



HAL
open science

Quelques contributions à l'analyse et au traitement des images numériques

Frédéric Sur

► **To cite this version:**

Frédéric Sur. Quelques contributions à l'analyse et au traitement des images numériques. Traitement du signal et de l'image [eess.SP]. Université de Lorraine, 2016. tel-01399973

HAL Id: tel-01399973

<https://hal.science/tel-01399973>

Submitted on 21 Nov 2016

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

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

Quelques contributions à l'analyse et au traitement des images numériques

Mémoire

présenté pour l'obtention d'une

Habilitation à Diriger des Recherches
(mention informatique)

par

Frédéric SUR

Mémoire soutenu publiquement le 16 novembre 2016

après avis des rapporteurs :

Isabelle BLOCH, Professeur à Télécom ParisTech

Fabrice HEITZ, Professeur à l'Université de Strasbourg

Lionel MOISAN, Professeur à l'Université Paris Descartes

devant le jury composé de :

Marie-Odile BERGER, Directrice de recherche Inria (garante scientifique)

Isabelle BLOCH, Professeur à Télécom ParisTech

Michel GRÉDIAC, Professeur à l'Université Clermont Auvergne

Lionel MOISAN, Professeur à l'Université Paris Descartes

Salvatore-Antoine TABBONE, Professeur à l'Université de Lorraine (président du jury)

Jean-Philippe THIRAN, Professeur à l'École polytechnique fédérale de Lausanne

Table des matières

1	Introduction	7
1.1	L'analyse et le traitement des images	7
1.2	Contributions scientifiques	8
1.2.1	Cheminement	9
1.2.2	Principaux travaux, plan du mémoire	10
1.2.3	Encadrement scientifique	14
1.2.4	Réalisations logicielles	15
1.2.4.1	Bibliothèques logicielles	15
1.2.4.2	Codes Matlab autonomes	16
1.3	Guide de lecture	17
I	Contributions à l'appariement de points d'intérêt entre deux images ou entre une image et un modèle de scène non-structuré	19
2	Appariement de points d'intérêt entre images	21
2.1	Introduction	21
2.1.1	Contexte	21
2.1.2	Contributions	24
2.2	Appariement de points d'intérêt sous contraintes photométriques et géométriques	24
2.2.1	Un modèle a-contrario pour les correspondances entre points d'intérêt sous contraintes photométriques et géométriques	25
2.2.2	Modèle a-contrario	26
2.2.3	Modélisation de la contrainte géométrique	28
2.2.4	Modélisation de la contrainte photométrique	28
2.2.5	Algorithme	29
2.2.6	Expérience illustrative	29
2.3	Comment rendre ASIFT davantage robuste à l' <i>aliasing</i> perceptuel?	30
2.3.1	ASIFT	30
2.3.2	Modification de ASIFT	32
2.3.3	Expériences illustratives	35
2.4	Prise en compte de l'incertitude de localisation des points d'intérêt	35
2.5	Conclusion et perspectives	39
3	Détermination de la pose d'une caméra	41
3.1	Introduction	41
3.1.1	Contexte	41
3.1.2	Contributions	42
3.2	Construction de mots visuels pour le calcul de pose	42

3.3	Simulation de points de vue pour l'appariement entre une image test et un modèle de la scène	43
3.3.1	Cas des homographies	46
3.3.2	Cas des transformations affines	46
3.3.3	Résumé de la méthode proposée	47
3.3.3.1	Génération des descripteurs synthétiques	47
3.3.3.2	Position des caméras virtuelles	48
3.3.3.3	Choix de la vue utilisée pour la simulation et extraction d'un descripteur SIFT	48
3.3.3.4	Correspondances image/modèle	48
3.3.3.5	<i>Perspective-n-Points</i>	48
3.3.4	Étude expérimentale	49
3.3.4.1	Protocole expérimental	49
3.3.4.2	Robustesse du calcul de pose aux changements de direction de vue	50
3.4	Conclusion et perspectives	58

