d'énergie intermédiaire (représentée en gris) entre le minimum et le maximum d'intensité est utilisée. Cette valeur intermédiaire peut être ajustée soit pour privilégier la qualité de la segmentation des franges (voir section suivante sur les traitements), soit pour augmenter le contraste vertical et rendre la classification plus robuste. Par défaut, une valeur fixée à 50% de l'intensité lumineuse maximale semble appropriée.

3.3.5 Choix de la couleur

Dans [Albitar 08], la couleur verte avait été choisie, car tout comme les spectres proches du rouge, le vert est bien réfléchi par les organes, mais à sa différence, le vert diffuse beaucoup moins (Fig. 3.12.d). Cependant, le problème principal, en vision endoscopique par lumière structurée, est la présence de bruit. Une intensité lumineuse maximale est donc souhaitable pour réduire le bruit. Ainsi, comme notre prototype endoscopique utilise un projecteur DLP, nous avons choisi d'employer de la lumière blanche.

3.4 Chaîne de traitements

Dans cette section, nous abordons les approches utilisées pour les traitements nécessaires jusqu'au décodage des primitives. Les traitements concernant la calibration et la reconstruction 3-D sont discutés dans le chapitre suivant. Si, dans la majorité des cas, les principes algorithmiques généraux sont donnés, certains indices d'implémentation ayant permis une exécution en-ligne de l'ensemble de la chaîne des traitements peuvent être ponctuellement précisés.

A travers la succession des différents algorithmes, nous montrons de nouveau la pertinence du choix des primitives et notamment de leurs contenus directionnels. Comme les algorithmes sont nombreux, nous n'entrons pas dans les détails d'implémentation, mais donnons les principes généraux.

Une approche mixte

Bien que les primitives utilisées soient positionnées par la discrétisation du domaine de l'image du projecteur en 2-D, leur continuité verticale autorise une reconstruction dense comme avec des franges.

Pour les traitements de scènes endoscopiques, nous supposons qu'à ce stade, la région d'intérêt correspondant au *cercle du champ endoscopique* est connue (voir son estimation, Chapitre 4).

A des fins de démonstration, les traitements sont illustrés sur un motif de résolution modeste :

- 30×30 primitives,
- $H_{Min} = 2$,
- 2 symboles,
- codage SPSM par alignement (voir Chapitre 2).

De façon générale, travailler dans l'image rectifiée offre plusieurs avantages.

- Le motif étant aligné sur la géométrie épipolaire, la taille des primitives selon la direction verticale de l'image rectifiée reste constante (voir section 3.4.4), quelle que soit la surface sur laquelle le motif est projeté, car elle est contrainte par la géométrie épipolaire. Ainsi,

Traitement	Principe
[Endo] Estimation du cercle endoscopique (offline, une seule fois)	Fonction de coût qui maximise le contraste entre l'intérieur du cercle (clair) et l'extérieur (foncé), (voir Ch. 4, Calibration)
Correction des distorsions géométriques	Utilise modèle de Brown, voir Ch. 4, Calibration
Rectification	En deux étapes. 1. Alignement horizontal des lignes épipolaires 2. Correction des distorsions horizontales créées à l'étape 1,
[Endo] Recalage épipolaire	1. Normalisation locale de l'intensité de l'image 2. Projection d'énergie horizontalement 3. Recalage depuis les deux lignes épipolaires centrales (codées)
Estimation des champs de déformation	Estimation des paramètres d'élongation et d'inclinaison par maximisation de l'auto-corrélation
Segmentation des franges	Adaptation morphologique du masque de corrélation à partir des champs estimés précédemment
Classification	Etude du sens du gradient vertical
Construction du graphe d'adjacence	Basée sur la continuité des franges, le parallélisme local des voisinages, et la minimisation de la distance de Hamming (décodage)
Décodage	1) Calcul de la distance de Hamming avec tous les codes possible pour cette ligne 2) Sélection de la meilleure correspondance sur une région 3) Vote pour les régions voisines 4) Propagation du décodage dans la zone d'adjacence.

FIGURE 3.15 – La chaîne de traitements.

la distance en pixels entre les lignes épipolaires dans l'image devient une métrique essentielle permettant de choisir a priori une taille de zone d'étude pour la segmentation et la classification.

- Les primitives étant groupées par lignes épipolaires, il y a beaucoup moins de lignes épipolaires à étudier (N au lieu de $M \times N$, pour un motif de $M \times N$ primitives avec M par ligne) que pour des motifs non alignés. Cela est souhaitable lorsqu'on vise à obtenir un traitement exécuté en temps-réel.
- Après rectification, les lignes épipolaires sont horizontales, et sont ainsi stockées de façon contiguë en mémoire, ce qui permet des accès plus rapides en favorisant la lecture dans les caches de mémoire, cela est d'autant plus important quand le flux d'images est en résolution Full-HD, comme c'est le cas pour nous.

Dans le cas où la qualité de l'image est bonne et que les scènes offrent peu de perturbations, la plupart de ces traitements ne sont pas nécessaires, une analyse plus simple des images basée

sur l'utilisation de seuils et de maxima locaux peut souvent suffir. Mais comme évoqué dans la section précédente, les images endoscopiques s'accompagnent de nombreuses perturbations.

Comme nous souhaitons traiter des images issues d'un flux Full-HD, et qu'avec le prototype actuel, il y a une présence de bruit important dans le flux d'images, il est nécessaire de projeter des primitives de grande taille. La segmentation de celles-ci pourra ensuite se faire par intégration de l'intensité lumineuse sur des zones de taille importante (correspondant à la taille des primitives). Tout ceci devra être réalisé en temps-réel. Il est donc nécessaire de disposer d'une méthode d'intégration sur des blocs d'image qui soit efficace. L'utilisation des images intégrales (voir Annexe B) est une réponse à cette problématique.

3.4.1 Alignement épipolaire du motif

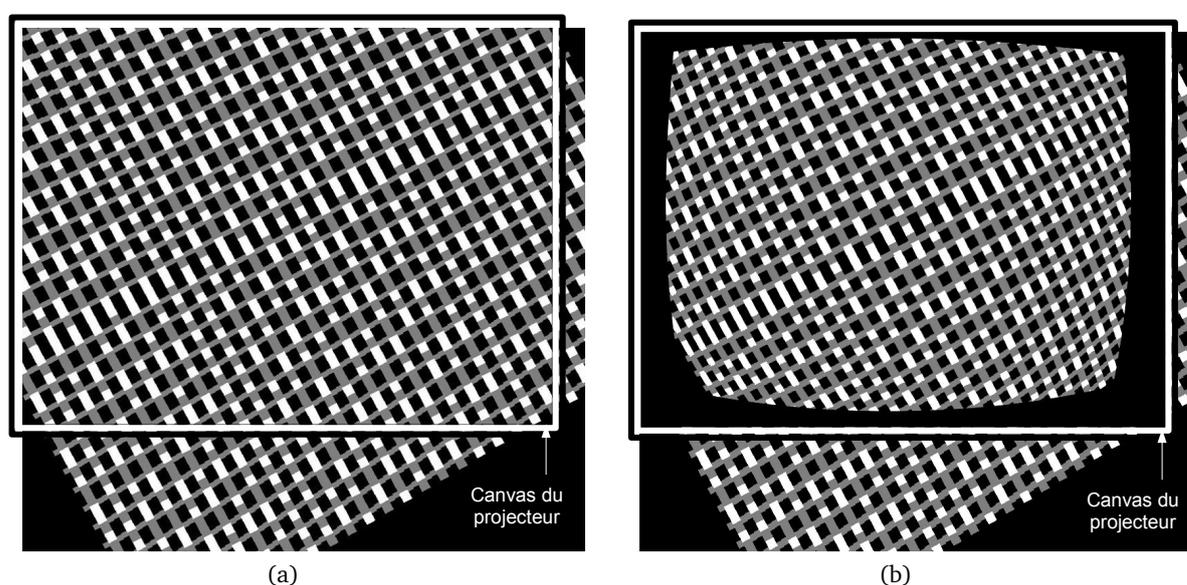


FIGURE 3.16 – Alignement épipolaire (a) et distorsion (b) du motif.

Le principe d'alignement épipolaire du motif (Fig. 3.16.a) a été décrit dans la section 3.1. Celui-ci repose sur une discrétisation de l'image du projecteur effectuée de telle sorte que l'ensemble des colonnes codées apparaissent dans l'image et que la distribution des primitives soit la plus homogène possible.

3.4.2 Distorsion du motif

La plupart du temps, les projecteurs que l'on trouve sur le marché possèdent une correction interne de la géométrie permettant de corriger partiellement ou totalement les distorsions géométriques créées par les optiques. Dans notre cas, des optiques endoscopiques supplémentaires viennent distordre le flux d'images en sortie du projecteur de façon significative. De plus des optiques d'interface (voir Chapitre 5 : développements matériel) ont été nécessaires pour la conception du projecteur endoscopique. L'ajout de cette chaîne d'optiques crée, en sortie, une distorsion importante du flux d'images projetées.

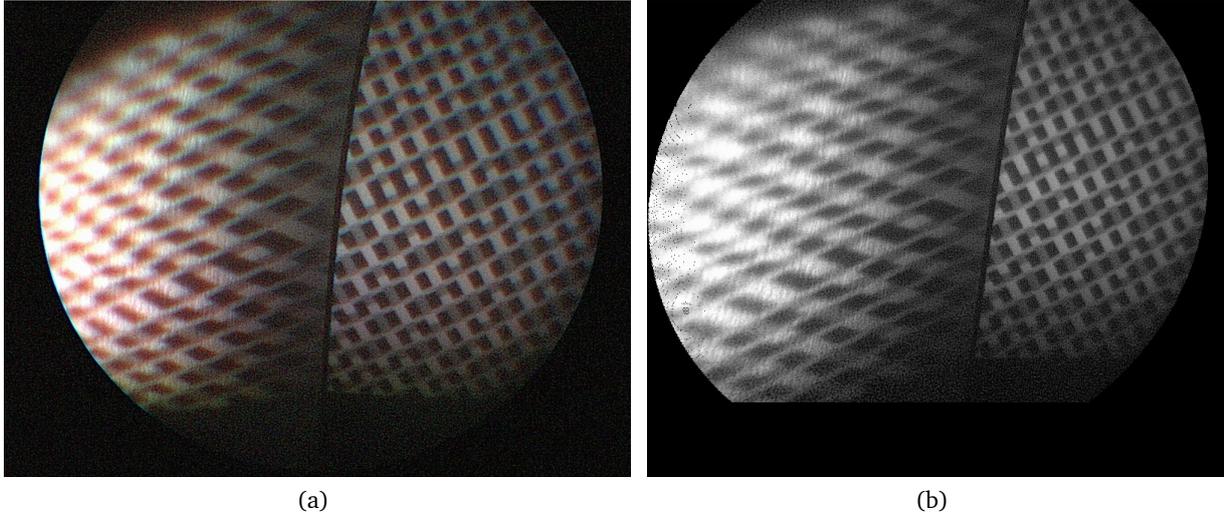


FIGURE 3.17 – (a) Image capturée. (b) Conversion en niveaux de gris, puis correction des distorsions créées par les optiques de la caméra.

Comme notre motif est aligné sur la géométrie épipolaire, il est nécessaire que le modèle sténopé puisse s'appliquer de la façon la plus précise possible et une correction a priori de ces distorsions géométriques est donc indispensable. La méthode utilisée pour estimer les paramètres de distorsion ainsi que le modèle de Brown [Brown66] sont décrits au chapitre 4, section étalonnage.

Soient $K1_p$, $K2_p$, x_{cp} , y_{cp} , $P1_p$, $P2_p$, les paramètres estimés du modèle de Brown pour le projecteur. Alors l'image distordue ($i(x_d, y_d)$) du motif aligné ($i(x_a, y_a)$) (Fig. 3.16.b) est obtenue par la formule :

$$\begin{aligned} x_d &= x_a + (x_a - x_{cp})(K1_p r^2 + K2_p r^4) + (P1_p(r^2 + 2(x_a - x_{cp})^2) + 2P2_p(x_a - x_{cp})(y_a - y_{cp})) \\ y_d &= y_a + (y_a - y_{cp})(K1_p r^2 + K2_p r^4) + (P2_p(r^2 + 2(y_a - y_{cp})^2) + 2P1_p(x_a - x_{cp})(y_a - y_{cp})) \end{aligned} \quad (3.5)$$

avec

$$r = \sqrt{(x_a - x_c)^2 + (y_a - y_c)^2}. \quad (3.6)$$

Les valeurs d'intensité finales sont calculées par interpolation bilinéaire.

Ce modèle permet d'estimer les distorsions géométriques qui sont créées par l'ensemble d'optiques du projecteur (Fig. 3.1 O_P). Ces dernières agiront alors comme des corrections de la distorsion du motif créée ici afin de pouvoir exploiter pleinement le modèle sténopé.

3.4.3 Correction de la distorsion de l'image

Comme l'information spectrale n'est pas utilisée dans la suite des traitements, l'image est tout d'abord convertie en niveaux de gris - ce qui permettra l'accélération des traitements ultérieurs (1 canal au lieu de 3) - puis les distorsions géométriques créées par les optiques de la caméra (Fig. 3.1 O_C) sont corrigées (Fig. 4.8) en utilisant également le modèle de Brown.

La transformation de l'image utilise la même formule (equ. 3.5, 3.6), mise à part l'inversion du sens de l'application : l'image corrigée est cette fois obtenue à partir de l'image distordue et les paramètres estimés, $K1_c$, $K2_c$, x_{cc} , y_{cc} , $P1_c$, $P2_c$ caractérisent cette fois l'optique de la caméra.

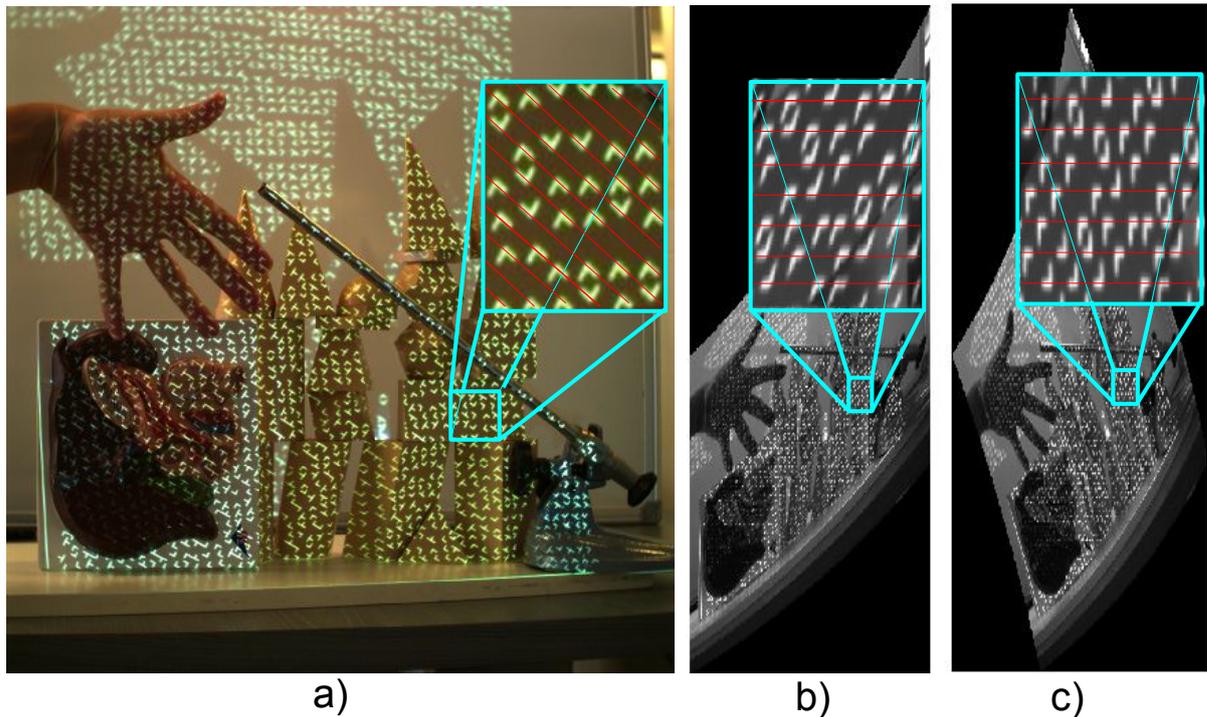


FIGURE 3.18 – a) Image acquise dans le PIC. b) Rectification 1 : les lignes épipolaires sont horizontales. c) Rectification 2 : la distorsion horizontale due à la rectification 1 est compensée. (le rapport de tailles n'est ici pas respecté, en réalité les images b) et c) sont plus grandes que a) car aucun pixel de l'image initiale ne subit de compression.

Cette formule est utilisée une fois pour toutes pour calculer la carte de correspondance entre les coordonnées des pixels de l'image non distordue et ceux des pixels de l'image distordue. La création de cette carte de correspondance, ainsi que l'estimation des paramètres mentionnés sont expliqués en détails dans le Chapitre 4 (section Étalonnage).

3.4.4 Rectification de l'image

La rectification de l'image consiste à opérer une transformation de celle-ci de telle sorte que les lignes épipolaires soient disposées horizontalement et ainsi, que la suite des traitements (segmentation, décodage) puisse se faire directement en analysant les lignes de l'image rectifiée.

Les propriétés désirées de l'image après rectification sont les suivantes :

- lignes épipolaires horizontales
- absence de compression de l'image (pas de perte d'information),
- conservation locale des angles.

La littérature est abondante et la question a déjà bien été étudiée. Ainsi nous ne ferons pas une étude mais nous désirons dégager dans cette section les éléments principaux qui nous ont amené à la solution choisie.

Une rectification des images acquises est d'abord effectuée pour disposer les lignes épipolaires (LEP) horizontalement. Cela permet d'imposer un espacement vertical constant entre les

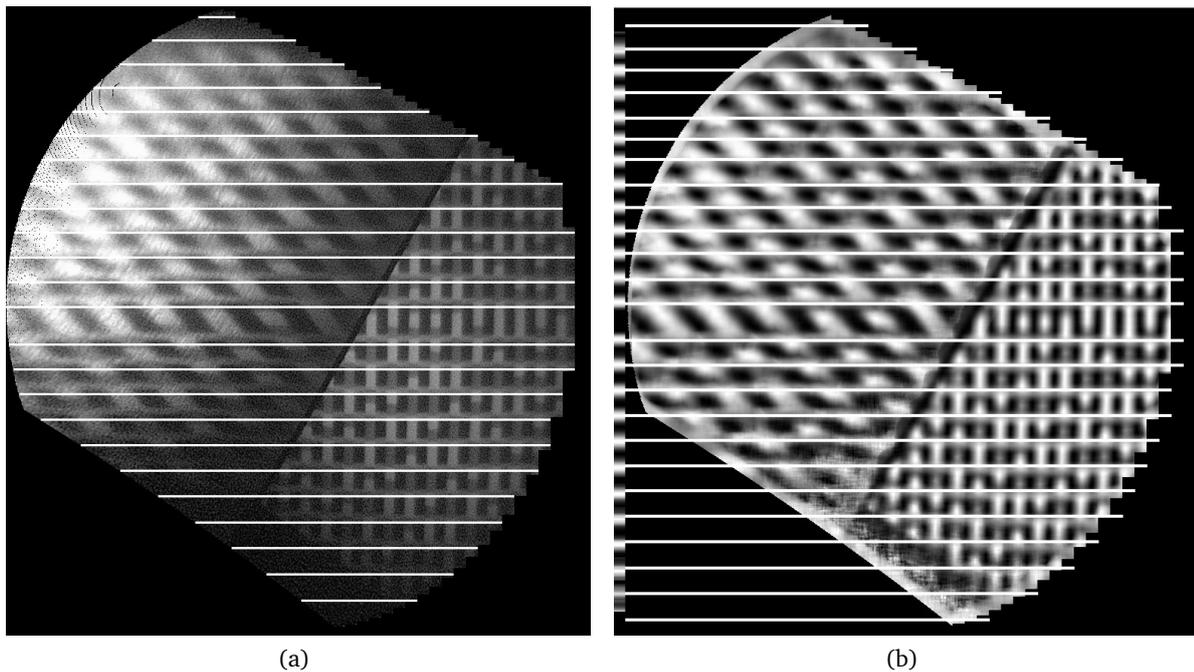


FIGURE 3.19 – (a) Léger déplacement vertical des lignes épipolaires dû à la manipulation du système stéréo pendant l’acquisition. (b) Les lignes épipolaires recalées (à gauche, les accumulateurs d’intensité le long des lignes épipolaires).

LEP (tous les 15 pixels Fig. 3.18.b/c) et de mieux distinguer les groupes de primitives alignées sur des LEP différentes. Par exemple sur la figure 3.18, les bras initialement alignés le long des LEP dans la figure 3.18.a se retrouvent tous horizontaux (Fig. 3.18.b/c) le long des LEP régulièrement espacée verticalement. L’approche choisie pour la rectification est inspirée de [Chen03]. Elle consiste à identifier les pixels de l’image les plus sujets à une compression (sur le bord de l’image le plus éloigné de l’épipole) et de s’en servir pour tracer les lignes épipolaires qui seront disposées horizontalement dans l’image résultante. Dans un deuxième temps, la distorsion horizontale créée lors de la première rectification est compensée (Fig. 3.18.c) pour conserver l’angle des bras des primitives dans l’image initiale (Fig. 3.18.a). Enfin, une carte de correspondance, cumulée avec la carte de correspondance créée pour corriger les distorsions est calculée (voir Ch. 4, sec. Étalonnage).

3.4.5 Recalage épipolaire (ENDO)

Si le système stéréo reste fixe (la position relative des dispositifs de projection/capture n’est pas modifiée), cette étape n’est pas nécessaire, car l’estimation de la géométrie épipolaire est suffisamment précise (voir Chapitre 4). Dans le où la géométrie projective du système stéréo est instable, un léger *recalage épipolaire* peut être nécessaire. En effet les endoscopes peuvent être assimilés à des tubes fins et allongés (1cm de diamètre pour environ 30 cm de longueur) et sont donc soumis à de petites déformations, celles-ci ayant un léger impact sur la géométrie du système. De plus, le système de contention du prototype, bien que présentant une bonne rigidité globale (voir Chapitre 6, développements matériel), ne peut garantir qu’il n’y ait aucun jeu, aussi minime soit-il, entre toutes les pièces utilisées. Sachant que la ligne de base est de 6 mm entre le centre optique de la caméra et celui du projecteur, une déformation infime du système

peut provoquer, dans l'image rectifiée, un déplacement non négligeable des lignes épipolaires par rapport à celles estimées lors de l'évaluation de la matrice Fondamentale (Fig. 3.19.a).

Soit $\overline{d_{Epi}}$, l'écart moyen entre les lignes épipolaires dans l'image rectifiée de la caméra. Pour prendre en compte des décalages épipolaires supérieurs à $\overline{d_{Epi}} \div 2$, nous avons effectué un codage global des lignes épipolaires en espaçant les deux lignes centrales du milieu par $\overline{d_{Epi}} \times 1.5$ pour éviter toute ambiguïté lors du recalage. Cette même technique de codage avait été utilisée par [Graebling02] pour référencer globalement les franges d'un système de métrologie utilisant la lumière structurée.

Une deuxième raison, non scientifique, de ce choix, est un hommage à Pierre Graebling, qui a été mon premier directeur de thèse et qui est décédé des suites d'un cancer après avoir encadré cette thèse pendant deux années. La vision de cet écart supplémentaire, au centre des motifs proposés, rappellera Pierre à nos bons souvenirs.

Avec cet écart supplémentaire, voici les différentes étapes du recalage :

1. Une élongation d'histogramme local (Fig. 3.19.b) est appliquée sur l'image. Cela permet de prendre en compte les inhomogénéités de réflectance de la lumière dans l'image, comme c'est le cas sur la figure 3.19.a, et ainsi, de ne pas donner un poids trop fort/faible aux zones sur/sous exposées. Pour ce faire, autour de chaque pixel, une région de taille $\overline{d_{Epi}} \times \overline{d_{Epi}}$, qui correspond à la taille moyenne approximative des primitives dans l'image, est étudiée. La valeur d'intensité minimum, i_{min} , et maximum, i_{max} , sont extraites. La dynamique d'intensité locale est alors $i_{dyn} = i_{max} - i_{min}$. L'intensité localement normalisée, i_{norm} , est alors exprimée en fonction de l'intensité initiale i_{init} par la relation :

$$i_{norm} = (i_{init} - i_{min}) \div i_{dyn} \quad (3.7)$$

2. La somme des intensités de chaque ligne est calculée et sauvegardée dans un accumulateur. Les primitives choisies (voir section 3.3) ayant statistiquement plus d'énergie au niveau de leur centre aligné sur les lignes épipolaires, des maxima locaux apparaissent alors dans l'accumulateur.
3. Après une élongation d'histogramme locale de l'accumulateur, ces maxima locaux apparaissent clairement (Fig. 3.19.b, bord gauche), et les lignes épipolaires peuvent être recalées à partir des deux lignes épipolaires centrales qui ont un écart 1,5 fois supérieur aux autres.

Une fois ce traitement effectué, la segmentation et la classification des primitives peuvent être réalisées. Le centre des primitives se trouvant à l'intersection des lignes épipolaires et des franges, il est à présent nécessaire de segmenter les franges. Cette étape permet en outre une reconstruction dense le long des franges.

3.4.6 Segmentation des franges

Les primitives choisies (section 3.3) émettent davantage d'énergie lumineuse le long des franges qu'entre ces dernières. Pour mettre en évidence celles-ci nous utilisons donc un masque de corrélation générique qui favorise l'énergie au milieu des franges (Fig. 3.20.a). Comme nous n'avons pas d'information a priori sur la déformation des primitives et de leur voisinage (qui dépendent des paramètres d'inclinaison et d'élongation évoqués plus tôt), excepté la hauteur moyenne des primitives ($\overline{d_{Epi}}$) qui reste constante, nous utilisons une taille $\overline{d_{Epi}} \times \overline{d_{Epi}}$ pour ce masque.

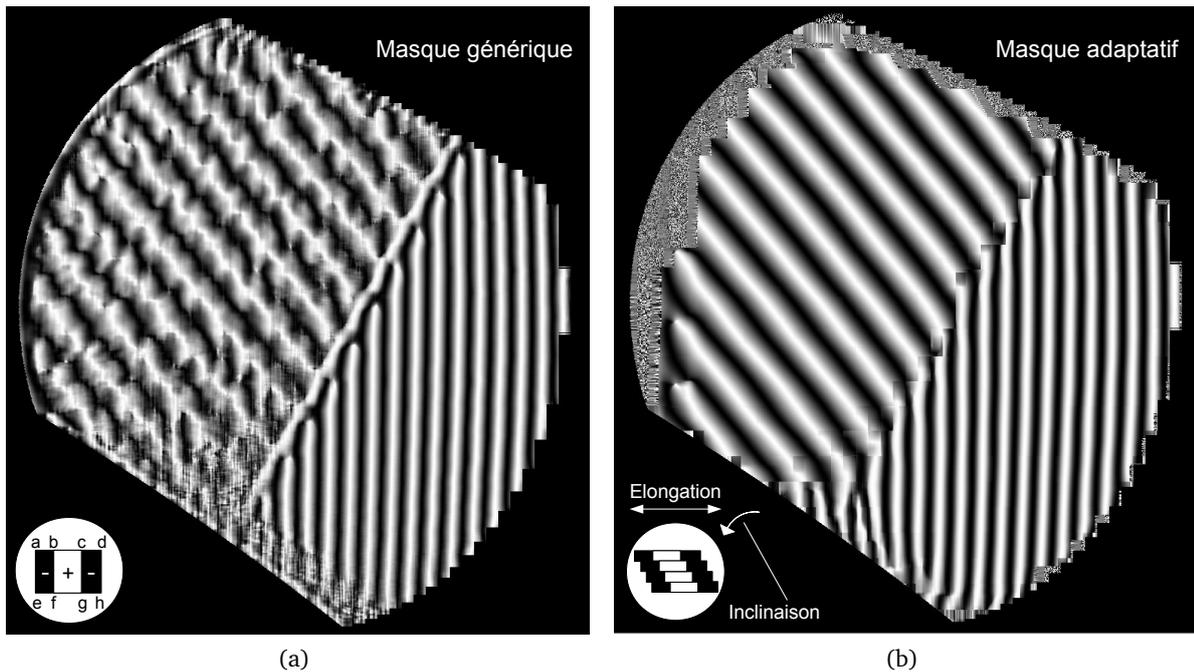


FIGURE 3.20 – (a) Franges segmentées avec un masque générique. (b) Franges segmentées avec un masque adaptatif.

Implémentation Pour une execution plus rapide, ce masque est appliqué sur l'image intégrale (Annexe B) de l'image rectifiée. Ainsi, chaque réponse, $RepKer$ au masque peut être calculée en 8 accès mémoire seulement quelle que soit la taille des primitives avec la formule suivante :

Soient b, c, f, g , les extrémités de la zone à maximiser, $ZoneMax$, et a, d, e, h , les extrémités extérieures du masque, $ZoneKer$, sur l'image intégrale, alors :

$$\begin{aligned}
 ZoneKer &= (h + a) - (e + d) \\
 ZoneMax &= (g + b) - (f + c) \\
 ZoneMin &= ZoneKer - ZoneMax \\
 RepKer &= ZoneMax - ZoneMin
 \end{aligned}
 \tag{3.8}$$

La figure (Fig. 3.20.a), montre clairement que cette approche ne permet pas de segmenter les franges de façon satisfaisante. En effet, de nombreux discontinuités apparaissent lorsque les franges sont inclinées et aboutissent à des imprécisions importantes lors du positionnement des primitives.

Pour corriger cela, il est donc nécessaire de prendre en compte l'inclinaison α et l'élongation S locales des franges. Pour ce faire, nous procédons par une estimation des champs de ces deux paramètres (expliqués dans le paragraphe suivant). Les résultats obtenus sont alors qualitativement bien meilleurs (Fig. 3.20.b). Les images intégrales étant adaptées pour lire des zones rectangulaires, nous avons subdivisé verticalement le masque par morceau (Fig. 3.20.b, en bas à gauche), l'écart horizontal entre les morceaux dépendant de la pente de l'inclinaison estimée. Le nombre de morceaux à utiliser dépend de l'équilibre rapidité/qualité désiré.

3.4.7 Estimation des champs de déformation

Pour estimer les champs de déformation, nous avons considéré l'élongation et l'inclinaison locales évoquées plus tôt. Comme ces valeurs varient par région, il est inutile de les étudier pour chaque pixel, nous avons donc échantillonné l'image par un champ de 60×60 pour l'exemple (Fig. 3.21). Bien sûr plus le motif utilisé est de résolution importante, et plus la scène comporte de discontinuités de courbure, plus il faudra augmenter le nombre d'échantillons pour capturer au mieux les paramètres de chaque zone.

Le principe général est de trouver les paramètres qui minimisent l'erreur de corrélation avec une zone autour de l'échantillon. Pour chaque point de l'échantillon, nous avons procédé comme suit pour estimer les deux paramètres :

1. Considérer une zone *patron* de $\overline{d_{Epi}} \times \overline{d_{Epi}}$ (en noir sur la figure 3.21.a) autour de l'échantillon. Une zone un peu plus large permettra de mieux prendre en compte les élongations importantes.
2. **Elongation** : à partir du centre de l'échantillon, déplacer le patron horizontalement vers la gauche et la droite (Fig. 3.21.a, en blanc vers S) et calculer l'erreur de corrélation. Dès que le premier minimum local est atteint, l'élongation correspond à la translation nécessaire depuis le centre de l'échantillon pour l'atteindre.
3. **Inclinaison** : à partir du centre de l'échantillon, déplacer le patron verticalement vers le haut et le bas de ΔY pixels et pour chaque déplacement déterminer la translation horizontale ΔX par minimisation de l'erreur de corrélation. L'angle d'inclinaison est alors donné par $\alpha = \arctan(\Delta Y \div \Delta X)$. La singularité $\Delta X = 0$, souvent présente, doit être gérée, l'angle est alors de $\pi/2$ rad.
4. Enfin, un filtre médian est appliqué pour lisser les éventuelles discontinuités dues à des inhomogénéités de réflectance tout en gardant un bon contraste entre les zones distinctes.

L'estimation de ces champs a quatre utilités :

- Ces paramètres permettront tout d'abord d'ajuster le masque de corrélation évoqué précédemment pour la segmentation des franges.
- D'autre part, ils permettront de connaître la taille précise des primitives donnant ainsi des indices précieux pour leur classification à venir.
- Ces deux champs permettront d'étudier les zones de contiguïté géométrique à travers la création d'un *graphe d'adjacence* et de donner des indices importants lors de la propagation du décodage en environnement bruité.
- Enfin ce graphe sera également utilisé lors de la création du maillage utilisé pour le rendu 3-D.

Sur l'exemple de la figure (Fig. 3.21.b), l'élongation varie entre 55 et 140 pixels (quasiment du simple au triple), alors que l'inclinaison varie de 55 à 90 (pas d'inclinaison) degrés. Il n'y a ici pas de zone de perte d'adjacence, car il n'y a pas de corrélation entre une faible élongation et une forte inclinaison. Mais si les mêmes paramètres avaient été inversement corrélés, l'indice d'adjacence serait bien inférieur à 1 (voir Fig. 3.11.a).

Quoi qu'il en soit, la présence des franges et leur segmentation nous assure que, dans ce cas, nous serions en mesure de retrouver les primitives du voisinage le long des franges quand l'approche par proximité aurait échoué.

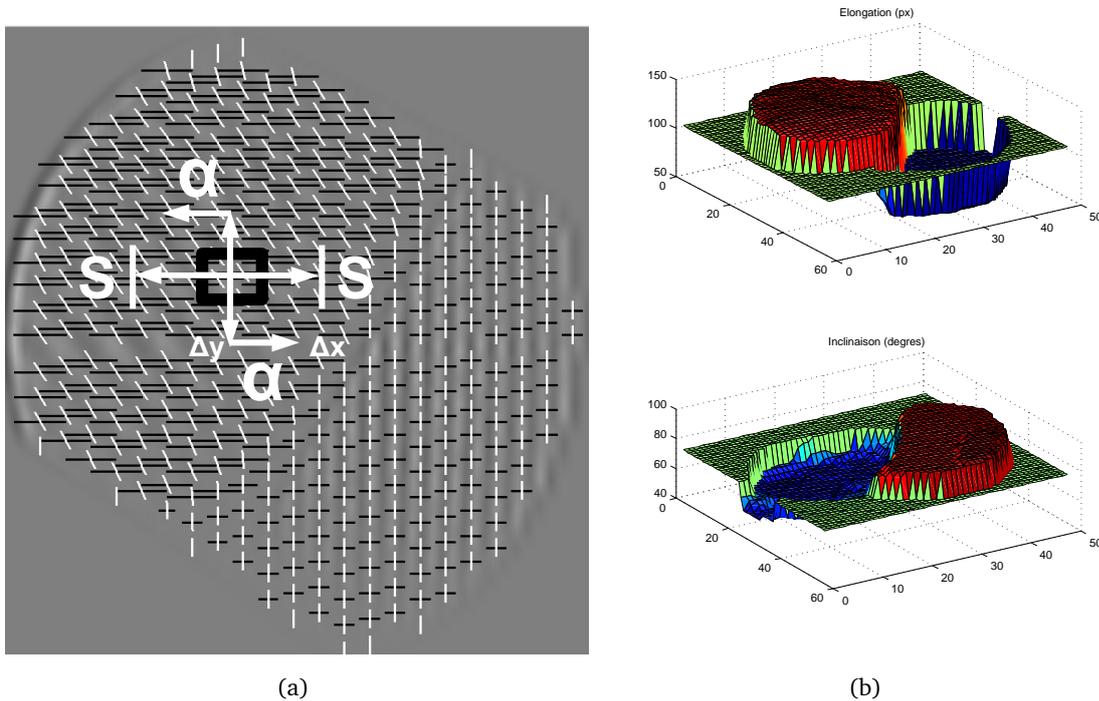


FIGURE 3.21 – Champs de déformation locale, a) en noir, l'élongation, en blanc l'inclinaison. b) En haut l'élongation, en bas l'inclinaison, les couleurs correspondant à l'amplitude du paramètre estimé.

Comme les lignes épipolaires et les franges sont segmentées, les primitives sont maintenant précisément localisées et leur taille est connue dans les deux dimensions. Nous pouvons procéder à leur classification.

3.4.8 Classification des primitives

La classification consiste à mettre en correspondance la primitive segmentée avec le symbole correspondant issu du codage du motif.

Dans notre cas, avec deux primitives différenciées par leur intensité relative d'énergie verticalement (voir section 3.3), la classification est très simple. Une fois le centre des primitives localisé et les paramètres α et S estimés, il suffit d'estimer, à partir du centre de la primitive et suivant la pente donnée par α si la moyenne d'intensité lumineuse est supérieure vers le haut ou vers le bas. La hauteur de la région étudiée est de $\overline{d_{Epi}} \div 2$ au-dessus et en-dessous du centre de la primitive. La largeur dépend de l'élongation S estimée précédemment.

3.4.9 Construction du graphe d'adjacence

Un graphe d'adjacence est ici un graphe qui met en relation chaque primitive segmentée avec ses 4 primitives directement voisines, en indiquant, pour chaque relation, si les deux primitives sont susceptibles d'appartenir ou non, à la même zone d'adjacence. Deux primitives appartiennent à la même zone d'adjacence si elles sont voisines à la fois dans l'image du projecteur et dans l'image de la caméra.

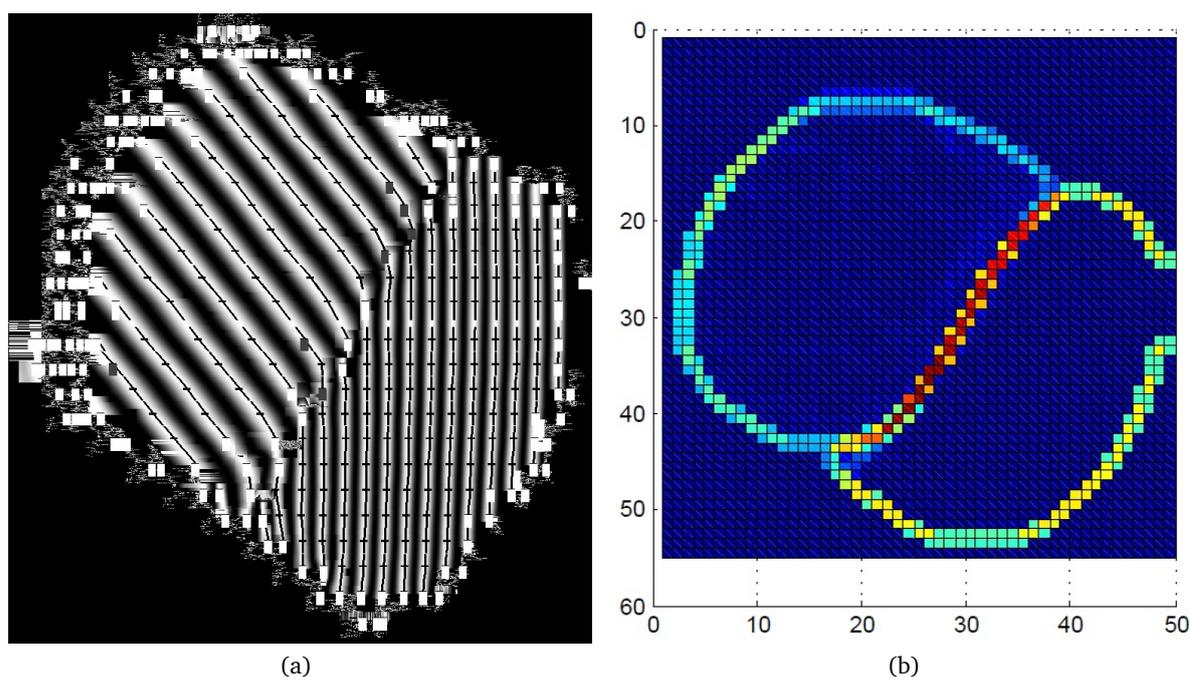


FIGURE 3.22 – A) Graphe d’adjacence, en trait noir, identification des primitives adjacentes verticalement. Carrés blancs et gris : frontières des zones d’adjacence. b) Amplitude du critère de non parallélisme, IP , des voisinages.

La construction d’un graphe d’adjacence est évoqué à plusieurs reprises dans la littérature (dans [Koninckx06], par exemple).

Le graphe d’adjacence est primordial dans l’approche proposée, car nous utilisons le voisinage spatial comme modalité de codage. Il est important que les bons voisinages soient retrouvés dans les images acquises. Bien souvent, la simple information de distance entre une primitive et ses voisines est utilisée pour retrouver le voisinage (voir section 3.2). C’est en effet le seul critère pour les motifs utilisant des primitives sans information de direction, comme les spots de couleurs ([Adán09], [Morano98], [Claes08], par exemple. Or nous avons vu précédemment que dans bien des cas, cette stratégie est mise en échec. L’utilisation d’un graphe d’adjacence permet de pallier ce problème.

Le graphe d’adjacence présenté ici est construit en deux temps et utilise deux critères géométriques (1ère étape) et un critère entropique, à travers le décodage (2ème étape) pour décider si des primitives voisines dans l’image acquise font parti d’un même voisinage dans le motif projeté.

Etape 1, critères géométriques :

1. La **continuité intra-frange**. Les franges étant segmentées et leur intensité normalisée (image *imgFranges*) dans une dynamique entre 0 (inter-frange) et 255 (centre d’une frange). S’il existe entre deux primitives voisines verticalement (le long d’une frange) des pixels dont l’intensité dans *imgFranges* est inférieure à 127, alors il y a là une discontinuité dans la frange et donc dans l’adjacence verticale.
2. Les variations d’**élongation** et d’**inclinaison** à l’intérieur d’un même voisinage, peuvent s’apparenter à un non parallélisme des franges. Comme il n’est pas souhaitable de multiplier les critères et les seuils, nous modélisons ici le non parallélisme de la façon suivante :

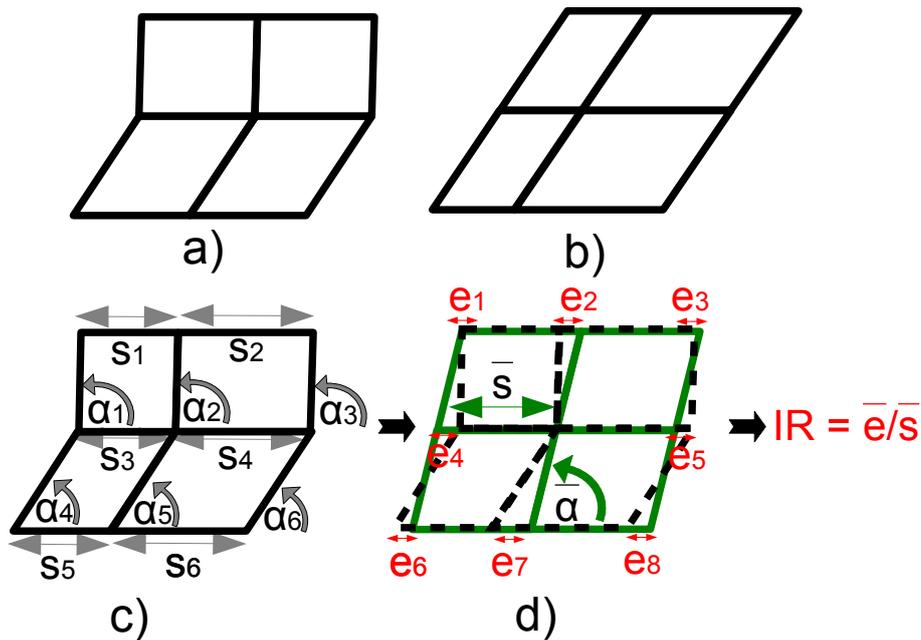


FIGURE 3.23 – a) Une variation d’inclinaison induit un non parallélisme du voisinage sans qu’une variation d’élongation n’intervienne. b) Une variation d’élongation induit une déformation du voisinage sans qu’une variation d’inclinaison n’intervienne. c), d) Un critère commun est utilisé, avec. Dans c) S_i : les écarts horizontaux entre les primitives du voisinage, α_i : les angles d’inclinaison. Dans d), \bar{S} : la moyenne des S_i , $\bar{\alpha}$: la moyenne des α_i et \bar{e} : la moyenne des écarts horizontaux e_i entre les positions des intersections de voisinage normalisé (en vert) et celles du voisinage segmenté (pointillé).

Une variation de α n’entraînant pas forcément une variation de S , et réciproquement, au sein d’un voisinage de primitives, ces deux paramètres ne sont pas toujours couplés (voir Fig. 3.23.a et 3.23.b). A la place, nous définissons un seul critère exprimant le degré de non régularité IR du voisinage (voir Fig. 3.23.c et 3.23.d). Pour cela, nous définissons un voisinage parallèle et sans variation d’élongation $Neigh_{\alpha,S}$ construit à partir des valeurs moyennes de α et de S des primitives du voisinage (en vert sur Fig. 3.23.d). IR est alors défini comme la moyenne des écarts entre les positions des primitives du voisinage $Neigh_{\alpha,S}$ et les positions réelles de ces primitives.

Ces deux premiers critères ne permettent pas de prendre en compte toutes les configurations géométriques possible. En particulier si deux surfaces voisines ont la même orientation vis-à-vis du système stéréo mais à des profondeurs différentes (exemple de deux plans parallèles). Alors il est probable que l’élongation et l’inclinaison soient très proches et à certaines profondeurs, que des franges distinctes dans le motif projeté s’alignent dans l’image capturée par la caméra. Aucun des critères précédents ne sera alors en mesure de caractériser la frontière entre les deux zones d’adjacence. Cependant le codage du motif peut encore être exploité.

Etape 2, critère entropique. Si, pour une primitive donnée, plusieurs appariements différents sont proposés lors du décodage (voir ci-après), celui qui correspondra à la plus faible distance de Hamming (entre le voisinage codé et celui décodé) avec l’appariement proposé, sera conservé.

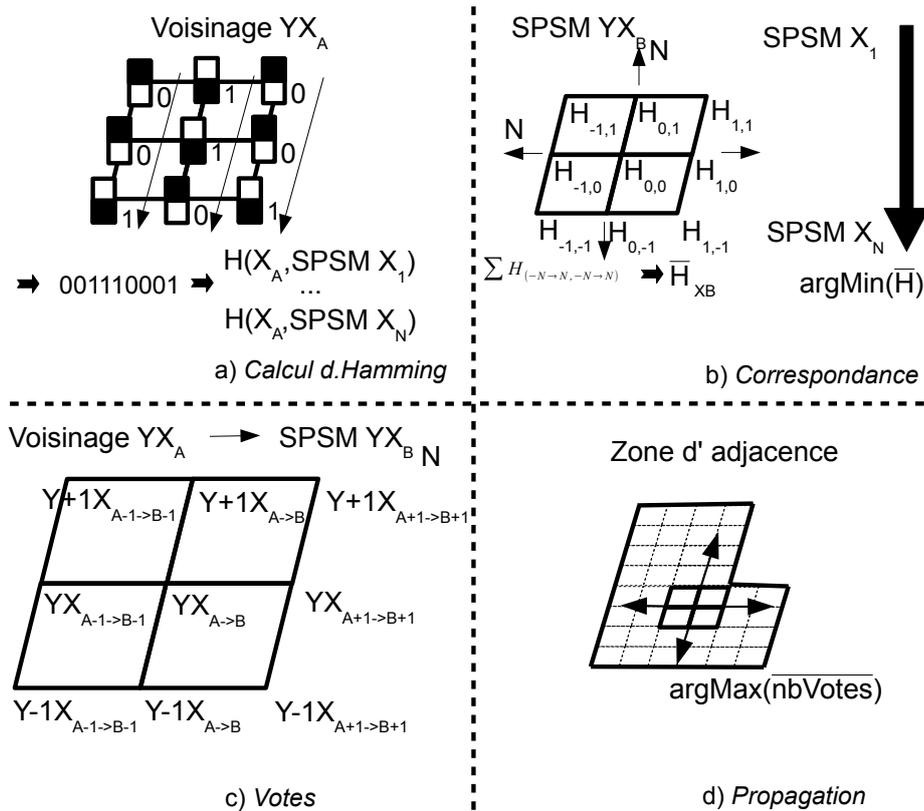


FIGURE 3.24 – Les étapes du décodage.

- Calcul de la distance de Hamming entre le code, X_A , de chaque voisinage segmenté sur la ligne épipolaire Y et tous les codes, de $SPSM X_1$ à $SPSM X_N$, associés aux primitives alignées sur Y dans le plan projecteur.
- Pour chaque primitive, X_B à décoder, calcul de la distance de Hamming moyenne, \bar{H}_{XB} (à partir de a)) sur une zone $N \times N$ voisinages autour de la primitive pour chaque appariement possible (de $SPSM X_1$ à $SPSM X_N$) et sélection de l'appariement qui obtient le plus faible \bar{H} .
- Pour chaque appariement effectué dans b), vote pour les 8 primitives voisines pour les appariements contigus.
- Les appariements qui ont obtenu le plus de votes à l'issue de c), ou un nombre de votes supérieur à un seuil sont considérés comme sûrs. A partir de eux, la propagation du décodage s'effectue de manière contiguë dans leur zone d'adjacence.

3.4.10 Décodage

Le décodage consiste à mettre en correspondance les primitives segmentées dans l'image de la caméra avec celles projetées dans l'image du projecteur.

Nous distinguons deux cas pour le décodage

- Soit la qualité des images est bonne et il y a peu de perturbations. On peut avoir confiance dans les classifications effectuées. Dans ce cas le traitement est largement simplifié. Il suffit pour le décodage d'identifier le code de cette ligne offrant la distance de Hamming minimale. Si cette distance est inférieure H_{Min} alors la primitive est décodée. La méthode des votes discutée plus loin peut également être employée pour rendre cette mise en correspondance plus robuste et mieux gérer les bords des objets.

- Si la qualité est mauvaise, il est nécessaire de prendre en compte un voisinage plus étendu (que 3×3). Le décodage se fait alors par propagation à partir des décodages les plus certains et suivant la méthode décrite ci-dessous.

Le décodage proposé est constitué de l'enchaînement successif des 4 étapes suivantes (voir Fig. 3.24) pour chaque voisinage segmenté :

- a) La distance de Hamming est calculée avec tous les codes associés à cette ligne épipolaire dans la SPSM.
- b) La meilleure correspondance (minimisant la distance de Hamming) est choisie sur une région de voisinages contigus. Plus cette région est petite, plus les petites discontinuités de courbure pourront être détectées. Plus elle est grande, plus le décodage sera robuste au bruit et aux mauvaises classifications par intégration.
- c) Une fois la meilleure correspondance trouvée pour ce voisinage, la méthode des votes (voir [Morano98]), reposant sur le principe de contiguïté locale, est utilisée. Chaque primitive mise en correspondance indique aux primitives voisines de prendre une mise en correspondance contiguë.
- d) Dans chaque zone de contiguïté, la mise en correspondance ayant reçu le plus de votes, ou ayant un nombre de votes supérieur à un seuil (7 pour un nombre maximal de 8 votes dans un voisinage 3×3), est identifiée. Cette mise en correspondance est propagée par contiguïté jusqu'aux frontières identifiées dans le graphe d'adjacence. Si pour un même voisinage, la propagation propose deux mises en correspondances incohérentes (deux zones d'adjacence différentes qui n'ont pu être discriminées dans le graphe), alors celle permettant de minimiser la distance de Hamming est conservée (critère entropique du graphe d'adjacence).

3.5 Conclusion

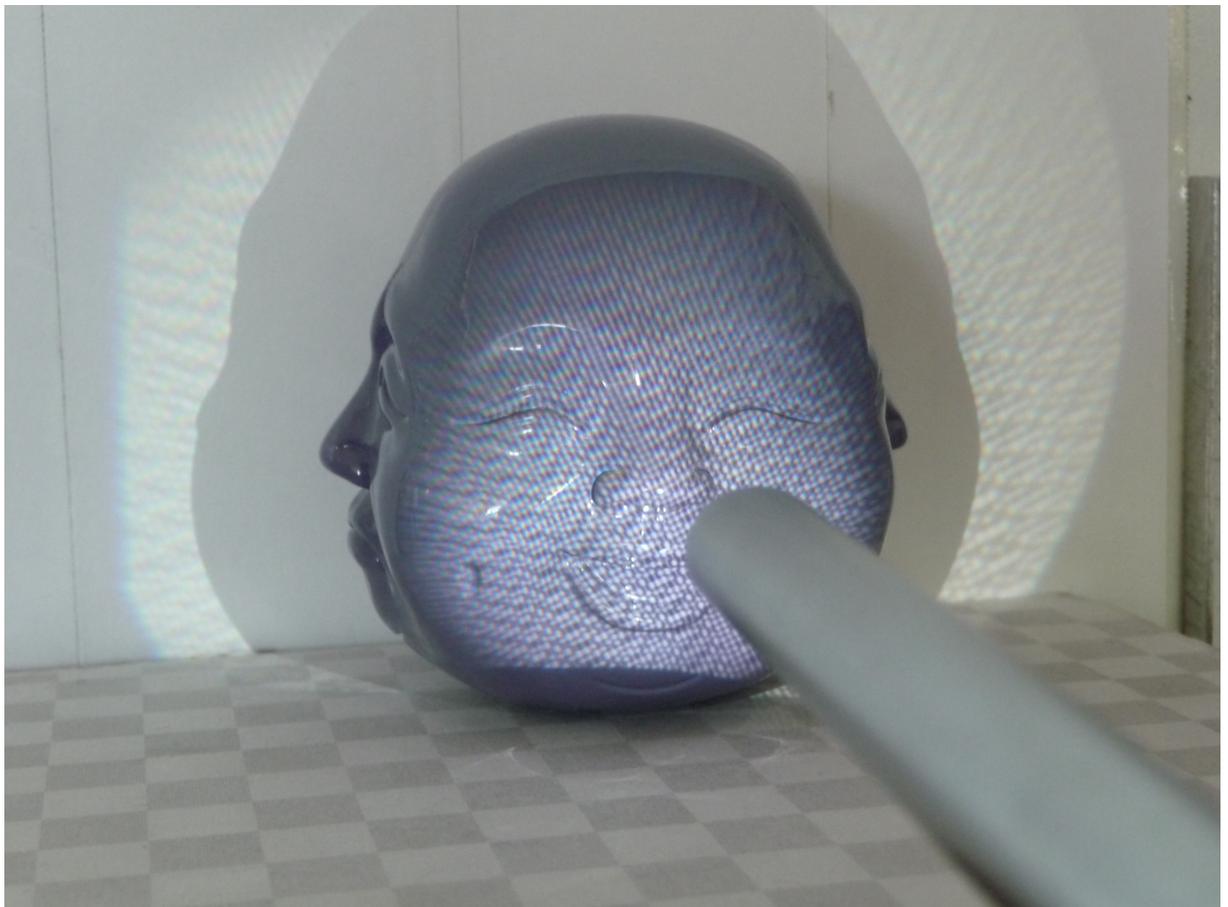
Dans ce chapitre, nous avons présenté la deuxième partie de la conception du motif de lumière structurée, celle qui consiste à lui donner une représentation physique. Nous avons ainsi étudié et motivé les choix effectués quant à l'organisation générale des primitives sur le motif ainsi qu'à leur modalité de discrimination qui exploite des formes différentes. Plusieurs solutions ont été évoquées quant au choix des primitives. Ces choix ont notamment été guidés par les résultats d'une étude sur la déformation des voisinages dans l'image rectifiée.

Enfin la chaîne des traitements que nous avons mise en place pour exploiter cette conception, dans des images bruitées, a été décrite, de la correction des distorsions des images à la mise en correspondance des primitives segmentées, en passant par la segmentation des franges et la création d'un graphe d'adjacence.

Une évaluation de cette conception et des traitements associés est donnée dans la deuxième partie du chapitre suivant, qui présente les reconstructions 3-D.

Chapitre 4

Reconstruction 3-D



Différents points de vue permettent, par un éclairage nouveau, de mieux mettre les choses en perspective.

Nous vivons dans l'ère de la 3-D, les jeux vidéos sont passés de la 2-D à la 3-D, le septième art propose de plus en plus de films dits 3-D, mais qui sont en réalité stéréo (2 points de vue), en conception industrielle, les modèles 3-D (CAO) sont omniprésents, même les derniers processeurs proposent une architecture 3-D de leurs transistors. Les applications de la 3-D sont très nombreuses. Cela est dû, en grande partie, au fait que l'être humain perçoit et se représente son environnement en trois dimensions. Les deux yeux forment, en effet, sur la rétine une image de la réalité selon deux points de vue qui permettent une appréciation de la distance de la scène, les cristallins jouant le rôle de lentilles optiques dans ce système stéréo. Cependant, les dispositifs de capture optique les plus utilisés, les caméras, sont majoritairement en deux dimensions, d'où le besoin d'effectuer une reconstruction 3-D.

La littérature traitant de la (les) reconstruction(s) 3-D est très vaste et toujours très active. Elle s'intéresse à différents types de **modèles de caméra** (projection orthographique/perspective-faible/ paraperspective/perspective). Elle s'intéresse à des reconstructions avec différents systèmes de capteurs : **mono, bi, tri, multi focaux**. Elle s'intéresse à la problématique de mise en correspondance en **vision passive** (composée de capteurs d'informations uniquement, le plus souvent des caméras) ou en **vision active** (composée également d'émetteurs d'informations, souvent des projecteurs). Elle s'intéresse également au **mouvement** des objets de la scène à partir de systèmes de capture/projection eux même **fixes** ou possédant une **mobilité**. En outre, elle s'intéresse aux reconstructions dans **différents espaces** (projectif, affine, métrique (ou similitude), euclidien) en exploitant les **invariants** présents dans chacun d'entre eux. Cette énumération n'est pas exhaustive.

4.1 Reconstruction 3D

Dans ce chapitre, l'objectif n'est pas de dresser un état de l'art des techniques de reconstruction 3-D. Pour cela, le lecteur est invité à consulter des ouvrages de référence comme ([Faugeras93], [Trucco98] etc...).

Nous nous contentons ici, depuis une description générale d'un système stéréo, de motiver et d'amener la discussion vers la méthode de reconstruction que nous avons choisie durant ce travail de thèse.

4.1.1 Modèles de caméra

Avant de considérer un système composé de plusieurs dispositifs optiques, il est nécessaire de s'intéresser, pour chacun des dispositifs à la façon dont l'image est formée à partir des points de l'espace 3-D.

A l'origine de la photographie, dans la deuxième moitié du 19^{ème} siècle, les premiers dispositifs de capture ne disposaient pas d'optiques. C'est le cas notamment du *sténopé*, qui est composé simplement d'un petit trou par lequel les rayons photoniques de la scène passent (Fig. 4.1.a). L'image inverse de la scène se créait alors sur une feuille photo-sensible placée dans une boîte totalement sombre de l'autre côté du trou. Mais le sténopé souffre d'un problème majeur : pour que l'image soit bien définie, il faut que le trou soit le plus petit possible, sinon des effets de flou apparaissent, dus à la diffraction. Mais plus le trou est petit, moins de photons passent et donc moins la luminosité de l'image obtenue est bonne.

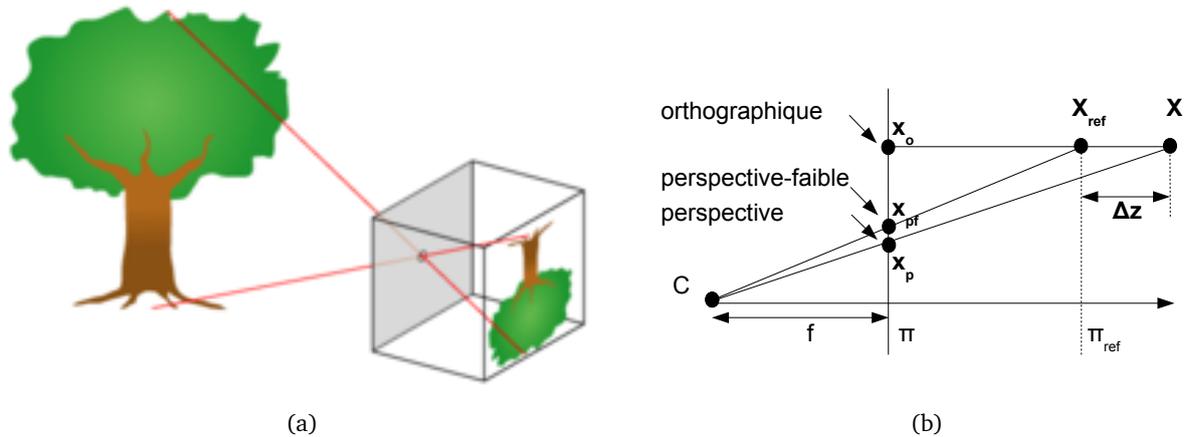


FIGURE 4.1 – (a) Image formée par un sténopé. (b) Différents modèles de projection

Plus tard, l'utilisation de lentilles optiques a apporté de grandes améliorations en permettant de capturer davantage de lumière par une focalisation des photons rendue possible par la courbure des lentilles. Ces lentilles, dans le modèle des lentilles minces, sont alors caractérisées par un centre de projection C (au centre de la lentille), situé à une distance focale f de l'image formée. f correspond à la distance à laquelle les rayons de la scène se focalisent en une image, lorsque que les rayons incidents (sur la lentille) sont parallèles. C'est le cas notamment lorsque la source de lumière (comme le soleil par exemple) est très éloignée de l'optique. Si l'utilisation de lentilles apporte un gain substantiel en termes de quantité d'énergie lumineuse dans l'image, la courbure de celles-ci a cependant pour effet de créer des distorsions géométriques qu'il faudra caractériser et corriger lors de l'utilisation du modèle de projection sténopé. Nous constaterons par ailleurs à la fin de ce chapitre et dans le chapitre suivant que les optiques utilisées pour les endoscopes sont nécessairement de taille réduite (quelques millimètres), l'objectif est de créer un système le plus miniaturisé possible (mini-invasif). L'effet de cette petite taille des optiques est alors, comme pour le sténopé initial, une réduction de la quantité de photons qui vont participer à la formation de l'image. En conséquence, une baisse non négligeable du rapport signal bruit est constatée par rapport à des optiques de plus grande taille.

Différents modèles de caméra peuvent être utilisés en vision par ordinateur. La figure 4.1.b en illustre trois. La projection perspective, issue du modèle sténopé est la plus utilisée aujourd'hui, car elle représente le mieux la réalité mais nécessite un étalonnage de l'optique. Il existe également, dans l'ordre d'approximation d'une projection perspective : la projection paraperspective, la projection perspective faible et la projection orthographique.

La projection d'un point 3-D \mathbf{X} de coordonnées (X, Y, Z) aura alors pour projection dans l'image, un point de coordonnées (x_p, y_p) dans le modèle perspectif (caméra projective), (x_{pf}, y_{pf}) dans le modèle perspectif-faible (caméra localement affine), (x_o, y_o) dans le modèle orthographique (caméra affine), tels que :

$$x_p = \frac{Xf}{Z}, y_p = \frac{Yf}{Z} \quad (4.1)$$

$$x_{pf} = \frac{Xf}{\bar{Z}}, y_{pf} = \frac{Yf}{\bar{Z}}, \quad (4.2)$$

avec \bar{Z} la profondeur moyenne des points de l'objet.

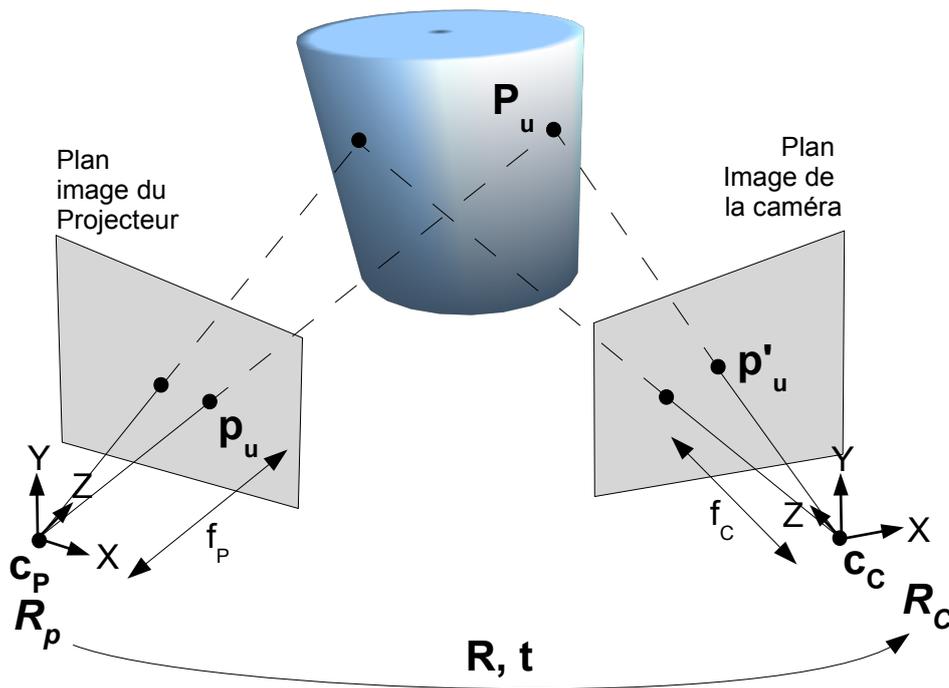


FIGURE 4.2 – Le modèle sténopé dans un système stéréo. C_p et C_c , les centres optiques du projecteur et de la caméra et f_p , f_c , les distances focales des dispositifs optiques respectifs. (R, t) , la transformation euclidienne reliant le repère du projecteur R_p au repère de la caméra R_c

$$x_o = X, y_o = Y \quad (4.3)$$

4.1.2 Stéréovision

La reconstruction 3-D consiste à retrouver, pour chaque point de la scène à reconstruire, les coordonnées du point étudié dans l'espace 3-D à partir des coordonnées 2-D des images de ces points dans les plans image des dispositifs de projection / capture utilisés. Dans le cas d'un système stéréo projecteur/caméra (bifocal) et d'un motif de lumière structurée projeté, ces coordonnées 2-D correspondent, pour le projecteur aux coordonnées du centre des primitives utilisées durant la conception du motif. Dans l'image de la caméra, ces coordonnées sont obtenues à l'issue de l'étape de segmentation des primitives. Avant de procéder à la reconstruction, il est nécessaire de mettre en correspondance les coordonnées obtenues pour les deux dispositifs composant le système stéréo. Par exemple, sur la figure 4.2, cela consiste à savoir que les points p_u , dans le plan image du projecteur, et p'_u , dans le plan image de la caméra, correspondent aux projections d'un même point P_u (de coordonnées initialement inconnues) dans l'espace 3-D.

Triangulation

Si la géométrie entre les deux dispositifs (paramètres extrinsèques) est connue, et que les caractéristiques modélisant chaque dispositif individuellement (paramètres intrinsèques) ont été estimés, il est possible de retracer en 3-D les deux rayons (C_p, p_u) et (C_c, p'_u) de chaque point

dans chacun des deux plans image et d'estimer leur intersection dans l'espace 3-D. Le point P_u est alors reconstruit par **triangulation** (voir Fig. 4.2). Dans la pratique les rayons ne s'intersectent pas dans le sens mathématique du terme. D'une part à cause du bruit présent dans l'image et des imprécisions de segmentation et d'autre part à cause de la précision, finie, du stockage des coordonnées dans les registres des mémoires informatiques. L'intersection 3-D retenue est alors celle qui minimise l'écart entre les deux rayons.

– **Paramètres intrinsèques.**

Chaque dispositif possède des paramètres intrinsèques pouvant être classés en deux groupes : les paramètres linéaires (distance focale (f), angle de non orthogonalité (s_θ) de la matrice d'éléments photo-sensibles de la caméra, taille métrique des colonnes / lignes de pixels (s_x, s_y)) et les paramètres non linéaires, qui sont issus des distorsions géométriques (radiales, tangentielles) créées par le mauvais alignement (distorsions tangentielles) sur l'axe optique et la courbure des lentilles (distorsions radiales), en particulier à leur périphérie.

Paramètres linéaires et projection perspective

Soit π , la matrice de projection (4×3) associée à la caméra, la projection perspective associée à cette caméra permet alors de trouver les coordonnées (u_d, v_d), distordus dans l'image par :

$$z_c \cdot \begin{pmatrix} u_d \\ v_d \\ 1 \end{pmatrix} = \pi \cdot x = \begin{bmatrix} f s_x & f s_\theta & u_c & 0 \\ 0 & f s_y & v_c & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \cdot \begin{pmatrix} x_c \\ y_c \\ z_c \\ 1 \end{pmatrix} \quad (4.4)$$

Les paramètres non linéaires permettant de trouver les coordonnées (u_{nd}, v_{nd}), non distordus, sont présentés dans la section étalonnage.

– **Paramètres extrinsèques.**

Les paramètres extrinsèques correspondent à la transformation géométrique euclidienne (R, t) reliant les repères du projecteur (R_p) et de la caméra (R_c). Avec R une matrice de rotation 3×3 et t le vecteur de translation 3×1

Changement de repère dans un espace euclidien

$$\begin{pmatrix} x_c \\ y_c \\ z_c \\ 1 \end{pmatrix} \equiv \begin{bmatrix} R & t \\ 0^T & 1 \end{bmatrix} \cdot \begin{pmatrix} x_p \\ y_p \\ z_p \\ 1 \end{pmatrix} \quad (4.5)$$

Pour utiliser un tel modèle, complet, il est cependant nécessaire d'estimer tous les paramètres intrinsèques et extrinsèques du système par une étape préalable d'étalonnage. La façon dont cet étalonnage est réalisé fait également l'objet d'une attention très importante dans la communauté scientifique.

4.1.3 Stratification

"Our results indicate that computer vision may have been slightly overdoing it in trying at all costs to obtain metric information from images." [Faugeras92].

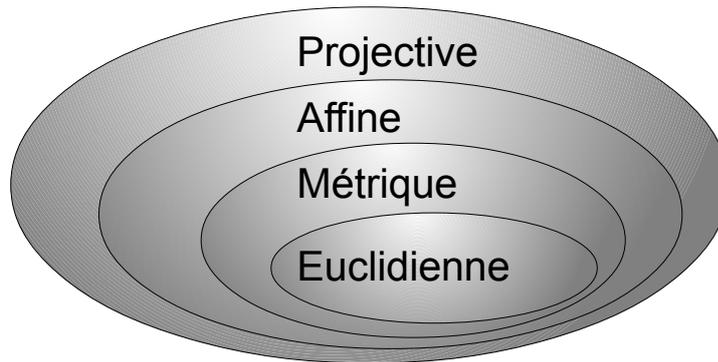


FIGURE 4.3 – Les différents espaces de transformations géométriques (stratification pour la reconstruction 3-D).

<i>Strate</i>	<i>DdL</i>	<i>MatriceTrans.</i>	<i>Distortion</i>	<i>Invariants</i>
Projective	15	P		Intersection, tangence de surface, birapport, géométrie épipolaire, points d'inflexion
Affine	12	$\begin{bmatrix} A & \\ 0^T & 1 \end{bmatrix}$		Parallélisme, plan à l'infini
Métrique (Similitude)	7	$\begin{bmatrix} sR & t \\ 0^T & 1 \end{bmatrix}$		Rapport de distances, angle, conique absolue
Euclidienne	6	$\begin{bmatrix} R & t \\ 0^T & 1 \end{bmatrix}$		Distance absolue

TABLE 4.1 – Caractéristiques des strates géométriques, tiré de [Pollefeys00]. La matrice P est une matrice 4×4 inversible. La matrice A est une matrice 3×4 . La matrice R , une matrice de rotation 3×3 , t un vecteur de translation 3-D et s un facteur d'échelle.

Dans [Faugeras92], Olivier Faugeras présente une méthode de reconstruction 3-D réalisée sans aucun étalonnage, à partir de simples correspondances de points. Cinq points non coplanaires sont choisis et définissent alors un système de coordonnées dans un espace projectif. Les reconstructions réalisées dans cette base sont alors définies à une transformation projective près et montrent qu'il n'est pas nécessaire d'étalonner l'ensemble du dispositif stéréo pour obtenir des reconstructions 3-D. Depuis, la littérature traitant des reconstructions à l'aide d'un système non

ou partiellement calibré s'est considérablement développée ([Mohr94], [Ach01], [Pollefeys98], [Bartoli01] par exemple) et s'est unifiée autour du concept de **stratification** (voir Fig. 4.3 et Tab. 4.1).

La stratification correspond au rapport hiérarchique des espaces dans lesquels sont effectuées les reconstructions, du plus général au plus spécifique (plus contraint). Ainsi, la strate projective contient la strate affine qui contient la strate métrique qui contient enfin la strate euclidienne. (Tab. 4.1). Il est possible de passer d'une strate plus générale à une strate plus spécifique en ajoutant des contraintes via l'exploitation des invariants présents dans la strate cible (Fig. 4.3). Un espace d'une strate possède ainsi les invariants des espaces des strates supérieures. Ce passage d'une strate à l'autre fait l'objet d'un nombre important de contributions ([Faugeras95],[Boufama93], [Hartley94], [Devernay96], [Fofi01], etc) et est un sujet trop vaste pour être développé ici plus en avant. Mais de façon générale, à partir d'une reconstruction projective, des informations sur la géométrie de la scène, ou sur le système stéréo, ou sur la géométrie du motif projeté sont exploitées pour contraindre les reconstructions et les représenter dans des strates inférieures.

Par exemple, dans [Fofi01] une reconstruction euclidienne est proposée avec un système non-calibré en exploitant la géométrie régulièrement grillagée d'un motif de lumière structurée qui illumine la scène. Le grillage permet en effet d'exploiter des contraintes euclidiennes (voir [Boufama93]) (orthogonalité, de parallélogramme, plans verticaux et horizontaux) (voir Fig. 4.4).

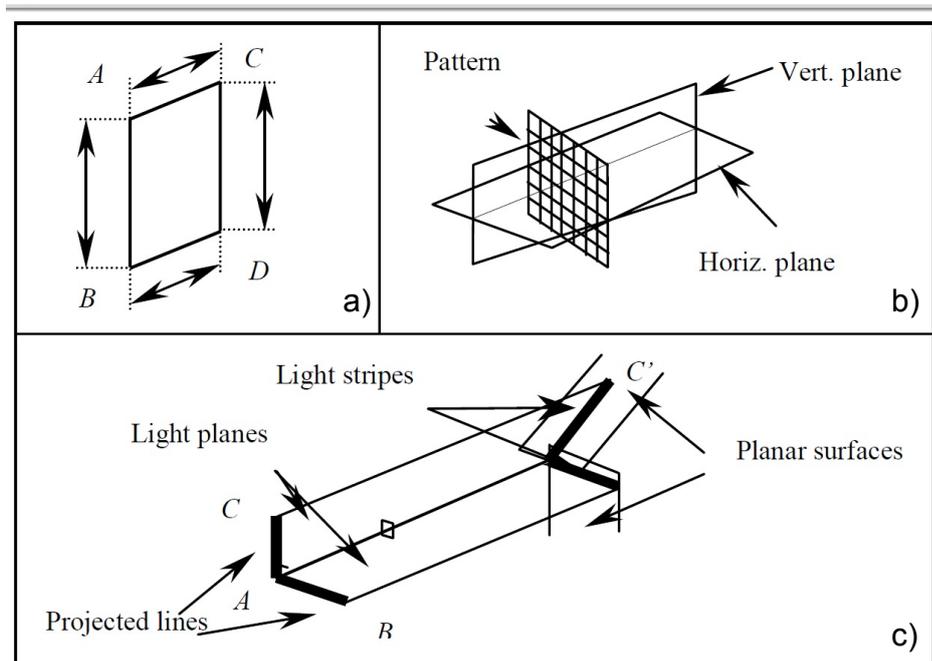


FIGURE 4.4 – Contraintes euclidiennes utilisées dans [Fofi01]. a) Contrainte du parallélogramme, b) plans vertical et horizontal, c) Orthogonalité.

Le choix d'un type de reconstruction va dépendre d'une part de l'application visée et ses besoins et d'autre part des informations disponibles sur la scène et la géométrie du système stéréo. Plusieurs contributions ont déjà traité de la reconstruction euclidienne pour un système

stéréo projecteur/caméra. Les approches présentées par [Chen87], [Zhou04] et [Albitar09b] peuvent notamment être utilisées.

La méthode que nous proposons ici donne lieu à des reconstructions dans l'espace projectif mais avec l'utilisation d'une contrainte de quasi-orthogonalité. Elle peut être adaptée à des applications ne nécessitant pas forcément une reconstruction euclidienne mais pour lesquelles cette quasi-orthogonalité permet de mieux établir (au moins visuellement), la relation entre la géométrie de la reconstruction 3-D et l'objet réel, ce qui n'est souvent pas le cas dans les reconstructions purement projectives ou purement affines, pour lesquelles l'objet peut apparaître très déformé par rapport à la réalité (voir Fig. 4.1). Ainsi cette reconstruction peut être exploitée par la suite aussi bien pour une simple visualisation que pour le guidage de robots ([Hespanha99], [Sebastián09]) ou pour des opérations de recalage de modèles 3-D utilisant des algorithmes adaptés à un espace non euclidien.

Avant de présenter cette reconstruction, il nous est nécessaire d'introduire la géométrie épipolaire dont elle va dépendre.

4.1.4 Matrice Fondamentale et géométrie épipolaire

Pour relier des points appariés entre deux vues, il existe deux situations, selon la connaissance ou non des paramètres intrinsèques :

- Si les paramètres intrinsèques sont inconnus, la matrice Fondamentale (F) met en relation les points d'une vue avec leur ligne de projection (épipolaire) dans l'autre vue. F est une matrice 3×3 de rang 2 et définie à un facteur d'échelle près. Elle met en relation les points des deux vues dans l'espace projectif selon la contrainte épipolaire :

$$x'^T F x \equiv 0. \quad (4.6)$$

où Fx correspond à la ligne épipolaire de x' . Les coordonnées des points en pixel peuvent directement être utilisés même si les pixels dans les deux images n'ont pas les mêmes dimensions.

- Si les paramètres intrinsèques sont connus (K pour une vue et K' pour l'autre), la matrice Essentielle (E) peut être utilisée pour exprimer la contrainte épipolaire avec la relation

$$E \equiv K'^T . F . K \quad (4.7)$$

Dans la figure 4.5, x représente les coordonnées homogènes du point dans le plan projecteur et x'_i les différentes projections de x dans le plan image de la caméra, en fonction de l'intersection (X_i) en 3-D avec les objets de la scène.

Ainsi les lignes épipolaires sont obtenues par la relation :

$$\begin{aligned} l'_x &= Fx \\ l_x &= F^T x \end{aligned} \quad (4.8)$$

Comme F est de rang 2, si le point x est confondu avec l'épipole (projection du centre optique d'une vue dans le plan image de l'autre), alors sa ligne épipolaire n'est pas définie. En effet, toutes les lignes épipolaires s'intersectent à l'épipole. Les coordonnées de l'épipole correspondent ainsi au noyau de F (e_p , noyau côté projecteur, e_c noyau côté caméra) et vérifient donc :

$$\begin{aligned} F^T e_c &= 0 \\ F e_p &= 0 \end{aligned} \quad (4.9)$$

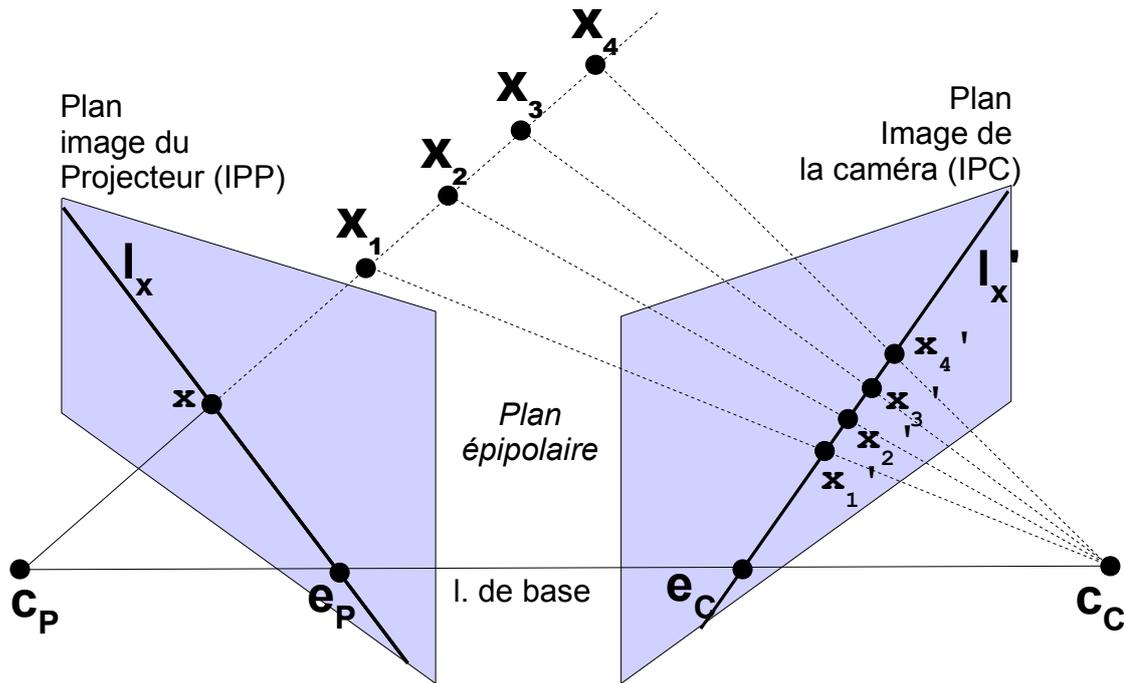


FIGURE 4.5 – La contrainte épipolaire. $x, x', X, C_c, et C_p$ sont coplanaires.

4.1.5 Une reconstruction 3-D projective avec contrainte de quasi-orthogonalité

Nous avons vu précédemment que la littérature propose des reconstructions 3-D euclidiennes avec un système non-calibré en exploitant la géométrie grillagée d'un motif de lumière structurée [Fofi01]. Cependant, dans notre cas, le motif pouvant être aligné sur la géométrie épipolaire, la grille de primitives se déforme et n'est plus orthogonale, à cause, entre autre, de l'angularité des lignes épipolaires. Cette approche ne peut donc pas être directement exploitée ici.

Principe de notre méthode

La géométrie apparente d'une reconstruction dans l'espace projectif dépend du choix des points qui définissent le système de coordonnées dans cet espace. Si ceux-ci sont choisis sur les arêtes d'un cube présent dans la scène 3-D, il est possible de créer un repère possédant des axes proches de l'orthogonalité et d'effectuer des reconstructions ayant une apparence proche de ce qui serait réalisé dans un espace métrique (similitude). Cependant il est nécessaire que la scène possède de tels objets, ce qui n'est pas courant dans des scènes intra-abdominales, à part, peut être, lorsque des chirurgiens y oublient des instruments après l'opération.

Par contre, il est plus facilement envisageable de trouver trois points (appariés) de la scène qui définissent un plan de référence (π_{Ref}), approximativement parallèle au plan image du projecteur (d'où le **quasi-orthogonal**). Grâce à un invariant projectif de la famille du birapport (présenté par la suite), généralement utilisé pour des tests de coplanarité, et grâce à la géométrie épipolaire, [Gros94] présente une méthode pour trouver la position des autres points de la scène dans le plan image de la caméra, tels qu'ils seraient projetés s'ils appartenaient à ce plan de référence. La figure 4.6 illustre ce principe. La distance (parallaxe) résultante entre cette position de coplanarité (pic' en violet dans la figure 4.6) et la position (pi') obtenue à l'issue

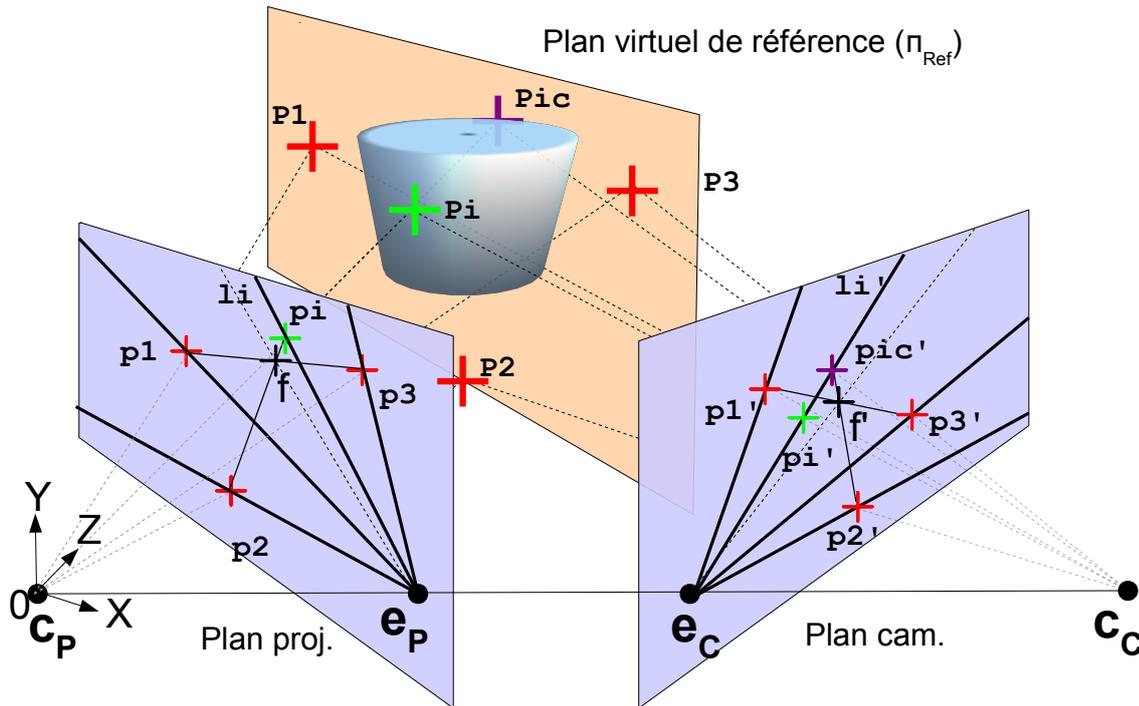


FIGURE 4.6 – Reconstruction proposée à l'aide de l'utilisation de 3 points (p_1, p_2, p_3 qui définissent un plan π_{Ref} approximativement parallèle au plan image du projecteur, créant une contrainte de quasi-orthogonalité)

de la segmentation des primitives, peut alors être directement interprétée comme la profondeur (voir [Sawhney94]), dans un espace projectif, de la primitive. L'origine du repère correspond alors à l'origine du repère du projecteur et les axes Y et X sont confondus avec ceux de ce repère.

Construction du point coplanaire au plan de référence

Nous proposons d'utiliser l'invariant projectif présenté dans [Gros94] (initialement dû à Roger Mohr et publié, sans tests, par [Faugeras92], dixit [Gros94]).

Cet invariant qui exploite la géométrie épipolaire (cf Fig. 4.6) peut être utilisé à la fois comme test de coplanarité pour des ensembles de points existants ou comme méthode de construction de points coplanaires, ce qui nous intéresse ici.

Contrairement à [Gros94] qui propose une méthode détournée pour retrouver la position du point pic' dans l'image de la caméra, nous proposons ici de retrouver ses coordonnées directement grâce à l'utilisation d'un invariant projectif CR dont l'expression est donnée par l'équation 4.10.

Le principe de construction du point pic' , projection dans le plan image de la caméra du point Pic appartenant au plan π_{Ref} est alors le suivant (voir Fig. 4.6) :

- Dans le plan projecteur, les droites (pip_2) et (p_1p_3) s'intersectent en un point f . L'invariant projectif CR (voir equ. 4.10) est alors calculé à partir des coordonnées de f , non de pi .
- Les coordonnées homogènes de f' dans le plan image de la caméra sont ensuite retrouvées à partir CR et des coordonnées des points p_1', p_2', p_3' et e_c dans le plan image de la

caméra.

- Les coordonnées de pic' sont alors données dans l'image de la caméra par l'intersection entre $p2'f'$ et la ligne épipolaire, li' associée à pi et définie par $Y = ax + b$.

Expression de l'invariant projectif CR dans [Gros94] :

Soient :

- $P1, P2, P3$ sont les 3 points qui définissent le plan de référence π_{Ref} dans l'espace 3-D. Ils se projettent dans le plan image du projecteur en $p1, p2, p3$ et dans le plan image de la caméra en $p1', p2', p3'$.
- pi est le point, dans le plan image du projecteur au centre de la primitive à reconstruire. Elle se projette en Pi dans l'espace 3-D, puis pi' dans le plan image caméra.
- pic' correspond à la projection dans le plan image de la caméra du point 3-D Pic , tel que $P1, P2, P3$, et Pic sont coplanaires, et que Pic se trouve sur la droite (C_p, Pi) .

Alors, on a :

$$CR = \frac{|epp1p3||epp2f|}{|epp1f||epp2p3|} = \frac{|ecp1'p3'||ecp2'f'|}{|ecp1'f'||ecp2'p3'|}, \quad (4.10)$$

avec

$$|P'P''P'''| = \det(X', X'', X''') = xy' + x'y'' + x''y - xy'' - x'y - x''y' \quad (4.11)$$

et $(x, y, 1)$, les coordonnées homogènes du point P dans l'image (projecteur / caméra).

Les projections des points sont supposées non confondues avec les épiholes et chacun des triplets de points $(p1, p3, e_p)$, $(p2, pi, e_p)$, $(p1', p3', e_c)$, et $(p2', pic', e_c)$ ne sont pas colinéaires.

Nous cherchons les coordonnées $(f'_x, f'_y, 1)$ du point f' . Dans [Gros94], le calcul n'était pas donné pour retrouver f' car une solution quadratique est seulement mentionnée. En appliquant la résolution suivante, nous n'avons pas rencontré un tel problème.

Solution :

Tout d'abord, la valeur numérique de CR est calculée à partir des coordonnées 2-D des points $p1, p2, p3, f$ et e_p , tous connus, dans le plan projecteur.

On note a et b , respectivement la pente et l'ordonnée à l'origine de la ligne épipolaire lf' passant par f' . lf' est obtenue depuis f et la matrice Fondamentale avec l'équation 4.8.

Comme l'expression de CR est symétrique entre les points p (dans le plan projecteur) et p' (dans le plan caméra), dans la suite, pour ne pas alourdir les notations, nous noterons momentanément p , au lieu de p' , et f , au lieu de f' , les points dans le plan caméra.

Ainsi on a $f_y = af_x + b$. Cette contrainte épipolaire peut être utilisée pour trouver la solution.

Ce qui donne, pour le numérateur de CR :

$$\begin{aligned} \text{Numérateur} &= |ecp1p3|. |ecp2f| \\ &= |ecp1p3|. [ecxp2y + p2xfy + fxe_{cy} - ecxfy - p2xe_{cy} - fxp2y] \\ &= |ecp1p3|. [fx(e_{cy} - p2y) + ecxp2y + fy(p2x - ecx) - p2xe_{cy}] \\ &= |ecp1p3|. [fx(e_{cy} - p2y) + ecxp2y + (af_x + b)(p2x - ecx) - p2xe_{cy}] \\ &= |ecp1p3|. [fx(e_{cy} - p2y + ap2x - aecx) + ecxp2y - p2xe_{cy} + b(p2x - ecx)] \end{aligned} \quad (4.12)$$

On note $A2 = (e_{cy} - p2_y + ap2_x - ae_{cx})$ et $B2 = e_{cx}p2_y - p2_x e_{cy} + b(p2_x - e_{cx})$

De la même façon pour le dénominateur de CR :

$$\begin{aligned} \text{Dénominateur} &= |e_cp2p3|. |e_cp1f| \\ &= |e_cp2p3|. [f_x(e_{cy} - p1_y + ap1_x - ae_{cx}) + e_{cx}p1_y - p1_x e_{cy} + b(p1_x - e_{cx})] \end{aligned} \quad (4.13)$$

On note $A1 = (e_y - p1_y + ap1_x - ae_x)$ et $B1 = e_x p1_y - p1_x e_y + b(p1_x - e_x)$

Ce qui donne :

$$CR = \frac{|e_cp1p3|. (f_x A2 + B2)}{|e_cp2p3|. (f_x A1 + B1)} \quad (4.14)$$

$$f_x \cdot A2 \cdot |e_cp1p3| + B2 \cdot |e_cp1p3| = f_x \cdot CR \cdot A1 \cdot |e_cp2p3| + CR \cdot B1 \cdot |e_cp2p3| \quad (4.15)$$

$$f_x \cdot A2 \cdot |e_cp1p3| - f_x \cdot CR \cdot A1 \cdot |e_cp2p3| = CR \cdot B1 \cdot |e_cp2p3| - B2 \cdot |e_cp1p3| \quad (4.16)$$

et finalement, en reprenant la notation initiale ($p \rightarrow p'$),

$$f'_x = \frac{CR \cdot B1 \cdot |e_cp2'p3'| - B2 \cdot |e_cp1'p3'|}{A2 \cdot |e_cp1'p3'| - CR \cdot A1 \cdot |e_cp2'p3'|} \quad (4.17)$$

Ainsi f'_x dépend des coordonnées homogènes des points $p1', p2', p3'$, de CR et de a, b , qui sont tous connus. f'_y est donnée par $f'_y = af'_x + b$. Notons bien qu'aucune approximation n'a été faite pour trouver f'_x et f'_y . Une autre solution, utilisant la matrice Fondamentale, consiste à calculer la ligne épipolaire lf' passant par f' . f'_x et f'_y sont alors donnés par l'intersection entre les lignes lf' et la droite $(p1', p3')$.

Enfin, pour chaque point, pi , à reconstruire dans le référentiel du projecteur, sa profondeur est donnée, à un facteur d'échelle près, par la distance $[pi', pic']$. X et Y correspondent alors à pi_x / Z_{piRef} et pi_y / Z_{piRef} . Z_{piRef} est un facteur d'échelle caractérisant l'éloignement du plan de référence par rapport à l'origine du repère. Dans le cas où $Z_{piRef} = 1$, cela correspondra à une projection orthographique.

Nous verrons dans les diverses reconstructions présentées à la fin de ce chapitre que projection non perspective est surtout visible quand les points de la scène ne sont pas suffisamment éloignés de l'origine du repère du projecteur par rapport à la profondeur des points choisis pour établir le plan de référence.

4.2 Estimation de la géométrie épipolaire et des distorsions géométriques

Parmi les motifs proposés, ceux utilisés pour les reconstructions ("motifs fins") sont alignés sur la **géométrie épipolaire** pour le codage (chapitre 2) et la disposition des primitives visuelles (chapitre 3). De plus, la reconstruction 3-D proposée (chapitre 4) fait également usage de celle-ci. Dans cette section nous présentons ainsi son estimation à travers l'évaluation de la matrice Fondamentale. Pour qu'elle puisse être estimée avec un minimum d'erreurs, il est également nécessaire de caractériser et de corriger les **distorsions géométriques** créées par les optiques de la caméra et du projecteur, qui ne sont pas prises en compte dans le modèle sténopé. Enfin, l'utilisation d'optiques endoscopiques a pour effet de réduire l'image effective à un cercle caractéristique que nous appelons **cercle endoscopiques**. La partie utile de l'image se trouvant à l'intérieur de ce cercle, il est également nécessaire de le localiser et d'estimer son rayon pour identifier la zone de l'image à traiter. Dans la suite de cette section, nous présentons donc successivement l'estimation du *cercle endoscopique*, des distorsions géométriques de la caméra, du projecteur, et enfin de la matrice Fondamentale.

4.2.1 Estimation du cercle endoscopique

L'estimation du *cercle endoscopique* définira la région d'intérêt qui sera utilisée pour tous les autres traitements, aussi bien pour l'estimation des distorsions, de la géométrie épipolaire, que les traitements menant aux reconstructions 3-D. C'est ainsi la première tâche à effectuer. Le cercle est caractérisé par trois paramètres : son centre, de coordonnées (C_x, C_y) et son rayon, C_R , tous exprimés en pixels dans l'image de la caméra.

Principe de l'estimation :

La zone extérieure au cercle endoscopique n'étant pas ou très faiblement éclairée, celle-ci est très sombre et contient surtout du bruit, à l'inverse de l'intérieur du cercle qui contient toute la lumière issue de la scène. Ceci est visible sur les différentes images endoscopiques présentées dans ce rapport. Il suffit d'éclairer fortement la scène pour obtenir suffisamment de lumière à l'intérieur du cercle. Ainsi l'estimation du cercle est basée sur la maximisation d'une fonction de coût qui quantifie le contraste entre l'intérieur et l'extérieur du cercle. La fonction d'optimisation associée procède par descente de gradient. Comme les deux zones considérées sont de tailles très importantes (environ la moitié de l'image) et que le contraste moyen entre l'intérieur et l'extérieur du cercle est significatif, le bruit ne crée pas d'apparition de minima locaux et l'optimisation converge sans problème, le pas du gradient étant choisi par affinements successifs. Comme conditions initiales, nous prenons (C_x, C_y) , le centre de l'image et C_R , la moitié de la hauteur de l'image. Nous avons effectué une estimation jusqu'à une précision de 0,1 pixel, ce qui est amplement suffisant pour ce type de données. Les paramètres trouvés pour notre prototype endoscopique (voir chapitre 5) sont : $C_x = 8.956e^{+02}$ px, $C_y = 4.562e^{+02}$ px, $C_R = 5.959e^{+02}$ px.

4.2.2 Distorsions géométriques de la caméra et du projecteur

Les distorsions géométriques sont présentes, de façon significative, aussi bien dans l'image de la caméra, que dans l'image du projecteur du système miniaturisé. En effet, la création du prototype endoscopique a mené à l'ajout d'optiques d'interface et d'optiques endoscopiques au projecteur, créant ainsi des distorsions inexistantes initialement (voir chapitre 5, développement

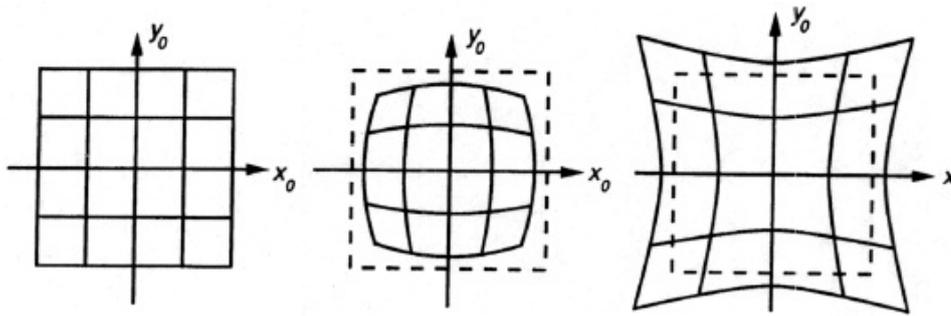


FIGURE 4.7 – Distorsions radiales d’une grille. De gauche à droite, grille non distordue, distorsion en barillet, distorsion en coussinet.

matériel). Pour que le modèle sténopé s’applique et que le motif puisse être aligné sur la géométrie épipolaire, il nous est donc nécessaire de corriger ces distorsions.

Les algorithmes de correction de distorsions peuvent être catégorisés d’une part selon le modèle mathématique qui les modélise, d’autre part, par selon la façon dont les paramètres de ce modèle sont estimés et enfin par la manière dont ces paramètres estimés sont utilisés (inversement du modèle, carte de correspondance) pour corriger les distorsions.

Modèles de distorsion

Différents modèles prenant en compte un nombre plus ou moins importants de types de distorsions sont présents dans la littérature. Ils sont généralement représentés sous forme de polynômes d’ordre croissant issus d’une expansion de Taylor qui approxime leur modélisation "parfaite". C’est le cas notamment des modèles de Tsai (1987) [Tsai87] et de Brown (1966) [Brown66], par exemple. Les coefficients de ces polynômes sont alors estimés pour modéliser au mieux les distorsions. En général seulement les polynômes d’ordre faible sont considérés, car l’expérience montre (voir [Tsai87]) qu’en pratique le premier terme est souvent suffisant. Les termes d’ordre plus élevés sont en effet trop sensibles au bruit et peuvent causer des instabilités numériques. Nous notons cependant qu’il existe d’autres modèles plus complexes, comme celui de Frits Zernike (1930), basé sur la représentation de fronts d’onde grâce aux *polynômes de Zernike* (voir [Bezdidko74]).

Nous avons exploité le modèle de Brown, ancien, mais toujours largement utilisé et bien connu dans la littérature. Ce modèle nous semble adapté car, contrairement à celui de [Tsai87], il permet de caractériser à la fois les distorsions radiales et tangentielles qui sont, comme nous le verrons, toutes les deux significatives dans le système stéréo endoscopique.

Le modèle de Brown

Dans le modèle de Brown, pour tout point de l’image distordue de coordonnées (u_d, v_d) , les coordonnées non distordues correspondantes (u, v) sont obtenues grâce aux relations suivantes :

$$\begin{aligned} u &= u_d + (u_d - u_c)(K_1 r^2 + K_2 r^4 + \dots) + (P_1(r^2 + 2(u_d - u_c)2) + 2P_2(u_d - u_c)(v_d - v_c))(\dots) \\ v &= v_d + (v_d - v_c)(K_1 r^2 + K_2 r^4 + \dots) + (P_2(r^2 + 2(v_d - v_c)2) + 2P_1(u_d - u_c)(v_d - v_c))(\dots), \end{aligned} \quad (4.18)$$

avec

$$\begin{aligned}
(u_c, v_c) & : \text{ le centre de distortion} \\
K_n & : \text{ le } n^{\text{ieme}} \text{ coefficient de distorsion radiale} \\
P_n & : \text{ le } n^{\text{ieme}} \text{ coefficient de distorsion tangentielle} \\
r & = \sqrt{(u_d - u_c)^2 + (v_d - v_c)^2} \\
\dots & : \text{ une série infinie}
\end{aligned} \tag{4.19}$$

Estimation des paramètres du modèle

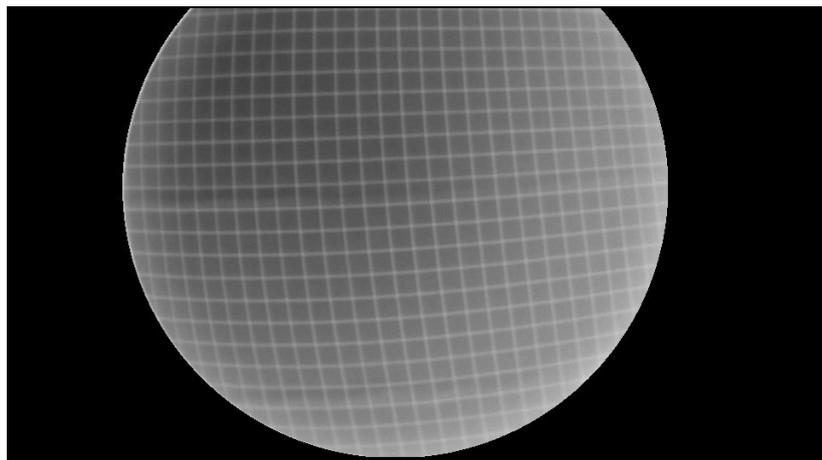
Il existe au moins deux grandes familles de méthodes pour estimer ces paramètres :

- **Estimation globale.** Les paramètres du modèle sont intégrés dans un processus d'estimation globale qui inclut également les paramètres extrinsèques et intrinsèques du système stéréo. L'optimisation s'effectue alors en utilisant des mires de calibration dont les dimensions sont connues en minimisant l'erreur de reprojection calculée à partir du modèle de projection présenté à la section précédente (equ. 4.4 et 4.5) et des équations du modèle de distorsion utilisé. C'est le cas dans [Tsai87], [Weng92], [Salvi98] et [Zhang00], par exemple.
- **Le principe *Plumb Line* et autres méthodes.** Si les paramètres intrinsèques et extrinsèques ne sont pas connus et qu'aucune information métrique n'est disponible sur la scène, le principe *Plumb Line* peut être utilisé. Celui-ci repose sur le fait qu'une ligne droite dans la scène 3-D doit se projeter en une ligne droite dans le plan image, ce qui n'est pas le cas lorsque des distorsions sont présentes. Pour se faire, les projections de lignes sont détectées dans l'image et l'optimisation se fait en maximisant une fonction de coût qui exprime la rectitude de ces lignes. De nombreuses contributions ([Devernay01], [Thormaehlen03], [Song10a]) traitent de cette problématique, dont [Sagawa05] où un motif de lumière structurée codée (bandes blanches et noires utilisant un code de Gray) est affiché sur un écran plat. D'autres méthodes non métriques existent, comme celle de [Kang00] qui est basée sur l'utilisation de contours déformables (*snakes*) et [Stein97] où des correspondances de points sont exploitées.

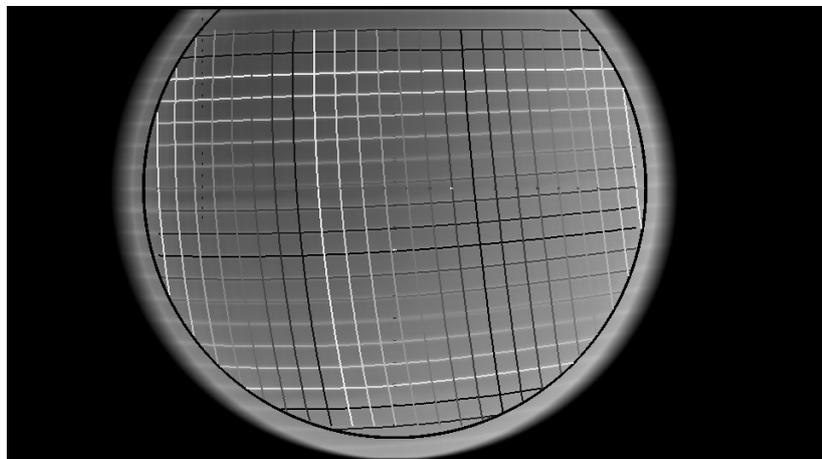
Nous proposons de corriger les distorsions de caméra et du projecteur avec le système non étalonné en utilisant le principe *Plumb-Line*. Mais contrairement à [Salvi98], où seul le paramètre K_1 est estimé lors d'une optimisation globale et où un couplage entre les distorsions du projecteur et de la caméra est envisageable, créant potentiellement des minima locaux, nous proposons d'estimer ces distorsions pour le projecteur et pour la caméra de façon indépendante.