II Contributions à l'étude de deux types de bruit dans les images numériques 59

4	Estimation des paramètres d'un capteur numérique	61
4.1	Introduction	61
4.1.1	Contexte	61
4.1.1.1	Modèle Poisson-Gauss	61
4.1.1.2	Estimation des paramètres du capteur	62
4.1.2	Contributions	64
4.2	Modélisation du bruit en présence d'une illumination scintillante et de micro-vibrations	65
4.2.1	Modèle cox-gaussien	65
4.2.2	Limites du modèle	66
4.3	Estimation des paramètres de la caméra	67
4.3.1	Estimation du scintillement relatif et élimination de son effet dans la variance	67
4.3.2	Estimation de l'amplitude des vibrations	69
4.3.3	L'effet « obturateur déroulant » (<i>rolling shutter</i>)	69
4.4	Résultats expérimentaux	69
4.4.1	Exemple illustratif	70
4.4.2	Estimation des paramètres de deux appareils photographiques	70
4.5	Estimation du bruit d'un capteur par analyse d'une série d'images de grilles pseudo-périodiques affectée par des vibrations	74
4.5.1	Approche proposée	74
4.5.2	Expériences	76
4.6	Conclusion et perspectives	79

5	Élimination du bruit périodique ou quasi-périodique	81
5.1	Introduction	81
5.1.1	Contexte	81
5.1.2	Contributions	82
5.2	Détection du bruit quasi-périodique comme déviation statistique des coefficients spectraux	83
5.2.1	Modèle d'imagerie	83
5.2.2	Algorithme	84
5.2.3	Exemples illustratifs	85
5.3	Détection a-contrario du bruit quasi-périodique	90
5.3.1	Exemple illustratif	92
5.3.2	Quelques expériences complémentaires	93
5.4	Interpolation des coefficients spectraux par minimisation sous contrainte de la variation totale	95
5.5	Conclusion et perspectives	98
III	Contributions à la mécanique des solides expérimentale	99
6	Quantification des performances métrologiques en mécanique des solides expérimentale	101
6.1	Introduction	101
6.1.1	Contexte	101
6.1.2	Contributions	103
6.2	Quantification des performances métrologiques de la méthode de la grille . . .	104
6.2.1	Modèle d'image de grille et estimation des champs de déplacement et déformation	104
6.2.2	Discussion du phénomène de repliement spectral	106
6.2.3	Biais de mesure, résolution de mesure, et résolution spatiale	107
6.3	Restauration des champs de déplacement et déformation	110
6.4	Quantification des performances métrologiques de la méthode basée sur la corrélation d'images (DIC)	113
6.5	Conclusion et perspectives	116
IV	Quelques perspectives de recherche	119
7	Conclusion et perspectives	121
7.1	Conclusion générale	121
7.2	Quelques perspectives de recherche	121
7.2.1	Vers une analyse spectrale non-locale	122
7.2.2	Étude et amélioration des performances métrologiques en mécanique des solides	123
	Publications personnelles	125
	Bibliographie	131

Introduction

Remarques préliminaires. Ce mémoire présente de manière synthétique une sélection de travaux, depuis ma thèse soutenue en septembre 2004 jusqu’au printemps 2016. Pour faciliter la lecture du document, les références bibliographiques dont je suis co-auteur sont indiquées de manière numérique, en caractères gras (**[1]**,**[3]**...), les autres références étant indiquées sous forme alphanumérique ([FB81],[Mal99]...). La liste de mes publications et la bibliographie générale figurent à la fin de ce mémoire. Pour limiter les risques d’ambiguïtés, certains termes techniques seront employés en langue anglaise, et apparaîtront alors en italique. Enfin, par abus de langage, le terme « caméra » fera généralement référence à un appareil photographique prenant des images sans continuité temporelle.

1.1	L’analyse et le traitement des images	7
1.2	Contributions scientifiques	8
1.3	Guide de lecture	17

1.1 L’analyse et le traitement des images

L’analyse d’images est la discipline scientifique qui vise à extraire de l’information à partir d’une ou plusieurs images. La première formalisation moderne de l’analyse des images vient sans doute des psychologues de la perception, en particulier des Gestaltistes comme Max Wertheimer [Wer23] qui définissent la notion de forme comme une structure émergeant d’un certain nombre de lois naturelles. Fred Attneave [Att54], dans les années 1950, étudie la perception sous l’angle de la théorie de l’information, et constate que l’information visuelle est particulièrement redondante, de la même manière que Claude Shannon mit en évidence la redondance de l’information dans la succession des lettres de l’anglais imprimé [Sha48]. Dans une expérience célèbre, Attneave montre qu’une image de chat réduite à trente-huit points de courbure maximale reliés par des segments suffit à la reconnaissance. L’analyse des images serait donc basée sur un nombre restreint de caractéristiques saillantes. Avec l’apparition des ordinateurs et l’essor du traitement numérique des images, de nombreux chercheurs ont contribué à développer la nouvelle discipline qu’était la vision artificielle, de manière à permettre l’analyse algorithmique des images à l’aide de l’outil informatique. Les premiers travaux se font en parallèle aux avancées des neuro-physiciens qui formalisent l’étude des très bonnes

performances du cortex visuel humain. David Marr [Mar82] formalise la perception visuelle comme une succession d'opérations allant de l'image (un champ d'intensité lumineuse) au « schéma fondamental » (*primal sketch*, fait de points caractéristiques particuliers provenant de détecteurs de contours ou d'extrema de courbure par exemple) puis au modèle dit 2D-1/2 de l'image représentant les surfaces visibles des objets composant la scène et leurs profondeurs relatives, le modèle 3D de la scène étant finalement composé de primitives volumiques placées les unes par rapport aux autres dans une organisation hiérarchique. Pendant de nombreuses années, la communauté scientifique a cherché à construire des outils algorithmiques permettant une approche pyramidale proche de celle formalisée par Marr. Les questions posées dans ce cadre sont diverses. Quels sont les meilleurs descripteurs de bas niveau pour tel ou tel problème ? Comment les hiérarchiser pour détecter des structures intermédiaires ? Comment en tirer une représentation sémantique du monde, ou en déduire une loi de commande d'un robot, ou détecter un colis suspect sur un quai ? Notons que cette approche ne semble pas rendue obsolète par l'explosion récente des performances des réseaux de neurones artificiels et de l'apprentissage profond : on peut supposer que la compréhension de ce qu'apprennent effectivement ces réseaux et des structures qui en émergent pourrait permettre d'affiner la modélisation de la pyramide perceptive.

Les travaux décrits dans ce mémoire s'inscrivent dans le cadre général de l'analyse des images numériques. La première partie traite de l'appariement de points d'intérêt et du problème dit de la structure et du mouvement. À partir d'une ou de plusieurs images, l'objectif est d'apparier des points d'intérêt extraits indépendamment dans les images, de manière à ce qu'ils soient la projection du même point de la scène. De ces appariements, il est possible de déterminer un modèle de la scène tridimensionnelle et de calculer la position des caméras ayant généré les images. Il s'agit, d'un certain point de vue, de réaliser le programme de Marr, même si nous nous limiterons à des modèles de scène non structurés. La seconde partie du mémoire porte sur un aspect inhérent à la formation des images, à savoir le bruit. La nature quantique des photons fait que l'intensité lumineuse mesurée en un photosite d'un capteur numérique présente des fluctuations aléatoires, générant du bruit dans l'image finale. Nous nous intéressons aussi au bruit périodique se traduisant comme une perturbation texturée superposée à l'image. Enfin la troisième partie présente une application pratique du traitement et de l'analyse des images, dans le cadre d'une collaboration interdisciplinaire. Nous proposons l'analyse de différentes méthodes utilisées en mécanique des solides expérimentale pour caractériser par un procédé d'imagerie les déformations d'un matériau soumis à un test de compression ou de traction. Le bruit dans les images numériques apparaît comme un des principaux éléments limitant la résolution des méthodes employées. Nous nous sommes attachés à caractériser les performances métrologiques de ces méthodes, et nous avons proposé l'étude de problèmes originaux de l'analyse et du traitement des images issues de ce problème pratique.

1.2 Contributions scientifiques

Après une courte description de mon parcours scientifique, les principaux travaux décrits dans ce mémoire sont résumés. Cette section comprend également une liste des thèses et stages encadrés, ainsi qu'une liste de réalisations logicielles distribuées publiquement.