Estimation des distorsions de l'image de la caméra

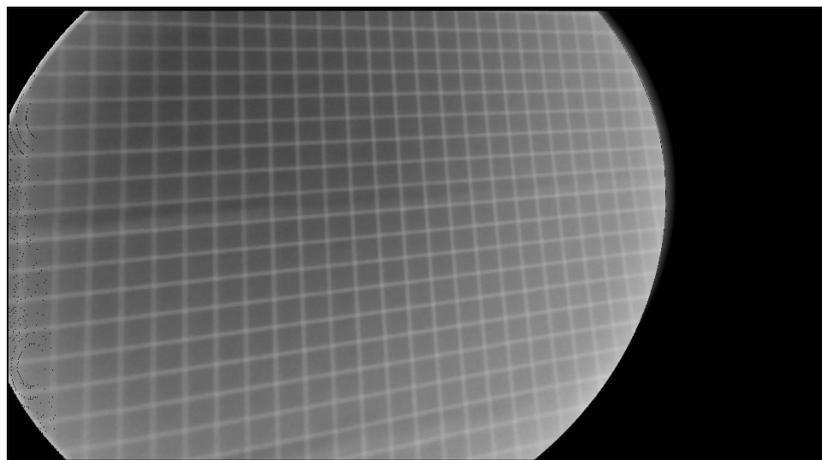
Les paramètres intrinsèques et extrinsèques n'étant pas estimés à ce stade, nous utilisons le principe *plumb line* pour corriger tout d'abord les distorsions de la caméra uniquement en capturant une grille imprimée sur un plan et donc non distordue dans l'espace 3D. Une acquisition est effectuée (Fig. 4.8.a) et les lignes et colonnes de cette grille sont alors segmentées, labellisées (Fig. 4.8.b), et l'optimisation est effectuée en minimisant la fonction de coût qui exprime la rectitude des points associés à une même labellisation de ligne. Une carte de correspondance est calculée suivant la méthode présentée plus loin et les distorsions de la caméra sont corrigées (Fig. 4.8.c).



(a)



(b)



(c)

FIGURE 4.8 – Correction des distorsions de la caméra avec l’algorithme *Plumb-Line*. a) La grille initialement distordue. b) Labellisation des lignes et colonnes de la grille. c) Correction des distorsions radiales et tangentielles, ces dernières provoquant une asymétrie caractéristique par rapport au centre de distorsions.

Estimation des distorsions de l'image du projecteur

Dans la plupart des situations, il n'est pas nécessaire d'estimer les paramètres de distorsion du projecteur. D'une part, parce que bien souvent, les projecteurs utilisés possèdent déjà une correction interne effectuée par le fabricant, qui permet de projeter des présentations non distordues et donc les distorsions résiduelles sont négligeables. Une deuxième raison est qu'il existe des méthodes d'étalonnage du système stéréo qui caractérisent, indépendamment, chaque rayon (correspondant à la projection du centre d'une primitive projetée) du projecteur, comme dans [Albitar09b] et donc il n'est pas nécessaire, dans ce cas, d'effectuer de correction globale des distorsions. Dans notre cas, le prototype créé a demandé l'ajout d'optiques d'interface aux optiques endoscopiques, ce qui crée d'importantes distorsions. D'autre part, notre conception (chapitres 2 et 3) dépend de la géométrie épipolaire et nécessite une estimation précise de celle-ci pour que le modèle sténopé s'applique. Il nous est donc nécessaire de corriger ces distorsions autant que possible.

Une fois les distorsions de la caméra corrigées, une grille est cette fois-ci projetée par le projecteur du système stéréo (Fig. 4.15.a) sur un plan dans l'espace 3-D (Fig. 4.15.b). Les lignes incurvées dans l'espace 3-D et donc dans l'image acquise par la caméra (après correction des distorsions de celles-ci) sont donc dues aux distorsions créées par l'optique du projecteur, toujours selon le principe *Plumb-Line*. Ces distorsions ne peuvent pas être corrigées directement avec une seule acquisition, comme pour la caméra, car les paramètres estimés dans le plan image de la caméra n'ont pas de lien direct (sans connaissance des paramètres intrinsèques et extrinsèques du dispositif) avec le plan image du projecteur. Pour contourner ce problème, nous avons développé et implémenté une nouvelle méthode d'optimisation itérative, procédant par affinements successifs. Celle-ci repose sur le principe suivant : à chaque appel de la fonction de coût qui doit évaluer la mesure de rectitude avec des valeurs de paramètres de distorsions qui doivent être testés, les opérations suivantes sont réalisées :

1. La fonction de coût est appelée par la méthode d'optimisation pour une nouvelle évaluation de la rectitude correspondant à un nouveau jeu de paramètres de distorsion du projecteur à évaluer.
2. La grille est distordue dans le plan du projecteur en appliquant les équations du modèle de Brown avec les valeurs des paramètres ($C_X, C_Y, K_1, K_2, P_1, P_2$) à tester.
3. Une acquisition de cette grille est réalisée : pour cela, l'acquisition doit être synchronisée avec la dernière grille projetée, sinon la rectitude d'une mauvaise grille sera évaluée et l'optimisation ne convergera pas vers la rectitude.
4. La segmentation, la labellisation et la mesure de rectitude des lignes de cette grille acquise sont effectuées.
5. La fonction de coût retourne l'erreur de rectitude à la méthode d'optimisation qui modifie les paramètres à estimer et passe à l'itération suivante.

Une fois que l'optimisation est terminée, après avoir atteint une erreur de rectitude minimum admissible, le motif affiché dans le plan projecteur est distordu (voir Fig. 4.15.c) avec le jeu de paramètres qui a fourni la meilleure rectitude, de façon à compenser a priori les distorsions qui vont s'opérer après sa projection dans l'espace 3-D, de telle sorte que les lignes et les colonnes de la grille soient bien rectilignes dans l'espace 3-D (Fig. 4.15.d) et donc également dans le plan image de la caméra.

Comme avec cette méthode, les distorsions de la caméra et du projecteur sont évaluées indépendamment, moins de variables interviennent simultanément qu'avec une méthode d'es-

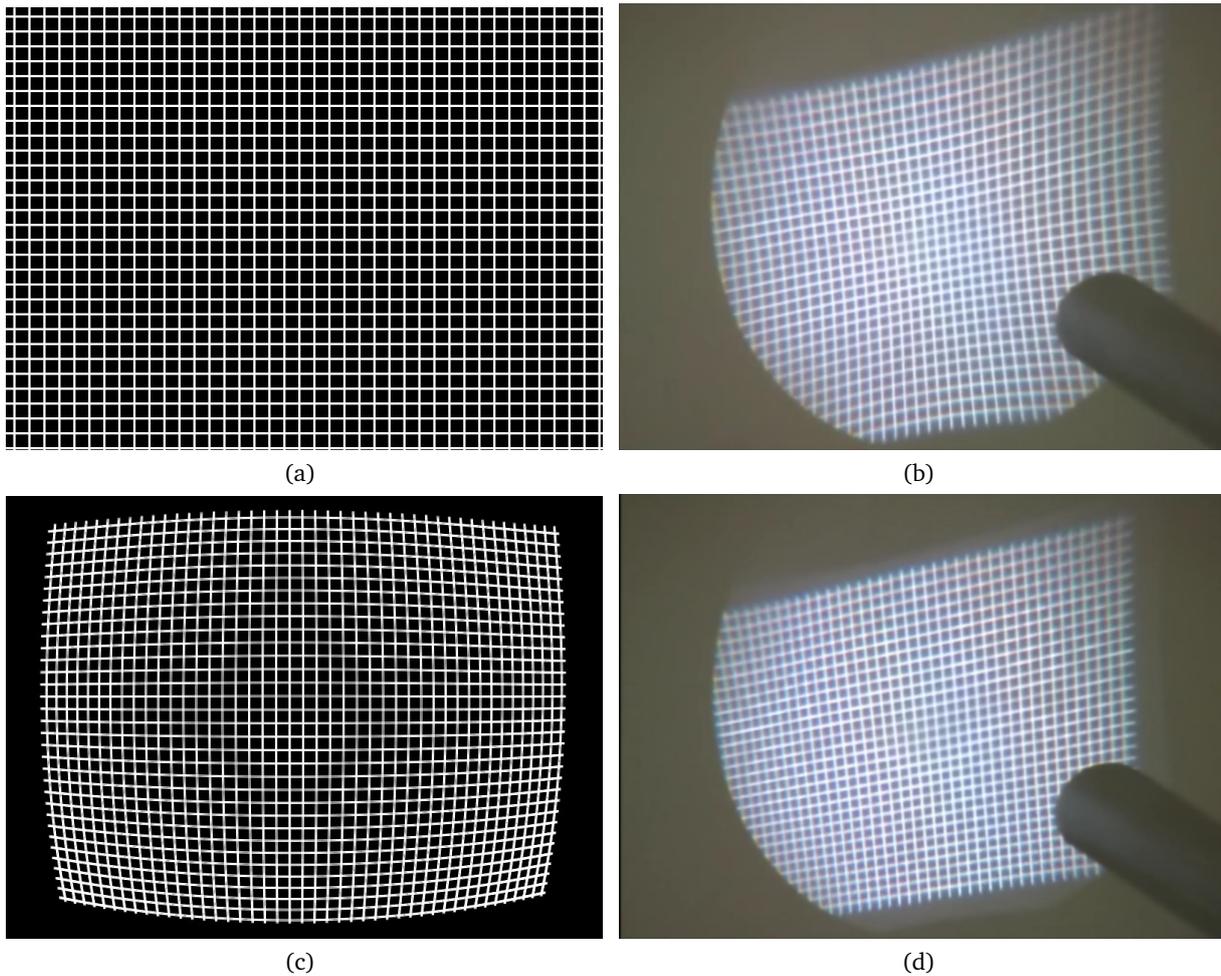


FIGURE 4.9 – (a) Les distorsions du projecteur ne sont pas corrigées dans le plan projecteur. (b) Elles sont visibles dans l'espace 3-D après projection d'une grille. (c) Les distorsions du projecteur sont corrigées a priori dans le plan projecteur par une distorsion compensatoire créée en amont. (d) Dans l'espace 3-D la grille est alors rectiligne et les distorsions ne sont plus présentes.

timisation globale. Les problèmes potentiels de couplage sont a priori évités, mais cela reste à démontrer.

Cette méthode d'optimisation demande plusieurs minutes de calcul, car la fonction la plus interne, la fonction de coût, requiert une nouvelle projection synchronisée avec une nouvelle acquisition à chaque appel et cette synchronisation, pour être réalisée de façon certaine, peut demander plusieurs dizaines, voire centaines de millisecondes, pour que la bonne image soit traitée. Une synchronisation matérielle permettrait sans doute de réduire ce temps de latence.

4.2.3 Résultats de l'étalonnage partiel

Distorsions :

Les résultats des estimations des paramètres de distorsion sont reportés sur le tableau 4.3. Pour chacun des dispositifs, les résultats sont reportés comme suit :

- La première ligne indique la valeur estimée du paramètre.

Paramètres	C_X	C_Y	K_1	K_2	P_1	P_2
Caméra	$1.125e^{+3}$	$3.119e^{+2}$	$1.654e^{-7}$	$2.931e^{-13}$	$-3.652e^{-5}$	$9.239e^{-6}$
Dép. indep. (px)			67.59	104.33	44.18	7.87
Dép. rad. tang.			radial	171.93	tang	45.82
Dép. moyen						196.97
Projecteur	$3.800e^{+2}$	$3.450e^{+2}$	$-9.200e^{-7}$	$9.600e^{-13}$	0.0	$-5.333e^{-5}$
Dép. indep. (px)			30.93	5.09	0.0	11.63
Dép. rad. tang.			radial	25.84	tang	11.63
Dép. moyen						29.98

TABLE 4.2 – Paramètres de distorsion radiales et tangentielles estimés pour la caméra (1920×1080) et le projecteur (800×600). Déplacement moyen (en pixels) induit individuellement (2ème ligne), puis déplacement groupé (K_1 et K_2 d'une part et P_1 et P_2 d'autre part (3ème ligne), et enfin le déplacement moyen global (Dev. moyenne), (4ème ligne).

- La deuxième ligne indique le déplacement moyen, en pixels, induit par ce paramètre seul (avec les C_X et C_Y estimés).
- La troisième ligne indique le déplacement moyen estimé conjointement par les deux paramètres de distorsion radiale (K_1 et K_2) d'une part et d'autre part par les deux paramètres de distorsion tangentielle (P_1 et P_2).
- Enfin, sur la dernière ligne est indiqué le déplacement moyen global créé par l'ensemble des paramètres estimés.

On remarque, d'une part l'importance des distorsions tangentielles dues à l'optique de la caméra (orientée de 30 par rapport à l'axe optique), créant un déplacement moyen de 45.82 pixels dans l'image visible. D'autre part, pour la caméra, les distorsions dues au paramètre K_2 (déviations moyennes de 104.33 pixels), sont supérieures à celles relatives au paramètre K_1 (déviations moyennes de 67.59 pixels), ce qui est inhabituel. En effet, comme noté précédemment, c'est généralement le paramètre de plus faible ordre qui permet de modéliser la majorité des distorsions.

Finalement, on remarque que les distorsions du projecteur sont non négligeables, avec un déplacement moyen de 29.98 pixels (calculé dans l'image de la caméra).

4.2.4 Création de la carte de correspondance

A l'issue de l'estimation des paramètres de distorsion, nous avons obtenu les coordonnées non distordues des points à partir des coordonnées distordues (*correspondance-avant* ou *forward mapping*). Mais l'inverse est ici désiré (*correspondance-arrière* ou *backward mapping*), car on souhaite obtenir la position des pixels dans l'image distordue à partir de celles des pixels dans l'image non distordue.

Un problème se pose alors. En effet, le modèle de brown présente des termes d'ordre élevé et n'est, de façon générale, pas inversible. Une solution consiste à considérer uniquement le terme d'ordre le plus faible : souvent K_1 est seulement utilisé car la distorsion radiale induite est celle qui produit généralement le plus de non linéarités géométriques et l'équation est alors inversible. Ainsi, la plupart du temps, dans la littérature, les autres paramètres sont négligés.

En ce qui nous concerne, toute la conception dépend de la géométrie épipolaire et nous avons donc besoin que le modèle sténopé puisse s'appliquer avec le minimum d'erreurs possibles. Or les

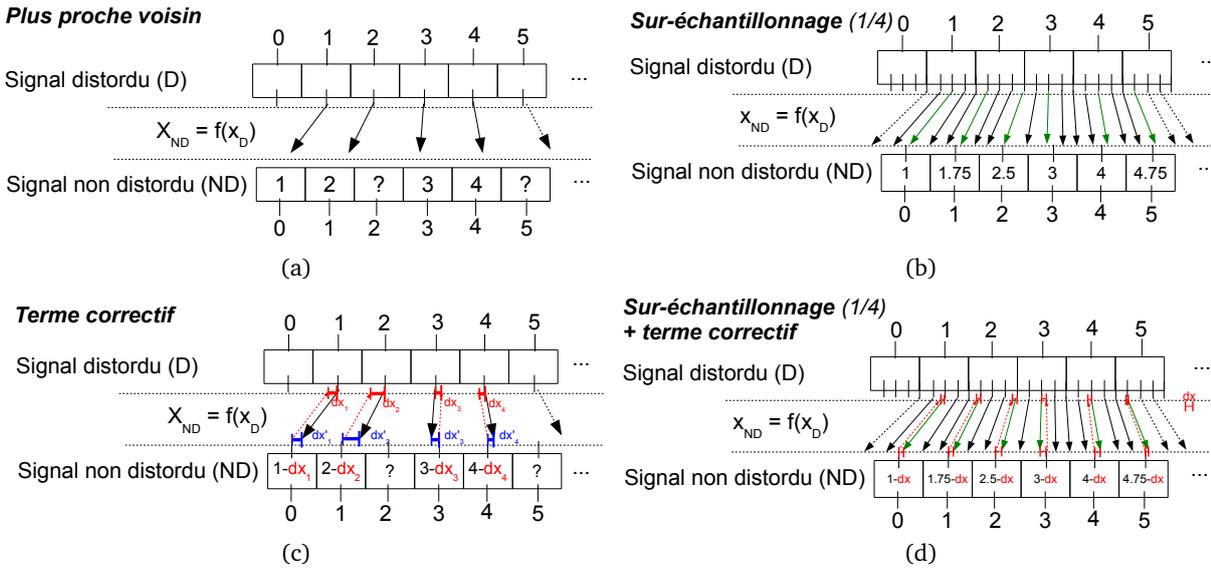


FIGURE 4.10 – Artefacts visibles dus aux pixels vides après la correction des distorsions avec une carte de correspondance réalisée à partir d’un échantillonnage de (a) 1 point par pixel, (b) 2 points par pixel, (c) 4 points par pixel, il ne reste plus de pixels vides

optiques endoscopiques peuvent contenir des angularités avec l’axe optique (30° dans notre cas). De plus notre prototype, conçu artisanalement, n’a pas forcément des optiques parfaitement alignées, il a donc été nécessaire (comme nous l’avons évoqué plus tôt), de prendre en compte les distorsions tangentielles (paramètres P_1 et P_2) et d’avantages de paramètres que K_1 pour les distorsions radiales, ainsi K_2 a également été estimée.

Comme les équations du modèle ne peuvent alors être inversées de façon unique, nous proposons d’établir une carte de *correspondance-arrière* (map_x et map_y) à partir des *correspondances-avant* directement. Avec une telle carte, les coordonnées dans l’image source, src , sont obtenues à partir de celles de l’image destination, dst , par la relation suivante :

$$dst(x, y) = src(map_x(x, y), map_y(x, y)), \quad (4.20)$$

où map_x , et map_y sont les cartes de correspondance pour, respectivement, les colonnes et les lignes de pixels.

La problématique de cette correspondance arrière est traitée notamment dans [Wolberg94]. Dans [Vass03], pour corriger le problème des pixels vides qui se pose alors, la valeur du plus proche voisin non vide est utilisée, mais cette méthode, bien que rapide, crée des effets de recouvrement. Le temps d’exécution nécessaire à la création de cette carte de correspondance inverse n’étant pas prioritaire dans notre cas, nous présentons ci-après d’autres méthodes qui fournissent une précision accrue.

La carte de correspondance que nous proposons de créer repose sur trois principes : la *correspondance-avant*, le *sur-échantillonnage* et un *terme de correction*, d_x , qui exploite une interpolation bilinéaire (Fig. 4.10). Bien que cette carte soit adaptée à des images, pour simplifier les explications, nous considérons, dans la suite, une correspondance 1-D, et donc une interpolation linéaire.

Avec les paramètres estimés précédemment pour la caméra, une *correspondance-avant* initiale (par plus proche voisin) mène à une erreur théorique maximale de 0.5 pixels et une erreur moyenne calculée de 0.37 pixels. De plus, à cause de l'augmentation de la surface de l'image, certains pixels n'obtiennent pas de correspondance, ce qui crée des artefacts (Fig. 4.10).

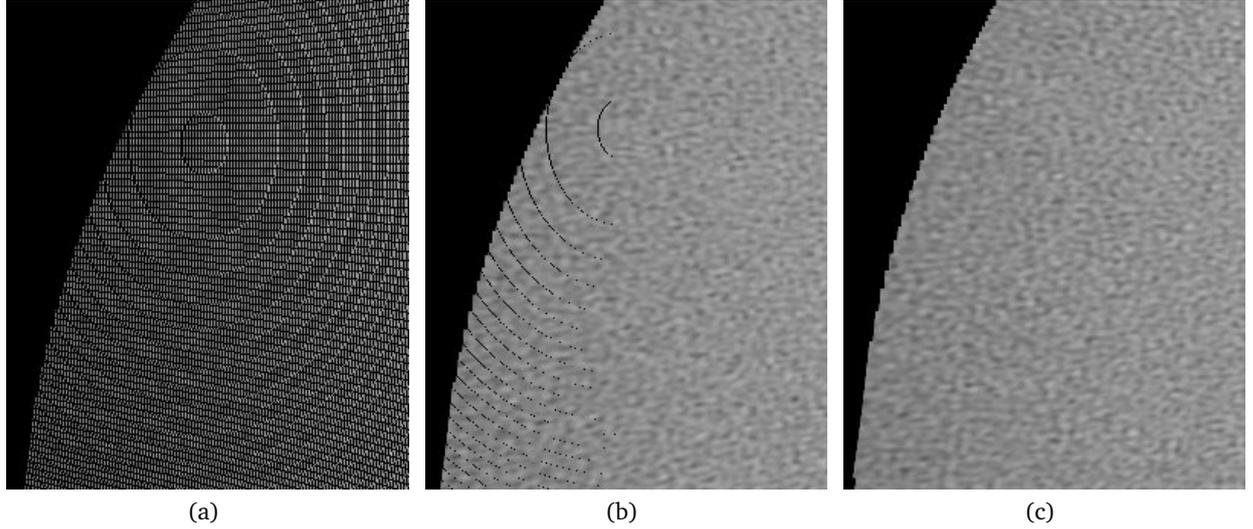


FIGURE 4.11 – Correspondance arrière depuis une correspondance avant, dans une zone de l'image fortement distordue, avec un échantillonnage de (a) 1 point par pixel, (b) 2 points par pixel, (c) 4 points par pixel.

Nous proposons deux mécanismes pour résoudre ce dernier problème et pour augmenter la précision :

1. La *correspondance-avant* sur-échantillonnée (donc avec plus de 1 échantillon par pixel dans le signal non distordu) qui fournit le plus faible écart avec le plus proche voisin est identifiée (en vert sur la figure 4.10.b pour un signal 1-D) et la coordonnée de l'échantillon associé à cette correspondance sur le signal distordu est alors utilisée pour la correspondance arrière.
2. Le terme de correction (dx, dy) est alors calculé par interpolation bilinéaire dans les images. Pour simplifier l'explication de son calcul, nous considérons une interpolation linéaire sur un signal 1-D dans la figure 4.10.c) pour le calcul de dx . A partir de l'écart dx' entre la coordonnée donnée par correspondance avant et le plus proche voisin sur le signal non distordu, l'écart dx correspondant sur le signal distordu est obtenu par la relation : $dx = k \cdot dx'$, où k est le facteur d'élongation local entre deux échantillons successifs. Soit f , l'application qui permet d'obtenir les coordonnées non distordues à partir des coordonnées distordues, e_X , la coordonnée de l'échantillon e dans le signal distordu, alors k est donné par :

$$k = \frac{de_{Dist}}{de_{NDist}}, de_{Dist} = \frac{1}{Ne}, de_{NDist} = f((e+1)_X) - f(e_X), \quad (4.21)$$

où Ne est le nombre d'échantillons par pixel, de_{Dist} correspond à l'écart entre deux échantillons successifs dans le signal distordu et de_{NDist} , son homologue dans le signal non distordu.

Erreur de correspondance La correspondance inverse qui est ainsi estimée est comparée à la correspondance donnée par les paramètres de distorsion estimés et l'erreur moyenne est reportée dans le tableau 4.3. Les erreurs sont alors reportées avec et sans l'utilisation du terme correctif et avec un nombre d'échantillons par pixels variant de 1 à 16 avec une progression géométrique.

On note une évolution linéaire de l'abaissement de l'erreur avec l'augmentation de la densité d'échantillonnage, ce qui veut dire qu'une précision très fine peut être obtenue si nécessaire. L'utilisation du terme correctif permet d'obtenir une précision en moyenne 3 fois plus élevée que sans. Au final, avec 16 échantillons par pixel et le terme correctif, il est possible d'atteindre une précision de l'ordre du centième de pixel. Ce qui est suffisant au regard de l'ordre de grandeur généralement constaté lors des processus d'étalonnage qui sont de l'ordre de quelques dixièmes de pixel.

D'un point de vue qualitatif, il est nécessaire, avec les paramètres estimés, d'échantillonner au moins 4 points par pixel pour que toutes les correspondances soient établies et qu'aucun artefact, ou pixel vide ne soit généré (Fig. 4.11).

Nb échantillons /pixel	1	2	4	8	16
Sans terme correctif	0.3796	0,2401	0,1205	0.0605	0.0328
Avec terme correctif	0.08380	0.0848	0.0427	0.0214	0.0117

TABLE 4.3 – Erreur de correspondance-arrière en pixels dans la carte de correspondance en fonction du niveau d'échantillonnage et de l'utilisation ou non du terme correctif.

En pratique, il existe deux cartes de correspondance, une relative à la correction des distorsions géométriques et une autre, relative à la rectification de l'image. Ces deux cartes sont cumulées en une seule. Ainsi une seule transformation avec la carte de correspondance cumulée sera nécessaire pour effectuer la correction des distorsions et la rectification d'image.

4.2.5 Estimation de la matrice Fondamentale

La matrice Fondamentale F est estimée à partir de correspondances de points 3-D non coplanaires, entre les plans images du projecteur et celui de la caméra. De façon générale, celle-ci est estimée de façon à minimiser l'écart moyen, $e(l_p, p')$, (en pixels) entre chaque ligne épipolaire l_p associée à un point p dans une image et son correspondant p' dans l'autre image du système stéréo. Plusieurs facteurs vont influencer la précision de son estimation. Parmi ceux-ci, nous notons :

- la validité du modèle sténopé,
- la précision de la segmentation des points mis en correspondance et la redondance de ceux-ci,
- le volume spatial dans lequel les correspondances sont effectuées,
- la méthode d'optimisation choisie.

Pour une meilleure estimation possible, il est nécessaire que tous ces facteurs soient pris en compte.

La validité du modèle sténopé

Comme nous l'avons vu, le modèle sténopé, basé sur la projection perspective, ne considère pas les distorsions géométriques dues aux optiques du système stéréo. Il est donc nécessaire que ces distorsions soient corrigées le plus précisément possible avant de pouvoir utiliser ce modèle.

Le volume spatial des correspondances de points

La matrice Fondamentale F et, de surcroît, la géométrie épipolaire permettent de mettre en relation un point d'une image avec sa ligne de projection dans l'autre image quelle que soit la scène 3-D sur laquelle cette projection est effectuée. Pour que F puisse prendre en compte l'aspect perspectif des correspondances, il est donc nécessaire que ces dernières soient présentes sur un large éventail de profondeurs. Plus les profondeurs des correspondances utilisées pour l'estimation de F sont diversifiées et représentatives du volume de travail désiré, meilleure sera la prise en compte de l'effet perspectif et donc sa validité dans l'ensemble du volume de travail.

La précision de la segmentation

Il est évident que la précision de la segmentation des points appariés joue un rôle dans la précision de l'estimation de la matrice Fondamentale, étant donné que les coordonnées des points sont les seules données d'entrée utilisées. Cependant cela est surtout vrai lorsqu'il y a relativement peu de points utilisés, car dans le cas inverse, quand un nombre important de correspondances peut être choisi, même avec une précision de l'ordre du pixel, une compensation statistique s'opérera, la plupart du temps, selon la loi des grands nombres. Notons également que les distorsions géométriques des optiques peuvent créer des courbures éloignant de plusieurs dizaines de pixels des points de la ligne moyenne (dans le cas d'une projection de grille) à laquelle ils sont sensés appartenir, et ce, de façon non homogène dans l'image. Il nous semble donc plus important de se concentrer sur cette dernière étape.

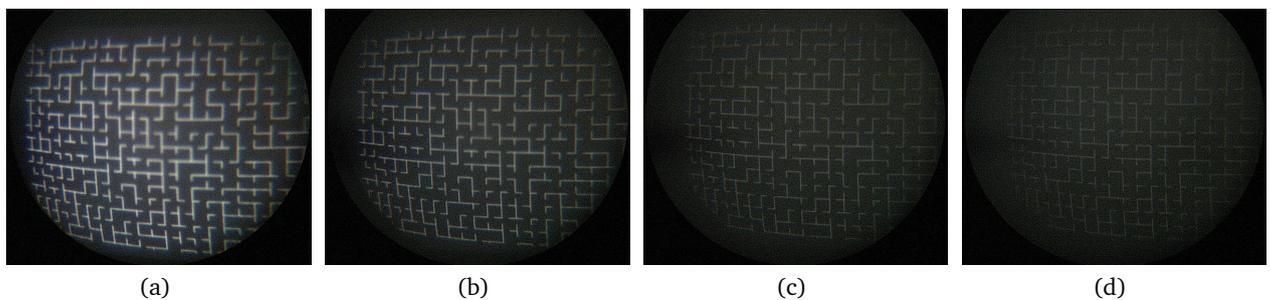


FIGURE 4.12 – Motif d'étalonnage à différentes profondeurs avant la correction des distorsions.

Pour répondre à cette problématique ainsi qu'à celle du volume spatial, nous proposons de projeter un "motif d'étalonnage" codé de type *PSM*, indépendant de la géométrie épipolaire, possédant une structure générale grillagée (voir Fig. 4.12). Contrairement aux mires de calibration à damier souvent utilisées, le codage *PSM* permet ici d'obtenir des correspondances directes même si une partie du motif n'est pas présente dans le champ de vue, ce qui arrive lorsque l'on rapproche/éloigne beaucoup le motif de la caméra pour couvrir une zone plus importante du volume de travail. De plus, les distorsions des optiques étant corrigées à cette étape, il est possible, pour la segmentation, d'intégrer le long de toutes les lignes et les colonnes de la structure grillagée au lieu de considérer chaque coins indépendamment. Ainsi davantage d'énergie lumineuse est utilisée pour la segmentation des intersections de la grille ce qui permet moins

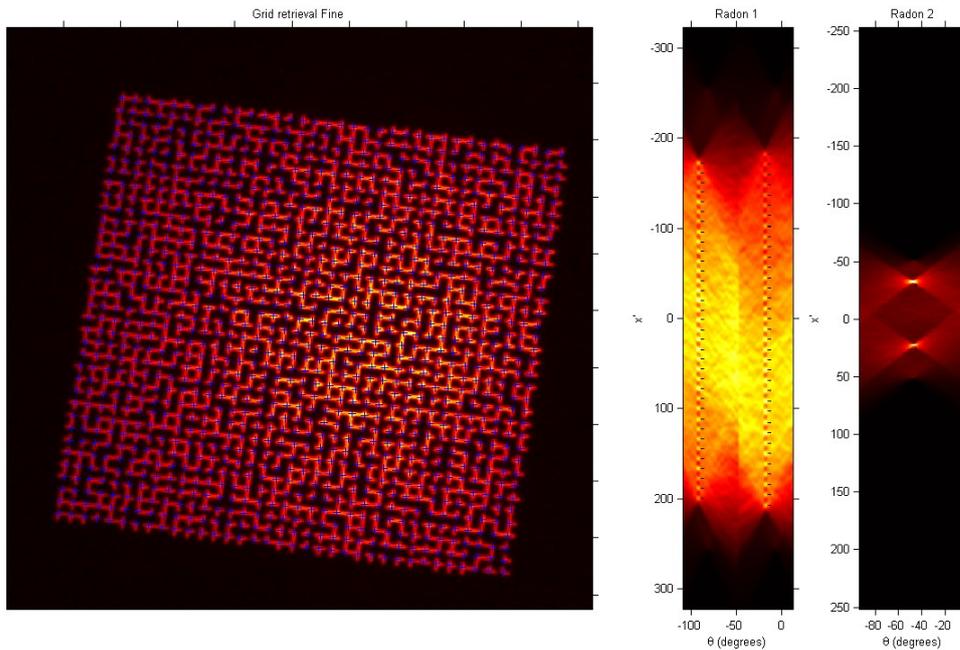


FIGURE 4.13 – Segmentation du Motif d’étalonnage (système non endoscopique) grâce à deux transformées de Radon successives.

de sensibilité au bruit que de considérer chaque intersection individuellement, avec un détecteur de coins par exemple.

Pour estimer les paramètres des lignes de la grille, des méthodes robustes existent comme la transformée de Hough ou celle de Radon que nous avons utilisée. Celles-ci procèdent par votes dans l’espace des paramètres des lignes. Cette robustesse est souhaitable dans notre cas comme le grillage du motif est discontinu pour permettre le codage par l’utilisation de primitives cunéiformes (voir chapitre 3, Choix des primitives).

Les méthodes de vote dans l’espace des paramètres peuvent être très longues à s’exécuter en fonction de la finesse de discrétisation de cet espace. Ainsi nous utilisons cette première étape pour donner les paramètres approximatifs de chaque ligne (à 1 pixel près) (voir Fig. 4.13). Dans une deuxième étape, ces paramètres sont affinés pour obtenir une précision finale de l’ordre de 0,1 pixels.

La méthode d’optimisation

Les méthodes d’optimisation permettant d’évaluer F sont basées sur la minimisation d’une fonction de coût qui évalue $\bar{e}(l_p, p')$, la valeur moyenne de $e(l_p, p')$, pour toutes les correspondances de points participant à l’estimation de F . Un grand nombre de telles méthodes a été proposé et testé dans la littérature. Il serait bien trop long de détailler ici toutes ces méthodes, le lecteur est invité à consulter les études comparatives disponibles, citons notamment [Zhang98] et plus récemment [Armangué03]. De façon générale, des méthodes linéaires, itératives et robustes sont testées. Elles procèdent toutes par la résolution d’un système d’équations homogènes possédant 7 paramètres linéairement indépendants (F est de rang 2) parmi les 9. Alors que les méthodes linéaires et itératives peuvent être robustes à la présence de bruit dans l’image si suffisamment de redondance (de correspondance de points) est fournie, les méthodes robustes permettent de gérer la présence de mauvais appariements (*outliers*) et de réduire leur influence.

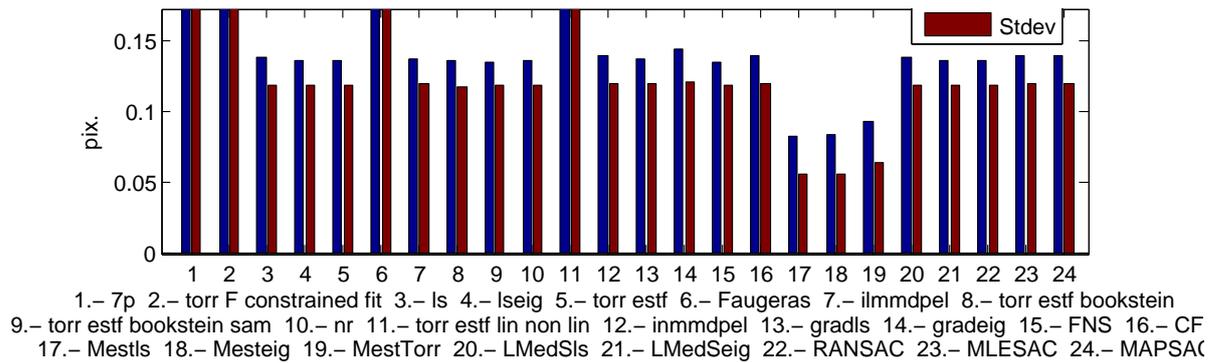


FIGURE 4.14 – Erreur d’estimation de la Matrice Fondamentale avec différentes méthodes d’optimisation. En bleu, l’erreur moyenne $\bar{e}(l_p, p')$, en bordeau, l’écart-type

Résultat pour le système stéréo non endoscopique Dans [Armangué03], un outil dans l’environnement *Matlab*, le *Fundamental Matrix Estimation Toolbox*, a été mis à disposition et nous a permis d’obtenir des résultats pour les différentes méthodes présentées. Pour un descriptif détaillé des différentes méthodes, nous invitons le lecteur à consulter la publication correspondante. Cet outil n’a été utilisé que pour les validations initiales sur un système stéréo non miniaturisé, dont les distorsions du projecteur, corrigées par le fabricant, n’ont ni été estimées ni corrigées. Les distorsions géométriques de la caméra ont, quant à elles, été corrigées au préalable. Les résultats obtenus avec le système stéréo non endoscopique (projecteur de résolution 1024×1024 et caméra 1000×1000 , détaillés dans la section suivante) sont représentés sur le graphique de la figure 4.14. Au mieux, une erreur moyenne $\bar{e}(l_p, p')$ de 0.08 pixels a été obtenue avec les M-estimateurs (méthodes robustes 17, 18 et 19 de la figure 4.14). Ce qui place ce résultat dans la fourchette basse (en terme d’erreurs) comparé aux résultats reportés dans [Armangué03] sur différents types de scènes. De plus, il est remarquable de constater que la majorité des autres méthodes (linéaires, itératives ou robustes, sauf 4 d’entre elles), forment un consensus autour d’une erreur de 0.14 pixels. Ce qui montre la qualité générale de la segmentation proposée. Les variations d’une méthode à l’autre sont bien plus importantes dans les résultats reportés dans [Armangué03] et varient de plusieurs $10^{\text{ième}}$ de pixels à plusieurs pixels selon les scènes testées. Une dizaine de captures du motif d’étalonnage projeté sur un plan à différentes profondeurs a été utilisée.

Résultat pour le système stéréo miniaturisé

Le système miniaturisé est composé d’un projecteur miniature de résolution 800×600 et d’une caméra endoscopique Full-HD (1920×1080), davantage détaillés dans le chapitre suivant. La matrice Fondamentale a été estimée à partir de la plate-forme logicielle intégrée, que nous avons développée, (voir chapitre 5) en utilisant la méthode robuste RANSAC proposée par la bibliothèque *OpenCV*. Une dizaine de captures du motif d’étalonnage projeté sur un plan à différentes profondeurs (allant de 1 à 20 cm des optiques endoscopiques) ont été utilisées (voir Fig 4.12).

L’erreur moyenne $\bar{e}(l_p, p')$ constatée est de 0.48 pixels dans le plan du projecteur et de 0.98 pixels dans le plan de la caméra. Cette différence s’explique par la différence de résolution des

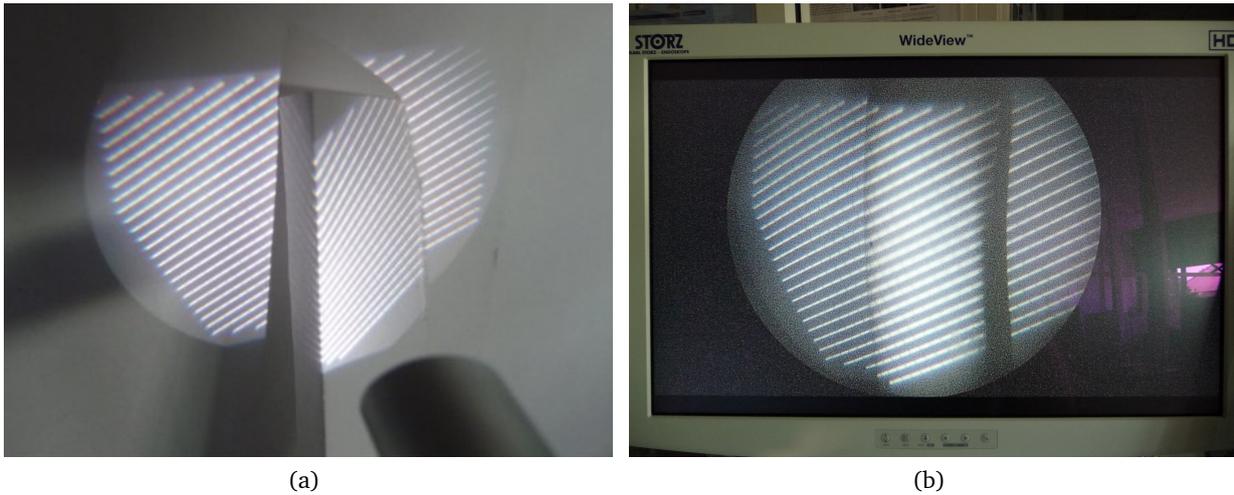


FIGURE 4.15 – (a) Projection de lignes épipolaires avec le système miniaturisé. (b) projection dans l'image de la caméra indépendante des discontinuités de courbures des objets de la scène.

deux dispositifs. L'erreur d'estimation de F est ici significativement plus élevée que pour le système non endoscopique. Le fort bruit présent dans les images acquises en est certes certainement l'une des causes, mais la redondance présente par le nombre de points mis en correspondances et le nombre d'images utilisées n'en font pas la cause principale. Nous pensons que si les distorsions géométriques ont été globalement corrigées pour le projecteur et la caméra, il en reste encore certains effets. Cependant, une projection de lignes épipolaires avec ce système (Fig. 4.15), montre que qualitativement, l'invariance de leur projection dans le plan image de caméra résiste bien aux discontinuités de courbure des objets présents dans la scène (Fig. 4.15.a). Dans l'image Fig. 4.15.b, les lignes sont courbées car les distorsions de l'image ne sont pas encore corrigées. Notre conception basée sur un alignement du motif sur la géométrie épipolaire peut donc être utilisée avec le système endoscopique.

4.3 Précision des reconstructions

Avec tout nouveau système de métrologie en général et de reconstruction 3-D en particulier, tel que celui présenté dans cette thèse, il est important de donner une estimation de la précision des mesures qui sont produites. D'une part, pour comparer les performances obtenues à d'autres systèmes existants, et d'autre part pour caractériser intrinsèquement les résultats obtenus et envisager ainsi les applications possibles. En effet, certaines applications peuvent nécessiter davantage de précision que d'autres. Par exemple, une simple visualisation sous différents points de vue peut être réalisée sans une précision aussi importante que ce qui est nécessaire pour des applications de métrologie, par exemple. Dans notre cas en particulier, la ligne de base du système stéréo endoscopique utilisé (voir Chapitre 5), est très faible, comparée à des systèmes plus classiques, avec ~ 6 mm seulement entre le centre optique du projecteur et celui de la caméra. Or, dans une étude récente, [Li10] a mis en avant que c'est la ligne de base qui a le plus d'influence sur la précision de la reconstruction finale, devant d'autres paramètres comme l'orientation des axes optiques des dispositifs composant le système stéréo. Ainsi si, dans notre système, l'endoscope droit exploite une optique orientée à 0° , et l'endoscope coudé à 30° , rien n'indique que cela suffise à compenser la faible ligne de base. Une estimation de la précision

obtenue permet d'en avoir une idée plus précise.

Un bon nombre de contributions utilisent une méthode qui mesure l'erreur de reconstruction de plans. Le motif de lumière structurée est projeté sur un plan, parallèle aux plans images des dispositifs du système stéréo et les centres des indices visuels qui sont projetés sur ce plan sont reconstruits. Dans une deuxième étape, l'équation du plan π_{LS} passant au mieux, dans le sens des moindres carrés (dans [Song10b], par exemple), par l'ensemble des points reconstruits, est estimée (Fig. 4.16). Il s'agit donc d'une régression linéaire. Dans une dernière phase, l'écart moyen, entre chaque reconstruction et π_{LS} , est calculé et rapporté à la profondeur estimée de la primitive reconstruite. Une mesure de précision relative (en pourcentage de la distance du plan) est ainsi obtenue. Si la profondeur métrique du plan est connue, la précision relative obtenue peut ensuite être rapportée en une mesure métrique. Le tableau 4.4, établi dans [Song10b], synthétise plusieurs résultats d'estimation de précision reportés dans la littérature en utilisant une reconstruction de plan.

Dans notre cas, cette méthode d'estimation est adaptée pour deux raisons. D'une part, nous ne disposons pas, pour l'instant, d'un dispositif complètement étalonné et, d'autre part, notre méthode de reconstruction, présentée plus tôt dans ce chapitre, fait usage d'un plan de référence parallèle au plan image du projecteur. Donc la reconstruction d'un tel plan peut tout à fait être utilisée. Pour avoir une idée de la précision sur l'ensemble du volume de travail, nous avons projeté un motif 50×50 (codage SPSM avec 2 symboles et $H_{Min} = 2$), sur des plans à des distances de 10,20,30,50,100,150 et 200 mm (Fig. 4.17). Les résultats obtenus à ces différentes profondeurs sont reportés dans le tableau 4.5. Ceux-ci montrent qu'une précision inférieure au demi-millimètre peut être obtenue sur l'ensemble de la plage de profondeurs considérée : jusqu'à 20 cm de profondeur. Comparés à plusieurs résultats reportés dans la littérature (Tab. 4.4), ces résultats, en valeur relative (environ 0.15%), sont semblables à la majorité des autres valeurs reportées. Mais pour la plage de profondeur qui nous intéresse, à une distance moyenne

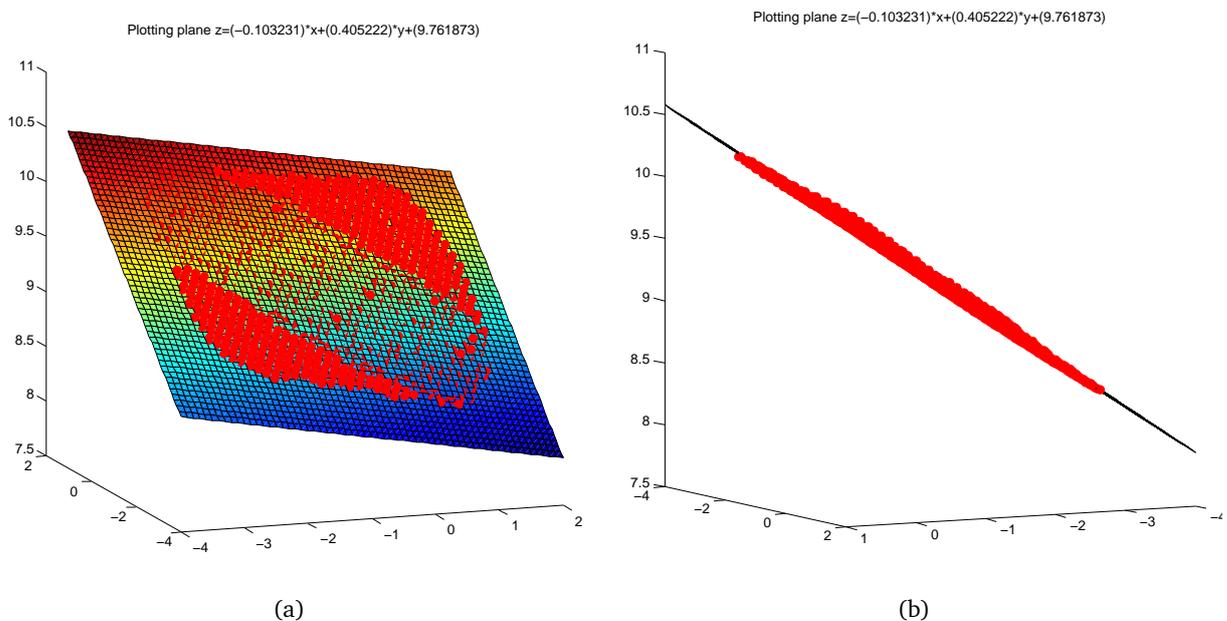


FIGURE 4.16 – a) Estimation du plan passant au mieux, au sens des moindres carrés, par les reconstructions. b) Vue de profil.

Référence	Motif	Erreur de régression de plan (% de distance)
[Song10b]	Code de Gray & Strip Shifting	0.11 - 0.13 mm ou 0.015%
[Chenb]	Motif 2-D	~ 2%
[Desjardins07]	Motif 2-D	2 mm (0.15% de 500mm)
[Fong05]	Motif 2-D	0.7 mm (0.11% de 650mm)
[Pagès05]	Franges colorées	0.3 – 0.6 mm (NA)
[Sadlo05]	Code de Gray & décalage de phase	~ 0.1 – 0.5 mm (NA)
[Zhang02]	Franges colorées	Ecart-type. 0.18 mm (NA)
[Guhring01]	Code de Gray & décalage de lignes	Ecart-type. 0.17 – 0.28 mm (NA)
[Chena]	Franges colorées	0.15 – 0.32 mm (NA)
[Kiyasu95]	Motif 2-D	1 mm (0.2% de 500 mm)
Nous	Motif 2-D	0.15 mm (0.15% de 100 mm)

TABLE 4.4 – Différents résultats de précision de reconstruction de plans par une projection de lumière structurée.

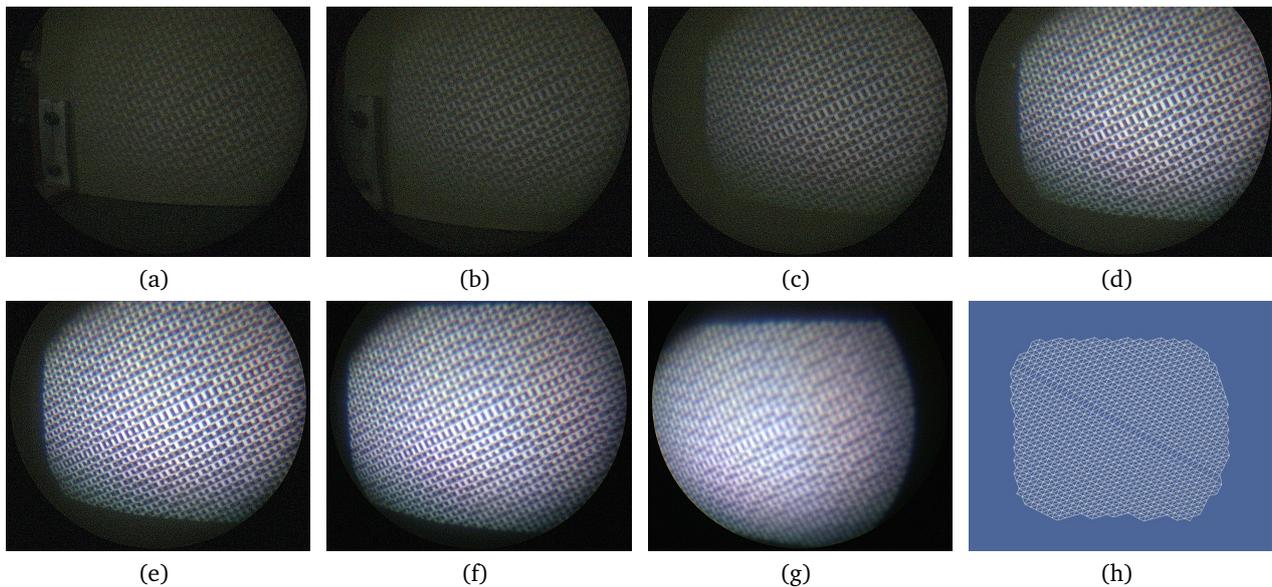


FIGURE 4.17 – Plans reconstruits de a) à g) respectivement, à ~ 200, 150, 100, 50, 30, 20, 10 mm de profondeur. h) Maillage d'un plan reconstruit.

de 10 cm, la précision correspondante de 0.15 mm est par conséquent plus élevée que ce que la majorité des autres méthodes obtiennent pour des scènes à une distance de 500 – 650 mm, qui doivent être considérées avec des systèmes stéréos plus classiques (non endoscopiques).

D'autre part, si on observe, pour chaque point reconstruit, l'erreur (dans la direction de l'axe optique) avec le plan π_{LS} estimé (Fig. 4.18.a/b), on remarque que la distribution des erreurs n'est pas répartie de façon homogène dans l'image. Donc, il y a de fortes chances que la composante principale de ces erreurs ne provienne pas d'imprécisions de segmentations dues au bruit, mais plutôt de distorsions géométriques résiduelles. Notons en effet, que ces estimations ont été réalisées après manipulation du système miniaturisé (après une séance in-vivo), donc il est possible que la géométrie du système ait été très légèrement modifiée.

Distance plan (mm)	10	20	30	50	100	150	200
Moyenne (%)	0.1785	0.2087	0.1722	0.1660	0.1529	0.1356	0.1199
Ecart-type (%)	0.1450	0.1647	0.1327	0.1296	0.117	0.1018	0.0883
Moyenne (mm)	0.0178	0.0417	0.0516	0.083	0.1529	0.2034	0.2398
Ecart-type (mm)	0.0145	0.0329	0.0398	0.0647	0.1170	0.1528	0.1765

TABLE 4.5 – Précision obtenue pour la reconstruction de plans à des profondeurs comprises entre 10 et 200mm.

Pts ref (20mm)	1000 A	1000 B	500 A	500 B	100 A	100 B
Moyenne (%)	0.2087%	0.2087%	0.2087%	0.2088%	0.2087%	0.2087%
Ecart-type (%)	0.1647%	0.1647%	0.1648%	0.1648%	0.1648%	0.1648%
Moyenne (mm)	0.04174	0.04173	0.04174	0.04174	0.04174	0.04174
Ecart-type (mm)	0.03295	0.03295	0.03296	0.03296	0.03295	0.03296

TABLE 4.6 – Précisions obtenues pour différents triplets de points définissant le plan de référence (20mm de profondeur).

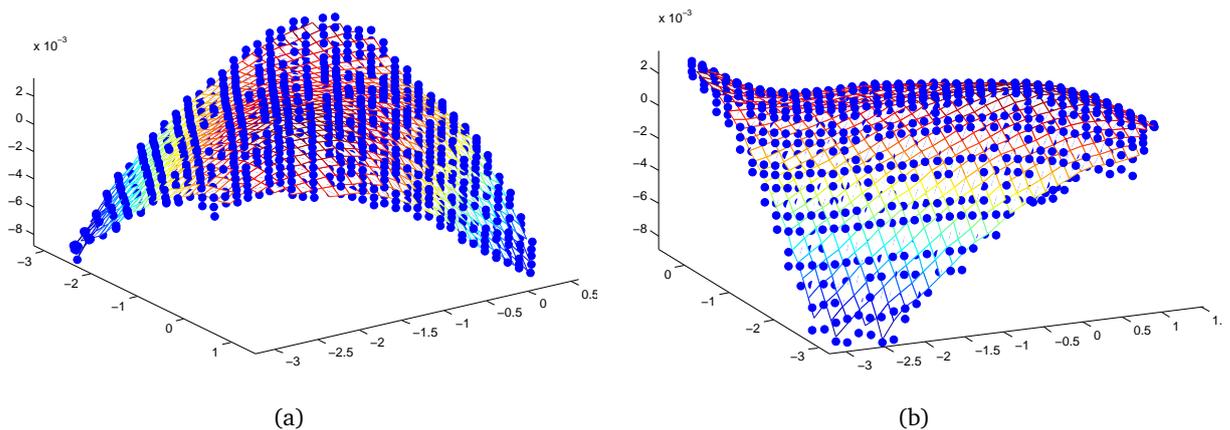


FIGURE 4.18 – (a),(b) Pour chaque point du plan reconstruit, l'erreur avec l'estimation du plan moyen π_{LS} est donnée sur l'axe vertical.

Comme la reconstruction 3-D utilisée exploite un invariant projectif de la famille du birapport et que ceux-ci sont connus pour leur forte sensibilité au bruit (voir [Morin93]), nous avons voulu vérifier si la précision obtenue varie en fonction du choix des 3 points utilisés pour définir le plan de référence (voir section 4.1). Pour cela, nous avons effectué des reconstructions de plans définis avec différents triplets de points (A,B dans la figure 4.6) dont la distance moyenne (entre les points) est de 1000, 500 et 100 pixels. Les résultats reportés sur le tableau 4.6, montrent que les variations obtenues sont négligeables.

Hormis les aspects quantitatifs d'estimation de la précision des reconstructions, on peut également remarquer, dans la figure (Fig. 4.17), la différence très significative du rapport signal/bruit entre l'image acquise avec un motif projeté sur un plan à une profondeur de 200 mm (Fig. 4.17.a) par rapport à une projection sur un plan à une profondeur de 10 mm (Fig. 4.17.f). Cela est dû à la dispersion de l'énergie lumineuse avec l'allongement du trajet des rayons lumineux. Cette observation renforce ce que nous noterons également plus tard, que le rapport

signal/bruit pourra être considérablement amélioré avec une meilleure efficacité optique du système ou avec une source de lumière plus puissante.

En conclusion, nous pouvons dire que comparés à d'autres résultats reportés dans la littérature, ceux obtenus pour une reconstruction de plan montre que la faible distance de base de 6 mm, n'est, a priori, pas un handicap pour notre système quant à la précision des reconstructions effectuées. En effet, la faible ligne de base est compensée par la combinaison de différents facteurs : résolution Full-HD de la caméra, faible profondeur des surfaces à reconstruire (environ 10 cm des centres optiques), différence d'orientation de 30° entre l'axe optique du projecteur et celui de la caméra. Cependant les valeurs de précision reportées ici sont à considérer avec précaution et ne reflètent, à notre avis, pas la précision réelle qui sera obtenue sur des surfaces intra-abdominales. Celles-ci possèdent en effet de nombreuses perturbations géométriques et spectrales qui viendront dévier l'énergie lumineuse des indices visuels projetés et causeront donc des imprécisions nombreuses qui n'ont pu être détectées ici. En effet, la segmentation d'un motif projeté sur un plan avec une réflectance homogène se fait en l'absence de telles imprécisions. De plus, la méthode d'estimation employée ici, qui utilise une régression de plan, n'est pas à même de capturer les erreurs d'étalonnage du système stéréo. Pour toutes ces raisons, il sera nécessaire, dans le futur, de refaire des estimations de précision avec un système complètement étalonné.

4.4 Résultats des reconstructions

Dans cette section, nous présentons les résultats de segmentation, de mise en correspondance et de reconstruction obtenus dans trois scénarii différents : en situation **non endoscopique**, **endoscopique EX-VIVO** et **endoscopique IN-VIVO (mini-invasif)**.

4.4.1 Reconstructions non endoscopiques

Réalisées en premier lieu, ces reconstructions avaient pour objectif de valider l'approche de conception du motif à l'aide d'un système stéréo classique. Celui-ci était composé d'une caméra couleur *Allied Pike* de résolution 1000×1000 associée à un projecteur couleur *Epson EB84* de résolution 1024×768 . Les premiers motifs projetés utilisaient un alignement local associé à des primitives cunéiformes avant de passer à un alignement global avec deux primitives discriminées par gradient vertical (voir Chapitre 3). La qualité des images étant bonne, avec très peu de bruit, toutes les étapes de segmentation décrites au chapitre 3, n'étaient pas nécessaires. En effet le système stéréo étant fixé sur un trépied tout au long des acquisitions et la géométrie épipolaire étant estimée précisément (voir résultats d'estimation de la matrice Fondamentale section 4.2), il n'a pas été nécessaire d'effectuer de recalage épipolaire. De même pour les franges, il n'y a pas eu besoin d'estimer des champs d'élongation et d'angle d'inclinaison des voisinages, une détection des maxima locaux successifs, correspondant à la présence des primitives, le long des lignes épipolaires associées était suffisante et très rapide. De plus, le bruit étant faible, des motifs de résolution importante ($\geq 150^2$ primitives), comparés à ceux utilisés dans l'état de l'art, ont pu être utilisés. Les diverses scènes reconstruites ont été choisies de façon à montrer la robustesse de l'approche tant vis-à-vis des perturbations géométriques (fortes discontinuités de courbure), que vis-à-vis des perturbations spectrales.

Ces reconstructions permettent de valider, en outre, la méthode proposée, qui bien que projective permet l'obtention de reconstructions très peu distordues grâce à la contrainte d'orthogonalité utilisée. Cela est valable ici car les variations de profondeur entre l'arrière plan, utilisé

comme plan de référence, et les objets, est faible par rapport à la distance entre cet arrière plan et la caméra. (rapport de 1/10 environ). Cela sera moins vrai pour les scènes endoscopiques.

Deux types de motifs ont été projetés, le premier utilise un alignement local avec un codage *SPSM* par proximité épipolaire (Figure 4.19) avec 4 symboles et $H_{Min} = 3$. Le second emploie un alignement global avec les mêmes propriétés de codage (Figures 4.20 et 4.21). Pour ces projections, des primitives cunéiformes (voir Chapitre 3) avaient été utilisées.

Les résultats qualitatifs montrent une bonne résistance globale aux différents types de perturbations géométriques et spectrales. La capacité du système à reconstruire des objets en mouvement est également illustrée par la figure 4.21. L'utilisation de voisinages de taille réduite (3×3) permet de distinguer les doigts dans différentes situation et même quand une seule phalange est correctement illuminée par le motif (Fig. 4.21.e). Ces reconstructions sont effectuées à une moyenne de 27 reconstructions par seconde.

4.4.2 Reconstructions endoscopiques Ex Vivo

Avec l'utilisation du système stéréo endoscopique miniaturisé (voir Chapitre 5), nous présentons tout d'abord quelques reconstructions Ex Vivo qui donnent l'occasion de mettre en évidence les capacités et certaines limites des traitements présentés au chapitre précédent. Les reconstructions présentées ici sont réalisées dans l'espace projectif avec la contrainte de quasi-orthogonalité présentée au chapitre 3. Les 3 points définissant le plan de référence utilisé pour ces reconstructions sont choisis de telle sorte qu'ils définissent un plan à peu près parallèle aux plan image du système stéréo. Pour toutes ces reconstructions un alignement global sur la géométrie épipolaire associé à codage *SPSM* par alignement (voir chapitre 2) a été utilisé. Pour chacune des reconstructions présentées, nous donnons de nombreuses informations qualitatives et quantitatives permettant d'apprécier ces résultats (voir Fig. 4.22, par exemple) et les traitements qui ont permis de les obtenir. Ces données seront les mêmes pour les reconstructions In Vivo présentées dans la section suivante. Elles sont composées des éléments suivants :

- L'image (a) correspond à l'image directement acquise par la caméra sans correction des distorsions ni rectification. Dans le coin inférieur gauche, nous indiquons les deux primitives visuelles utilisées.
- L'image (b) correspond aux traitements de l'image(a) après correction des distorsions et rectification. Elle contient de nombreuses informations :
 - La segmentation des franges (continuité verticale des primitives) est représentée par les bandes blanches.
 - Les primitives localisées sur ces bandes sont représentées par des carrés de couleur ou noir et blanc.
 - Les primitives en noir et blanc correspondent aux frontières du graphe d'adjacence présenté au chapitre 3, elles définissent les régions dans lesquelles un décodage peut se propager de proche en proche. Les primitives en blanc correspondent à des discontinuités des franges et celles en noir à un non parallélisme local des franges sur des voisinages 3×3 .
 - Les primitives représentées par un carré vert ont pu être décodées avec une forte confiance, car la méthode des votes évoquée au chapitre 3 a fourni un score suffisant. Les paramètres utilisés pour définir une primitive décodée avec une forte certitude sont indiqués ci-après.
 - Les primitives représentées par un carré bleu ont été décodées par propagation à partir des primitives vertes de la même zone d'adjacence.
 - Enfin, le carré supérieur gauche représente l'état des primitives du motif dans le plan

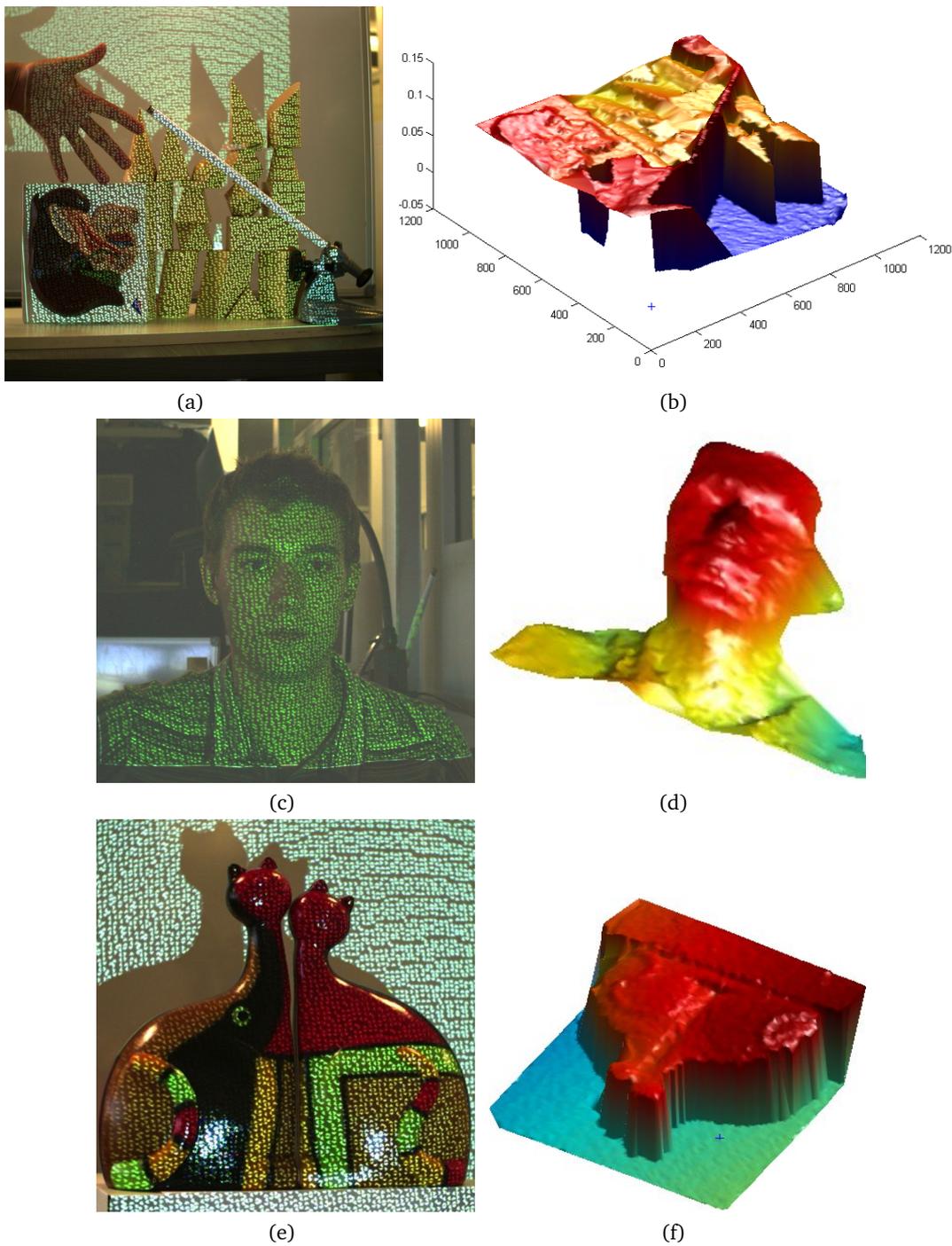


FIGURE 4.19 – Reconstructions 3-D. Motif 100×100 utilisant un alignement local, un codage *SPSM* par proximité épipolaire. a,b) Scènes possédant de nombreuses discontinuités de courbure (4 symboles, $H_{Min} = 4$). c,d) Relief d'un buste humain avec des vêtements striés (4 symboles, $H_{Min} = 4$). e,f) Scène possédant de nombreuses perturbations spectrales (2 symboles, $H_{Min} = 1$).

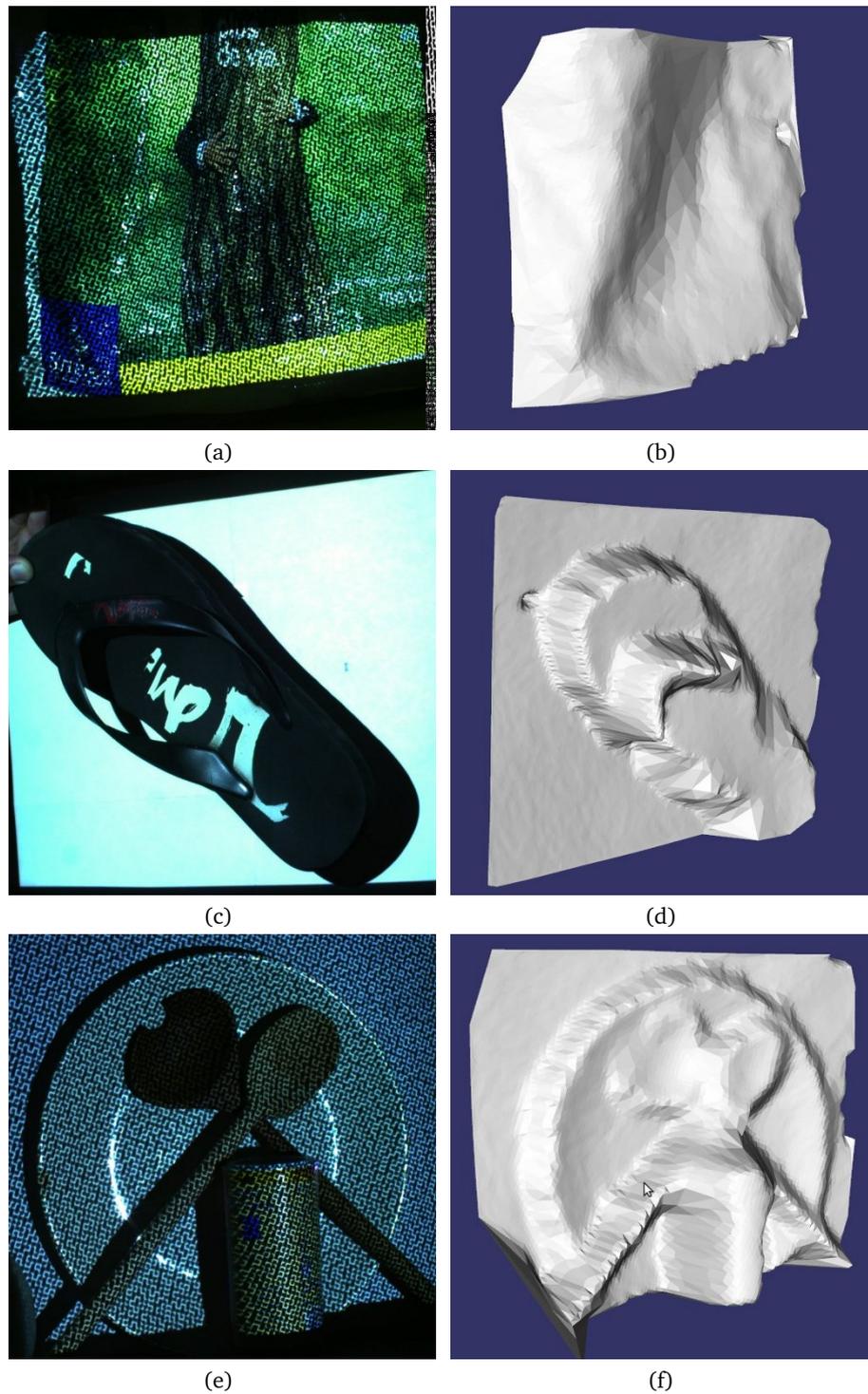


FIGURE 4.20 – Reconstructions 3-D. Motif 100×150 utilisant un alignement global, un codage *SPSM* par proximité épipolaire et 4 primitives cunéiformes, ($H_{Min} = 4$). a,b) Scènes possédant de nombreuses perturbations spectrales. c,d,e,f) Scènes possédant de nombreuses discontinuités de courbure.

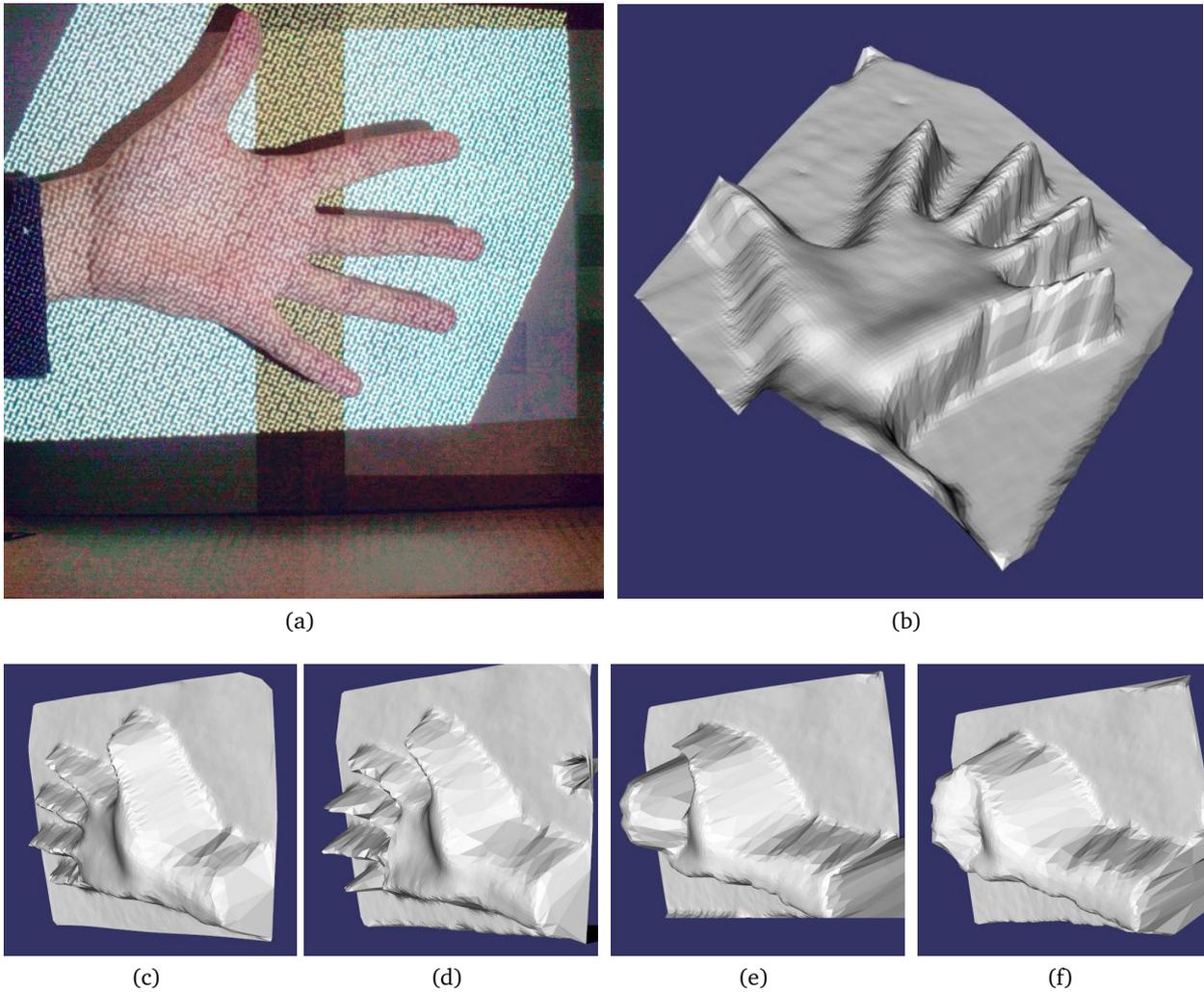


FIGURE 4.21 – Reconstructions 3-D. Motif 100×150 utilisant un alignement global, un codage *SPSM* par proximité épipolaire et 4 primitives cunéiformes, ($H_{Min} = 4$). a - f) La scène montre une main en mouvement

projecteur, il a la résolution de la matrice *SPSM* codée et donne donc une indication sur la mise en correspondance. Le vert et le bleu ont la même signification que dans le plan image de la caméra. Les primitives grises ont été codées mais non projetées et les primitives rouges ont été projetées mais non décodées. L'inclinaison de l'interface entre les primitives rouges et grises est due à l'orientation des lignes épipolaires.

- L'image (c) montre les reconstructions 3-D correspondantes sous différents points de vue (haut/bas) et rendus (éclairage selon l'orientation des normales des sommets à gauche) et rendu, en fil de fer, à droite.
- Le tableau (d) donne diverses informations sur le codage, certains paramètres utilisés et des statistiques sur le décodage, dans l'ordre :
 - "**Taille codage**" : taille de la matrice *SPSM* codée.
 - "**Nb symboles**" : le nombre de symboles/primitives utilisés.
 - " H_{Min} " : la distance de Hamming minimale imposée.
 - "**Dans proj**" : le nombre de primitives dans le plan image du projecteur.

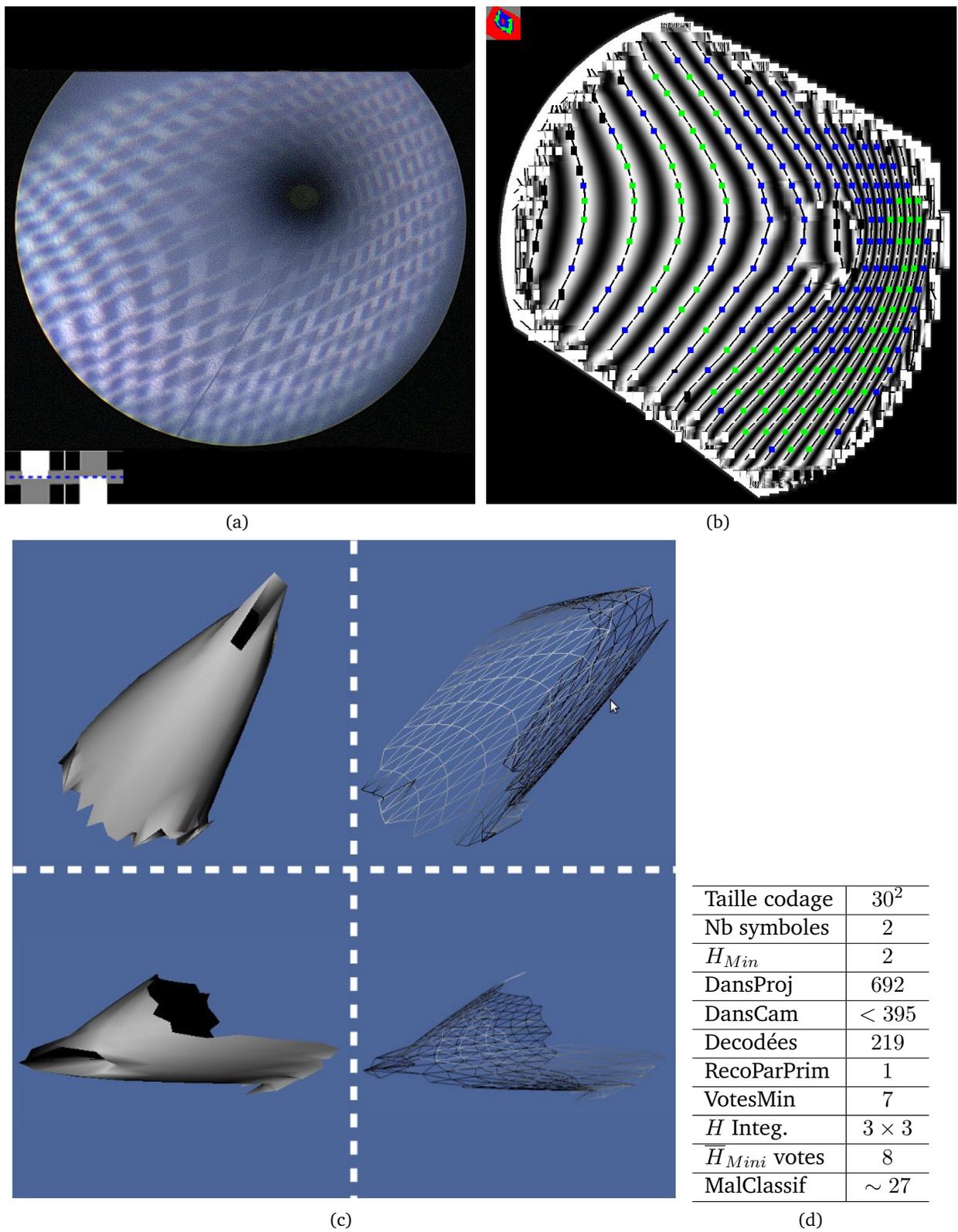


FIGURE 4.22 – Reconstruction d'un cylindre de papier vu de l'intérieur.

- "**Dans cam**" : la limite supérieure sur le nombre de primitives dans l'image acquise (en général 57% de "Dans proj"). Pour estimer ce rapport une projection sur un plan avec le champ de la caméra rempli de primitives a été effectué et le nombre de primitives dénombré et rapporté aux primitives présentes dans le plan image du projecteur.
- "Décodées" : le nombre total de primitives décodées (en vert et bleu sur l'image (b)).
- "**RecoParPrim**" : le nombre de points reconstruits le long des franges pour chaque primitive segmentée.
- "**VotesMin**" : le nombre minimal de votes de mise en correspondance qu'une primitive doit recevoir des ses primitives voisines (par contiguïté locale des mises en correspondances, voir chapitre 3) pour être considérée comme une primitive décodée avec une forte certitude (en vert dans l'image (b)).
- "**H Integ.**" : les dimensions de la zone de primitives considérées pour déterminer la meilleure mise en correspondance (correspondant à la plus faible distance de Hamming moyenne \bar{H} sur cette zone). C'est à partir de cette mise en correspondance que les votes pour les correspondances des 8 primitives voisines s'opère (voir chapitre 3).
- Cette mise en correspondance n'est utilisée pour les votes que si le coefficient de confiance correspondant ($9 - \bar{H}$) est supérieur ou égal à " \bar{H}_{Mini} votes".
- "**Mal classif**" : le nombre estimé de primitives mal classifiées. Ce nombre correspond aux primitives décodées dont la classification dans l'image de la caméra ne correspond pas à celle dans l'image du projecteur. Cette estimation n'est bonne que pour les décodages correctement réalisés.

Les reconstructions présentées ici (Figures 4.22 à 4.24) permettent de mettre en évidence plusieurs résultats intéressants.

Tout d'abord, la figure 4.22 montre clairement que les algorithmes de segmentation prennent en compte d'importantes déformations géométriques des voisinages grâce à l'estimation des champs de déformation (élongation et inclinaison) de ceux-ci. Il est donc possible de reconstruire des scènes possédant des surfaces très fortement inclinées par rapport aux plans image du système stéréo. Cette situation peut se présenter très fréquemment dans des scènes intra-abdominales. C'est aussi le cas pour des organes à conformation tubulaire (oesophage, trachée, côlon) qui sont souvent explorés en chirurgie transluminale à l'aide d'endoscopes flexibles. Cette figure correspond à la projection du motif à l'intérieur d'un cylindre de papier présenté au chapitre 3, lors de l'étude de la déformation des voisinages. La reconstruction 3-D correspondante met également en évidence le fait que le modèle de projection est orthographique. En effet ici les points sont régulièrement distribués en profondeur à partir de l'origine du repère utilisé et l'approximation du modèle de projection fait alors apparaître le cylindre comme un cône.

Les reconstructions présentées sur les figures 4.23 et 4.24 mettent en évidence la capacité du système à mettre en correspondance et reconstruire les points d'un motif de résolution comparable (70×70) au motifs utilisés avec des systèmes stéréo classiques (voir chapitre 1). De plus, les détails observables sur ces reconstructions montrent que malgré la très faible ligne de base (6 mm) du système stéréo endoscopique (voir chapitre 5) et la présence importante de bruit dans les images, des détails de scène peuvent être capturés.

Nous pouvons enfin déjà noter les effets de bord de la propagation du décodage sur la reconstruction de la figure 4.23. On peut alors observer sur la reconstruction (c), à l'interface droite entre l'arrière plan et la masse, que le décodage s'est propagé sur des zones ombrées dans l'image. Cela arrive quand un seuil trop large est utilisé pour déterminer les frontières du graphe d'adjacence. Avec un seuil plus strict, cette frontière est bien marquée dans le graphe

d'adjacence (image (b)).

Note sur le bruit des images (a) Sur certaines images (a), le bruit semble plus élevé que sur d'autres. C'est le cas, en particulier, sur la figure 4.24.a entre l'image du haut et celle du bas. En effet un filtre matériel est présent sur le contrôleur de la caméra endoscopique, mais celui-ci, pour des raisons pratiques, n'a pu être utilisé pour toutes les acquisitions.

4.4.3 Reconstructions mini-invasives In Vivo

Nous présentons ici les traitements et reconstructions effectués en situation mini-invasive In Vivo grâce au prototype de système endoscopique réalisé (voir chapitre 5). Pour obtenir un bon retour d'expérience, différents motifs ont été projetés sur de nombreuses parties intra-abdominales de porc. Ces expériences ont été réalisées au **bloc opératoire de l'IRCAD de Strasbourg**. Le porc possède une anatomie intra-abdominale très proche de celle de l'homme et est ainsi utilisé pour la formation des praticiens à la chirurgie mini-invasive. Les expérimentations effectuées sont donc représentatives de ce qui peut être obtenu chez l'humain. Quatre séances in-vivo de 1 heure et demi à 2 heures chacune ont été réalisées, donnant lieu à des retours d'expérience permettant, au fur et à mesure, d'améliorer les traitements (segmentation / décodage). Nous précisons que l'animal est endormi pendant l'expérience et que les quatre porcs sur lesquels les expérimentations ont été effectuées avaient servi au préalable à la formation de chirurgiens, ces expérimentations ayant eu lieu à la fin des journées de formation. Aucun animal supplémentaire n'a donc été utilisé exclusivement pour ces expérimentations.

La mise en place générale du système est présentée sur la figure 4.25. Grâce à la plate-forme logicielle mise en place (voir chapitre 5), des motifs ont pu être générés à la demande en une fraction de seconde pour essayer différentes configurations. A cause des écarts de température à l'intérieur et à l'extérieur de l'abdomen, de la buée s'est régulièrement formée sur les optiques endoscopiques qui ont alors été nettoyées.

Des reconstructions des différentes parties intra-abdominales : (intestin grêle, gros intestin, foie, vésicule biliaire, estomac et paroi intra-abdominale) ont ainsi pu être réalisées. Celles-ci sont visibles sur les figures 4.26 à 4.45. Pour chacune de ces reconstructions, les mêmes données que celles détaillées à la section précédente sont reportées.

Des motifs codés à une résolution de 30×30 , 50×50 et 70×70 primitives ont été utilisées. Trois types de formes de primitives (présentées au chapitre 3) ont été testées, ainsi que différents paramètres de décodage.

Ces reconstructions ont été effectuées à un taux d'exécution moyen de 12 reconstructions par seconde, soit à peu près deux fois moins rapidement que les reconstructions non endoscopiques présentées précédemment. Cette diminution est due aux traitements supplémentaires (estimation des champs de déformation des voisinages, segmentation adaptative des franges, établissement d'un graphe d'adjacence etc...) qui n'avaient pas été nécessaires pour le système non endoscopique, car la qualité des images acquises était bien meilleure, que le système stéréo n'était pas manipulé et que nous n'étions pas allés aussi loin dans la prise en compte des déformations des voisinages.

Certaines divergences mineures peuvent être notées entre les images (b) (les traitements), et les reconstructions 3-D. D'une part certaines mauvaises reconstructions sont sorties du champ de vue OpenGL (*Frustrum*) et ne sont donc pas représentées. D'autre part, dans certains cas,

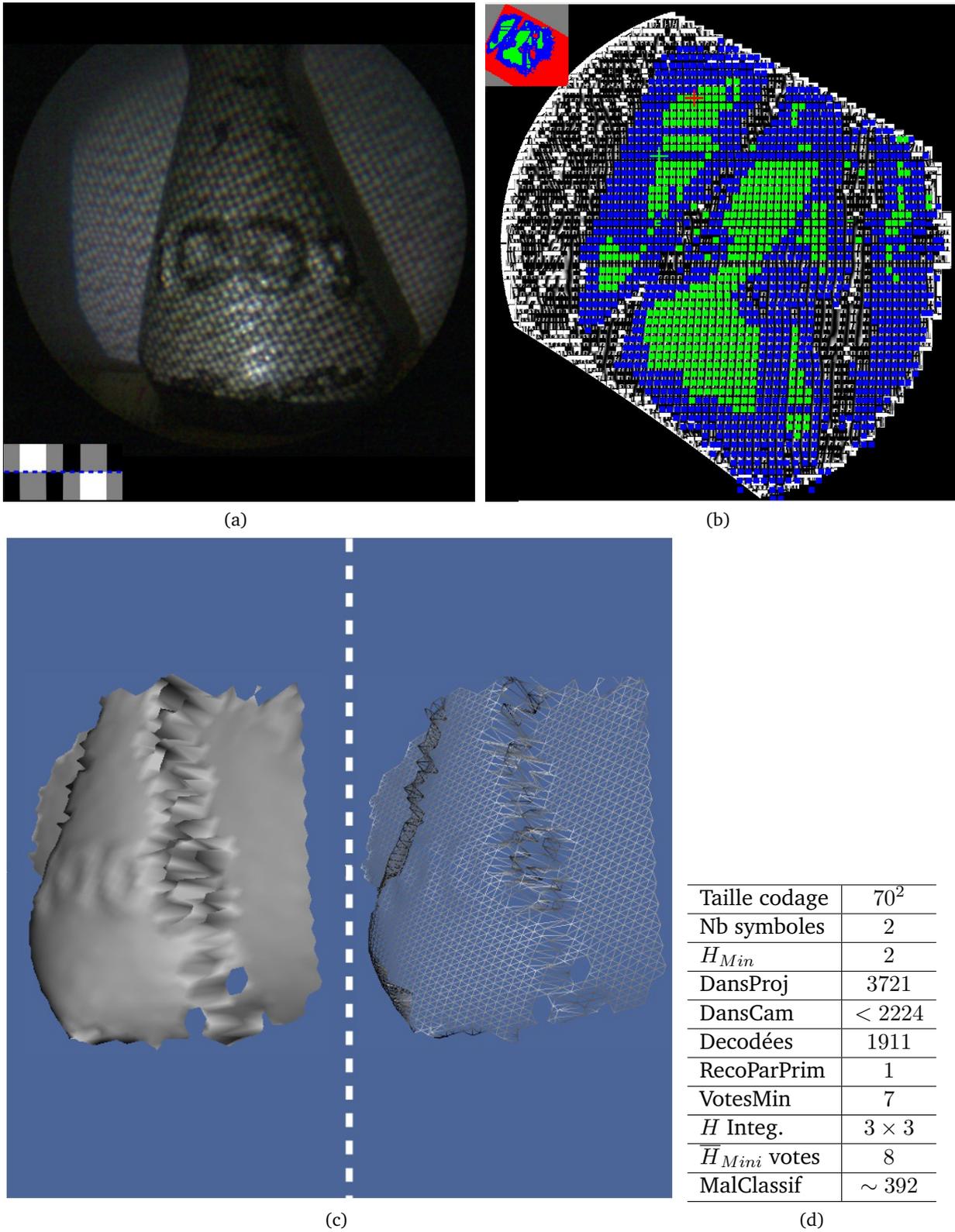


FIGURE 4.23 – Reconstruction d’une masse avec détails sur son poids.

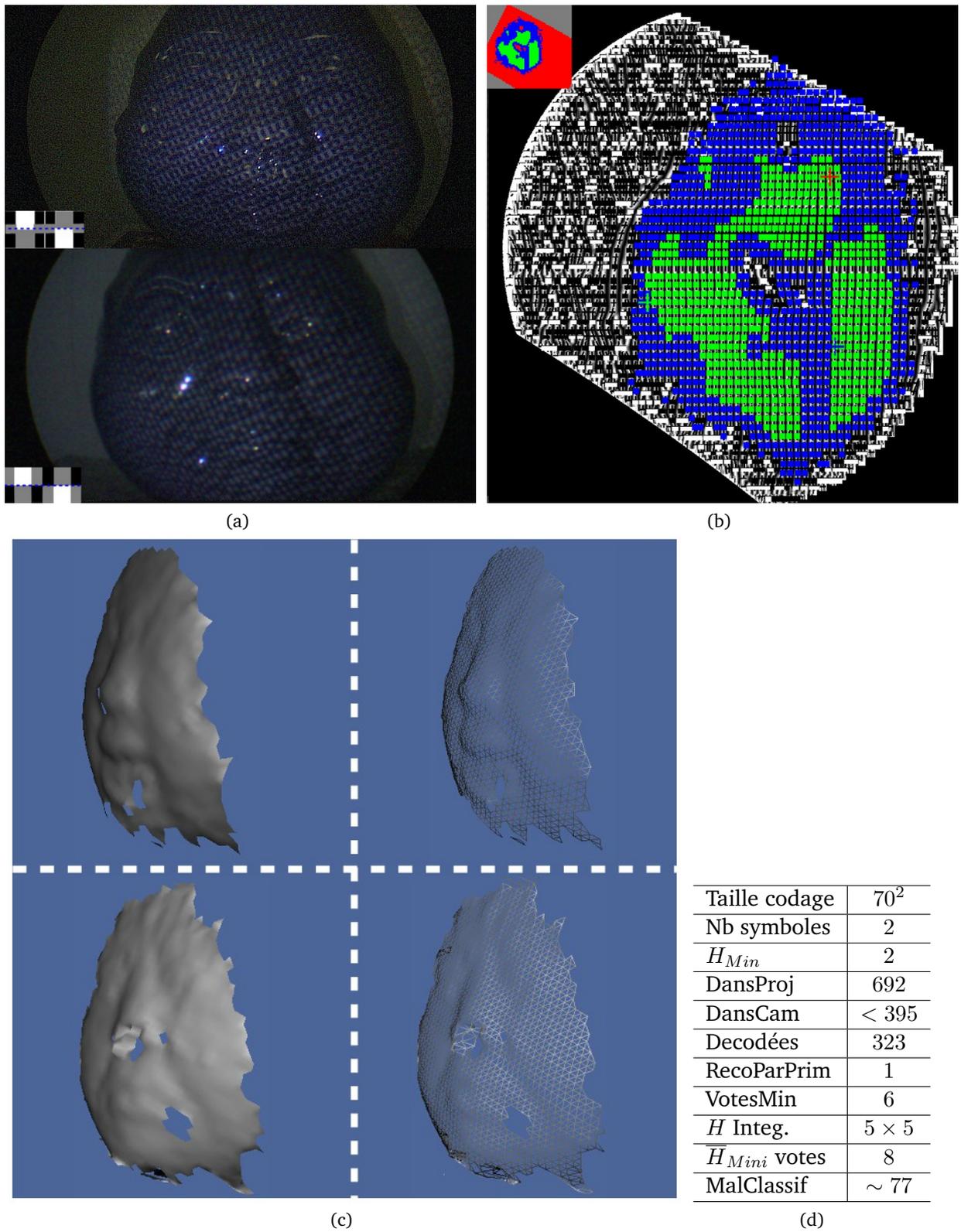


FIGURE 4.24 – Reconstructions de deux expressions (souriante, maussade) d'une figurine.



FIGURE 4.25 – a) Configuration du système stéréo manipulé pendant l'une des expérimentations In Vivo. Ici la main gauche manipule le système endoscopique stéréo, la main droite insère un autre endoscope dans le champ pour tester sa reconstruction. b) Le système endoscopique miniaturisé

les paramètres définissant les frontières des zones d'adjacence dans le graphe d'adjacence ont été ajustés (à travers le seuil portant sur l'indice IP exprimant le parallélisme intra-voisinage, voir chapitre 3) pour montrer les frontières dans l'image (b) et pour montrer une continuité des reconstructions l'image (c). Dans tous les cas, les statistiques données par le tableau (d) correspondent aux traitements montrés par l'image (b).

A partir des résultats présentés par les reconstructions, nous pouvons tirer un certain nombre de remarques, générales et spécifiques.

Analyse générale

Au vu des différentes reconstructions présentées sur les diverses surfaces intra-abdominales, nous pouvons écrire, de façon générale, que le système permet d'effectuer des reconstructions 3-D de scènes intra-abdominales et que l'approche proposée est globalement valide, avec une robustesse aux nombreuses perturbations présentes, mais un certain nombre de remarques doivent être précisées.

En effet, des reconstructions avec des motifs de résolutions 30×30 , 50×50 ou 70×70 ont pu être obtenues sur les parois intra-abdominales (Figures 4.27, 4.42, 4.32, 4.38, 4.28), sur la vésicule biliaire (Figures 4.35, 4.44, 4.45,), sur le foie (Figures 4.30, 4.35, 4.45, 4.33), l'estomac (Figures 4.38, 4.29) et diverses parties du côlon (Fig. 4.35, 4.31, 4.36). Le système est également capable de reconstruire des parties d'un instrument (un endoscope) inséré dans un second trocart (Figures 4.40, 4.41), rendant possible sa détection et son suivi par recalage à partir d'un modèle cylindrique par exemple. Au vu des images (b) illustrant les décodages et des statistiques du tableau (d), il est évident qu'il y a, de façon générale un nombre important de primitives mal

classifiées, largement au-delà des capacités de correction du codage, dont la redondance est, de plus, faible ici ($H_{Min} = 2$) car seulement 2 symboles sont utilisés pour limiter le nombre de mauvaises classifications. La qualité des reconstructions serait bien faible si seulement les primitives en vert (dans l'image (b)) étaient reconstruites. Cela rend l'approche par propagation du décodage indispensable et souligne donc également l'importance de la segmentation des franges et de la construction du graphe d'adjacence. Comme H_{Min} est faible (bien que supérieure à la majorité des motifs proposés en lumière structurée, pour lesquels $H_{Min} = 1$, voir chapitre 1), il est nécessaire qu'un fort coefficient de confiance (8 ou 9 sur une échelle de 0 à 9) soit utilisé pour définir les correspondances à partir desquelles la propagation du décodage va être effectuée. Des valeurs de " \bar{H}_{Mini} votes" inférieures à 8 menaient inévitablement à de nombreux mauvais décodages.

Points forts de l'approche

Tout d'abord, nous notons que le bruit important dans les images a pu être maîtrisé d'une certaine façon en choisissant de projeter des motifs de résolution modérée par rapport aux motifs de résolution supérieure à 100×100 utilisés avec un système stéréo plus classique. Cela nous a donné la possibilité de projeter des primitives de taille plus importante et donc d'apporter une redondance au signal transmis qui permet, in fine, d'obtenir un rapport signal/bruit suffisant pour la segmentation. Ainsi l'utilisation du motif 30×30 a souvent montré la capacité de correctement reconstruire les scènes dans leur ensemble, même si bien sûr, elles seront plus précisément reconstruites avec un motif de résolution plus élevée. Cela est visible sur les figures 4.27, 4.32 et 4.29, par exemple. Pour ces scènes, la quasi totalité des primitives présentes dans l'image ont pu être décodées et visiblement bien reconstruites.

D'autre part, une certaine robustesse aux nombreuses perturbations spectrales (veines, spécularités souvent présentes, et de façon générale, toutes les fortes variations de réflectance) peut être mise en avant. De nombreuses images illustrent cela, par exemple dans les scènes des Figures 4.29, 4.30. Cela est dû notamment à l'utilisation du filtre médian sur les champs de déformation (élongation et inclinaison) des voisinages estimés et à l'élongation locale d'histogramme utilisée pour normaliser localement (sur une zone qui dépend de l'écart entre les franges) l'intensité lumineuse aux abords des franges. Par exemple dans les reconstructions de la vésicule biliaire (Figures 4.44 et 4.45), les spécularités et le réseau sanguin, très fortement présents, ne viennent perturber la segmentation des franges que très faiblement et la propagation du décodage peut s'opérer sans difficulté majeure.

Points faibles de l'approche

Des effets de bord de la propagation du décodage sont visibles dans plusieurs des reconstructions présentées (Fig. 4.33 et 4.34 par exemple). Dans ces cas, la propagation se fait dans des zones non adjacentes qui n'ont pas pu être détectées lors de l'établissement du graphe d'adjacence et des appariements ne correspondant pas à la réalité sont effectués et conduisent alors à des reconstructions ne correspondant à rien dans la scène, soit des artefacts. Nous avons en effet noté au chapitre 3 que certaines configurations de discontinuités de profondeur de voisinages ne pouvaient être détectées par des déformations géométriques dans l'image et qu'il fallait alors utiliser la propagation du décodage qui permet d'obtenir la distance de Hamming la plus faible. Cependant cela n'est valable que si une des primitives appartenant à la bonne zone d'adjacence a pu être décodée parmi les appariements "sûrs" (en vert dans les images (b)) pour servir de point de départ à une propagation dans cette zone, ce qui n'est pas toujours le cas. Nous notons cependant que ces effets de bord n'apparaissent que dans un nombre limité des reconstructions présentées.

Le meilleur motif

Il est difficile de dresser une conclusion générale et définitive quant au choix de la meilleure résolution de motif ou des meilleures formes de primitives à utiliser. Tout dépendra bien sûr de l'application souhaitée. Nous notons cependant qu'utiliser un motif permettant la projection de primitives de taille important (comme les motifs 30×30 ou 50×50 , dans lesquels chaque primitive occupe une zone de près de 30^2 pixels), apportera le plus de robustesse au bruit et aux perturbations spectrales dans la durée. Avec le motif de taille 70×70 , la plupart des tests ne donnait pas de bons résultats, les reconstructions montrées ici avec un tel motif étaient réalisées sur la paroi latérale interne de l'abdomen ou la vésicule biliaire, qui possèdent une bonne réflectance par rapport à d'autres organes. Pour pouvoir utiliser des motifs de résolution supérieure sur la majorité des surfaces présentes dans l'abdomen, il serait nécessaire d'améliorer tout d'abord le rapport signal/bruit par des modifications sur le prototype. Concernant le choix final des primitives, nous laissons la question ouverte. Les reconstructions effectuées avec un motif 30×30 avec différents ensembles de primitives ne permet pas de faire ressortir de façon évidente un ensemble de primitives qui aurait un taux de mauvaises classifications inférieur à d'autres, de plus toutes les primitives utilisées ont permis une segmentation satisfaisante des franges.

4.5 Conclusion

Dans ce chapitre, nous avons traité les différents aspects relatifs aux problématiques de reconstruction 3-D, d'estimation et de correction des distorsions géométriques pour la caméra et le projecteur, et d'estimation de la matrice Fondamentale.

Tout d'abord, les modélisations exploitées, modèles de projection, vision stéréo, la stratification ont été présentés. Ceux-ci nous ont mené à proposer une reconstruction 3-D réalisée dans l'espace projectif et exploitant une contrainte de quasi-orthogonalité. Celle-ci exploite un invariant projectif et la géométrie épipolaire pour retrouver une position de coplanarité des points à reconstruire par rapport à un plan de référence défini par trois points appariés choisis dans la scène. La profondeur est alors retrouvée grâce à la parallaxe entre cette position de coplanarité et la position obtenue à l'issue de la segmentation des primitives. Des reconstructions peu distordues par rapport à des reconstructions euclidiennes peuvent alors être effectuées. Puis les différentes phases menant à l'estimation de la géométrie épipolaire, par le biais de la matrice Fondamentale F ont été décrites. En particulier, une **nouvelle méthodologie** permettant d'estimer individuellement les paramètres de distorsion de la caméra et du projecteur a été proposée. Les résultats de l'estimation de la matrice Fondamentale, pour un système stéréo non endoscopique, montrent que, comparé à l'état de l'art, la précision obtenue est, avec 0.08 pixels, au niveau des meilleurs résultats. Pour le système endoscopique, les résultats sont moins bons, mais les distorsions (radiales et tangentielles) ont pu être suffisamment corrigées pour que l'invariance projective des lignes épipolaires puisse être exploitée pour l'alignement épipolaire du motif avec le système stéréo endoscopique.

Enfin, de nombreuses reconstructions 3-D réalisées dans diverses configurations (système non endoscopique / endoscopiques ex-vivo / endoscopique in-vivo) ont été présentées. Pour chacune d'entre elles, de nombreuses informations liées aux traitements, à la segmentation, au décodage, ont été données. Elles nous ont permis de dégager les points forts et les points faibles de l'approche proposée. Ces reconstructions mettent en évidence la robustesse générale

de notre approche face à des perturbations géométriques et spectrales. Finalement, différents motifs, réalisés avec des primitives différentes et possédant des résolutions différentes, ont été testés sur des scènes In-vivo en situation mini-invasive. Des reconstructions ont pu être obtenues sur la majorité des surfaces des organes (foie, estomac, intestin, paroi latérale, vésicule biliaire) présents dans la cavité intra-abdominale de porc, malgré la présence importante de bruit dans les images. La vitesse moyenne actuelle de l'ensemble des traitements est de 12 reconstructions par seconde.

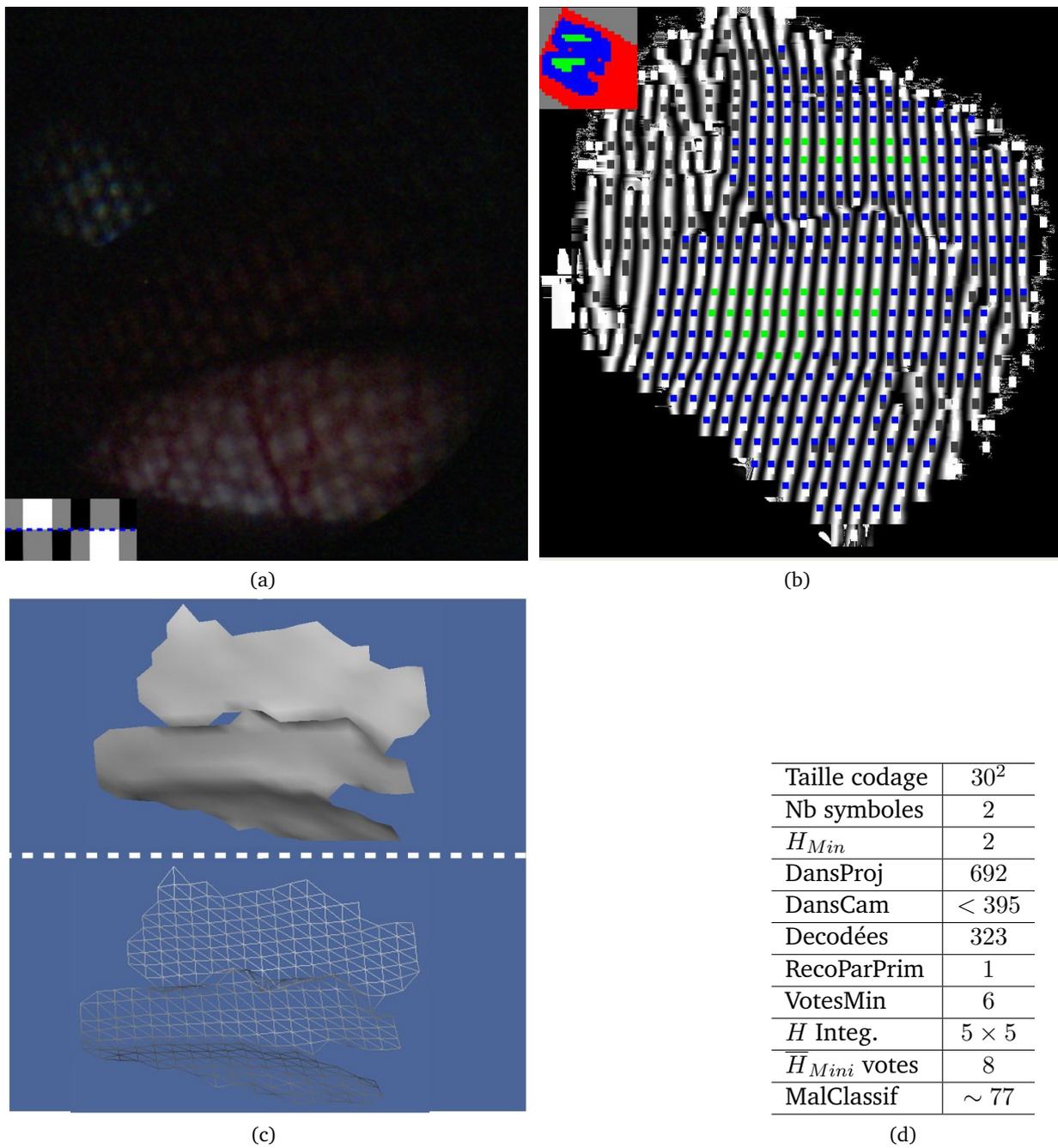
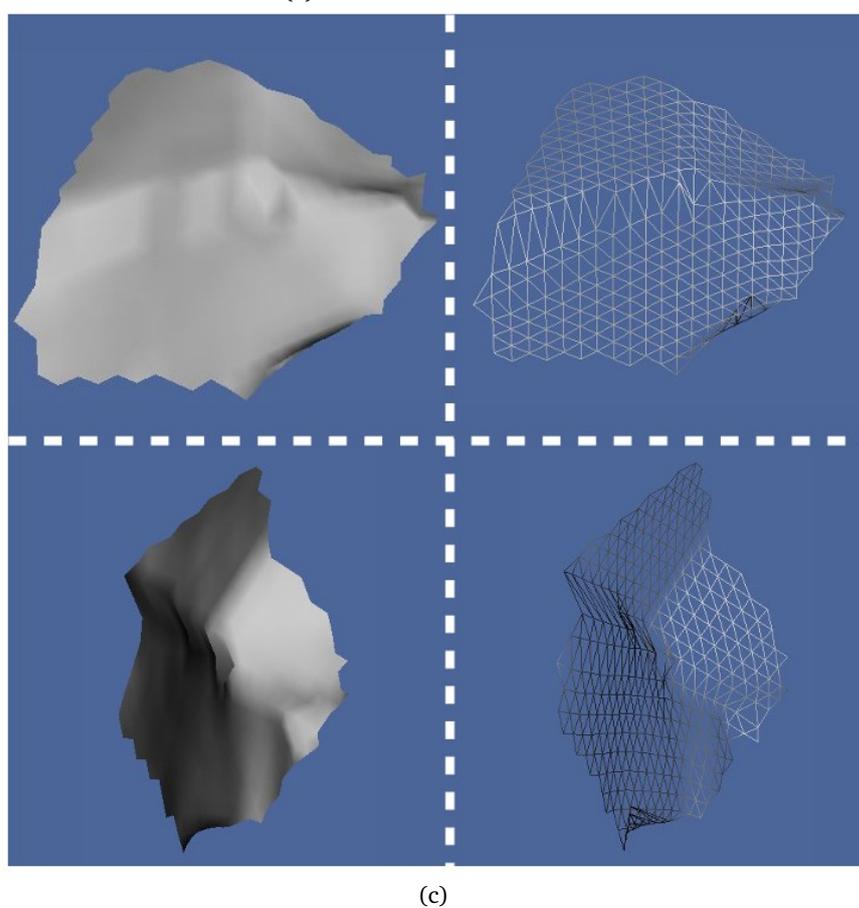
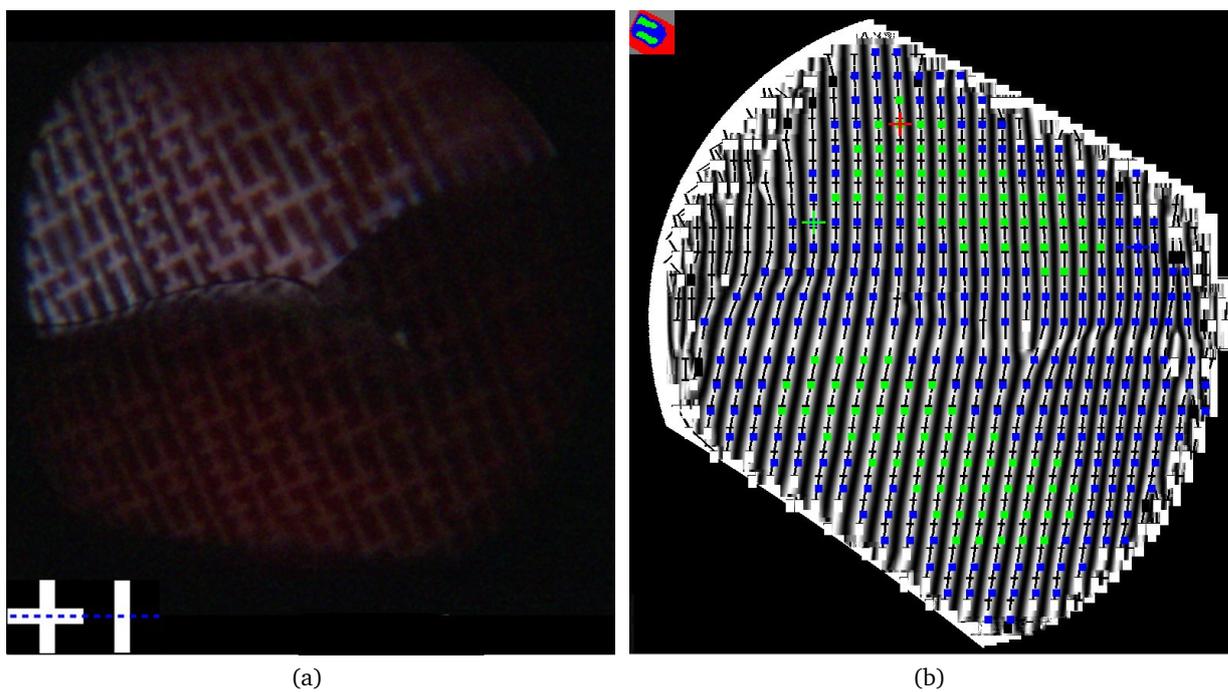
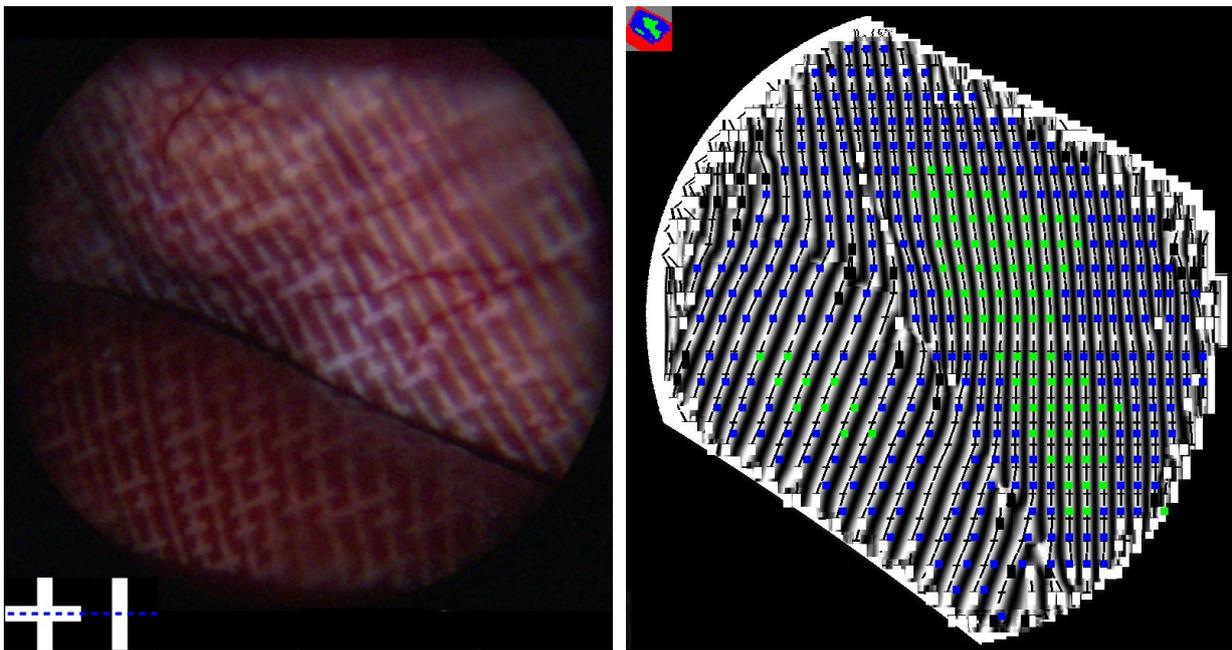


FIGURE 4.26 – De bas en haut : l'estomac, le foie et la paroi.



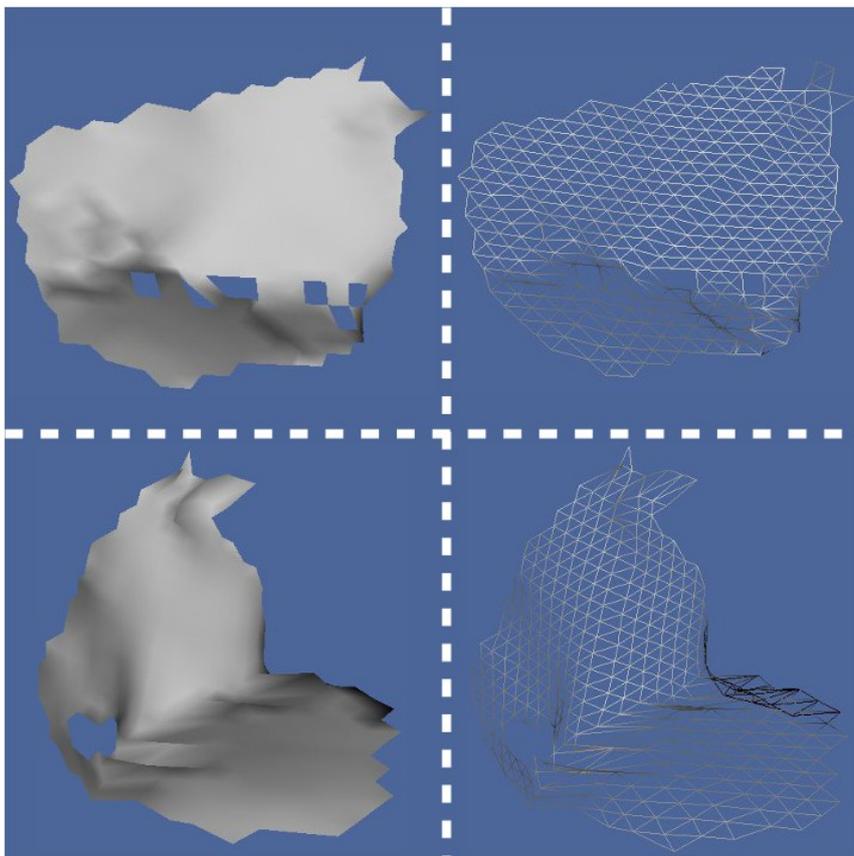
Taille codage	30^2
Nb symboles	2
H_{Min}	2
DansProj	692
DansCam	< 395
Decodées	339
RecoParPrim	1
VotesMin	7
H Integ.	3×3
\bar{H}_{Mini} votes	8
MalClassif	~ 29

FIGURE 4.27 – De bas en haut : le foie et la paroi.



(a)

(b)

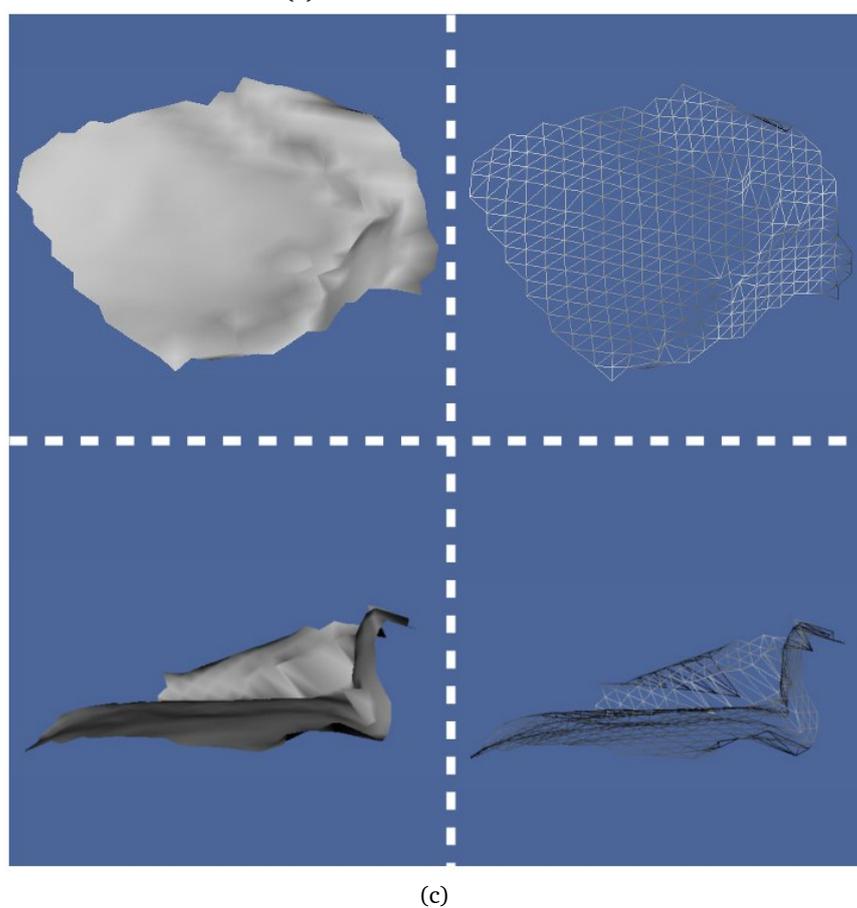
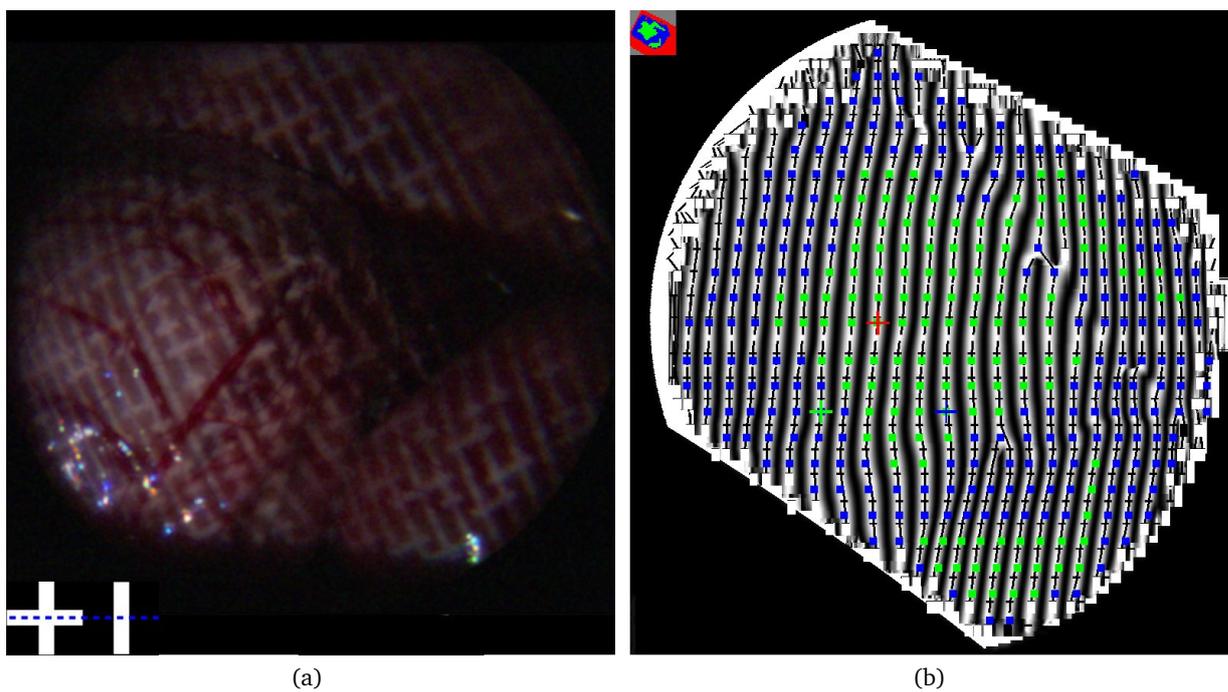


(c)

(d)

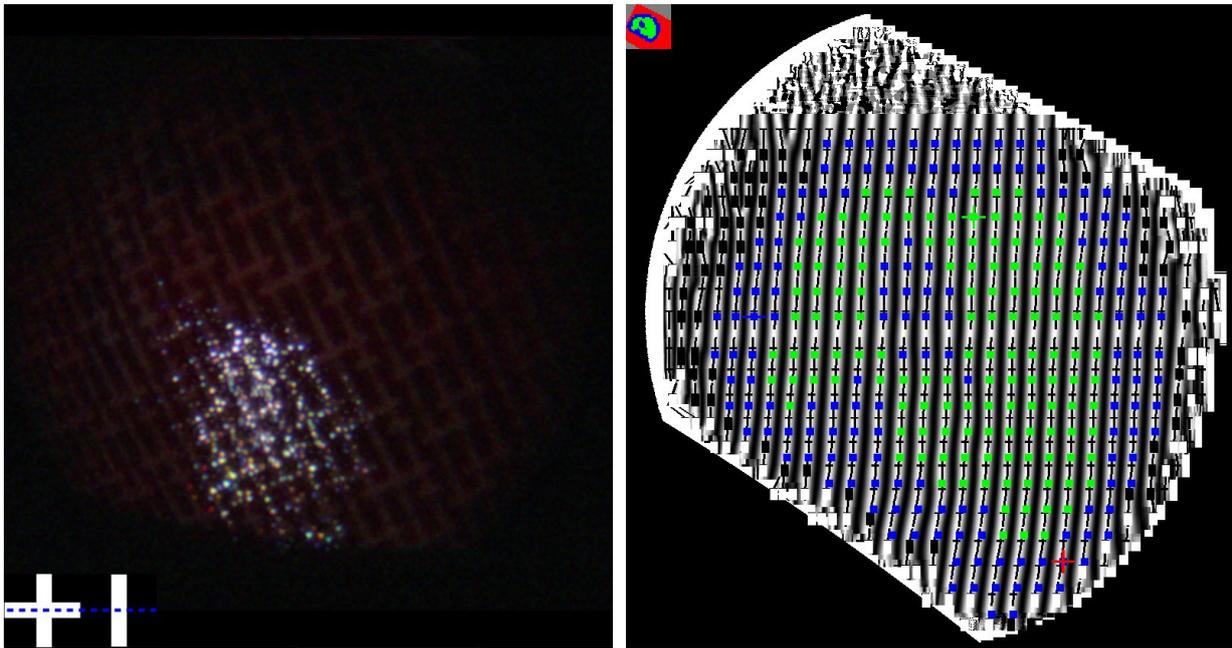
Taille codage	30^2
Nb symboles	2
H_{Min}	2
DansProj	692
DansCam	< 395
Decodées	364
RecoParPrim	1
VotesMin	7
H Integ.	3×3
\overline{H}_{Mini} votes	8
MalClassif	~ 49

FIGURE 4.28 – De bas en haut : le foie et la paroi.



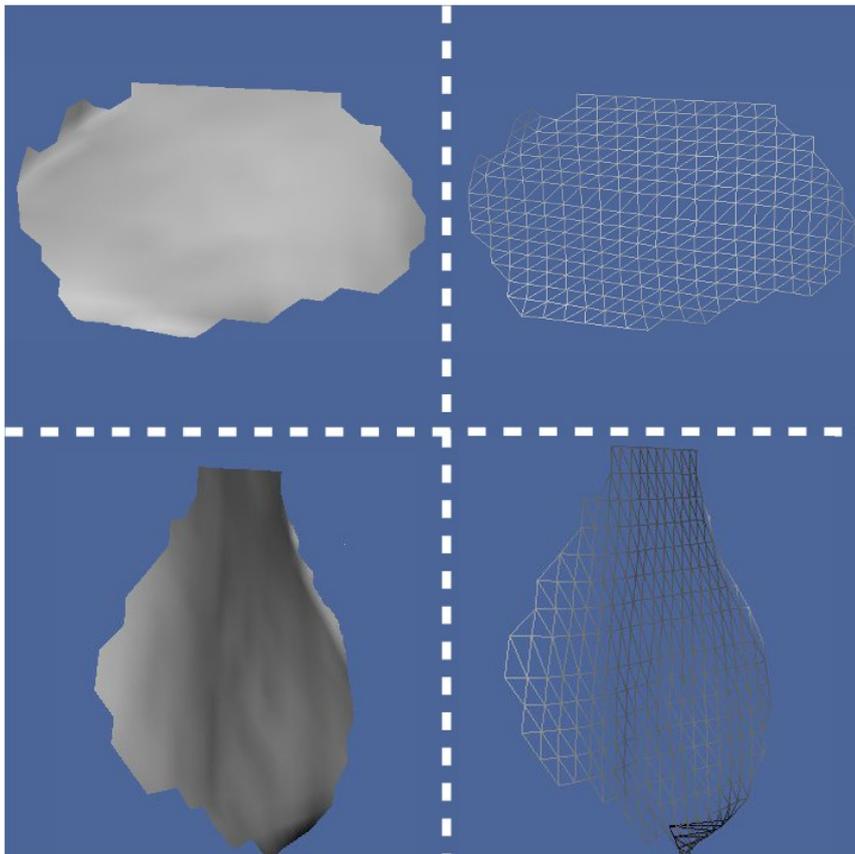
Taille codage	30^2
Nb symboles	2
H_{Min}	2
DansProj	692
DansCam	< 395
Decodées	354
RecoParPrim	1
VotesMin	7
H Integ.	1×1
\overline{H}_{Mini} votes	8
MalClassif	~ 63

FIGURE 4.29 – L'estomac (sur la gauche) et l'intestin.



(a)

(b)

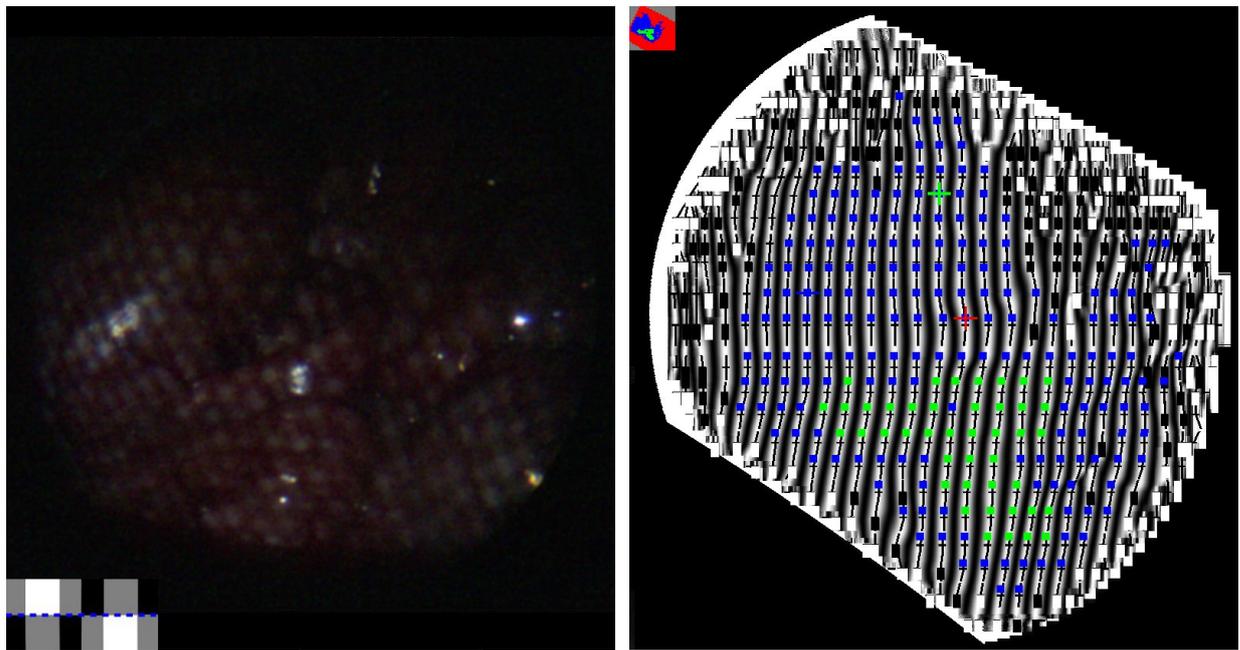


(c)

(d)

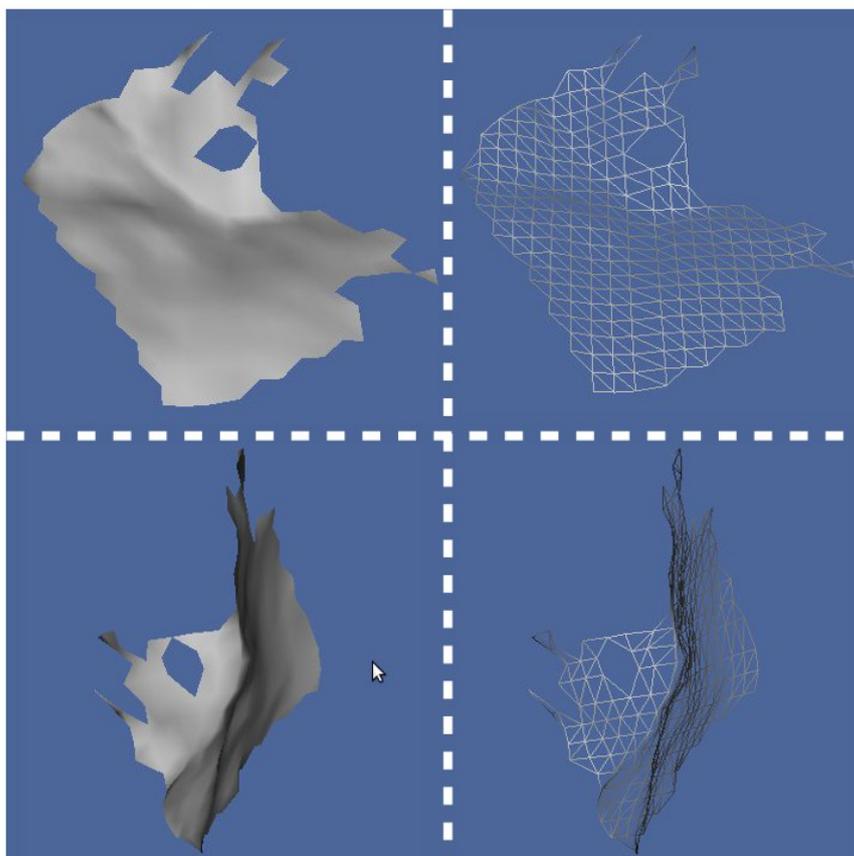
Taille codage	30^2
Nb symboles	2
H_{Min}	2
DansProj	692
DansCam	< 395
Decodées	288
RecoParPrim	1
VotesMin	7
H Integ.	3×3
\bar{H}_{Mini} votes	8
MalClassif	~ 17

FIGURE 4.30 – Le foie.



(a)

(b)

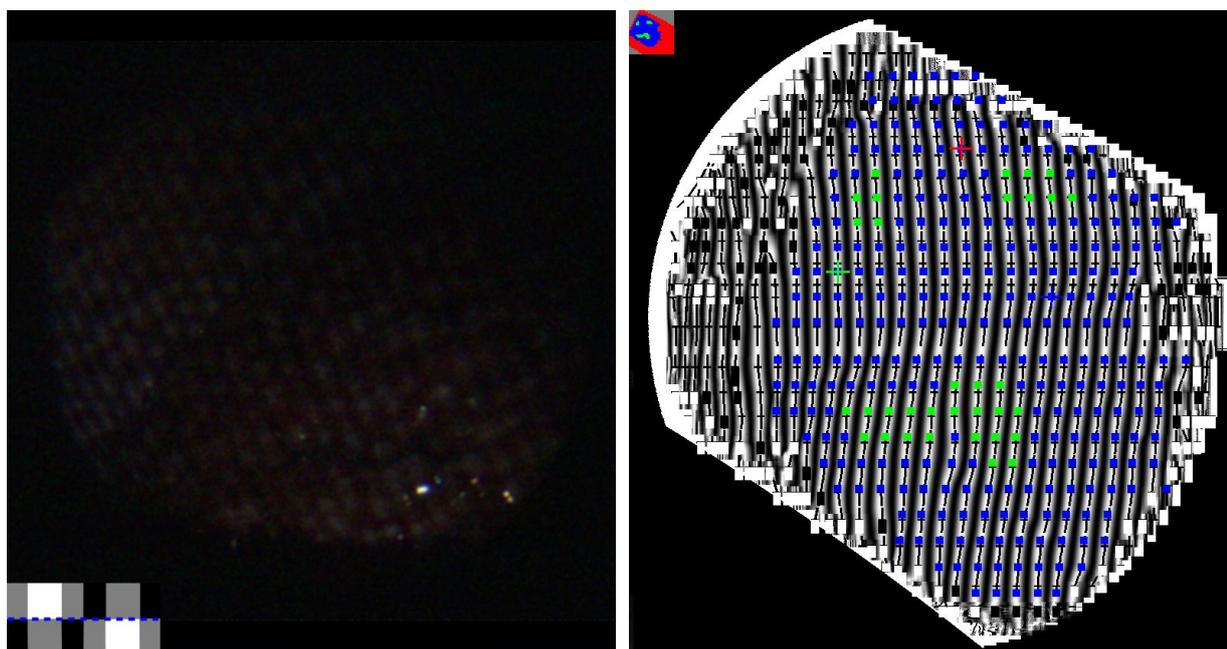


(c)

(d)

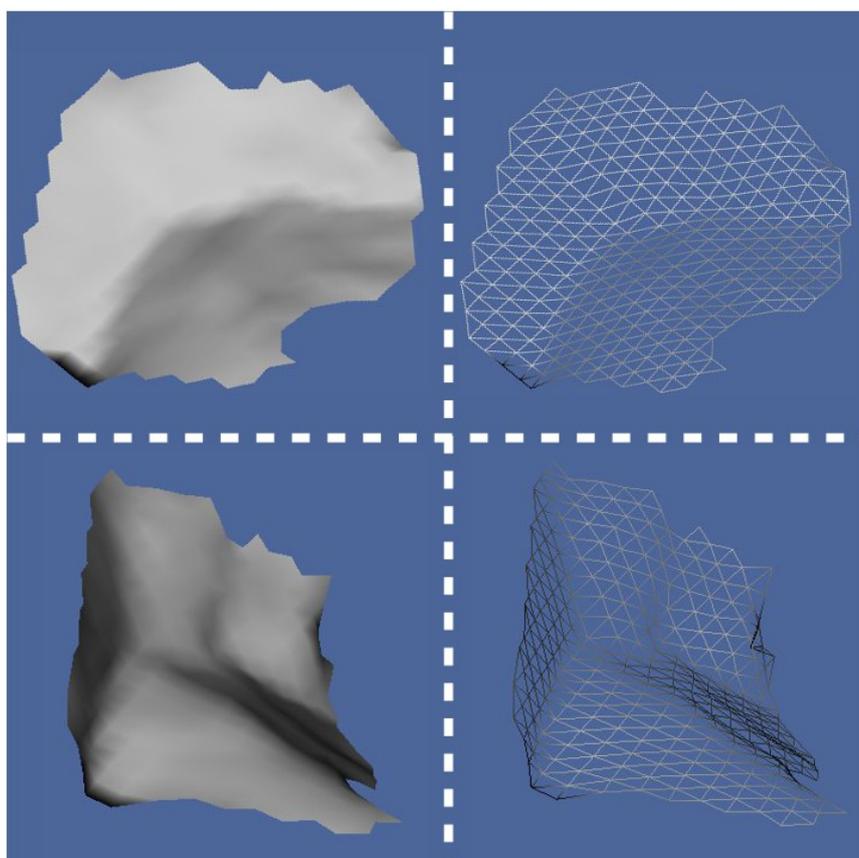
Taille codage	30^2
Nb symboles	2
H_{Min}	2
DansProj	692
DansCam	< 395
Decodées	230
RecoParPrim	1
VotesMin	4
H Integ.	3×3
\bar{H}_{Mini} votes	8
MalClassif	~ 66

FIGURE 4.31 – L'intestin grêle.



(a)

(b)



(c)

(d)

Taille codage	30^2
Nb symboles	2
H_{Min}	2
DansProj	692
DansCam	< 395
Decodées	302
RecoParPrim	1
VotesMin	7
H Integ.	3×3
\bar{H}_{Mini} votes	8
MalClassif	~ 51

FIGURE 4.32 – L'intestin grêle et la paroi.

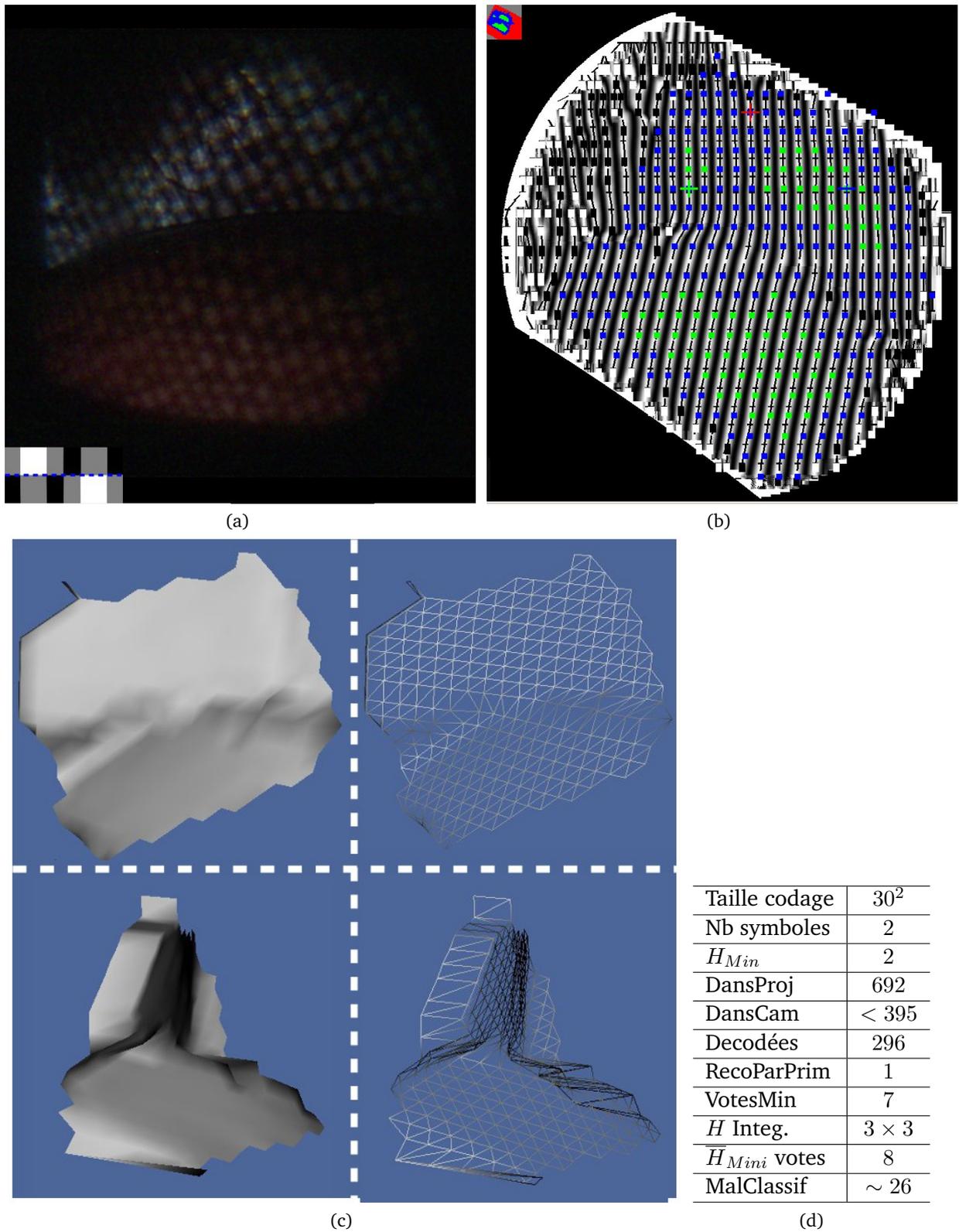
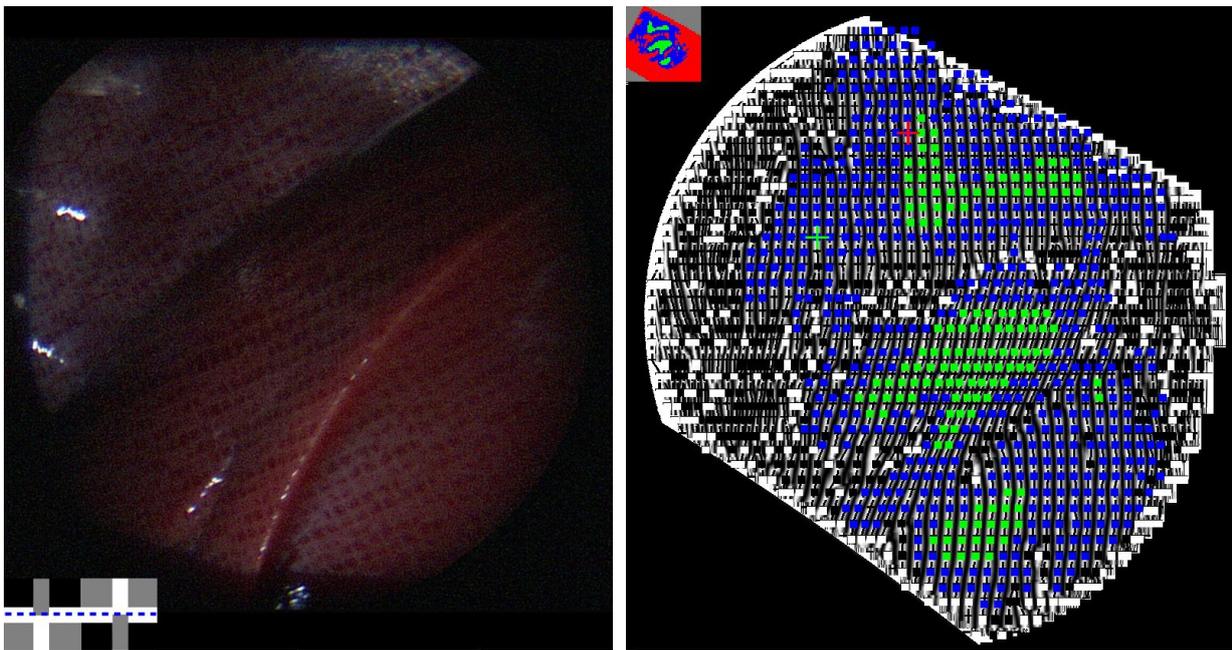
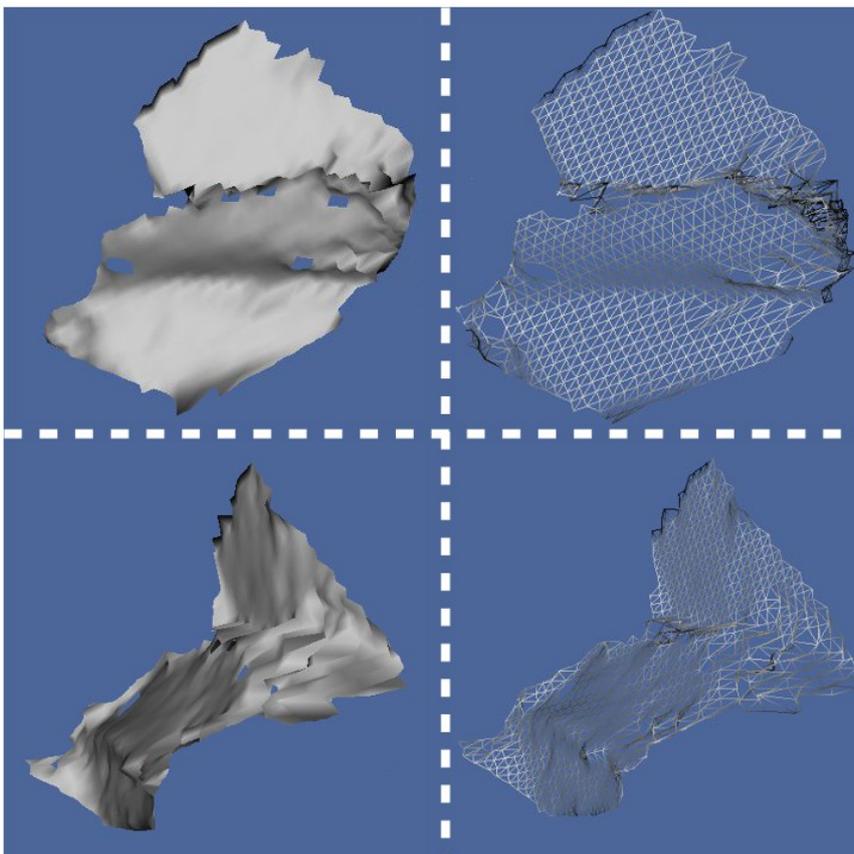


FIGURE 4.33 – Le foie et la paroi.



(a)

(b)

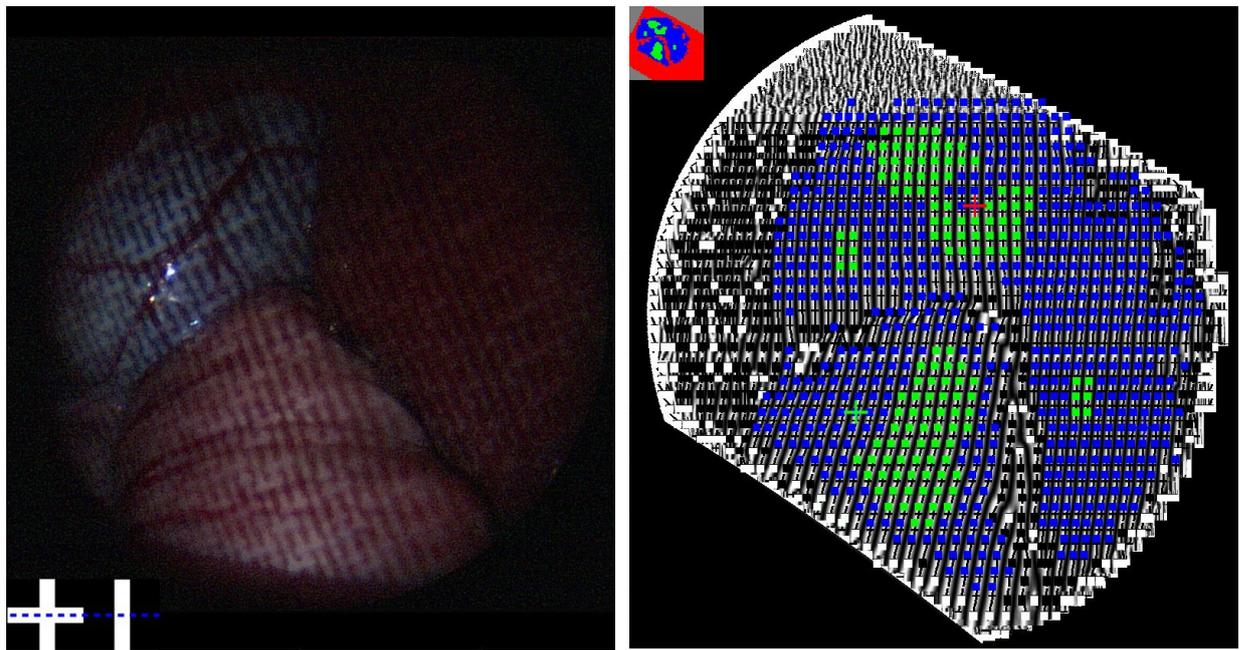


(c)

(d)

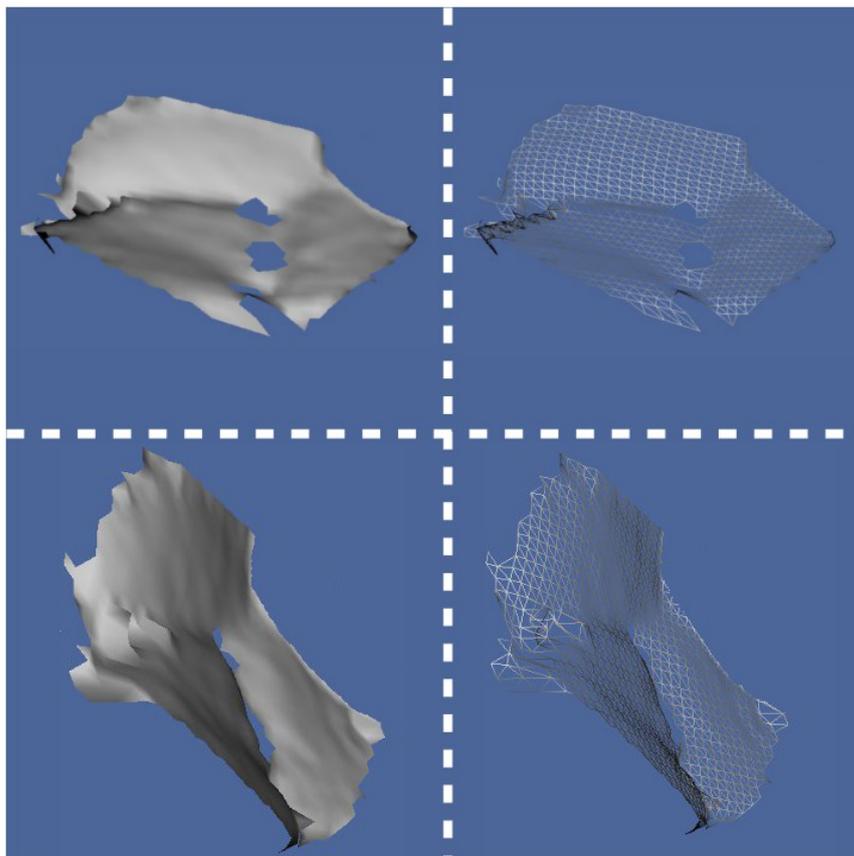
Taille codage	50^2
Nb symboles	2
H_{Min}	2
DansProj	1905
DansCam	< 1086
Decodées	730
RecoParPrim	1
VotesMin	7
H Integ.	3×3
\bar{H}_{Mini} votes	8
MalClassif	~ 148

FIGURE 4.34 – Le foie et la paroi.



(a)

(b)

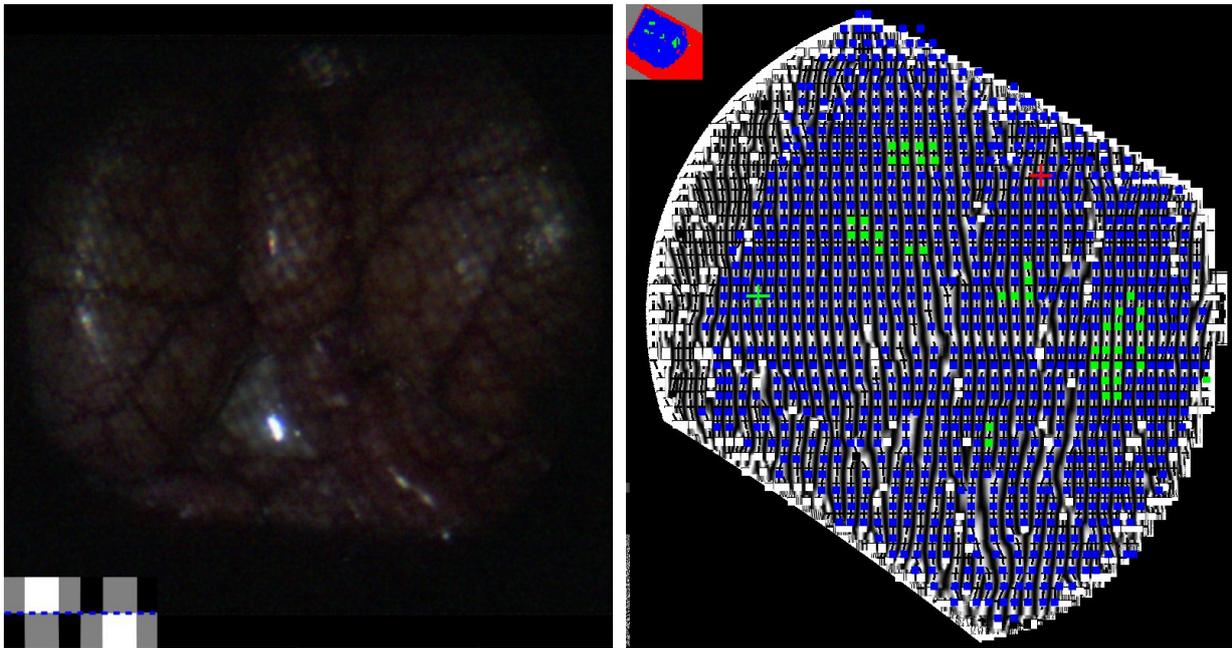


(c)

(d)

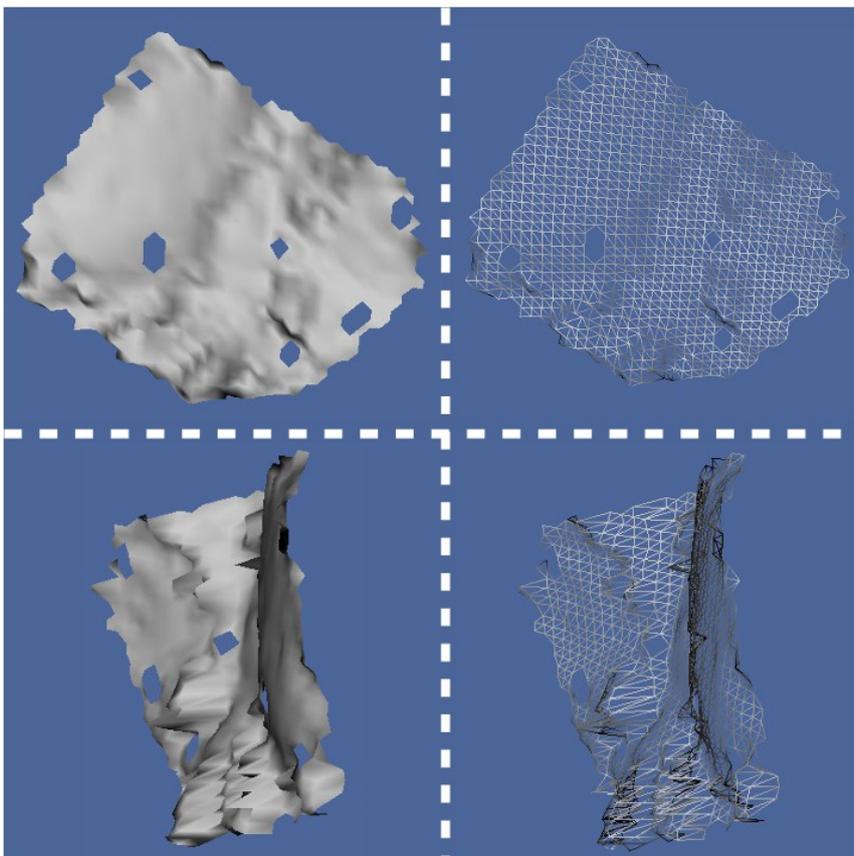
Taille codage	50^2
Nb symboles	2
H_{Min}	2
DansProj	1905
DansCam	< 1086
Decodées	841
RecoParPrim	1
VotesMin	5
\bar{H}_{Mini} votes	8
H_{Min} Integ.	3×3
MalClassif	~ 217

FIGURE 4.35 – La vésicule biliaire (en bleu), l'estomac (côté inférieur) et le foie (côté droit).



(a)

(b)

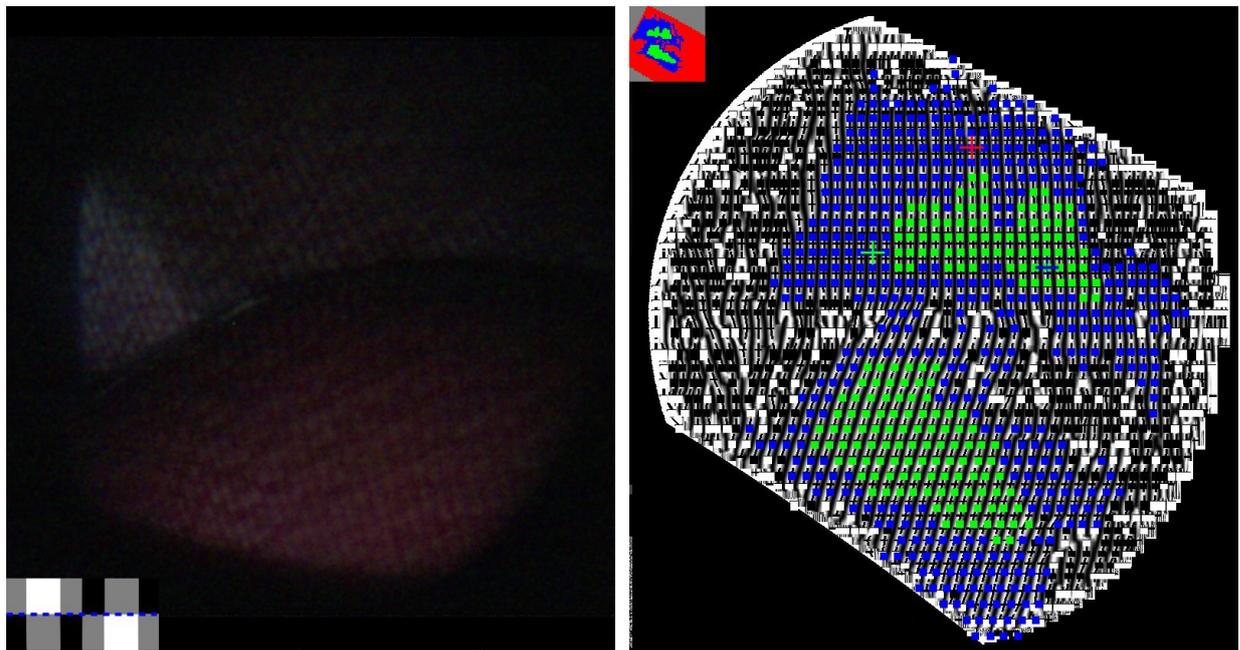


(c)

(d)

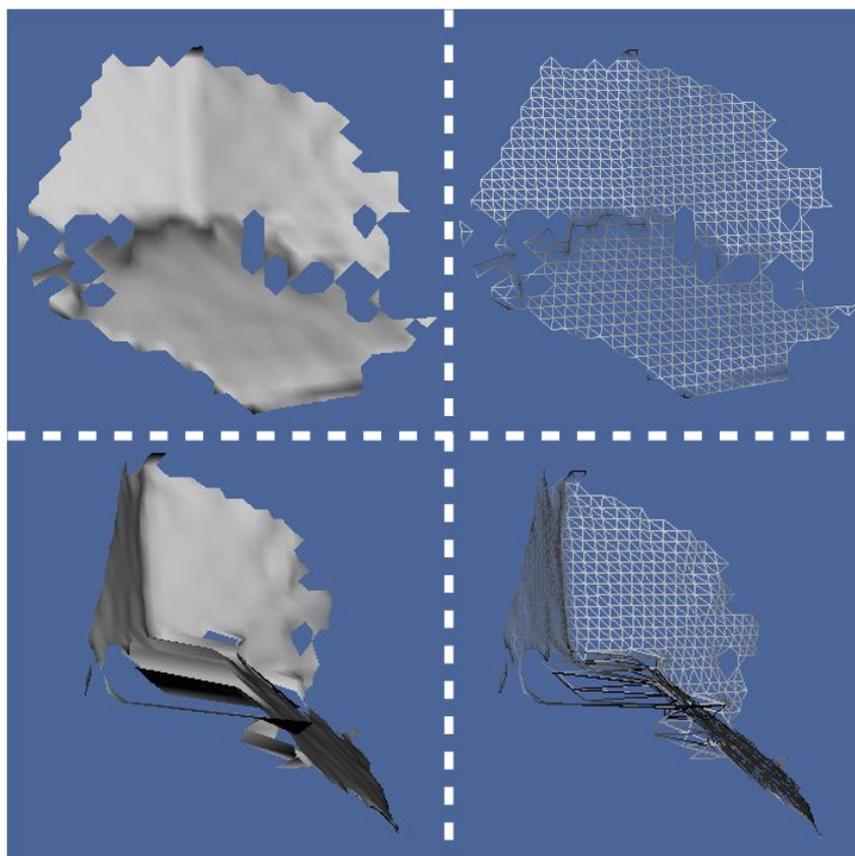
Taille codage	50^2
Nb symboles	2
H_{Min}	2
DansProj	1905
DansCam	< 1086
Decodées	1022
RecoParPrim	1
VotesMin	7
H Integ.	3×3
\bar{H}_{Mini} votes	8
MalClassif	~ 275

FIGURE 4.36 – L'intestin.



(a)

(b)

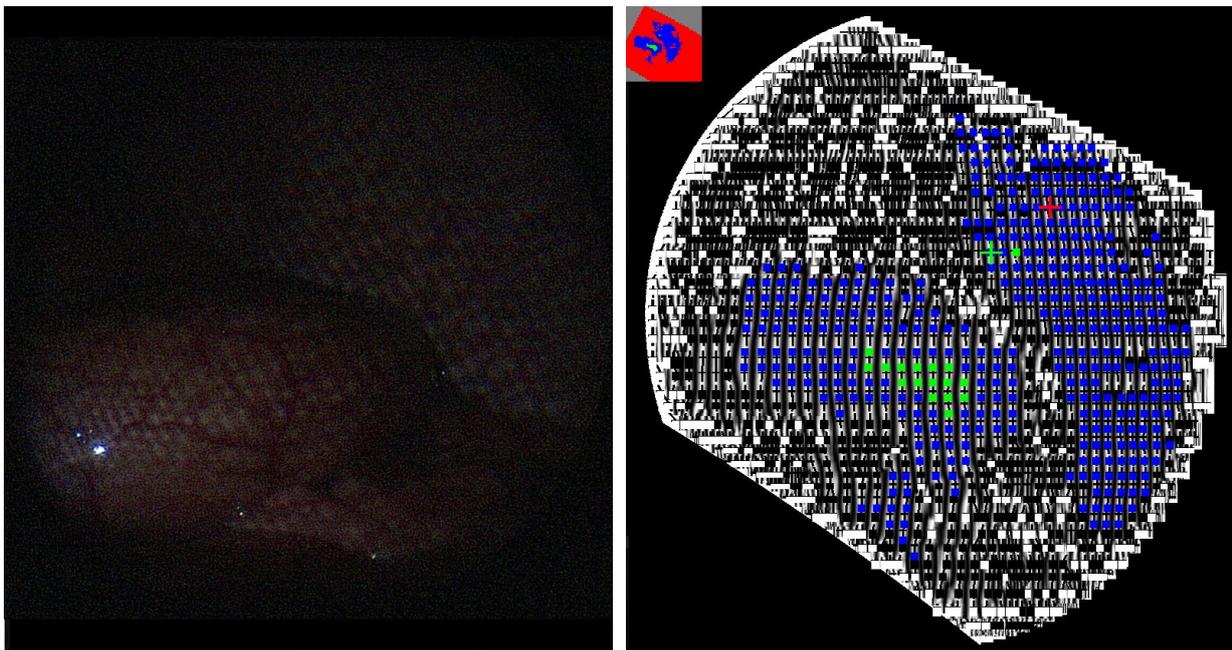


(c)

(d)

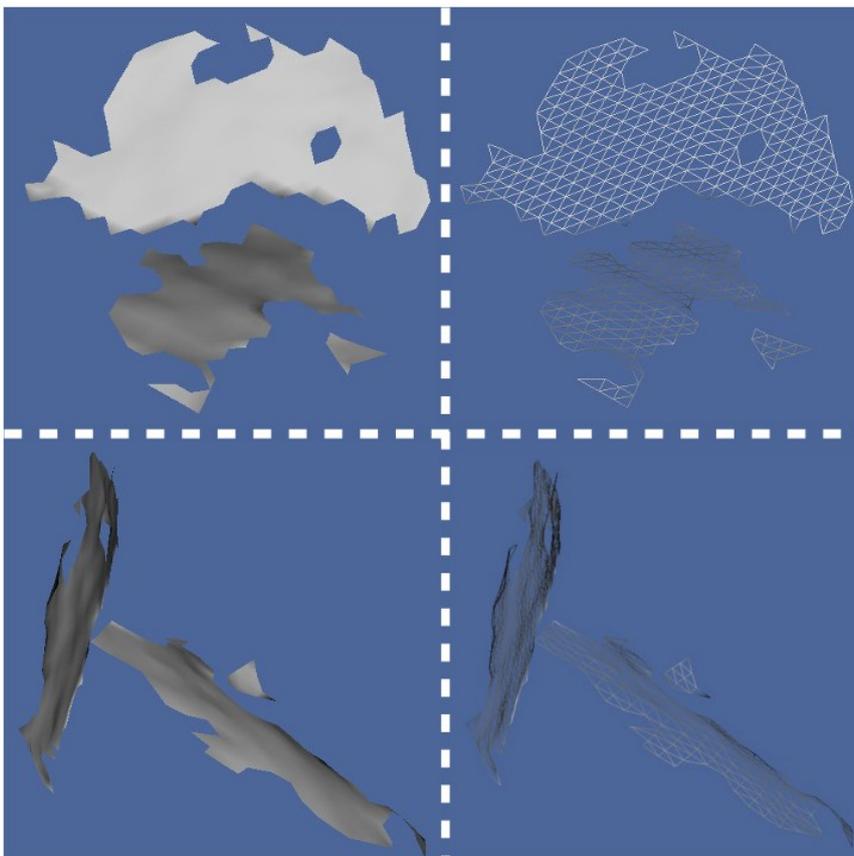
Taille codage	50^2
Nb symboles	2
H_{Min}	2
DansProj	1905
DansCam	< 1086
Decodées	678
RecoParPrim	1
VotesMin	4
H Integ.	3×3
H_{Mini} votes	8
MalClassif	~ 138

FIGURE 4.37 – Le foie et la paroi.



(a)

(b)

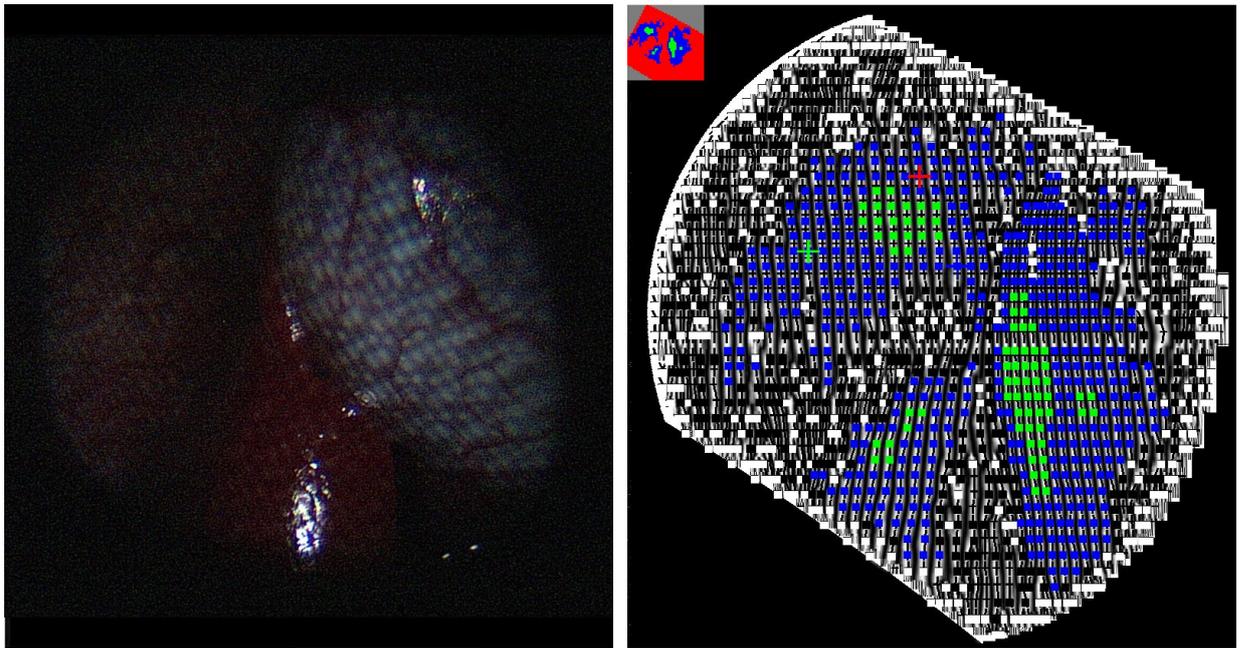


(c)

Taille codage	50^2
Nb symboles	2
H_{Min}	2
DansProj	1905
DansCam	< 1086
Decodées	410
RecoParPrim	1
VotesMin	7
H Integ.	3×3
\bar{H}_{Mini} votes	8
MalClassif	~ 115

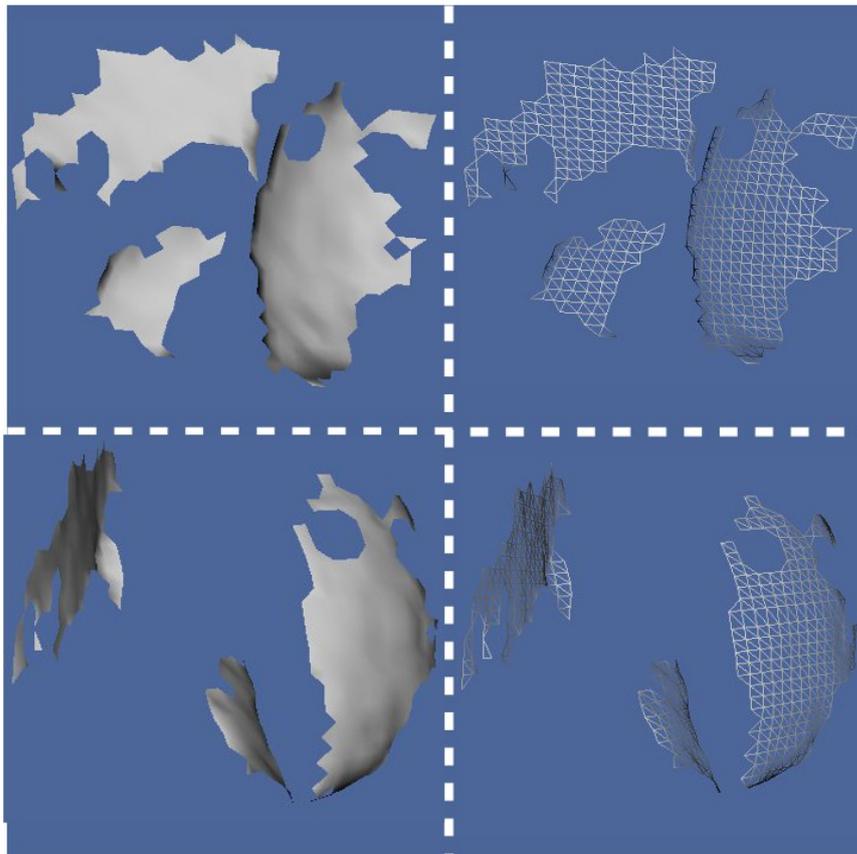
(d)

FIGURE 4.38 – L'estomac et la paroi.



(a)

(b)



(c)

(d)

Taille codage	50^2
Nb symboles	2
H_{Min}	2
DansProj	1905
DansCam	< 1086
Decodées	516
RecoParPrim	1
VotesMin	4
H Integ.	3×3
H_{Mini} votes	8
MalClassif	~ 149

FIGURE 4.39 – La paroi (côté gauche), la vésicule biliaire (en bleu), et le foie.

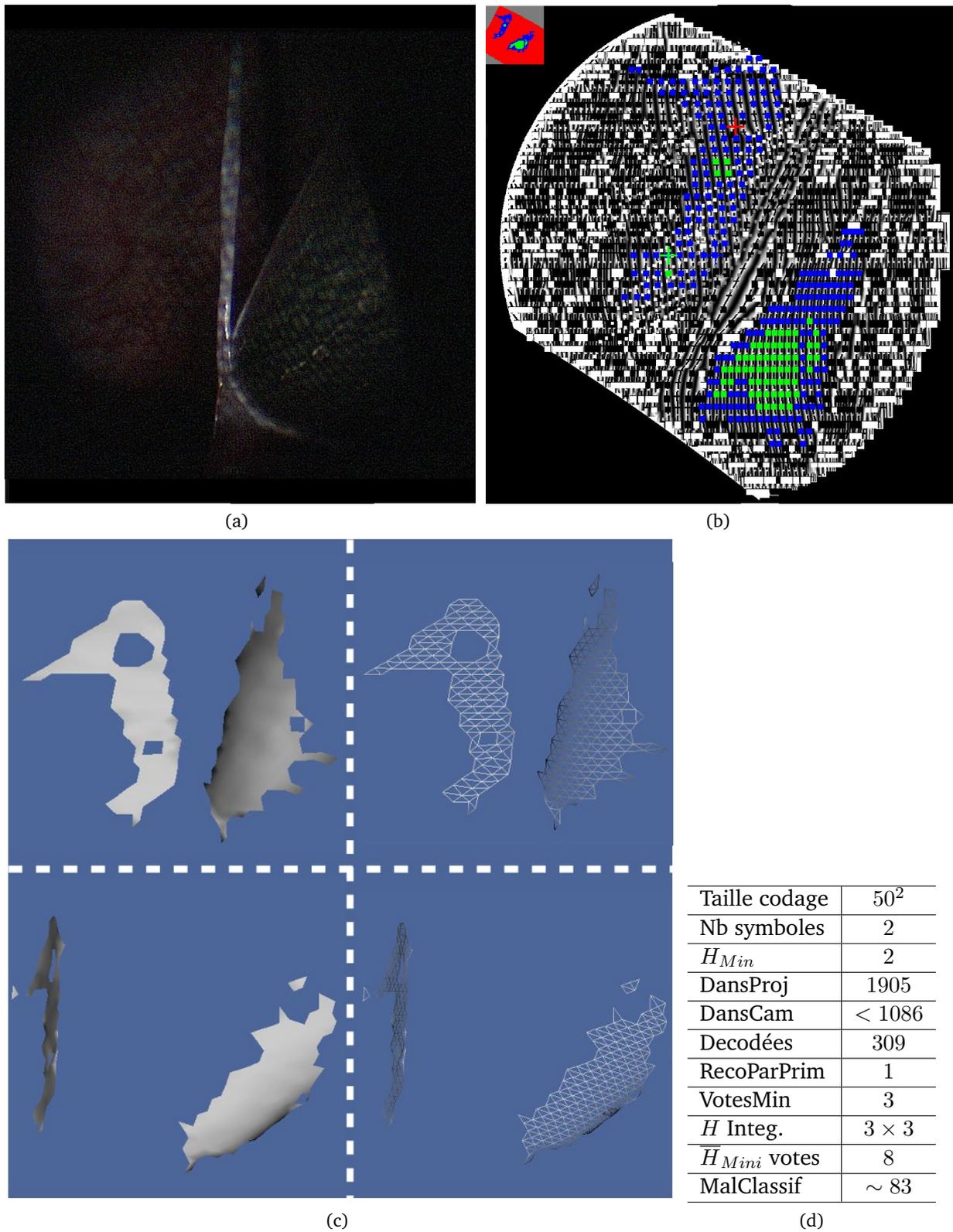


FIGURE 4.40 – La paroi et un endoscope inséré dans un deuxième trocart.

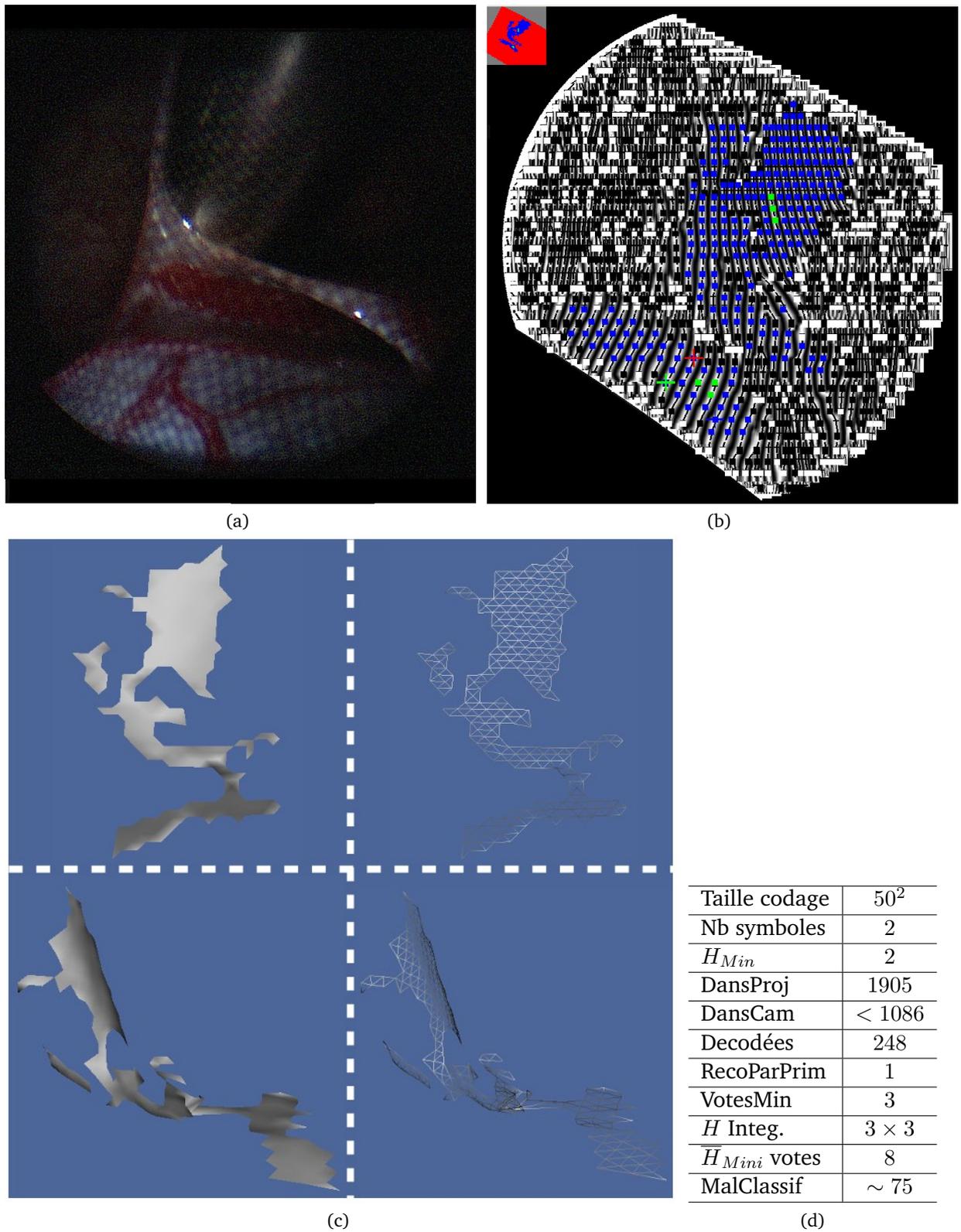
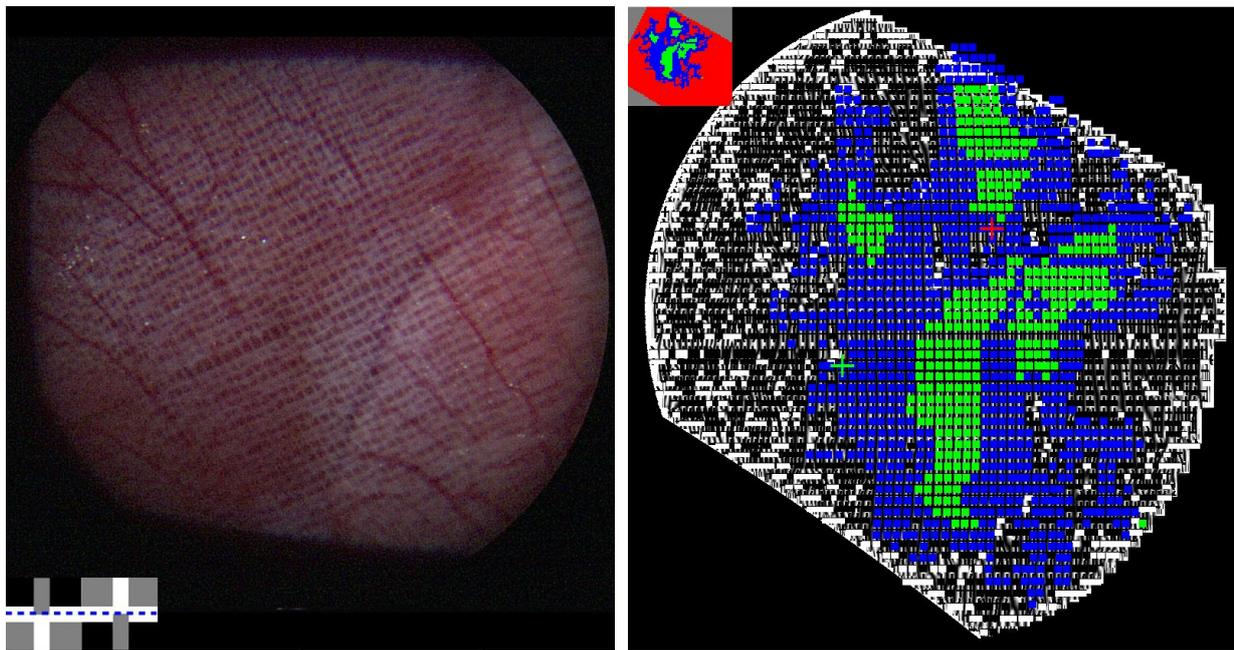
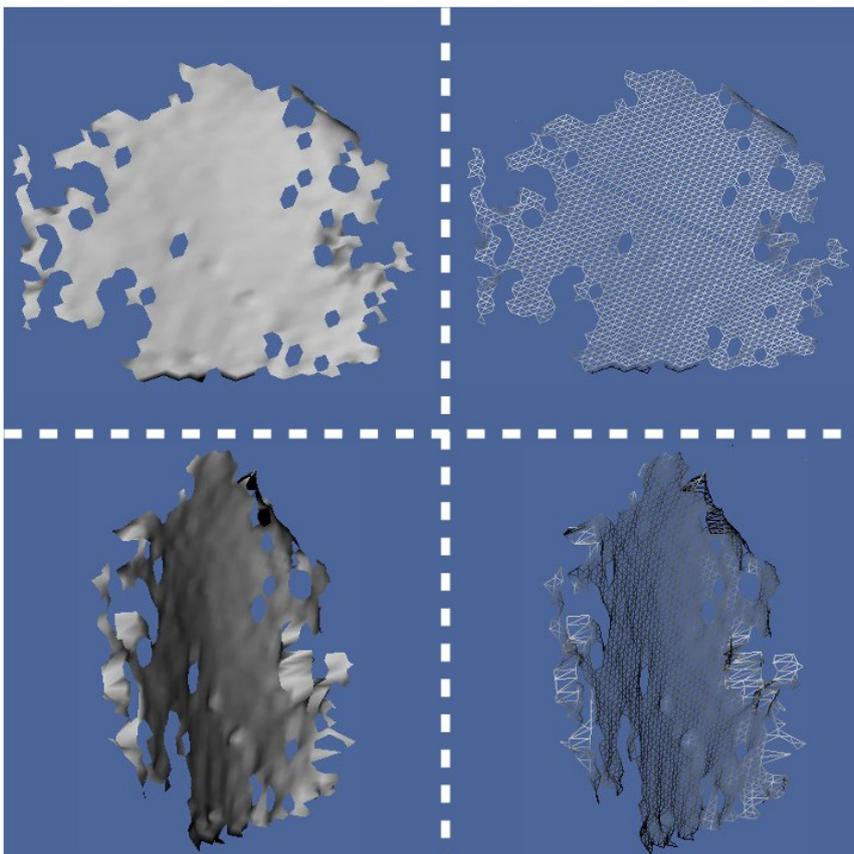


FIGURE 4.41 – La vésicule biliaire et un endoscope inséré dans un deuxième trocart.



(a)

(b)



(c)

(d)

Taille codage	70^2
Nb symboles	2
H_{Min}	2
DansProj	3721
DansCam	< 2121
Decodées	1203
RecoParPrim	1
VotesMin	7
H Integ.	3×3
\bar{H}_{Mini} votes	8
MalClassif	~ 115

FIGURE 4.42 – La paroi.

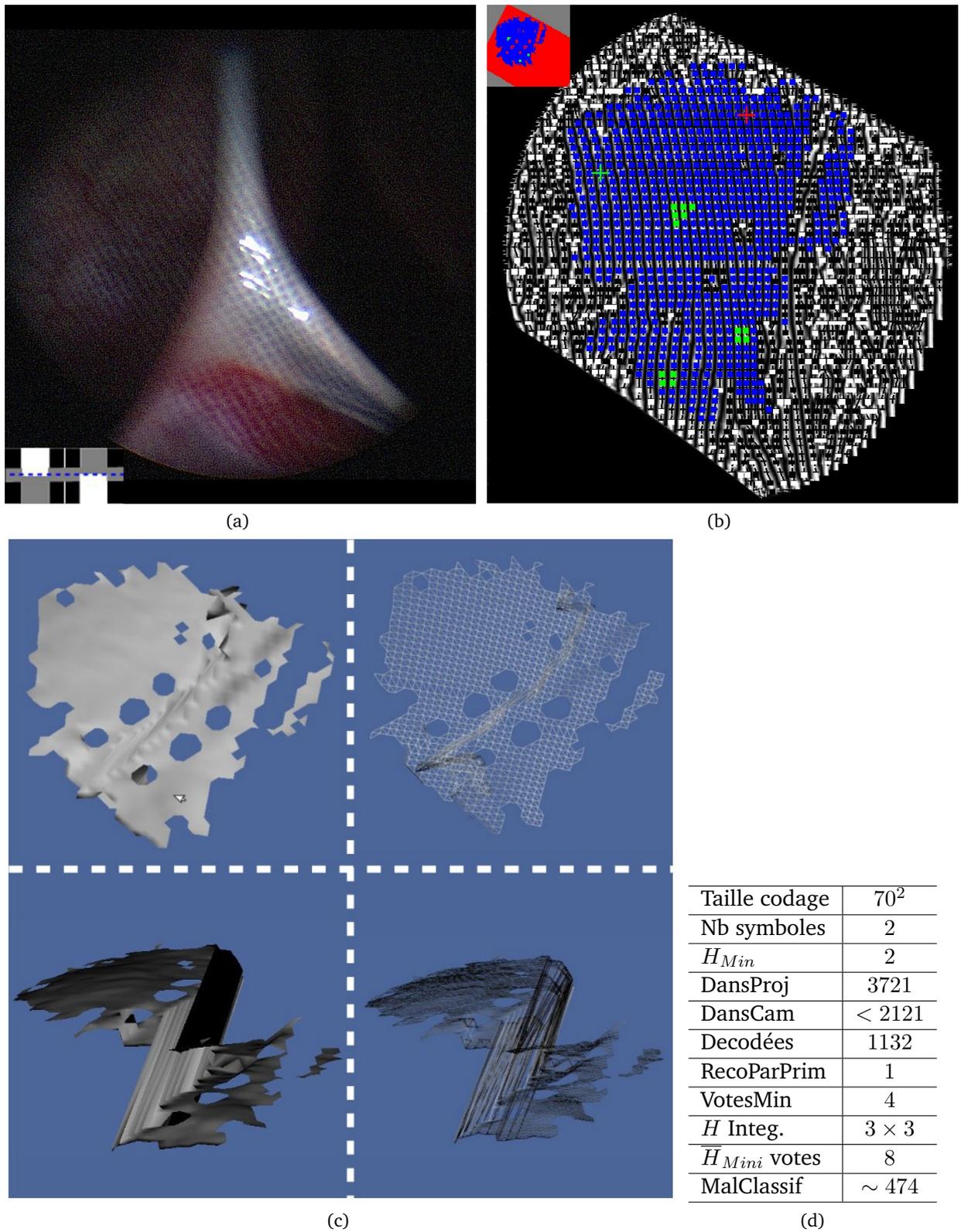
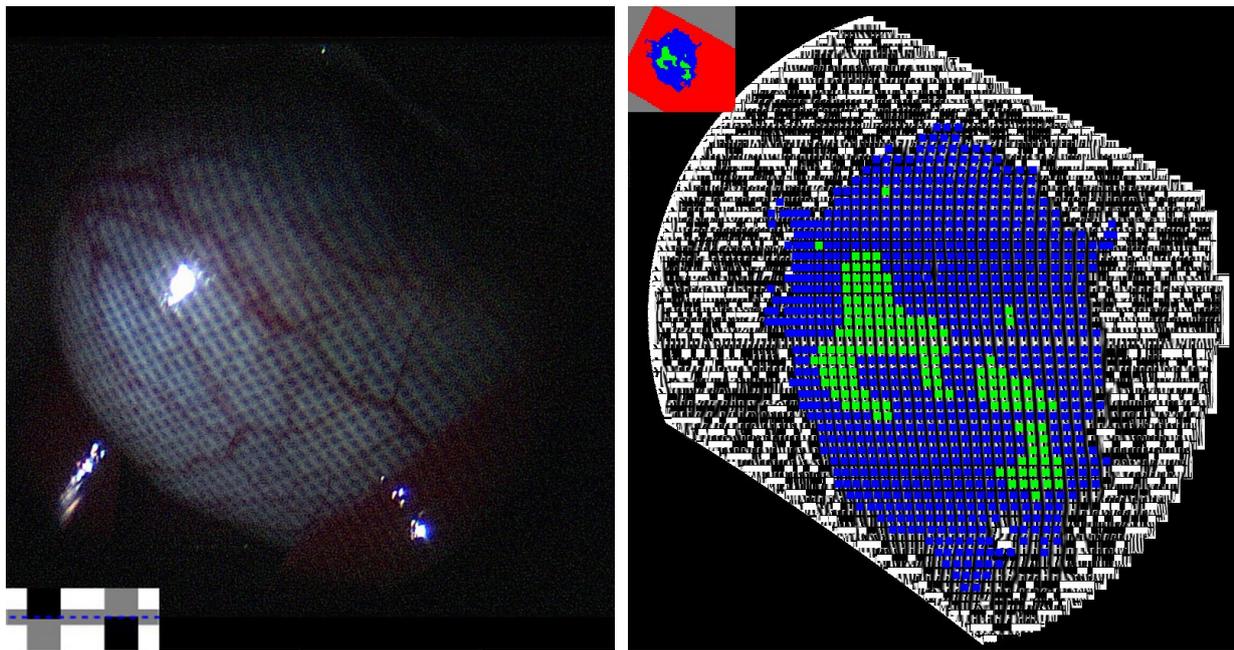
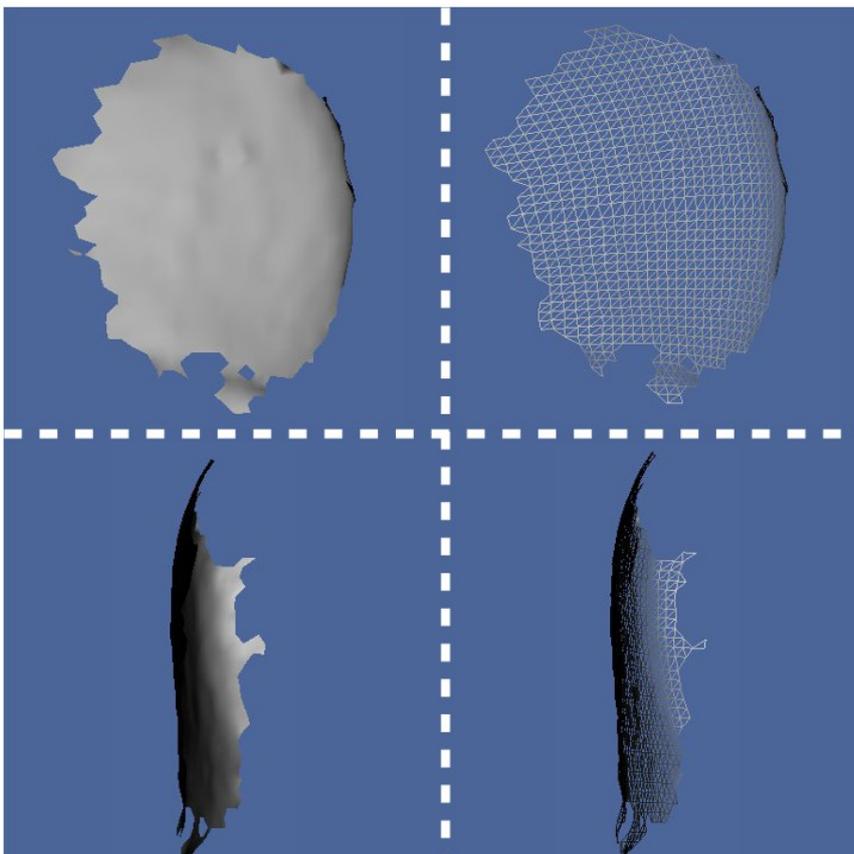


FIGURE 4.43 – Reconstruction effectuée avant vérification de l'unicité des appariements dans le sens caméra \rightarrow projecteur, quelques appariements multiples sont ici visibles.



(a)

(b)

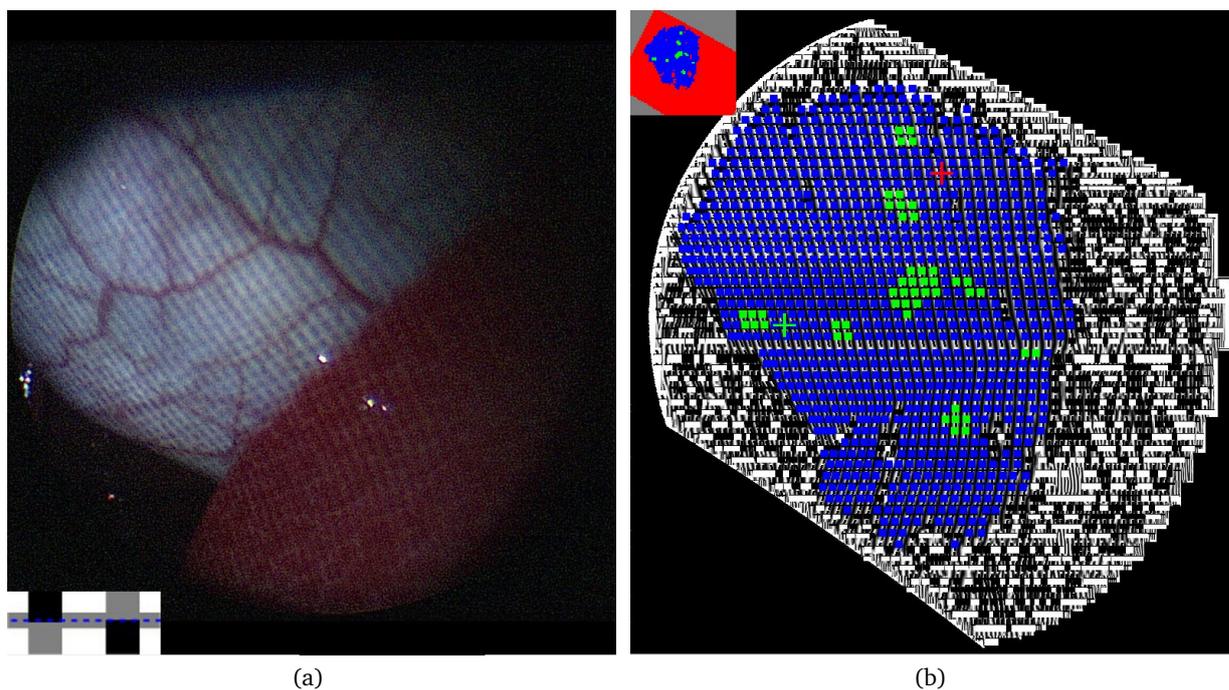


(c)

(d)

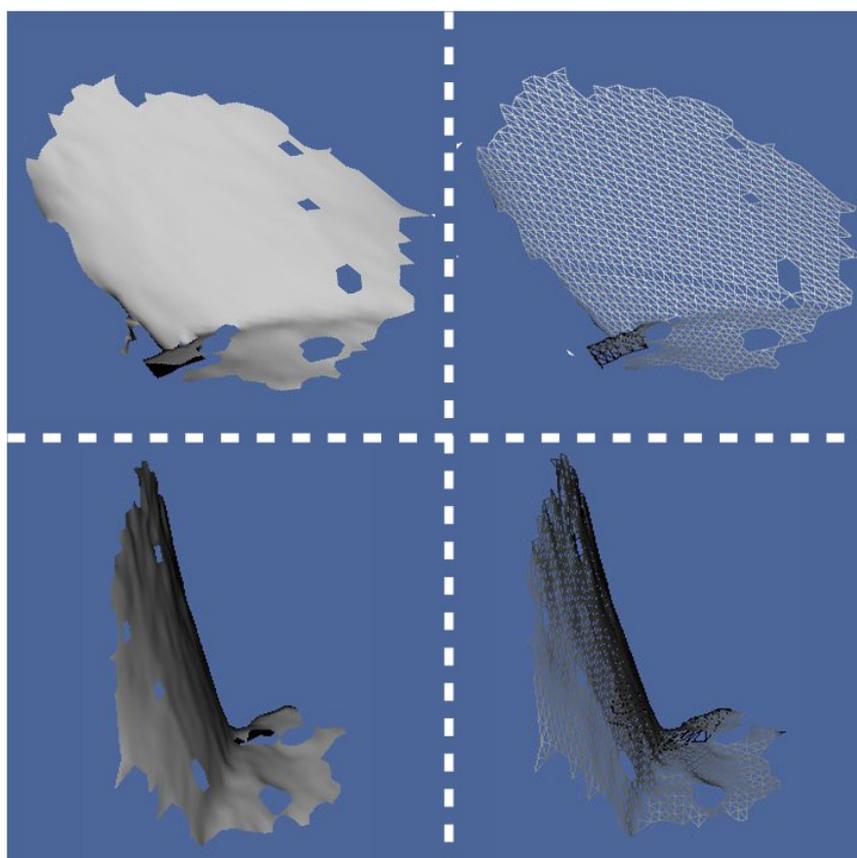
Taille codage	70^2
Nb symboles	2
H_{Min}	2
DansProj	3721
DansCam	< 2121
Decodées	884
RecoParPrim	1
VotesMin	6
H Integ.	3×3
\bar{H}_{Mini} votes	8
MalClassif	~ 180

FIGURE 4.44 – La vésicule biliaire.



(a)

(b)



(c)

(d)

Taille codage	70^2
Nb symboles	2
H_{Min}	2
DansProj	3721
DansCam	< 2121
Decodées	1024
RecoParPrim	1
VotesMin	5
H Integ.	3×3
H_{Mini} votes	8
MalClassif	~ 301

FIGURE 4.45 – La vésicule biliaire et le foie.

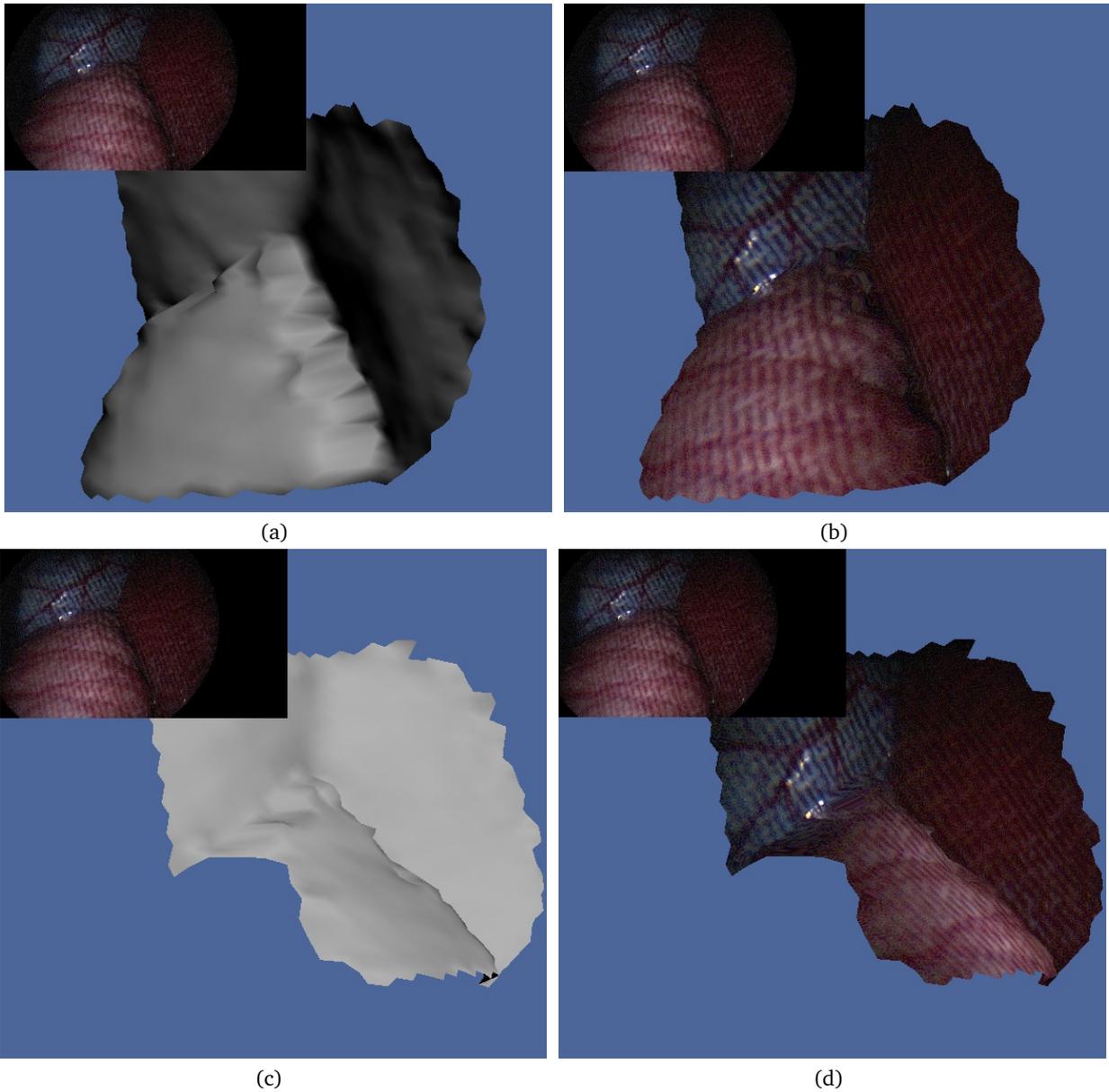


FIGURE 4.46 – Exemple de reconstruction d’une scène intra-abdominale avec le système endoscopique mini-invasif. Divers organes sont reconstruits : la vésicule biliaire (en bleu), le foie (à droite) et l’estomac. La scène est rendue sous deux points de vue (a,b) et (c,d). (en haut à gauche), l’image acquise, (a,c) la surface reconstruite, (b,d) reconstruction avec placage de texture.

Chapitre 5

Développements système

La part de développement système, aussi bien matériel que logiciel, a représenté un investissement significatif de ce travail de thèse.

D'une part, la réalisation d'un prototype de projection/acquisition vidéo mono-trocart miniaturisé a été nécessaire. En effet, nous avons comme objectif final d'effectuer des reconstructions in-vivo en situation mini-invasive et aucun dispositif de ce type n'est aujourd'hui disponible sur le marché. La littérature propose des systèmes utilisant des sources laser, ou des fibres optiques (voir chapitre 1) mais à notre connaissance aucune réalisation concrète de ce type n'a été effectuée avec un projecteur DLP. Cela a donc été un véritable défi. Une chaîne d'acquisition du signal vidéo Full-HD a également été mise en place pour acquérir le flux vidéo endoscopique en temps-réel.

D'autre part, il était utile d'intégrer toute la chaîne de traitements (conception des motifs, étalonnage, acquisition, segmentation, mise en correspondance, reconstruction) dans une interface unifiée pour fournir un outil pérenne qui propose des reconstructions 3-D en temps-réel et qui puisse être réutilisé par la suite pour exploiter celles-ci.

Dans ce chapitre, nous présentons ces développements système, **matériel** puis **logiciel**, qui, bien que possédant une dimension moins "scientifique" que les sujets abordés dans les chapitres précédents, n'ont pas moins été indispensables pour la mise en place, la validation et la pérennité du travail. En particulier, ces développements ont permis d'effectuer des projections et acquisitions IN-VIVO avec de nombreux motifs différents.

5.1 Développement matériel pour la vision endoscopique

Dans cette section, nous traitons des éléments matériels qui ont été mis en place pour réaliser la chaîne complète de l'information, de son émission (motif de lumière structurée), à son acquisition au sein des mémoires (centrale, graphique) du système. Nous présentons ainsi le prototype de projection/acquisition vidéo endoscopique réalisé puis la chaîne d'acquisition du flux depuis la capture par la caméra endoscopique.

5.1.1 Un système de projection/acquisition vidéo endoscopique mono-trocart

Par rapport à l'utilisation de sources laser, l'utilisation de projecteurs vidéo offre de multiples avantages. Il est notamment beaucoup plus facile de modifier la géométrie et les couleurs du motif de lumière structurée projeté pour tester différentes configurations. A contrario, les sources laser collimatées offrent l'avantage d'une projection nette sur une grande profondeur de

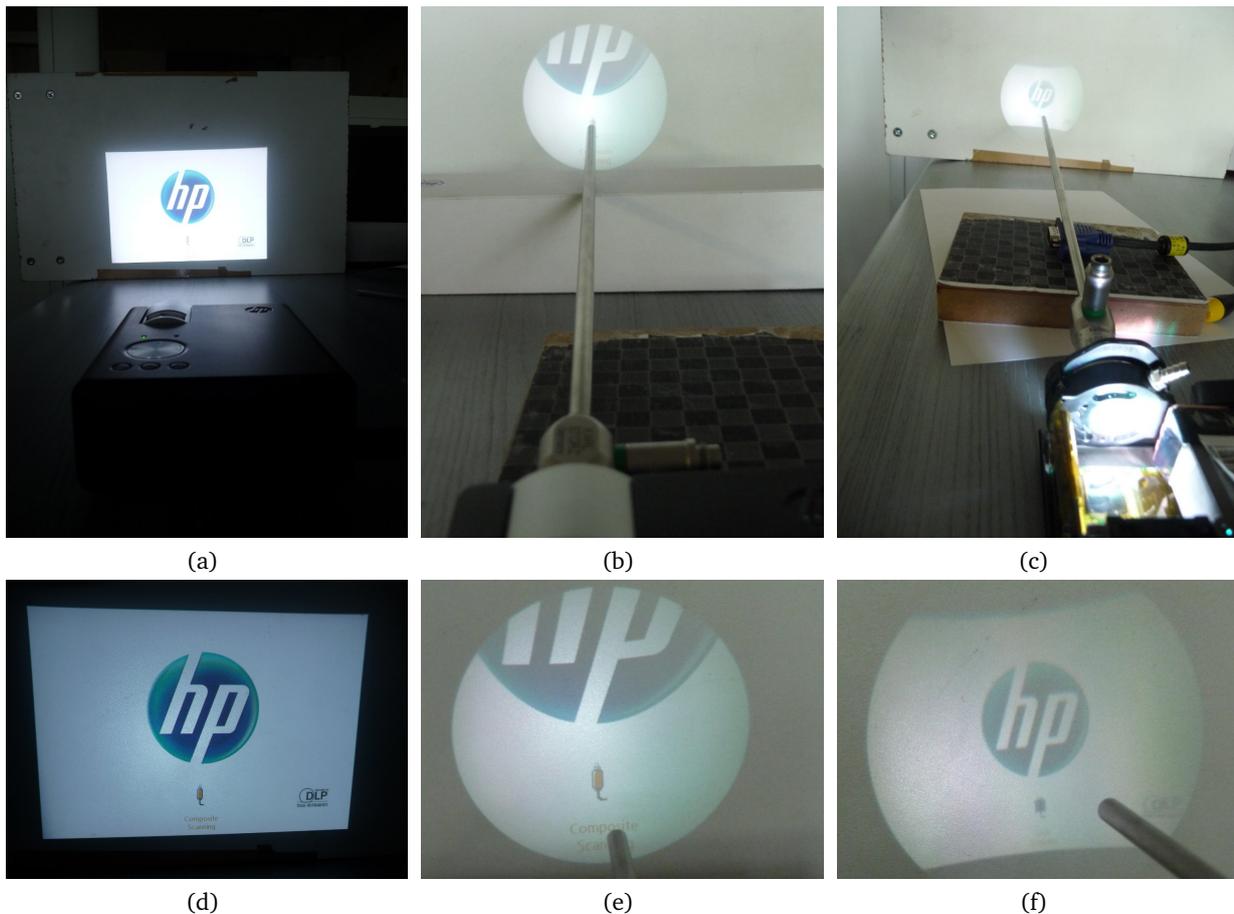


FIGURE 5.1 – a,d) La projection initiale du mini-projecteur, non distordue. e,b) la projection obtenue en accolant directement un endoscope à la sortie de l'optique du mini projecteur. c,f) Le champ obtenu après ajout des optiques d'interface.

champ. De plus ces sources sont plus facilement intégrables dans un système miniaturisé que des projecteurs vidéo souvent lourds et volumineux, et sont donc souvent plus adaptées par leur taille et leur poids réduits à la conception de systèmes manipulables par des bras robotisés par exemple. C'est ainsi que la littérature propose plusieurs systèmes stéréo endoscopiques basés sur des projections laser ([Hasegawa01], [Krupa03],[Chan03], [Albitar09a]).

Depuis 2008, l'industrie des projecteurs vidéo miniaturisés est en plein essor. Elle suit en effet le développement général de l'industrie des objets mobiles. Parmi ces mini-projecteurs (ou *pico projecteurs*), il existe notamment des projecteurs vidéo basés sur une technologie laser n'utilisant aucune optique et qui permettraient de cumuler les avantages du laser et de la vidéo énoncés précédemment. Nous notons par exemple le *ShowWX+* de *Microvision* et le *L1 v2* de *AAXA*. Cependant de tels projecteurs souffrent pour l'instant d'une puissance lumineuse très faible : entre 10 et 100 lumens contre plusieurs milliers de lumens pour les projecteurs classiques plus volumineux. Pour pallier cet inconvénient, il est possible, sur certains de ces mini-projecteurs, de déporter la source de lumière grâce à un guide de lumière, solution que nous avons retenue. Nous avons ainsi sélectionné le *HP Notebook Projection Companion AX325AA* qui offre une résolution de 800×600 et était facilement démontable pour intégrer les optiques d'interface et raccorder le guide de lumière.

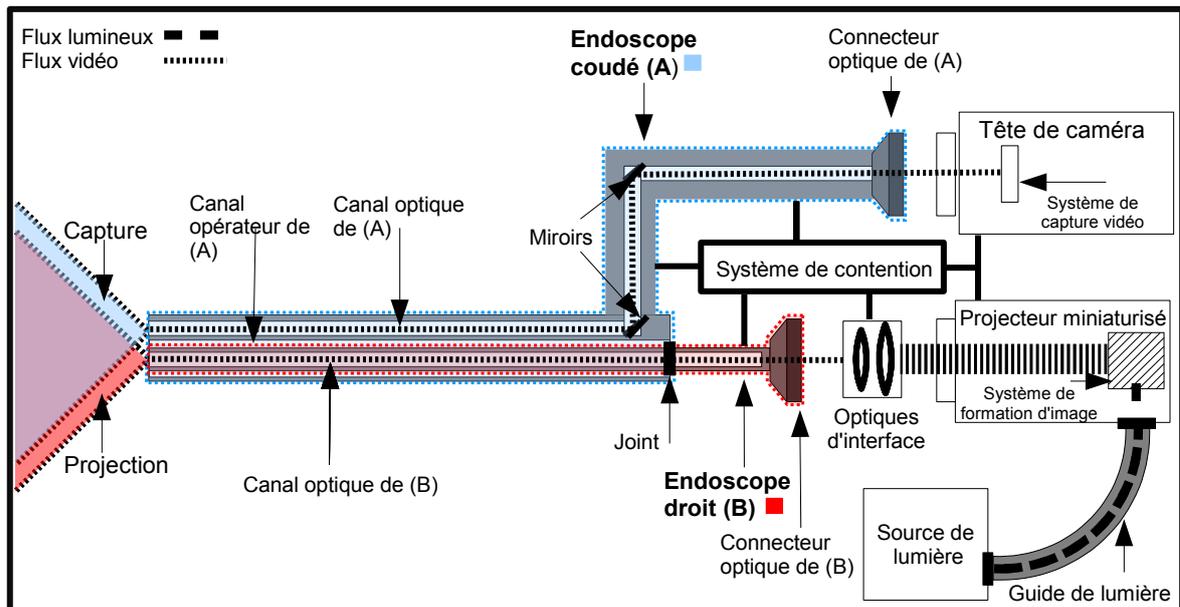


FIGURE 5.2 – Prototype de projection/acquisition vidéo endoscopique.

Une autre problématique qui s'est posée était la connection déportée du projecteur et de la caméra au système endoscopique binoculaire. Sur les endoscopes stéréo tels que ceux utilisés dans le système *Da-Vinci* (voir Chap. 1), l'espace est très restreint entre les extrémités proximales des deux canaux optiques. La connection d'une tête de caméra à l'un des canaux et du projecteur à l'autre ne peut se faire sans dévier l'un des canaux avec un système de miroirs, comme proposés dans [Chan03] ou [Keller03], par exemple. Une solution simple que nous proposons pour résoudre cette difficulté est d'insérer un endoscope droit de 5 mm de diamètre à l'intérieur du canal opératoire (de 6 mm de diamètre) d'un endoscope coudé de 10 mm de diamètre (canal d'acquisition). Nous détournons ainsi le rôle initial de ce canal opératoire qui est d'accueillir des instruments d'opération (couper, brûler des zones cancéreuses etc...) en un second canal optique. Un système stéréoscopique est alors créé dans un seul endoscope de 10 mm avec embouts déportés pour accueillir d'une part la tête de caméra et d'autre part, le mini-projecteur. Enfin, le flux vidéo en sortie du projecteur n'étant optiquement pas adapté à une entrée dans un endoscope, il a été nécessaire de joindre au système des optiques d'interface, permettant d'une part de faire converger le faisceau dans l'endoscope droit pour obtenir une résolution satisfaisante en sortie de l'endoscope et, d'autre part, pour obtenir une zone de netteté suffisante pour des scènes endoscopiques situées à une profondeur comprise entre 2 et 20 cm de la proximité distale du système endoscopique. Les évolutions du champ projeté sont visibles sur la figure 5.1. Deux problèmes apparaissent cependant par rapport à la projection initiale : d'une part des distorsions géométriques importantes apparaissent et d'autre part une importante perte de luminosité est constatée (Fig 5.1c,f). Un joint en silicone a finalement été ajouté à l'interface entre les deux endoscopes pour éviter le reflux du gaz intra abdominal.

Ainsi le système endoscopique proposé est composé des éléments suivants (voir Fig. 5.2) :

- Un endoscope rigide coudé (A) avec canal opératoire. Son optique est orientée à 30° par rapport à l'axe principal de l'endoscope.
- Un endoscope rigide droit (B), d'un diamètre inférieur au diamètre du canal opératoire de

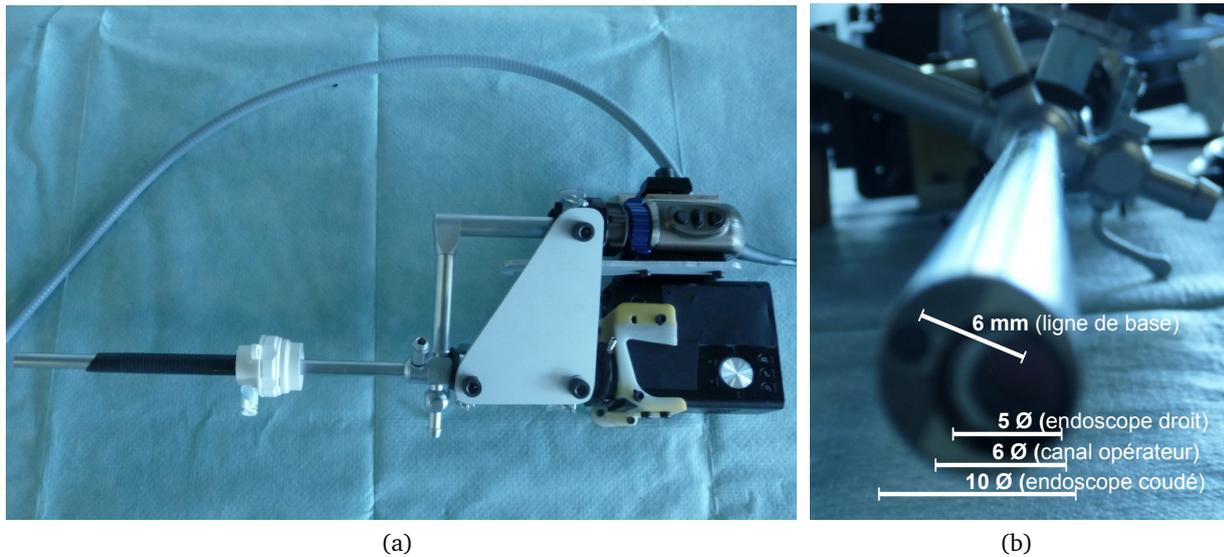


FIGURE 5.3 – a) Le prototype réalisé, sur le bord supérieur, le guide de lumière vient se connecter à une source externe. b) Gros plan sur le système optique binoculaire

(A). Cet endoscope vient s'insérer dans le canal opérateur de (A) et y est relié rigidement par un système de contention. Son optique est alignée avec l'axe principal de l'endoscope.

- Une tête de caméra est reliée directement au connecteur optique de (A ou de B). Elle sert à l'acquisition du flux vidéo.
- Un projecteur vidéo miniaturisé est relié au connecteur optique de l'autre endoscope via un jeu d'optiques d'interface permettant d'ajuster le champ et la focale du flux vidéo. Il sert à la création du flux vidéo.
- Un système de contention permet de relier rigidement tous les éléments (les deux endoscopes, la tête de caméra, le projecteur, les optiques d'interface) entre eux.
- Une source de lumière externe (Storz Xenon de 300 watts).
- Un guide de lumière flexible, dont le rôle est de transférer l'énergie lumineuse de la source externe à la matrice DLP du projecteur.
- Un joint en silicone entre les deux endoscopes (A et B) permet de boucher la circulation de gaz dans le canal opérateur et éviter le reflux du gaz intra-abdominal à l'extérieur de l'abdomen.

La réalisation concrète de ce dispositif est observable sur les photos de la figure 5.3.

La ligne de base de 6 mm (Fig. 5.3.b), est extrêmement faible comparée à celles généralement utilisées dans des systèmes stéréo plus classiques pour lesquels l'écart entre les centres optiques des deux périphériques est souvent de plusieurs dizaines de centimètres. Cependant, ce problème est a priori compensé d'une part par la faible distance de la scène par rapport aux centres optiques du système stéréo (en général moins de 10 cm), d'autre part par la résolution élevée du dispositif de capture (Full-HD) et enfin par le non parallélisme des axes optiques, l'optique de l'endoscope droit étant orientée à 0 degrés et celle de l'endoscope coudé à 30° degrés, ce qui accentue la parallaxe du système stéréo.

Le dispositif pèse 1,6 kg pour un volume de $55 \times 19 \times 8 \text{ cm}^3$. Ce qui en fait un objet tout à fait manipulable par un bras robotisé.

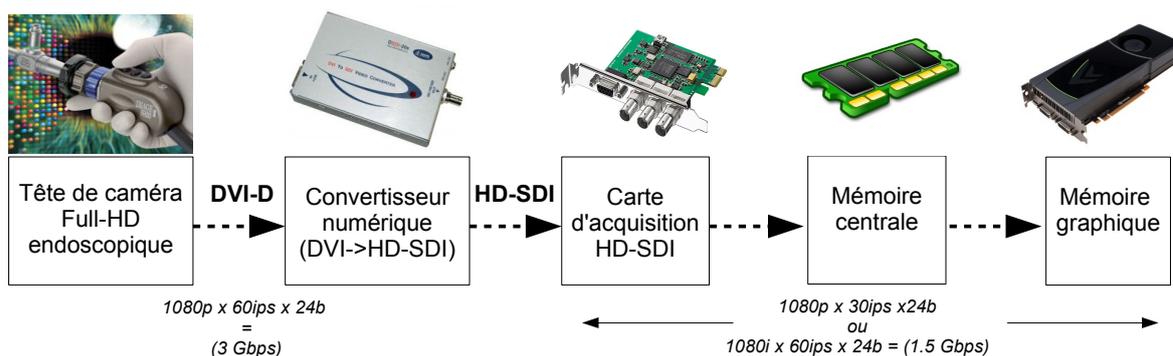


FIGURE 5.4 – Chaîne d'acquisition du signal vidéo.

5.1.2 Chaîne d'acquisition du signal vidéo

La chaîne d'acquisition du signal Full-HD est composée de 3 intermédiaires entre le contrôleur de la caméra endoscopique et la mémoire du processeur graphique. Ces éléments sont illustrés par la figure 5.4. La chaîne mise en place permet un débit maximum théorique de 1.5 Gbits par secondes ou 1080p (progressif) à 30ips (images par secondes) ou 1080i (interlacé) à 60 ips.

Celle-ci est composée, dans l'ordre, des éléments suivants :

- Tête de caméra endoscopique Storz et son contrôleur (Image 1 HD), flux capturé 1080p à 60 ips.
- Convertisseur *Doremi DSDI 20S*. Ce convertisseur numérique permet d'obtenir un signal HD-SDI (1.5 Gbps) à partir du signal DVI-D provenant du contrôleur de la caméra.
- Carte d'acquisition HD-SDI *BlackMagic DeckLink SDI*. Elle nécessite un port PCI Express 2X voire 4X pour permettre un débit maximal.
- Mémoire centrale de l'ordinateur (4Go DDR2). Le transfert entre la carte d'acquisition et la mémoire centrale se fait par accès direct DMA (Direct Memory Access).
- Mémoire de la carte graphique *NVIDIA GeForce GTX 470*. Cette carte exploite pleinement la bande passante d'un bus PCI Express 16X .

Nous notons qu'il existe des cartes d'extension permettant l'acquisition du flux HD-SDI directement en mémoire GPU et à 3Gbps, par exemple la *NVIDIA Quadro SDI Capture*. Mais le coût de ces cartes s'élève à plus de 2000 euros, contre moins de 200 euros pour la Decklink SDI.

Notons qu'il serait possible d'obtenir un flux final à 1080p 60 ips (au lieu de 30) en choisissant un convertisseur et une carte d'acquisition supportant ce débit, mais ceci demeure plus onéreux.

5.1.3 Conclusion

Nous avons pu mettre en place durant cette thèse, un système complet de projection et d'acquisition vidéo endoscopique mono-trocart. Celui-ci a été suffisant pour effectuer les premières validations in-VIVO (voir chapitre 4). Cependant, le système matériel ainsi mis en place peut être amélioré à plusieurs niveaux. Nous avons vu au chapitre précédent que la présence importante de bruit est causée principalement par le manque de luminosité globale du système. Ce problème avait également été mentionné dans [Keller00].

Il serait donc nécessaire de retravailler à ce niveau non seulement sur la puissance nominale

de la source de lumière, qui avec 300W Xenon est déjà très élevée, mais surtout sur l'efficacité optique des différents éléments mis en place. Le flux lumineux doit en effet traverser 4 systèmes optiques successifs : optiques du projecteur, optiques d'interface, optiques de l'endoscope en projection, optiques de l'endoscope en acquisition provoquant des pertes non négligeables à chaque intermédiaire. Un travail d'optimisation à ce niveau pourrait améliorer significativement le rapport signal/bruit des images acquises.

De plus, comme nous l'avons évoqué, des dispositifs matériels sont aujourd'hui disponibles sur le marché pour obtenir une chaîne d'acquisition à 3 Gbps et ainsi exploiter les caméras HD à leur maximum de potentiel.

5.2 Développement logiciel pour la vision endoscopique

Les contributions et la méthodologie proposées pendant cette thèse ouvrent la porte à de nombreuses variantes de motifs générés et de traitements associés, comme en attestent les reconstructions présentées au chapitre précédent. Nous n'avons pas proposé un motif unique avec des propriétés (nombre de symboles, résolution, distance de Hamming minimale) fixes comme c'est souvent le cas dans littérature. De plus, de nombreux traitements sont nécessaires pour étalonner (partiellement pour l'instant) le système stéréo, caractériser les distorsions géométriques de la caméra et du projecteur, reconstruire et mailler les points mis en correspondance.

5.2.1 Une plate-forme toute intégrée

Pour permettre davantage de flexibilité au système et le rendre plus facilement accessible et intuitif à utiliser, une intégration de tous les traitements a été effectuée dans une plate-forme unique, ne nécessitant pas d'échanges de fichiers avec des applications tierces. Celle-ci a été très utile notamment lors des sessions IN-VIVO, effectuées en temps limité (2 heures maximum) aux cours desquelles une mise en place rapide du système et des différents motifs projetés était nécessaire.

Les différents outils logiciel utilisés pour la réalisation de la plate-forme sont les suivants :

- **Système** : *Linux (Ubuntu)*
- **Langages** : *C/C++*, *Nvidia CUDA*
- **Vision, graphismes** : *OpenCV*, *OpenGL*
- **Acquisition du flux vidéo** : *LibDC1394* (caméra FireWire), *Decklink SDK* (caméra endoscopique, signal HD-SDI)
- **Interface graphique** : *WxWidgets* associé à *wxFormBuilder*.

L'environnement Matlab a également été utilisé en début de thèse pour les premières validations, mais n'intervient plus dans la plate-forme présentée ici.

OpenGL a servi à la fois pour le rendu 3-D des reconstructions et pour la génération graphique des motifs. Le rendu de ceux-ci est réalisé automatiquement après chaque nouveau codage en prenant en compte la géométrie épipolaire et les distorsions géométriques du projecteur. La librairie *OpenCV* a été utilisée pour estimer la matrice Fondamentale à partir des correspondances de points obtenues avec le motif d'étalonnage. La majorité de ces outils étant utilisable sous environnement *Linux* ou *Windows*, ces choix assurent une bonne portabilité à la plate-forme. Une fois que le motif qui doit être projeté est généré, sa fenêtre est détachée de l'interface et est glissée sur le deuxième écran, celui du projecteur, et redimensionnée en plein écran.

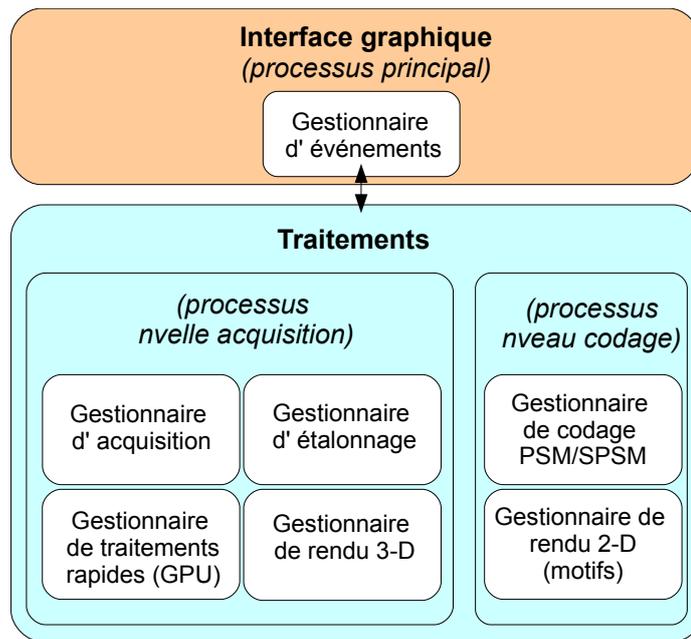


FIGURE 5.5 – Architecture logicielle utilisée

Architecture logicielle

La plate-forme a été implémentée avec une architecture modulaire basée sur l'utilisation de gestionnaires implémentés sous forme de classes *Singleton* (*une seule instance présente dans l'application*) (voir Fig. 5.5). Cette modularité simplifiera la maintenance et permettra une évolution plus aisée. Les gestionnaires traitent des problématiques suivantes :

- Gestion des événements de l'interface
- Gestion du codage (PSM, SPSM)
- Gestion de l'étalonnage du système : caractérisation des paramètres du modèle utilisé.
- Gestion des acquisitions (choix et configuration des sources : HD-SDI, Firewire, séquence de fichiers).
- Gestion du rendu 2-D des motifs générés
- Gestion du rendu 3-D des reconstructions
- Gestion des traitements rapides (interface et exécution sur le processeur graphique).

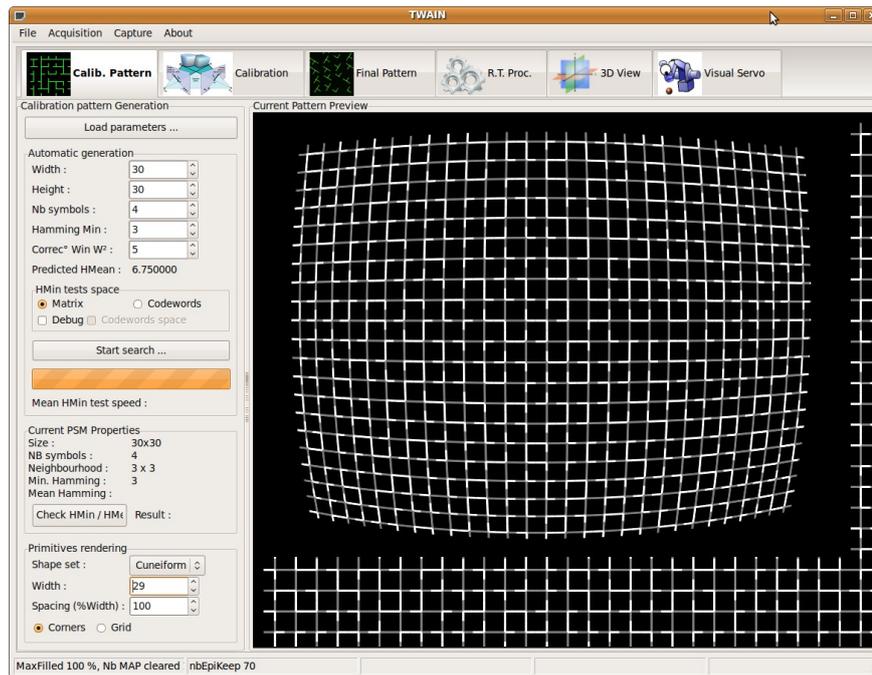
Le code relatif aux traitements est séparé de celui de l'interface graphique selon l'approche *document-vue* utilisée en ingénierie logiciel.

Trois processus sont utilisés :

- Le processus principal qui gère les événements générés par l'interface graphique.
- Un processus qui correspond aux traitements réalisés à partir de chaque nouvelle acquisition.
- Un processus qui se lance à chaque nouveau codage demandé.

Ces deux derniers processus permettent de ne pas bloquer toute l'interface lorsque les traitements nécessitent beaucoup de ressources de calcul.

Le nombre de lignes codées s'élève à un peu plus de 50000 (somme des lignes des fichiers sources codés) sans prendre en compte le code généré automatiquement pour l'interface graphique par *wxFormBuilder*, ni les interfaces des classes C++. Environ 15000 de ces lignes de code s'exécutent sur le processeur graphique.



(a)

FIGURE 5.6 – Panneau de conception du motif d'étalonnage.

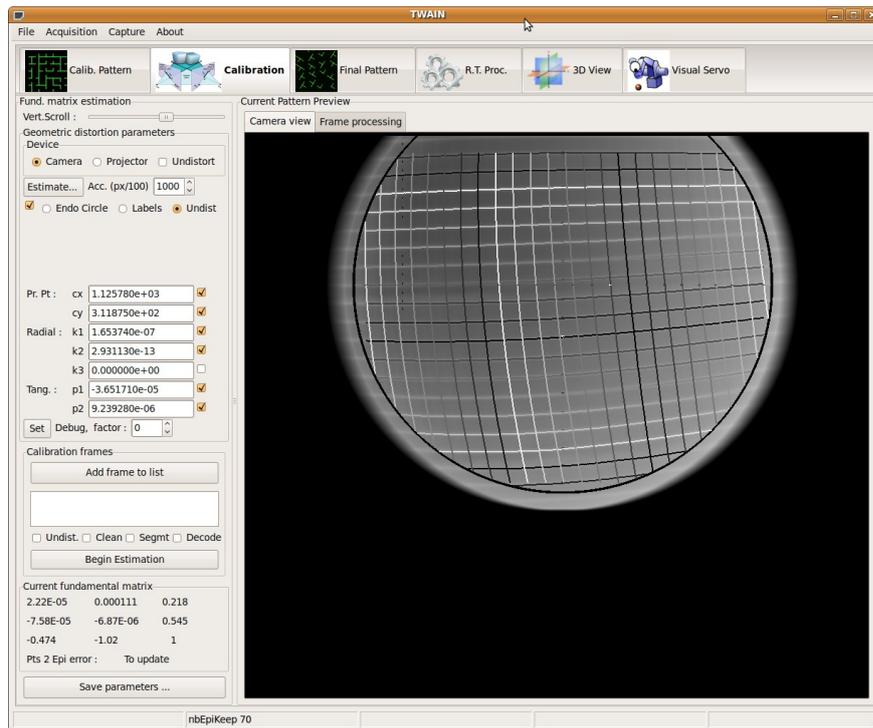
5.2.2 Interface graphique

L'interface graphique se subdivise 5 panneaux, chacun d'entre eux possédant trois parties principales, visibles sur les figures 5.6 à 5.10 :

- Sur le bord supérieur, la liste des panneaux relatifs à la tâche à effectuer.
- Un panneau de contrôle sur le bord gauche pour choisir les paramètres et les actions désirés.
- Une fenêtre de visualisation centrale affichant les images, les résultats et informations des traitements.

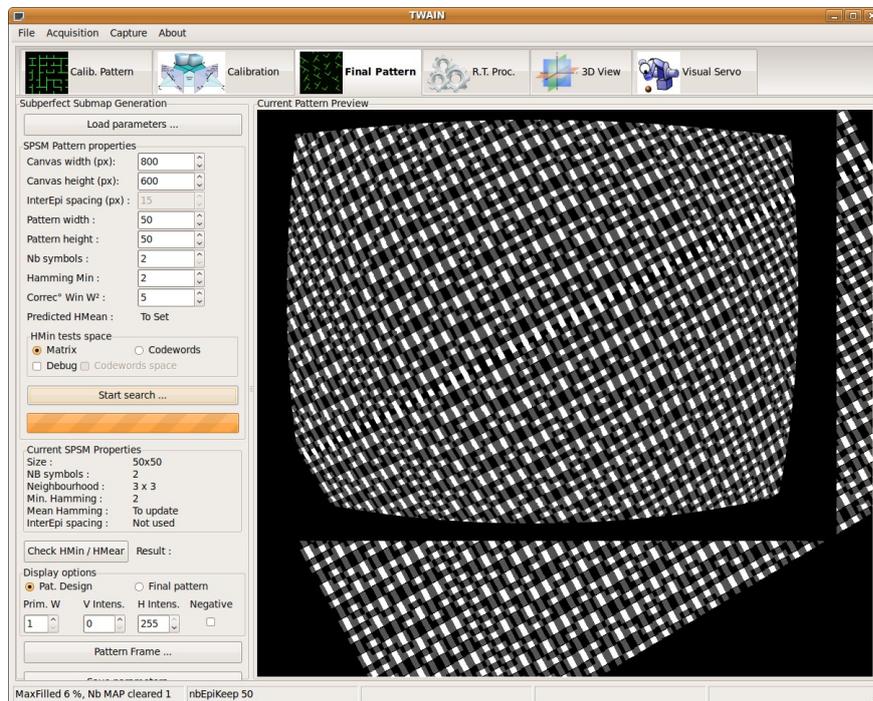
Les fonctionnalités se répartissent entre les panneaux comme suit :

- Le panneau de **conception du motif d'étalonnage** (Fig. 5.6) présente les paramètres de codage (*PSM*) et de rendu du motif utilisé pour l'étalonnage du dispositif stéréo.
- Sur le panneau d'**étalonnage** (Fig.5.7), il est possible de choisir le périphérique (caméra ou projecteur) dont les distorsions géométriques vont être estimées et corrigées. Puis les images contenant la projection du motif de calibration à différentes profondeurs qui vont servir à estimer la matrice Fondamentale peuvent être sélectionnées. La fenêtre de visualisation permet alors de vérifier que la l'étiquetage des lignes est correct.
- Le panneau de **conception du motif final** (Fig.5.8) présente les paramètres de codage (*SPSM*) et de rendu du motif finalement utilisé pour les reconstructions.
- Le panneau des **traitements rapides** (Fig.5.9) a pour principal objectif de déboguer chaque étape de traitement à la réception d'une nouvelle image d'une scène à reconstruire. Ces étapes présentées au chapitre 2, vont du simple affichage à la reconstruction 3-D. Pour chacune de ces étapes un panneau inférieur permet d'ajuster certains paramètres. Par exemple pour la reconstruction 3-D, il est nécessaire de sélectionner dans l'image les 3 points qui



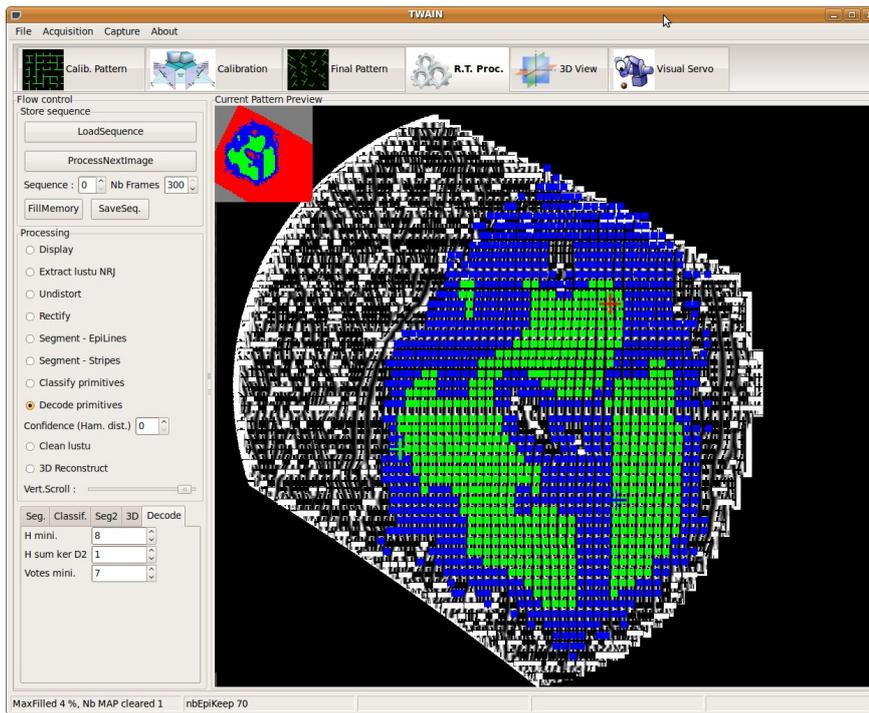
(a)

FIGURE 5.7 – Panneau d'étalonnage.



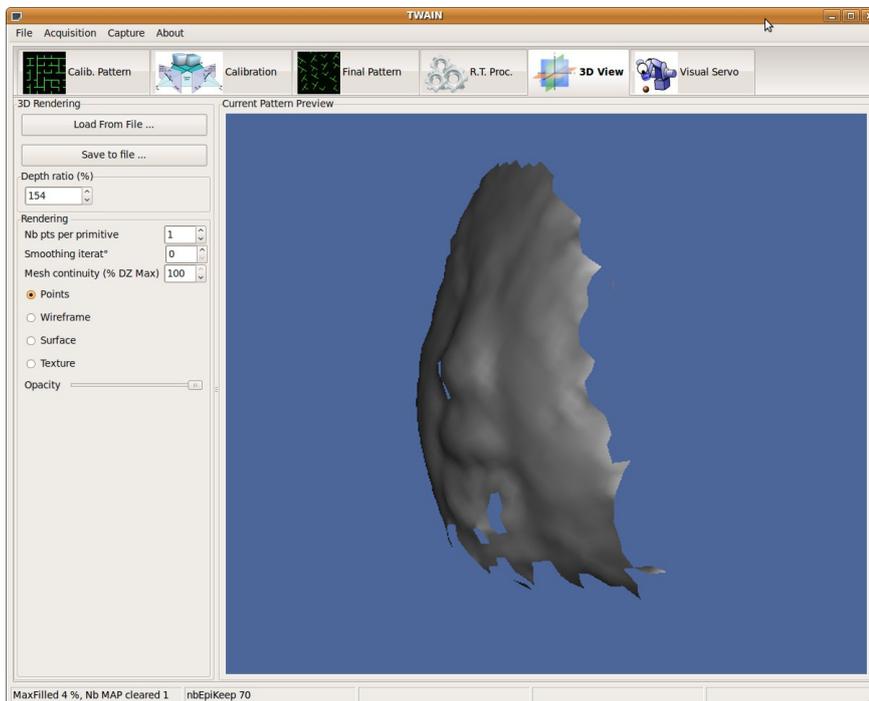
(a)

FIGURE 5.8 – Panneau de conception du motif final.



(a)

FIGURE 5.9 – Panneau des traitements temps-réel.



(a)

FIGURE 5.10 – Panneau des reconstructions 3-D.

vont définir le plan de référence à partir duquel la reconstruction projective va s'opérer (voir chapitre 4).

- Le panneau des **reconstructions 3-D** (Fig.5.10) permet de choisir un facteur d'échelle selon l'axe optique, le type de rendu 3-D (points, fil de fer, surface pleine, surface texturée (à implémenter)), un niveau de lissage 3D et d'opacité du rendu, pour visualiser les parties cachées. Le nombre de points à reconstruire entre deux primitives successives sur une même frange (reconstruction dense) peut également être choisi. Le panneau de visualisation répond à la souris pour modifier le point de vue et le facteur d'échelle global de la reconstruction.
- D'autres panneaux peuvent être ajoutés en fonction de l'exploitation désirée des données 3-D.

Enfin le menu permet de choisir et de paramétrer la source d'acquisition : FireWire, HD-SDI, ou séquence de fichiers image pré-enregistrés et d'effectuer des captures d'écran.

5.2.3 Exploitation du processeur graphique

Effectuer toutes les étapes de traitements évoquées au chapitre 3, en temps-réel ou en ligne sur un flux d'images Full-HD, 3 canaux, avec un processeur ordinaire, bi ou quadri coeurs serait bien difficile. D'après nos essais initiaux, la rectification seule de l'image demandait plusieurs dizaines de milli-secondes sur un *Pentium 4 Dual Core* en utilisant la librairie *OpenCV*, pourtant optimisée pour exploiter les instructions SIMD (MMX, SSE) (Single Instruction Multiple Data) des processeurs récents. Ainsi, pour les traitements critiques, une implémentation sur processeur graphique (*GPU*) a été préférée. En effet, l'architecture des processeurs graphiques, dont le développement a été fortement poussé par l'industrie des jeux vidéo, est optimisée pour effectuer des calculs en parallèle (à l'origine sur chaque pixel/fragment d'une image) et en temps-réel. Elles possèdent ainsi plusieurs centaines d'unités de calculs pouvant s'exécuter simultanément. Depuis l'avènement des architectures unifiées sur les cartes graphiques, ces dernières sont de plus en plus utilisées pour les applications de GPGPU (General Purpose GPU), à côté des jeux vidéo et de la CAO, et deviennent facilement programmables (en témoignent les nombreux articles scientifiques depuis quelques années sur la proposition de versions parallélisées d'algorithmes de traitements pour des applications très variées : aussi bien le traitement d'images que les prédictions météorologiques, le calcul de repliement de protéines ainsi que tous les calculs complexes qui peuvent être parallélisés (souvent linéaires, mais pas toujours)). Les résultats obtenus avec de telles implémentations sont souvent impressionnants et permettent régulièrement un gain à 2 voire 3 chiffres par rapport aux implémentations *CPU*.

Des langages récents comme le *CUDA* (*Compute Unified Device Architecture*) de Nvidia , très proches du *C*, sont simples à prendre en main et ont largement contribué à démocratiser l'utilisation de processeurs graphiques au sein de la communauté scientifique. Cependant si le langage paraît familier, l'architecture de ces processeurs graphiques, différentes de celle des CPU, doit être bien maîtrisée pour effectuer des traitements efficaces. C'est le cas en particulier des différents niveaux de mémoire (globale, locale, partagée, constante), des modalités d'accès à ces mémoires et de la communication entre les unités de calculs.

Il n'est pas question d'étudier ici tout cela en détail, la littérature sur le sujet est déjà très abondante par ailleurs. Nous désirons simplement pointer l'approche globale employée dans cette thèse pour paralléliser au mieux les différents traitements présentés au chapitre 3.

Implémentation GPU proposée

Le lecteur aura remarqué sur les différentes illustrations de ce manuscrit que les traitements

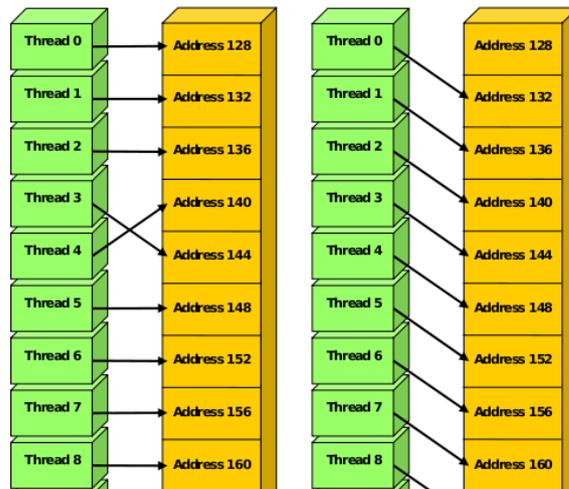


FIGURE 5.11 – Accès en mémoire GPU, à gauche un chargement mémoire par processus est effectué de façon sérialisée à cause de la non contiguïté des accès et du non alignement (tous les 128 bits). A droite lecture *coalesce*, un seul chargement mémoire pour le groupe de processus GPU.

sont effectués sur des images rectifiées. Or la rectification demande la mise en oeuvre d'une transformation d'image supplémentaire, et il est tout à fait possible d'analyser les lignes épipolaires dans l'image non rectifiée. Cependant, d'une part l'utilisation d'une carte de correspondance qui cumule la correction des distorsions et la rectification (voir chapitre 4) permet d'effectuer ces deux transformations en une seule lecture, et comme la correction des distorsions est nécessaire, la rectification ne demande pas davantage de traitements. D'autre part, travailler sur des images rectifiées offre un avantage majeur pour une implémentation GPU : comme les lignes épipolaires sont disposées horizontalement dans l'image, elles peuvent être stockées de façon contiguë en mémoire graphique. Ainsi en associant les unités de calcul successives à des pixels successifs en mémoire (Fig. 5.11), il est possible d'effectuer des lectures optimisant les accès mémoire grâce au mécanisme *coalesce* (*lecture mémoire en union*) qui permet à un groupe d'unités de calcul de lire les informations contiguës en mémoire GPU avec une seule instruction (*SIMD*). Cette modalité d'accès mémoire, très optimisée, est toujours recherchée lors d'implémentations GPU. Nous l'utilisons pour pratiquement tous les traitements, c'est le cas également de la version parallélisée du calcul de l'image intégrale que nous avons proposée au chapitre 3.

Conclusion générale

Dans cette thèse, une solution a été proposée pour la reconstruction 3-D en-ligne de surfaces intra-abdominales dans un contexte mini-invasif. Diverses applications, permettant d'améliorer l'acte chirurgical, sont visées : l'amélioration du guidage de robot, la visualisation 3-D selon différents points de vue, le recalage de modèles d'organes préexistants, la métrologie.

Notre solution fait appel à la projection d'un motif de lumière structurée codée monochromatique permettant la capture de la profondeur à partir d'une seule image. La problématique phare a été la robustesse du motif et des traitements associés, nécessaires pour effectuer des appariements malgré les nombreuses perturbations géométriques et spectrales présentes dans les scènes intra-abdominales. A cette fin, deux concepts ont été exploités : l'**invariance** de l'homographie épipolaire et la **redondance** du codage. Contrairement à de nombreuses contributions en lumière structurée, nous ne proposons pas un motif avec des propriétés fixes (résolution, nombre de symboles utilisés, distance de Hamming minimale imposée, taille des voisinages), mais une méthodologie de conception qui peut amener à la création de motifs différents selon les besoins applicatifs. Ainsi certains motifs possédant une plus grande résolution avec plus de symboles seront adaptés à des scènes possédant beaucoup de discontinuités de courbure mais peu de perturbations spectrales et d'autres, de résolution plus faible mais avec une distance de Hamming minimale plus élevée et moins de symboles seront plus indiqués avec des scènes possédant des caractéristiques inverses.

Pour ce faire, différentes contributions ont été apportées aussi bien au niveau de la conception numérique que physique du motif, et des traitements utilisés pour l'exploiter. Un résultat remarquable est la proposition d'une nouvelle classe de motifs : les **motifs alignés sur la géométrie épipolaire**.

Analyses et contributions

Dans le premier chapitre, nous avons étudié les travaux existants en vision endoscopique puis en lumière structurée. Dans un premier temps, les contributions faisant usage d'un système monoculaire, binoculaire actif, binoculaire passif et Time-Of-Flight ont été étudiés. Ceci nous a permis de motiver l'usage de la vision active. Par la suite, les principales approches utilisées en lumière structurée ont été appréciées, leurs avantages et inconvénients mis en avant. Notamment différents travaux exploitant les principales stratégies de codage du motif : 1-D, 2-D, par multiplexage-temporel ont été considérés. Nous avons mis en avant les deux grandes approches pour effectuer ces codages : les approches analytiques, basées, pour la grande majorité, sur les séquences de De-Brujin et les codes de Gray et les algorithmes *brute-force*. Si les approches analytiques produisent un codage optimal instantanément, les algorithmes *brute-force* permettent davantage de flexibilité en offrant la possibilité d'une part d'imposer des contraintes lors du codage, comme une distance de Hamming Minimale supérieure à 1, l'invariance par rotation des mots de code, la symétrie centrale du motif, et d'autre part de relaxer les contraintes en n'assurant l'unicité des codes qu'entre un sous-ensemble des codes du motif par exemple.

Enfin nous avons introduit la motivation géométrique, exploitant l'invariance projective de l'homographie des lignes épipolaires, qui nous a conduit à une conception de motifs alignés sur la géométrie épipolaire.

Le chapitre 2 présente nos différentes contributions sur le codage 2-D utilisant des algorithmes *brute-force*. Tout d'abord, une discussion est menée sur les aspects théoriques du codage. Différentes limites théoriques (Hamming Bound, Singleton Bound) sont présentées. Les matrices parfaites (*PM*) sont introduites, ainsi que les sous-matrices parfaites (*PSM*) et une nouvelle catégorie : les **sous-matrices sous-parfaites (SPSM)**. Dans cette dernière catégorie, l'unicité n'est pas assurée entre tous les mots de code présents dans le motif mais seulement entre des sous-ensembles de ceux-ci. Les justifications géométriques sont présentées : la course limitée des projections des primitives sur leur ligne épipolaire dans le plan image de la caméra et l'homographie épipolaire pour les motifs alignés. Nous proposons également une formule permettant de connaître **a priori** la distance de Hamming moyenne que posséderont les mots de code d'un futur motif. Celle-ci dépend du nombre de symboles utilisés et de la taille des mots de code et pose l'hypothèse, raisonnable, d'une distribution homogène des symboles dans la matrice codée. Les estimations ont montré une très bonne corrélation avec les résultats obtenus et mesurés après codage. La distance de Hamming moyenne donne une mesure statistique de la robustesse générale, numérique, du codage du motif et a mis en évidence le fait que le gain en robustesse évolue de façon décroissante avec l'augmentation du nombre de symboles utilisés. Ce qui montre qu'en terme de conception, au-delà d'un certain nombre de symboles (3, 4, pour un voisinage 3×3), le gain est faible comparé aux difficultés de classification qui vont apparaître avec plus de symboles.

Dans un deuxième temps, quatre contributions améliorant le codage de type brute-force sont présentées et évaluées :

- La recherche d'un **équilibre entre exploration et exploitation** dans le comportement de recherche dans l'espace des solutions.
- La possibilité d'effectuer le **test d'unicité dans l'espace des mots de code** plutôt que dans celui de la matrice. À cette occasion, une formule de dénombrement de mots de code est présentée. Celle-ci permet d'étudier la complexité algorithmique de cette approche et de la comparer à celles procédant par tests dans l'espace de la matrice.
- Le codage par **proximité épipolaire**, adapté lorsque l'épipole est proche du centre de l'image. Cette approche donne lieu à la création de sous-matrices sous-parfaites, car l'unicité ne doit être assurée qu'entre les codes associés à des primitives se trouvant proches d'une même ligne épipolaire.
- Le codage par **alignement épipolaire**, adapté lorsque l'épipole est éloigné du centre de l'image. Cette approche donne lieu à la création de sous-matrices sous-parfaites dans lesquelles le problème de codage est encore moins contraint que par proximité épipolaire, car seules les H premières lignes (pour un voisinage de taille $H \times W$) doivent être codées, les codes des autres lignes étant obtenus par une copie directe.

Ces deux dernières contributions produisent ce que nous appelons un **codage 1,5D**. En effet, pour de tels codages, les voisinages formant les mots de code exploitent les deux dimensions, mais l'unicité n'est assurée que sur une des deux dimensions. Ainsi nous cumulons l'avantage du codage 2D : la taille importante des mots de codes qui mène à davantage de redondance et l'avantage du codage 1D qui réside dans le faible nombre de voisinages entre lesquels l'unicité doit être assurée. Cela mène à un problème de codage moins contraint et donc des propriétés plus intéressantes (résolution plus élevée, moins de symboles, distance de Hamming Minimale (redondance) plus importante). Ceci est mis en évidence dans les différents résultats reportés

à la fin du chapitre. Nous avons également mis en avant que le test d'unicité dans l'espace des mots de code n'est intéressant que dans le cadre des codages de matrices PSM lorsque la distance de Hamming minimale imposée est faible et que la taille de la matrice est importante. Ou de façon absolue, lorsque la distance de Hamming minimale est de 1, car alors le test peut être effectué par un unique accès direct. Enfin, nous avons montré qu'avec la diminution de la complexité du problème de codage, il est possible de produire des codes optimaux, trouvés par une recherche exhaustive, et possédant une taille autorisant une utilisation en pratique.

Dans le chapitre 3, les aspects d'organisation spatiale et de représentation des primitives visuelles sont étudiés. Les traitements associés à cette conception de motif sont également présentés. Mais tout d'abord, nous présentons une **étude sur la déformation des voisinages** observée dans l'image de la caméra. En effet, le décodage de voisinages 2-D suppose l'identification des primitives voisines. Nous montrons que les déformations des voisinages peuvent mettre en échec, dans plus de 40% des configurations étudiées, l'approche classique de reconstitution des voisinages utilisant le seul critère de proximité. Cela motivera, par la suite, le choix de primitives possédant une continuité directionnelle verticale. Nous présentons ensuite deux structures d'organisation des primitives sur l'image du projecteur : l'une par alignement local, associée à un codage par proximité épipolaire et adaptée lorsqu'une angularité non négligeable existe entre les lignes épipolaires et l'autre, par un alignement global, associée à un codage par alignement épipolaire et adaptée lorsque les lignes épipolaires sont quasiment parallèles. Enfin nous avons présenté la chaîne de traitements utilisée pour l'exploitation de cette conception. Les traitements présentés vont de la compensation a priori des distorsions du motif pour que le modèle sténopé s'applique, au décodage des primitives, en passant par la segmentation des franges et le recalage des lignes épipolaires. Ces traitements sont adaptés à un niveau de bruit élevé, et, pour le décodage, à un taux élevé de primitives mal classifiées. Ils incluent en particulier l'évaluation des champs de déformation des voisinages, de façon à adapter la conformation d'un masque de corrélation basé sur les images intégrales pour segmenter les franges. Les traitements incluent également la création d'un graphe d'adjacence qui fait usage d'un critère de parallélisme des franges pour définir les frontières de zones d'adjacences. Il est utilisé pour la propagation du décodage à partir des primitives décodées avec le plus de certitude vers les autres primitives se trouvant dans la même zone d'adjacence. La distance de Hamming sert alors de critère de base pour évaluer la confiance dans les appariements effectués.

Le chapitre 4 est consacré aux reconstructions 3-D, à la géométrie épipolaire et à la correction des distorsions géométriques pour la caméra et pour le projecteur. Les modèles classiques, sténopé, géométrie épipolaire et triangulation utilisés en stéréovision ont été tout d'abord présentés. Une nouvelle reconstruction 3-D, effectuée dans l'espace projectif avec une contrainte d'orthogonalité a été proposée. Avec celle-ci sont réalisées des reconstructions visuellement moins distordues que ce qui est généralement observé avec des reconstructions purement projectives, sans la nécessité d'étalonner les paramètres linéaires internes et externes du système stéréo. La méthode repose sur l'utilisation d'un invariant projectif et de la géométrie épipolaire. La correction des distorsions radiales et tangentielles pour la caméra et le projecteur a été abordée. Celle-ci fait appel au modèle de Brown avec l'approche "Plumb-Line". Une **nouvelle méthode d'estimation** a été présentée pour estimer ces paramètres **pour le projecteur**. Celle-ci repose sur une nouvelle projection/acquisition d'un motif grillagé distordu a priori en fonction des paramètres de distorsion à estimer au sein même de la fonction de coût. Cette fonction quantifie la rectitude des lignes et colonnes de la grille projetée. Enfin, les distorsions étant corrigées pour le projecteur et la caméra, l'estimation de la matrice Fondamentale a été présentée. Celle-ci

emploie un motif de lumière structurée codé de type *PSM* ce qui permet de segmenter et d'ap-
parier des indices visuels sur une grande profondeur de champ. Des résultats de précision ont
été reportés pour un système stéréo classique et pour le système stéréo endoscopique développé.

Finalement, dans le chapitre 5, les différents développements système, matériel et logiciel,
ont été décrits. Nous avons en effet été capables de concevoir, réaliser et utiliser en pratique le
premier dispositif endoscopique de projection/acquisition vidéo mono-trocart. La quasi-totalité
du champ de l'image du projecteur a pu être transmise à travers le système endoscopique avec
une plage de netteté suffisante pour les distances généralement constatées entre les optiques
et les surfaces intra-abdominales. Cependant le prototype souffre, pour l'instant, d'une faible
puissance lumineuse en sortie, un défaut qui est à l'origine de l'important niveau de bruit pré-
sent dans les images acquises avec ce système. Puis, les différents éléments composant la chaîne
d'acquisition vidéo numérique Full-HD temps-réel ont été décrits. Dans un deuxième temps,
les développements logiciel sont abordés. Ceux-ci exploitent une architecture modulaire ainsi
qu'une interface ergonomique, dans une plate-forme logicielle toute intégrée. Enfin nous avons
abordé l'approche générale utilisée pour l'implémentation avec le processeur graphique.

Résultats

De nombreux résultats, tant qualitatifs que quantitatifs ont été reportés au cours de cette
thèse.

Tout d'abord les différentes contributions sur le codage ont permis de générer des codes
possédant des propriétés qui dépassent de loin (en termes de résolution, de redondance, de
nombre de symboles utilisés), ce qui était réalisé auparavant. Les gains les plus importants ont
été obtenus avec l'alignement épipolaire. Pour effectuer celui-ci, il est simplement nécessaire
de connaître la position des épipoles. Or la matrice Fondamentale, de laquelle ils peuvent être
extraits, peut être calculée avec des simples correspondances de points et est déjà utilisée dans
la majorité des systèmes stéréo pour effectuer la mise en correspondance des primitives. C'est
donc une information qui était déjà à disposition des systèmes antérieurs et qui est simplement
utilisée ici en amont, lors de la conception du motif. De plus la réduction de la complexité al-
gorithmique du problème de codage a rendu possible la génération en temps-réel de motifs de
dimension suffisante pour une utilisation pratique. Cela ouvre la voie à des motifs dynamiques
dont le codage (sa densité, sa robustesse) peut être modifié en temps-réel pour adapter le motif
aux spécificités de la scène observée.

Différents résultats partiels ont ensuite été donnés quant à la précision de l'estimation de
la matrice Fondamentale et des reconstructions effectuées avec un recalage de plan. Ceux-ci
montrent que pour la plage de profondeur considérée (environ 10 cm) avec le système minia-
turisé, une précision millimétrique peut être obtenue, malgré la faible ligne de base de 6 mm.
Ce qui est suffisant pour une large palette d'applications. Cependant les valeurs obtenues sont à
considérer avec circonspection, car la segmentation de la grille sur un plan sans perturbations
spectrales est forcément bien plus précise que lorsque des spécularités ou le réseau veineux
par exemple, viennent dévier l'énergie lumineuse des franges le long des lignes épipolaires. Ces
résultats ne sont, à notre avis, pas représentatifs de la précision réellement obtenue. Il s'agit
simplement d'une indication partielle qui a pu être mise en regard d'autres résultats reportés
dans la littérature. De plus, ces résultats ne prennent pas en compte les erreurs d'étalonnage du
système. Ainsi d'autres validations seront nécessaires dans le futur avec un système complète-
ment étalonné.

Finalement de nombreux résultats de reconstructions 3-D ont été présentés. Des reconstructions ont été effectuées à l'aide d'un système stéréo classique, du système miniaturisé sur des scènes ex-vivo et enfin dans des conditions mini-invasives in-vivo. Les premiers résultats ont montré la robustesse globale aux perturbations spectrales du système proposé et sa capacité à reconstruire des surfaces avec de fortes discontinuités de courbures. Des motifs de résolution importante (150×100), par rapport aux motifs de type M-Array utilisés dans la littérature, ont pu être utilisés. Avec le système miniaturisé, les résultats obtenus sont prometteurs, mais à cause de la forte quantité de bruit dans l'image endoscopique, des motifs de résolution inférieure, projetant des indices visuels de plus grande taille ont été utilisés. Plusieurs reconstructions ont pu être réalisées dans un contexte mini-invasif sur différents types de surfaces organiques présentes dans la cavité intra-abdominale. Des motifs de résolution 30×30 , 50×50 et 70×70 codés avec seulement 2 primitives, ont pu être testés et ont mené à des reconstructions de scènes comportant toutes les perturbations spectrales (spécularités, inter-réflexions, passage du réseau sanguin etc...) que l'on peut rencontrer dans le corps. Cela a été rendu possible, en dépit du niveau élevé de bruit, grâce à l'exploitation de l'invariance épipolaire, à la redondance du codage et aux traitements avancés proposés. Si les résolutions utilisées ici dans les conditions in-vivo sont inférieures à celles que nous avons pu utiliser avec le système stéréo classique, celles-ci sont du même ordre que celles utilisées dans les travaux précédents utilisant des matrices pseudo-aléatoires (voir chapitre 1).

Perspectives

Différents travaux sont envisageables pour améliorer et compléter le système mis en place. Dans la conception même du motif, il serait intéressant d'étudier à court terme, quand les évolutions du dispositif miniaturisé permettront d'utiliser les couleurs, l'opportunité d'exploiter une discrimination spectrale conjointement à la discrimination par la forme actuellement utilisée. En effet l'utilisation de deux couleurs seulement réduit les risques de mauvaises classification malgré les perturbations spectrales présentes dans les scènes. Un étalonnage complet du dispositif pourrait également être réalisé et rendre possible des applications de métrologie des scènes intra-abdominales. Le prototype actuel a été réalisé avec des éléments du marché non conçus à l'origine pour être associés dans un même système. La conception d'un nouveau système, utilisant la même approche qu'actuellement, mais pour lequel l'efficacité optique serait optimisée, mènerait à un gain important de la quantité de lumière projetée dans la scène et augmenterait de façon significative le rapport signal/bruit. Cela autoriserait l'utilisation, in-vivo, de motifs de plus grande résolution, donc avec des primitives de plus petite taille afin de capturer davantage de discontinuité de courbure sur des surfaces plus éloignées des optiques que les résultats reportés au chapitre 4.

A plus long terme, des motifs dynamiques, s'adaptant en temps-réel aux propriétés spectrales et géométriques de la scène peuvent être envisagés pour maximiser la robustesse et/ou la densité d'indices visuels projetés. Il serait également intéressant d'étudier les différentes possibilités pour rendre le motif invisible. Cela pourrait être utile pour utiliser conjointement les reconstructions 3-D avec les nombreuses informations 2D de texture des surfaces existant dans les scènes intra-abdominales.

Enfin, les résultats obtenus peuvent sans doute déjà être exploités dans diverses applications nécessitant une information 3-D. Nous pensons notamment au guidage robotisé par asservissement visuel, dont la convergence serait alors assurée. Pour des applications de réalité augmentée, un recalage d'organes dont les modèles sont préexistants est également envisageable à moyen terme si les quelques améliorations proposées plus tôt sont mises en oeuvre et que l'infor-

mation de texture peut être utilisée conjointement aux reconstructions. Pour une visualisation simultanée sous plusieurs points de vue, il serait nécessaire au préalable, de rendre le motif invisible. Des traitements de plus haut niveau pourraient également être envisagés, comme la cartographie complète de la cavité intra-abdominale réalisée par un recalage de vues obtenues au fur et à mesure des acquisitions.

Annexe A

Codes optimaux

Dans cette annexe nous donnons quelques résultats remarquables obtenus lors de la génération de motifs alignés sur la géométrie épipolaire. Les codages optimaux sont issus d'une recherche exhaustive menée à son terme : toutes les combinaisons ont été testées. Le nombre de colonnes maximal absolu a donc été atteint pour le nombre de symboles (k), la distance de Hamming minimale (H_{Min}) données et la taille de voisinages utilisés (3×3).

Code optimal, $k = 2$, $H_{Min} = 5$, 7 colonnes

```
1 1 0 0 1 0 1
0 1 1 1 0 0 0
0 1 0 1 1 0 0
```

Code optimal, $k = 2$, $H_{Min} = 4$, 18 colonnes

```
0 0 1 0 0 0 1 1 0 1 0 0 0 0 1 1 1 1
1 1 1 0 1 1 0 0 0 0 0 1 0 0 1 0 1 1
1 0 0 0 0 1 1 0 0 1 1 1 0 1 0 1 1 0
```

Code optimal, $k = 2$, $H_{Min} = 3$, 39 colonnes

```
0 0 0 1 0 0 0 1 0 1 0 0 1 1 0 0 0 1 1 1 1 0 1 1 0 1 1 1 1 0 1 0 1 1 0 0 0 1 0 0
0 0 0 1 1 1 1 1 0 0 1 1 0 0 0 1 0 0 0 0 1 0 1 0 0 1 1 0 1 0 1 0 1 1 0 1 1 0 0
0 0 1 1 0 0 1 0 1 0 1 0 1 0 0 1 1 1 1 1 0 1 0 0 1 1 0 0 0 0 0 1 1 0 0 1 1 0 1 0 1
```

Code optimal, $k = 3$, $H_{Min} = 7$, 8 colonnes

```
0 1 0 2 1 2 0 1
0 1 0 2 1 2 0 1
1 1 0 0 2 2 1 1
```

Code optimal, $k = 3$, $H_{Min} = 6$, 26 colonnes

```
0 1 0 1 0 0 0 2 1 0 0 2 2 0 1 1 1 1 2 2 0 2 1 2 0 1
0 2 2 1 1 2 0 0 0 0 1 1 2 1 0 1 2 0 2 0 2 2 2 1 0 2
0 0 1 1 2 1 0 1 2 0 2 0 2 2 2 1 0 2 2 1 1 2 0 0 0 0
```

Code optimal, $k = 4$, $H_{Min} = 7$, 17 colonnes

```
0 1 0 0 2 3 2 1 0 3 2 2 1 3 1 0 1
0 1 1 2 0 0 2 2 2 3 1 3 0 3 3 0 1
```

1 2 0 1 3 1 0 3 2 0 2 2 0 3 1 1 2

Code quasi-optimal (511 voisinages sur 512 possibles). $k = 2$, $H_{\text{Min}} = 1$, Dans ce cas la recherche exhaustive n'a pas été menée à son terme, à cause de l'explosion combinatoire due au nombre de voisinages. Mais cependant le maximum théorique (512) est quasiment atteint, à 1 voisinage près. Par doute nous avons re-vérifié, après coup, l'unicité des voisinages avec une routine très simple qui se contente de vérifier l'unicité en comparant tous les couples de mots de code possibles.

Ligne 1 :

```
1 0 0 0 0 0 1 0 1 1 0 1 0 0 0 1 1 0 1 0 1 0 1 0 1 0 1 0 0 1 0 1 0 0 0 1 0 0 1 1 1 0 0 0 1 1 0 0
1 1 0 0 1 1 1 0 1 1 0 1 1 1 1 1 0 0 0 1 1 1 1 0 0 0 1 0 1 0 0 1 0 0 0 0 0 1 0 0 1 1 0 1 1 0 1 0 0 0 0
1 1 0 1 1 1 0 0 0 1 1 1 1 0 0 0 1 0 1 0 0 1 1 0 1 0 1 1 1 0 1 1 1 1 0 1 1 0 1 0 0 0 0 1 0 0 0 1 1 1
0 1 1 0 1 0 0 0 1 0 1 1 0 1 1 0 0 0 1 1 1 1 0 0 0 0 0 0 1 1 0 1 1 0 1 0 0 1 1 1 0 1 1 1 1 0 1 0 0 1
0 0 1 0 0 1 0 0 0 1 0 1 1 0 0 1 0 1 1 1 0 0 1 0 1 0 0 0 1 0 0 1 0 0 1 0 0 1 0 0 0 0 0 0 0 0 0 1 0 0
0 1 1 1 1 1 1 0 1 0 1 1 0 0 0 1 1 0 0 1 1 0 1 1 1 1 1 1 0 1 0 1 0 1 1 0 0 0 0 0 1 0 0 0 0 0 1 1 1 1
1 1 0 1 0 1 1 0 1 0 0 0 1 1 1 1 0 1 1 1 0 0 1 0 1 1 1 0 1 0 0 1 0 1 1 0 0 1 0 0 0 0 1 0 0 1 0 0 1 0
0 0 1 1 1 0 0 1 1 0 0 1 0 1 0 1 0 0 1 0 0 0 1 1 0 1 0 1 1 1 1 1 0 0 0 0 1 1 0 1 1 1 0 1 0 1 1 1 0 0
0 1 0 1 0 1 1 1 1 0 0 0 0 1 1 0 0 0 1 0 0 1 0 0 1 1 1 0 1 1 0 1 1 1 1 1 0 1 1 1 1 1 1 1 0 1 1 1 1 1
0 0 0 0 0 0 0 0 1 1 0 1 0 1 1 1 1 0 0 1 0 0 0 1 1 1 1 1 1 1 0 1 1 0 0 0 1 1 1 0 0 0 1 0 1 1 0 1 1 0
0 0 0 1 0 1 1 0 0 0 0 1 1
```

Ligne 2 :

```
0 0 1 1 1 0 1 1 0 1 0 1 1 0 1 0 1 0 1 0 1 1 1 1 0 0 0 0 1 1 0 0 0 1 0 0 1 1 1 1 1 1 0 1 1 1 1 0 1 0
1 0 0 0 1 0 0 0 1 1 1 1 1 0 1 1 1 1 0 0 0 1 0 1 1 0 0 0 0 1 0 1 1 1 0 1 0 1 1 1 1 0 0 1 0 1 0 0 0 1
1 0 1 1 0 1 1 1 0 1 0 1 0 1 0 1 1 0 0 0 0 0 0 1 1 1 0 0 0 1 1 1 1 1 0 1 0 0 0 0 1 0 0 0 0 1 1 1 1 0
0 0 1 0 0 0 0 0 1 0 0 1 1 1 1 1 0 0 1 1 0 0 1 0 0 1 1 0 1 1 1 0 0 0 1 0 0 0 0 0 1 0 0 1 1 0 0 1 1 0
0 0 0 1 0 0 0 1 0 1 1 0 0 0 0 0 1 0 0 0 1 1 0 1 1 0 0 1 0 1 1 1 0 1 0 0 1 1 1 1 0 0 1 1 0 0 0 1 1 0
1 1 0 0 0 1 1 0 1 0 1 1 0 0 0 0 0 0 0 0 0 0 1 1 1 0 1 0 1 1 1 0 0 1 0 0 1 0 0 1 0 1 0 1 0 0 0 0 0 1
0 0 1 0 0 0 0 0 1 1 1 1 0 0 1 1 1 1 1 1 0 1 0 0 0 0 1 1 0 0 1 1 0 1 1 1 1 0 1 0 1 1 1 0 0 1 1 1 0
1 0 1 0 1 1 1 0 1 1 0 1 0 0 1 1 0 1 1 0 0 0 1 0 1 0 1 1 0 0 1 0 0 0 0 1 0 1 1 1 0 1 0 0 1 0 1 0 1 0
1 0 1 0 0 1 1 1 1 0 1 1 1 0 0 0 1 0 0 1 0 0 1 0 0 1 0 0 0 1 0 1 1 0 1 1 1 0 1 0 0 0 1 0 0 0 1 1 0 0
0 0 1 0 1 1 1 0 0 0 1 0 0 0 1 1 1 0 1 1 1 0 0 1 1 1 0 0 1 1 1 1 0 1 1 0 0 1 1 1 1 1 1 1 0 1 1 1 1 0
1 1 1 0 1 0 1 0 0 0 0 1 0
```

Ligne 3 :

```
0 1 1 0 0 0 0 0 0 0 0 1 1 1 0 0 1 1 1 1 0 1 1 0 1 1 1 0 1 0 1 1 1 1 0 0 1 0 0 1 0 0 1 1 0 1 1 1 0 0
1 1 1 1 0 0 1 1 0 1 0 1 1 0 0 1 1 0 0 1 0 0 1 0 0 1 0 1 1 0 1 1 1 0 0 1 0 1 1 0 1 1 1 1 0 1 0 1 1 0
1 0 0 1 1 1 0 0 1 0 1 1 1 1 1 1 0 0 1 0 1 1 1 0 1 0 0 0 0 0 0 1 0 0 1 0 0 0 0 1 1 0 0 1 1 1 0 1
1 0 1 0 1 0 0 0 1 1 0 0 0 1 0 1 0 0 0 0 0 0 1 1 1 1 1 0 1 0 0 1 0 1 0 0 0 0 0 1 1 0 1 1 0 1 1 0 1 1
1 0 0 0 1 0 1 0 1 1 0 0 0 1 1 1 1 1 0 1 0 0 0 1 1 1 0 0 0 0 0 1 1 1 1 0 1 1 1 1 1 0 1 0 1 0 1 0 0 1
0 0 1 1 0 1 0 0 0 1 0 0 1 0 0 1 0 0 1 0 0 0 1 1 0 0 0 1 0 1 1 0 1 1 1 0 0 0 0 0 1 0 1 1 0 1 1 1 1 0
1 0 0 0 1 0 1 0 0 1 1 1 1 1 1 1 0 0 1 1 1 0 0 0 0 1 1 0 0 1 1 1 1 1 0 1 0 1 0 1 0 0 1 0 0 0 1 0 0 0
1 1 1 0 1 0 1 0 0 1 1 0 1 0 1 0 1 1 0 0 1 0 1 0 1 1 1 1 0 0 1 0 0 0 1 1 0 1 1 0 1 0 0 1 0 1 0 0 1 0
1 1 0 0 0 0 1 0 0 0 0 1 0 1 1 1 1 1 0 0 0 1 1 0 1 1 0 1 1 0 1 1 1 1 0 0 0 0 1 1 0 0 0 0 0 1 0 1 1 1
0 1 0 0 0 0 1 1 1 0 1 0 0 0 1 1 1 0 0 0 0 0 1 1 0 1 0 1 0 0 1 0 0 0 1 0 0 0 0 0 0 0 0 1 1 1 1 1 1 1
0 1 1 0 0 1 1 1 1 1 0 0 0
```

Annexe B

Image intégrale

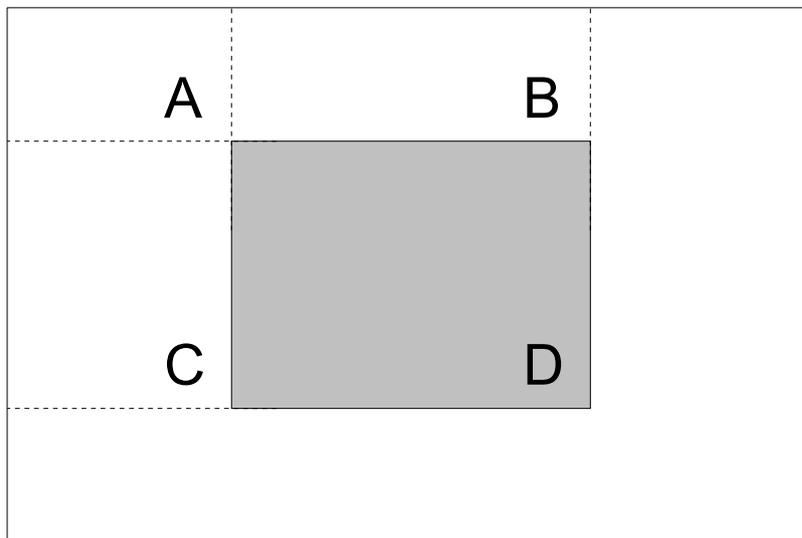


FIGURE B.1 – Une image intégrale. L'aire en gris est calculée en 4 accès mémoire à partir de A, B, C, D.

Avec les images intégrales (Fig. B.1), il est possible de calculer l'aire d'une région rectangulaire avec seulement 4 accès en mémoire, quelle que soit la taille de la région.

Calcul de l'image intégrale.

Soient, ii , l'image intégrale, de coordonnées (x, y) , et i , l'image source de coordonnées (x', y') , ii et i sont de même dimensions. A chaque emplacement de l'image intégrale correspond la somme du rectangle supérieur gauche de l'image source, tel que :

$$ii(x, y) = \sum_{x' \leq x} \sum_{y' \leq y} i(x', y'), \quad (\text{B.1})$$

Calcul de l'aire ABCD en temps constant (4 accès mémoire) :

$$\text{Aire}_{ABCD} = (D + A) - (C + B), \quad (\text{B.2})$$

Une implémentation récursive, n'effectuant qu'une lecture de l'image existe. Elle est donnée par :

$$s(x, y) = s(x, y - 1) + i(x, y) \quad (\text{B.3})$$

$$ii(x, y) = ii(x - 1, y) + s(x, y), \quad (\text{B.4})$$

où $s(x, y)$ est la somme cumulée de la ligne x jusqu'à la colonne y .

L'implémentation récursive classique (equ. B.3 et B.4) étant adaptée à un calcul séquentiel, par CPU, nous proposons l'implémentation parallélisée suivante sur le GPU. Cette implémentation est optimisée pour effectuer des lectures mémoire en union ou "Coalesce" sur le GPU. Chaque unité de calcul (UC) traite 1 colonne en parallèle, les données sont rangées en mémoire de façon contiguës pour chaque ligne, pour une image de taille initiale $X \times Y$. La somme, $sumCols$, est d'abord calculée de façon indépendante sur les colonnes (equ. B.5) puis sur les lignes (colonnes du résultat précédent transposé) (equ. B.6), dans le même esprit que les filtres séparables. La transposée de l'image intégrale ii^T est alors obtenue par la relation suivante :

$$sumCols(UC_x, y) = \sum_{y' \leq Y} i(UC_x, y'), \text{ avec } x \in [0, X[\quad (\text{B.5})$$

$$ii^T(UC_x, y) = \sum_{y' \leq X} sumCols^T((UC_x, y'), \text{ avec } x \in [0, Y[, \quad (\text{B.6})$$

où l'image initiale est de dimensions (X, Y) et UC_x correspond aussi bien à l'indice x des colonnes des images qu'à l'unité de calcul UC qui lui est associée.

Les images intégrales peuvent donc servir de méthode de moyennage, mais non moyennageuse puisque, bien qu'issue du domaine du graphisme par ordinateur ([Crow84]), son utilisation n'a été popularisée dans la communauté vision qu'en 2001 par Viola and Jones [Viola01] et est largement utilisée aujourd'hui, notamment pour la détection d'objets par la méthode du *boosting* qui utilise des classifieurs basés sur les fonctions de Haar (*Haar features*) (voir Papageorgiou et al. [Papageorgiou98]).

Bibliographie

- [Ach01] «A new approach to 3D reconstruction without camera calibration». *Pattern Recognition*, Vol. 34, n° 12, pages 2467 – 2476, 2001.
- [Adan04] A. Adan, F. Molina, et L. Morena. «Disordered Patterns Projection For 3D Motion Recovering». Dans «Proceedings of the 3D Data Processing, Visualization, and Transmission, 2nd International Symposium», 3DPVT '04, pages 262–269. IEEE Computer Society, Washington, DC, USA, 2004. ISBN 0-7695-2223-8.
- [Adán09] A. Adán, A. Vázquez, C. Cerrada, et S.Salamanca. «Moving surface extraction based on unordered hexagonal perfect submaps projection : Applications to 3D feature tracking». *Image Vision Computing*, Vol. 27, n° 8, pages 1083–1096, 2009. ISSN 0262-8856.
- [Albitar07] C. Albitar, P. Graebing, et C. Doignon. «Robust structured light coding for 3D reconstruction». Dans «Int'l Conf. on Computer Vision», Rio de J., Brazil, 2007.
- [Albitar09a] C. Albitar. *Reconstruction 3D par vision active codée. Application à la vision endoscopique..* P. graebing (directeur),c. doignon (co-directeur), m. mouaddib (rapp. externe), e. marchand(rapp. externe), j.m. dischler (rapp. interne), d. wolf (examinateur), Université de Strasbourg, Mar 2009.
- [Albitar09b] C. Albitar, C. Doignon, et P. Graebing. «Calibration of vision systems based on pseudo-random patterns». Dans «IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2009)», Oct 2009.
- [Altschuler81] M. Altschuler, B. Altschuler, et J. Taboada. *A laser electro-optic system for rapid 3-D topographic mapping of surfaces*. Medical Image Processing Group technical report. State University of New York at Buffalo, Dept. of Computer Science, 1981.
- [Armangué03] X. Armangué et J. Salvi. «Overall view regarding fundamental matrix estimation». *Image and Vision Computing*, Vol. 21, pages 205–220, 2003.
- [Atzpadin04] N. Atzpadin, P. Kauff, et O. Schreer. «Stereo analysis by hybrid recursive matching for real-time immersive video conferencing». *Circuits and Systems for Video Technology, IEEE Transactions on*, Vol. 14, n° 3, pages 321 – 334, march 2004. ISSN 1051-8215.
- [Bartoli01] A. Bartoli, P. Sturm, et R. Haraud. «Projective structure and motion from two views of a piecewise planar scene». Dans «Computer Vision, 2001. ICCV 2001. Proceedings. Eighth IEEE International Conference on», Vol. 1, pages 593 –598 vol.1. 2001.

- [Batlle98] J. Batlle, E. Mouaddib, et J. Salvi. «Recent progress in coded structured light as a technique to solve the correspondence problem : a survey». *Pattern Recognition*, Vol. 31, n° 7, pages 963–982, 1998.
- [Benveniste10] R. Benveniste et C. U andnsalan. «A Color Invariant Based Binary Coded Structured Light Range Scanner for Shiny Objects». Dans «Pattern Recognition (ICPR), 2010 20th International Conference on», pages 798–801. aug. 2010.
- [Bezdidko74] S. Bezdidko. «The use of Zernike polynomials in optics». *Sov. J. Opt. Technol.*, Vol. 41, pages 425–429, 1974.
- [Boufama93] B. Boufama, R. Mohr, et F. Veillon. «Euclidean Constraints for Uncalibrated Reconstruction». Dans «International Conference on Computer Vision», 1993.
- [Brown66] D. Brown. «Decentering Distortion of Lenses». *Photogrammetric Engineering*, Vol. 32, n° 3, pages 444–462, 1966.
- [Caban04] J. J. Caban et B. W. Seales. «Reconstruction and Enhancement in Monocular Laparoscopic Imagery». Dans «Proc. of Medicine Meets Virtual Reality», Vol. 12. 2004.
- [Carrhill85] B. Carrhill et R. A. Hummel. «Experiments with the intensity ratio depth sensor». *Computer Vision, Graphics, and Image Processing*, Vol. 32, n° 3, pages 337–358, 1985.
- [Chan03] M. Chan, W. Lin, C. Zhou, et J. Y. Qu. «Miniaturized Three-Dimensional Endoscopic Imaging System Based on Active Stereovision». *Applied Optics*, Vol. 42, n° 10, pages 1888–1898, 2003.
- [Chena] C.-S. Chen, Y.-P. Hung, C.-C. Chiang, et J.-L. Wu. «Range data acquisition using color structured lighting and stereo vision.» *Image Vision Comput.*
- [Chenb] S. Y. Chen, Y. F. Li, et J. Zhang. «Vision Processing for Realtime 3-D Data Acquisition Based on Coded Structured Light.» *IEEE Transactions on Image Processing*.
- [Chen87] C. Chen et A. Kak. «Modeling and calibration of a structured light scanner for 3-D robot vision». *IEEE International Conference on Robotics and Automation, ICRA*, 1987.
- [Chen03] Z. Chen, C. Wu, et H.-T. Tsui. «A new image rectification algorithm». *Pattern Recognition Letters*, Vol. 24, n° 1-3, pages 251–260, 2003.
- [Chen10] Y. Chen, K. Wang, W. Zuo, et Q. Wu. «Sub-Pixel Coded Structured Light Stripe Boundary Location Using Weighted Centroid Method». Dans «Biomedical Engineering and Computer Science (ICBECS), 2010 International Conference on», pages 1–4. april 2010.
- [Cheng10] F.-H. Cheng, C.-T. Lu, et Y.-S. Huang. «3D Object Scanning System by Coded Structured Light». Dans «Proceedings of the 2010 Third International Symposium on Electronic Commerce and Security», ISECS '10, pages 213–217. IEEE Computer Society, Washington, DC, USA, 2010. ISBN 978-0-7695-4219-5.
- [Chien09] H.-J. Chien, C.-Y. Chen, C.-F. Chen, et Y.-M. Su. «Adaptive pixel classifier for binary structured light : A probabilistic kernel approach». Dans

- «Image and Vision Computing New Zealand, 2009. IVCNZ '09. 24th International Conference», pages 367–372. nov. 2009. ISSN 2151-2205.
- [Claes07] K. Claes et H. Bruyninckx. «Robot positioning using structured light patterns suitable for self calibration and 3D tracking». Dans «Int'l Conf. on Advanced Robotics», Jeju, Korea, 2007.
- [Claes08] K. Claes. «Structured Light Adapted to Control a Robot Arm». Dans «PhD thesis, Kath. Univ. Leuven», 2008.
- [Clancy11] N. T. Clancy, D. Stoyanov, G.-Z. Yang, et D. S. Elson. «An endoscopic structured lighting probe using spectral encoding». Dans «SPIE-INT SOC OPTICAL ENGINEERING», Munich, Allemagne, Mai 2011.
- [Cobelli09] P. Cobelli, A. Maurel, V. Pagneux, et P. Petitjeans. «Global measurement of water waves by Fourier transform profilometry». *Experiments in Fluids*, Vol. 46, pages 1037–1047, 2009. ISSN 0723-4864. 10.1007/s00348-009-0611-z.
- [Crow84] F. C. Crow. «Summed-area tables for texture mapping». *SIGGRAPH Comput. Graph.*, Vol. 18, pages 207–212, January 1984. ISSN 0097-8930.
- [D. Fofi01] e. J. S. D. Fofi, et E.M. Mouaddib. «DÃ©codage d'un motif structurant codÃ© par la couleur». *Colloque sur le traitement du signal et des images No18*, , n^o 14, Septembre 2001.
- [Dao09] V. N. Dao et M. Sugimoto. «A robust and dynamic scene geometry acquisition technique for a mobile projector-camera system». Dans «ACM SIGGRAPH ASIA 2009 Sketches», SIGGRAPH ASIA '09, pages 27 :1–27 :1. ACM, New York, NY, USA, 2009.
- [deBruijn46] N. G. de Bruijn. «A combinatorial problem». Dans «Nederlandske Akad. van Wetenschappen», pages 758–764. 1946.
- [Desjardins07] D. Desjardins et P. Payeur. «Dense Stereo Range Sensing with Marching Pseudo-Random Patterns». Dans «Proceedings of the Fourth Canadian Conference on Computer and Robot Vision», pages 216–226. IEEE Computer Society, Washington, DC, USA, 2007. ISBN 0-7695-2786-8.
- [Devernay96] F. Devernay et O. Faugeras. «From projective to Euclidean reconstruction». Dans «Proceedings of the 1996 Conference on Computer Vision and Pattern Recognition (CVPR '96)», CVPR '96, pages 264–. IEEE Computer Society, Washington, DC, USA, 1996. ISBN 0-8186-7258-7.
- [Devernay01] F. Devernay et O. Faugeras. «Straight lines have to be straight : automatic calibration and removal of distortion from scenes of structured environments». *Mach. Vision Appl.*, Vol. 13, pages 14–24, August 2001. ISSN 0932-8092.
- [Doignon08] C. Doignon, D. Knittel, et X. Maurice. «A Vision-Based Technique for Edge Displacement and Vibration Estimations of a Moving Flexible Web». *Instrumentation and Measurement, IEEE Transactions on*, Vol. 57, n^o 8, pages 1605–1613, aug. 2008.
- [Dupont10] P. Dupont, J. Lock, B. Itkowitz, et E. Butler. «Design and Control of Concentric-Tube Robots». *Robotics, IEEE Transactions on*, Vol. 26, n^o 2, pages 209–225, april 2010. ISSN 1552-3098.

- [Durrani95] A. F. Durrani et G. M. Preminger. «Three-dimensional video imaging for endoscopic surgery». *Comput Biol Med*, Vol. 25, n° 2, pages 237–47, 1995.
- [Espinal10] J. Espinal, M. Ornelas, H. J. Puga, J. M. Carpio, et J. A. Munoz. «3D Object Reconstruction Using Structured Light and Neural Networks». Dans «Proceedings of the 2010 IEEE Electronics, Robotics and Automotive Mechanics Conference», CERMA '10, pages 74–79. IEEE Computer Society, Washington, DC, USA, 2010. ISBN 978-0-7695-4204-1.
- [Etzion88] T. Etzion. «Constructions for perfect maps and pseudorandom arrays». *IEEE Int'l Trans. on Information Theory*, Vol. 34, n° 5/1, pages 1308–1316, 1988.
- [Faugeras92] O. D. Faugeras. «What can be seen in three dimensions with an uncalibrated stereo rig». Dans «Proceedings of the Second European Conference on Computer Vision», ECCV '92, pages 563–578. Springer-Verlag, London, UK, 1992. ISBN 3-540-55426-2.
- [Faugeras93] O. Faugeras. *Three-dimensional computer vision : a geometric viewpoint*. MIT Press, 1993.
- [Faugeras95] O. Faugeras. «Stratification of 3-D vision : projective, affine, and metric representations». *Journal of the Optical Society of America A*, Vol. 12, pages 46548–4, 1995.
- [Fechteler08] P. Fechteler et P. Eisert. «Adaptive color classification for structured light systems». Dans «Computer Vision and Pattern Recognition Workshops, 2008. CVPRW '08. IEEE Computer Society Conference on», pages 1 –7. june 2008.
- [Fofi01] D. Fofi, J. Salvi, et E. Mouaddib. «Uncalibrated vision based on structured light». Dans «Robotics and Automation, 2001. Proceedings 2001 ICRA. IEEE International Conference on», Vol. 4, pages 3548 – 3553 vol.4. 2001. ISSN 1050-4729.
- [Fofi04] D. Fofi, T. Sliwa, et Y. Voisin. «A Comparative Survey on Invisible Structured Light». *proceedings of SPIE on Machine vision applications in industrial inspection XII*, Vol. 5303, pages 90–98, 2004.
- [Fong05] P. Fong et F. Buron. «High-resolution three-dimensional sensing of fast deforming objects». Dans «Intelligent Robots and Systems, 2005. (IROS 2005). 2005 IEEE/RSJ International Conference on», pages 1606 – 1611. aug. 2005.
- [Geng11] J. Geng. «DLP-based structured light 3D imaging technologies and applications». *Proc. SPIE 7932, 79320B*, janvier 2011.
- [Graebling02] P. Graebling, A. Lallement, D. Y. Zhou, et E. Hirsch. «Optical high-precision three-dimensional vision-based quality control of manufactured parts by use of synthetic images and knowledge for image-data evaluation and interpretation». *Applied Optics*, Vol. 41, n° 14, pages 2627–2643, 2002.
- [Griffin92] P. M. Griffin, L. S. Narasimhan, et S. R. Yee. «Generation of uniquely encoded light patterns for range data acquisition». *Pattern Rec.*, Vol. 25, n° 6, pages 609–616, 1992.

-
- [Gros94] P. Gros. «How to Use the Cross Ratio to Compute Projective Invariants from Two Images». Dans «Proceedings of the Second Joint European - US Workshop on Applications of Invariance in Computer Vision», pages 107–126. Springer-Verlag, London, UK, 1994. ISBN 3-540-58240-1.
- [Guhring01] J. Guhring. «Dense 3-D surface acquisition by structured light using off-the-shelf components». Dans «Proc. Videometrics and Optical Methods for 3D Shape Measurement», pages 220–231. 2001.
- [Hamming50] R. Hamming. «Error Detecting and Error Correcting Codes». *Bell System Technical Journal*, Vol. 26, n° 2, pages 147–160, 1950.
- [Hartley94] R. I. Hartley. «Euclidean Reconstruction from Uncalibrated Views». *Proceedings of the Second Joint European - US Workshop on Applications of Invariance in Computer Vision*, pages 237–256, 1994.
- [Harvent08] J. Harvent, F. Bugarin, J.J. Orteu, M. Devy, P. Barbeau, et G. Marin. «Inspection of aeronautics parts for shape defect detection using a multi-camera system». Dans «International Symposium on Optical Metrology in Industrial, Medical and Daily Life Applications», Orlando, June 2008.
- [Hasegawa01] K. Hasegawa et Y. Sato. «Endoscope system for high-speed 3D measurement». *Systems and Computers in Japan*, Vol. 32, n° 8, pages 30–39, 2001.
- [Hespanha99] J. Hespanha, Z. Dodds, G. Hager, et A. Morse. «What Tasks can be Performed with an Uncalibrated Stereo Vision System?» *International Journal of Computer Vision*, Vol. 35, pages 65–85, 1999. ISSN 0920-5691. 10.1023/A :1008111128520.
- [Hu07] M. Hu, G. Penney, P. Edwards, et D. Hawkes. «3D Reconstruction of internal organ surfaces for minimal invasive surgery». *MICCAI*, pages 68–77, 2007.
- [Hu09] Z. Hu, Q. Guan, S. Liu, et S. Chen. «Robust 3D Shape Reconstruction from a Single Image Based on Color Structured Light». Dans «Artificial Intelligence and Computational Intelligence, 2009. AICI '09. International Conference on», Vol. 3, pages 168 –172. nov. 2009.
- [Ji09] X. Ji, K. Wang, et Z. Wei. «Structured Light Encoding Research Based on Sub-Pixel Edge Detection». Dans «Information Engineering and Computer Science, 2009. ICIECS 2009. International Conference on», pages 1 –4. dec. 2009.
- [Jia11] X.-j. Jia, Z.-j. Zhang, et Q.-c. Yu. «Construction for M-arrays and application in structured light». *Journal of Shanghai University (English Edition)*, Vol. 15, pages 63–68, 2011. ISSN 1007-6417. 10.1007/s11741-011-0693-2.
- [Kallem09] V. Kallem et N. Cowan. «Image Guidance of Flexible Tip-Steerable Needles». *Robotics, IEEE Transactions on*, Vol. 25, n° 1, pages 191 –196, feb. 2009. ISSN 1552-3098.
- [Kang00] S. B. Kang. «Radial Distortion Snakes». *IEICE Trans. Inf. and Syst*, pages 1603–16112, 2000.

- [Keeratitivittayanun11] S. Keeratitivittayanun, T. Kondo, P. Sira-uksorn, T. Phatrapornant, et M. Sato. «3D scan of a color object using a color structured light pattern». Dans «Signal Processing and its Applications (CSPA), 2011 IEEE 7th International Colloquium on», pages 460–463. march 2011.
- [Keller00] K. Keller et J. Ackerman. «Real-time Structured Light Depth Extraction». *proceedings of SPIE on Three Dimensional Image Capture and Applications III*, Vol. 3958, pages 11–18, 2000.
- [Keller03] K. Keller, J. Ackerman, M. Rosenthal, H. Fuchs, et A. State. «Methods and systems for real-time structured light depth extraction and endoscope using real-time structured light depth extraction». 2003.
- [Kiyasu95] S. Kiyasu, H. Hoshino, K. Yano, et S. Fujimura. «Measurement of the 3-D shape of specular polyhedrons using an M-array coded light source». *Instrumentation and Measurement, IEEE Transactions on*, Vol. 44, n° 3, pages 775–778, jun 1995. ISSN 0018-9456.
- [Koninckx06] T. Koninckx et L. V. Gool. «Real-Time Range Acquisition by Adaptive Structured Light». *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 28, n° 3, pages 432–445, March 2006.
- [Krupa03] A. Krupa, J. Gangloff, C. Doignon, M. de Mathelin, G. Morel, L. Soler, J. Leroy, et J. Marescaux. «Autonomous retrieval and 3D positioning of surgical instruments in robotized laparoscopic surgery». *IEEE Int'l Trans. on Robotics and Automation*, Vol. 19, n° 5, pages 842–853, October 2003.
- [Li10] X. Li et L. Yaping. «Optimization design for structured light system». Dans «Intelligent Computing and Intelligent Systems (ICIS), 2010 IEEE International Conference on», Vol. 3, pages 602–605. oct. 2010.
- [Li11] Q. Li, M. Biswas, M. Pickering, et M. Frater. «Two-shot sparse depth estimation using adaptive structured light». *Electronics Letters*, Vol. 47, n° 13, pages 745–746, 23 2011. ISSN 0013-5194.
- [Lu10] S. Lu, F. Guo, K. Wang, B. Fei, et Y. Li. «One Method for Extracting Sub-pixel Light Stripe Centerlines Applied to the Structured Light». Dans «Intelligent Systems (GCIS), 2010 Second WRI Global Congress on», Vol. 3, pages 343–347. dec. 2010.
- [Maurice09] X. Maurice, P. Graebing, et C. Doignon. «Génération pouet automatique pouet de motifs symétriques pseudo-aléatoires pour la vision active codée». *ORASIS'09 Treizième congrès francophone des jeunes chercheurs en vision par ordinateur, Trégastel, France*, june 2009.
- [Maurice11] X. Maurice, P. Graebing, et C. Doignon. «A pattern framework driven by the Hamming distance for structured light-based reconstruction with a single image.» *CVPR 2011, IEEE International conference on Computer Vision and Pattern Recognition*, Avril 2011.
- [Mitchell95] C. J. Mitchell. «Aperiodic and semi-periodic perfect maps». *IEEE Transactions on Information Theory*, Vol. 41, n° 1, pages 88–95, 1995.
- [Modrow07] D. Modrow, C. Laloni, G. Doemens, et G. Rigoll. «3D face scanning systems based on invisible infrared coded light». Dans «Proceedings of the 3rd international conference on Advances in visual computing - Volume

-
- Part I», ISVC'07, pages 521–530. Springer-Verlag, Berlin, Heidelberg, 2007. ISBN 3-540-76857-2, 978-3-540-76857-9.
- [Mohr94] R. Mohr, B. Boufama, et P. Brand. «Accurate Projective Reconstruction». Dans «Proceedings of the Second Joint European - US Workshop on Applications of Invariance in Computer Vision», pages 257–276. Springer-Verlag, London, UK, 1994. ISBN 3-540-58240-1.
- [Morano98] A. Morano, C. Ozturk, R. Conn, S. Dubin, S. Zietz, et J. Nissanov. «Structured Light Using Pseudo-random Codes». *IEEE Int'l Trans. on Pattern Analysis and Machine Intelligence*, Vol. 20, n° 3, pages 322–327, 1998.
- [Morin93] L. Morin. «Quelques contributions des invariants projectifs à la vision par ordinateur». Dans «Thèse de doctorat, Institut National Polytechnique de Grenoble», 1993.
- [Morita88] H. Morita, K. Yajima, et S. Sakata. «Reconstruction Of Surfaces Of 3-D Objects By M-Array Pattern Projection Method». *ICCV*, Vol. 88, pages 468–473, 1988.
- [Pagès03] J. Pagès, J. Salvi, et C. Matabosch. «Robust Segmentation and Decoding of a Grid Pattern for Structured Light». Dans «IbPRIA», pages 689–696. 2003.
- [Pages04] J. Pages, J. Salvi, et J. Forest. «A new optimised De Bruijn coding strategy for structured light patterns». Dans «Pattern Recognition, 2004. ICPR 2004. Proceedings of the 17th International Conference on», Vol. 4, pages 284 – 287 Vol.4. aug. 2004. ISSN 1051-4651.
- [Pagès05] J. Pagès, J. Salvi, C. Collewet, et J. Forest. «Optimised De Bruijn patterns for one-shot acquisition». *Image and Vision Computing*, Vol. 23, pages 707–720, 2005.
- [Pan10] B. Pan, Q. Guan, X. Wang, et S. Chen. «Strategies of Real-Time 3D Reconstruction by Structured Light». Dans «Pattern Recognition (CCPR), 2010 Chinese Conference on», pages 1 –5. oct. 2010.
- [Papageorgiou98] C. P. Papageorgiou, M. Oren, et T. Poggio. «A General Framework for Object Detection». Dans «Proceedings of the Sixth International Conference on Computer Vision», ICCV '98, pages 555–. IEEE Computer Society, Washington, DC, USA, 1998. ISBN 81-7319-221-9.
- [Patterson94] K. G. Patterson. «Perfect Maps». *IEEE Int'l Trans. on Information Theory*, Vol. 40, n° 3, pages 743–753, 1994.
- [Penne09] J. Penne, K. Höller, M. Stürmer, T. Schrauder, A. Schneider, R. Engelbrecht, H. Feussner, B. Schmauss, et J. Hornegger. «Time-of-Flight 3-D Endoscopy». Dans «Proceedings of the 12th International Conference on Medical Image Computing and Computer-Assisted Intervention : Part I», MICCAI '09, pages 467–474. Springer-Verlag, Berlin, Heidelberg, 2009. ISBN 978-3-642-04267-6.
- [Pollefeys98] M. Pollefeys, R. Koch, M. Vergauwen, et L. V. Gool. «Metric 3D surface reconstruction from uncalibrated image sequences». Dans «3D STRUCTURE FROM MULTIPLE IMAGES OF LARGE SCALE ENVIRONMENTS. LNCS SERIES», pages 138–153. Springer-Verlag, 1998.

- [Pollefeys00] M. Pollefeys. «Tutorial on Visual 3d modeling from images». University of North Carolina, Chapel Hill, USA, 2000.
- [Posdamer82] J. Posdamer et M. Altschuler. «Surface measurement by space-encoded projected beam systems». *Computer Graphics and Image Processing*, Vol. 18, n° 1, pages 1 – 17, 1982. ISSN 0146-664X.
- [Pribanic10] T. Pribanic, H. Dzapo, et J. Salvi. «Efficient and Low-Cost 3D Structured Light System Based on a Modified Number-Theoretic Approach». *EUR-ASIP J. Adv. Sig. Proc.*, Vol. 2010, 2010.
- [Reed60] I. S. Reed et G. Solomon. «Polynomial Codes Over Certain Finite Fields». *Journal of the Society for Industrial and Applied Mathematics*, Vol. 8, n° 2, pages 300–304, 1960.
- [Reichinger11] A. Reichinger. «Kinect Pattern Uncovered». <http://azttm.wordpress.com/2011/04/03/kinect-pattern-uncovered/>, août 2011.
- [Ribo05] M. Ribo et M. Brandner. «State of the art on vision-based structured light systems for 3D measurements». *International Workshop on Robotic and Sensor Environments*, pages 2–6, 2005.
- [Röhl11] S. Röhl et al. «Real-time surface reconstruction from stereo endoscopic images for intraoperative registration». *Spie Medical Imaging*, Mars 2011.
- [Roman10] C. Roman, G. Inglis, et J. Rutter. «Application of structured light imaging for high resolution mapping of underwater archaeological sites». Dans «OCEANS 2010 IEEE - Sydney», pages 1 –9. may 2010.
- [Sadlo05] F. Sadlo, T. Weyrich, R. Peikert, et M. Gross. «A Practical Structured Light Acquisition System for Point-Based Geometry and Texture». *Proceedings of the Eurographics Symposium on Point-Based Graphics*, 2005.
- [Sagawa05] R. Sagawa, M. Takatsuji, T. Echigo, et Y. Yagi. «Calibration of Lens Distortion by Structured-Light Scanning». Dans «Proc. 2005 IEEE/RSJ International Conference on Intelligent Robots and Systems», pages 1349–1354. Edmonton, Canada, August 2005.
- [Sagawa09] R. Sagawa, Y. Ota, Y. Yagi, R. Furukawa, et N. Asada. «Dense 3D reconstruction method using a single pattern for fast moving object». Dans «Int'l Conf. Computer Vision», Kyoto, Japan, 2009.
- [Salvi98] J. Salvi, J. Batlle, et E. Mouaddib. «A robust-coded pattern projection for dynamic 3d scene measurement». *Pattern Recognition Letters*, Vol. 19, pages 1055–1065, 1998.
- [Salvi04] J. Salvi, J. Pagès, et J. Batlle. «Pattern codifications strategies in structured light systems». *Pattern Recognition*, Vol. 37, pages 827–849, 2004.
- [Salvi10] J. Salvi, S. Fernandez, T. Pribanic, et X. Llado. «A state of the art in structured light patterns for surface profilometry». *Pattern Recogn.*, Vol. 43, pages 2666–2680, August 2010. ISSN 0031-3203.
- [Sambongi00] M. Sambongi, M. Igarashi, T. Obi, M. Yamaguchi, N. Oyama, M. Kobayashi, Y. Sano, S. Yoshida, et K. Gono. «Analysis of spectral reflectance of mucous membrane for endoscopic diagnosis». In *Proceedings of the 22nd Annual Engineering in Medicine and Biology Society International Conference*, pages 1026–1029, 2000.

- [Sawhney94] H. S. Sawhney. «3D Geometry From Planar Parallax». *IEEE Conference on Computer Vision and Pattern Recognition*, pages 929–934, juin 1994.
- [Schmalz10a] C. Schmalz et E. Angelopoulou. «Belief propagation for improved color assessment in structured light». Dans «Proceedings of the 32nd DAGM conference on Pattern recognition», pages 403–411. Springer-Verlag, Berlin, Heidelberg, 2010. ISBN 3-642-15985-0, 978-3-642-15985-5.
- [Schmalz10b] C. Schmalz et E. Angelopoulou. «A graph-based approach for robust single-shot structured light». Dans «Computer Vision and Pattern Recognition Workshops (CVPRW), 2010 IEEE Computer Society Conference on», pages 80–87. juin 2010.
- [Sebastián09] J. M. Sebastián, L. Pari, L. Ángel, et A. Traslosheros. «Uncalibrated visual servoing using the fundamental matrix». *Robotics and Autonomous Systems*, Vol. 57, n° 1, pages 1–10, 2009.
- [Shpunt10a] A. Shpunt. «Depth mapping using multi-beam illumination». Dans «US Patent Application 201002820078», 2010.
- [Shpunt10b] A. Shpunt et B. Pesach. «Optical Pattern Projection». Dans «US Patent 20100284082», 2010.
- [Singleton64] R. Singleton. «Maximum distance q-nary codes». *IEEE Trans. on Information Theory*, Vol. 10, pages 116–118, 1964.
- [Song10a] G.-Y. Song et J.-W. Lee. «Correction of Radial Distortion Based on Line-Fitting». *International Journal of Control, Automation, and System*, Vol. 8, n° 3, pages 615–621, 2010.
- [Song10b] Z. Song et C.-K. Chung. «Determining Both Surface Position and Orientation in Structured-Light-Based Sensing». *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, Vol. 32, n° 10, pages 1770–1780, oct. 2010. ISSN 0162-8828.
- [Srinivasan85] V. Srinivasan, H. C. Liu, et M. Halioua. «Automated phase-measuring profilometry : a phase mapping approach». *Appl. Opt.*, Vol. 24, n° 2, pages 185–188, Jan 1985.
- [Stein97] G. P. Stein. «Lens distortion calibration using point correspondences». Dans «Proceedings of the 1997 Conference on Computer Vision and Pattern Recognition (CVPR '97)», CVPR '97, pages 602–. IEEE Computer Society, Washington, DC, USA, 1997.
- [Sutton98] R. S. Sutton et A. G. Barto. *Reinforcement Learning : An Introduction (Adaptive Computation and Machine Learning)*. The MIT Press, march 1998. ISBN 0262193981.
- [Tajima90] J. Tajima et M. Iwakawa. «3-D data acquisition by Rainbow Range Finder». Dans «Pattern Recognition, 1990. Proceedings., 10th International Conference on», Vol. i, pages 309–313 vol.1. jun 1990.
- [Takeda83] M. Takeda et K. Mutoh. «Fourier transform profilometry for the automatic measurement of 3-D object shapes». *Appl. Opt.*, Vol. 22, n° 24, pages 3977–3982, Dec 1983.
- [Thormaehlen02] T. Thormaehlen, H. Broszio, et P. Meier. «Three-Dimensional Endoscopy», 2002.

- [Thormaehlen03] T. Thormaehlen, H. Broszio, et I. Wassermann. «Robust Line-Based Calibration of Lens Distortion from a Single View». Dans «Proceedings of Mirage 2003», pages 105–112. 2003.
- [Trucco98] E. Trucco et A. Verri. *Introductory Techniques for 3-D Computer Vision*. Prentice Hall, 1998.
- [Tsai87] R. Tsai. «A versatile camera calibration technique for high-accuracy 3D machine vision metrology using off-the-shelf TV cameras and lenses». *IEEE Journal on Robotics and Automation*, Vol. 3, n° 4, pages 323–344, août 1987.
- [Vass03] G. Vass et T. Perlaki. «Applying and removing lens distortion in post production». Dans «The Second Hungarian Conference on Computer Graphics and Geometry», 2003.
- [Viola01] P. Viola et M. Jones. «Robust Real-time Object Detection». Dans «International Journal of Computer Vision», 2001.
- [Wei10] Z. Wei, S. Jiang, et X. Ji. «3-D reconstruction based on structured light of composite colorful stripes». Dans «Advanced Computer Theory and Engineering (ICACTE), 2010 3rd International Conference on», Vol. 6, pages V6–123 –V6–125. aug. 2010. ISSN 2154-7491.
- [Weng92] J. Weng, P. Cohen, et M. Herniou. «Camera Calibration with Distortion Models and Accuracy Evaluation». *IEEE Trans. Pattern Anal. Mach. Intell.*, Vol. 14, pages 965–980, October 1992. ISSN 0162-8828.
- [Wolberg94] G. Wolberg. *Digital Image Warping*. IEEE Computer Society Press, Los Alamitos, CA, USA, 1st édition, 1994. ISBN 0818689447.
- [Woodland57] N. Woodland et B. Silver. «Classifying apparatus and method». *US Patent 2612994 (A)*, October 1957.
- [Xu07] Y. Xu et D. G. Aliaga. «Robust pixel classification for 3D modeling with structured light». Dans «Proceedings of Graphics Interface 2007», GI '07, pages 233–240. ACM, New York, NY, USA, 2007. ISBN 978-1-56881-337-0.
- [Xu10] J. Xu, N. Xi, C. Zhang, et Q. Shi. «Real-time 3D shape measurement system based on single structure light pattern.» Dans «ICRA'10», pages 121–126. May 2010.
- [Yamazaki11] S. Yamazaki, A. Nukada, et M. Mochimaru. «Hamming Color Code for Dense and Robust One-shot 3D Scanning». In *Proc. British Machine Vision Conference*, aout 2011.
- [Yanjun10] C. Yanjun. «Medical endoscope containing structured light three-dimensional imaging system». *Invention No. 101836852*, Septembre 2010.
- [Yee94] S. Yee et P. Griffin. «Three-dimensional Imaging System». *Optical Engineering*, Vol. 33, n° 6, pages 2070–2075, 1994.
- [Young07] M. Young, E. Beeson, J. Davis, S. Rusinkiewicz, et R. Ramamoorthi. «View-point coded structured light». Dans «Int'l Conf. on Computer Vision and Pattern Recognition», Minneapolis, MN, USA, June 2007.

-
- [Zhang98] Z. Zhang et T. Kanade. «Determining the Epipolar Geometry and its Uncertainty : A Review». *International Journal of Computer Vision*, Vol. 27, pages 161–195, 1998.
- [Zhang00] Z. Zhang. «A Flexible New Technique for Camera Calibration». *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 22, pages 1330–1334, 2000. ISSN 0162-8828.
- [Zhang02] L. Zhang, B. Curless, et S. Seitz. «Rapid shape acquisition using structured light and multi-pass dynamic programming». Dans «Int'l Symp. on 3D Data Processing, Visualization and Transmission», pages 24–36. 2002.
- [Zhou04] F. Q. A. Zhou et G. J. Zhang. «Complete calibration of a structured light stripe vision sensor through planar target of unknown orientations». *Image and Vision Computing*, Vol. 23, n° 1, pages 59–67, January 2004.

Liste des publications

– Brevet

1. Procédés de codage et de correction et système stéréo de projection et de capture les mettant en oeuvre .
Xavier Maurice and Christophe Doignon. *Brevet provisionnel américain*, ref : 61/563,649 , USA, 25 Novembre 2011.

– Revues Internationales

1. Real-time structured light coding for adaptive pattern.
Xavier Maurice, Pierre Graebing and Christophe Doignon. *Journal of Real-Time Image Processing (JRTIP)*, (à paraître)
2. A subject-specific software solution for the modeling and the visualization of muscles deformations. [Hors-thèse]
Xavier Maurice, Anders Sandholm, Nicolas Pronost, Ronan Boulic and Daniel Thalmann. *The Visual Computer : International Journal of Computer Graphics* , 25 :835-842, juillet 2009.
3. A vision-based technique for edge displacement and vibration estimations of a moving flexible web. [Hors-thèse]
Christophe Doignon, Dominique Knittel and Xavier Maurice. *IEEE Transactions on Instrumentation and Measurement* , 57(8) :1605-1613, août 2008.

– Conférences Internationales

1. A pattern framework driven by the Hamming distance for structured light-based reconstruction with a single image.
Xavier Maurice, Pierre Graebing and Christophe Doignon. *In IEEE International Conference on Computer Vision and Pattern Recognition (CVPR)*, Colorado Springs, Colorado, avril 2011.
2. Epipolar based structured light pattern design for 3-D reconstruction of moving surfaces.
Xavier Maurice, Pierre Graebing and Christophe Doignon. *In IEEE International Conference on Robotics and Automation (ICRA)*, Shanghai, Chine, mai 2011.
3. Real-time structured light coding with subperfect submaps.
Xavier Maurice, Pierre Graebing and Christophe Doignon. *In SPIE Photonic Europe, (Best Student Paper Award)*, Bruxelles, Belgique, avril 2010.
4. On-line vibrations characterization of a moving flexible web with a fast structured lighting stereovision system. [Hors thèse]
Xavier Maurice, Christophe Doignon and Dominique Knittel. *In IEEE Instrumentation and Measurement Technology Conference (IMTC)*, Varsovie, mai 2007.

– Conférences Nationales

1. Méthodologie de conception de motifs dirigée par la distance de Hamming : application à la reconstruction de surfaces à partir d'une seule image.
Xavier Maurice et Christophe Doignon. *Congrès francophone sur la Reconnaissance des Formes et l'Intelligence Artificielle (RFIA'12)*, Lyon, janvier 2012.
2. Conception et appariement de motifs géométriques pour une projection de lumière structurée.
Xavier Maurice, Pierre Graebbling et Christophe Doignon. *Congrès francophone sur la Reconnaissance des Formes et l'Intelligence Artificielle (RFIA'10)*, Caen, janvier 2010.
3. Génération automatique de motifs symétriques pseudo-aléatoires pour la vision active codée.
Xavier Maurice, Pierre Graebbling et Christophe Doignon. *Congrès francophone des jeunes chercheurs en vision par ordinateur (ORASIS'09)*, Trégastel, juin 2009.