1.2.1 Cheminement

Après un cursus en mathématiques à l'École Normale Supérieure de Cachan et à son antenne de Rennes (à présent École Normale Supérieure de Rennes), j'ai soutenu une thèse en septembre 2004 au Centre de Mathématiques et Leurs Applications (CMLA, UMR 8536) à l'ENS de Cachan, sous la direction de Jean-Michel Morel. Mon travail de thèse portait sur la formalisation d'un problème de l'analyse des images, à savoir la définition de la notion de forme géométrique et l'appariement des formes entre images. Plus précisément, il abordait les problèmes de la détection des formes comme des lignes de niveau significativement contrastées [18], d'appariement de descripteurs semi-locaux de formes invariants par transformations affines [17, 20], et de détection de groupes de formes appariées dans une classification hiérarchique [16], dans le cadre de la théorie statistique de détection a-contrario [DMM08]. Une présentation complète de ces travaux ainsi que diverses extensions sont disponibles dans le livre [1].

Les principales publications relatives à ces travaux sont :

[1] F. Cao, J.L. Lisani, J.-M. Morel, P. Musé, and F. Sur. *A theory of shape identification*. Number 1948 in *Lecture Notes in Mathematics*. Springer, 2008. 264 pages.

[16] F. Cao, J. Delon, A. Desolneux, P. Musé, and F. Sur. A unified framework for detecting groups and application to shape recognition. *Journal of Mathematical Imaging and Vision*, 27(2):91–119, 2007.

[17] P. Musé, F. Sur, F. Cao, Y. Gousseau, and J.-M. Morel. An a contrario decision method for shape element recognition. *International Journal of Computer Vision*, 69(3):295–315, 2006.

[18] F. Cao, P. Musé, and F. Sur. Extracting meaningful curves from images. *Journal of Mathematical Imaging and Vision*, 22(2-3):159–181, 2005.

[20] P. Musé, F. Sur, and J.-M. Morel. Sur les seuils de reconnaissance des formes. *Traitement du Signal*, 20(3):279–294, 2003.

Bien que certains de ces travaux ont été publiés plusieurs années après la fin de ma thèse, je ne les aborderai pas dans le présent document. J'ai ensuite effectué un stage post-doctoral au Laboratoire Lorrain de Recherche en Informatique et ses Applications (LORIA, UMR 7503) à Nancy, sous la direction de Yann Guermeur, en 2004-2005. Ce stage portait sur différents problèmes d'apprentissage automatique (*machine learning*) et en particulier sur les machines à noyau [Vap00]. Si j'ai travaillé sur des problèmes de l'apprentissage non-supervisé (classification, ou *clustering*) pendant ma thèse, l'apprentissage automatique est pour sa part largement supervisé. L'apprentissage non-supervisé consiste à identifier des classes d'observations similaires à partir d'une base de données préalablement acquises. Dans l'apprentissage supervisé, la base de données a été préalablement étiquetée (les classes d'appartenance sont connues a priori), et le problème est d'entraîner à l'aide de cette base un algorithme pouvant prédire la classe d'une nouvelle observation avec de bonnes capacités de généralisation. Dans les méthodes dites « à noyau », le noyau joue le rôle d'un produit scalaire dans un espace de Hilbert où les observations sont virtuellement plongées. Il permet de comparer les observations et est adapté à leur nature. J'ai travaillé à la construction d'un noyau basé sur un modèle de Markov caché pour une machine à vecteurs supports (*Support Vector Machine*, SVM), dans le but de prédire la structure secondaire de protéines. J'ai aussi contribué à l'élaboration d'une borne du risque pour des machines à vecteurs supports multiclassées [38].

[38] Y. Guermeur, M. Maumy, and F. Sur. Model selection for multi-class SVMs. In *Proceedings of the International Symposium on Applied Stochastic Models and Data Analysis (ASMDA)*, pages 506–517, Brest, France, 2005.

Par souci de concision, ces travaux ne seront pas non plus détaillés dans ce document.

Après avoir été ATER au département d'informatique de l'École des Mines de Nancy (INPL, à présent Université de Lorraine) en 2005-2006, je suis depuis septembre 2006 maître de conférences au département Génie Industriel et Mathématiques Appliquées de cette école, et je suis membre du projet Inria Magrit, dirigé par Marie-Odile Berger au LORIA (UMR 7503, CNRS, Inria, Université de Lorraine). Ce sont mes travaux récents au sein du projet Magrit qui constituent les contributions à l'analyse et au traitement des images numériques décrites dans ce mémoire.

1.2.2 Principaux travaux, plan du mémoire

Ce mémoire s'articule en trois parties relativement indépendantes.

La **première partie** traite d'un problème à la base de nombreuses applications de la vision par ordinateur, à savoir l'appariement de points d'intérêt. Lorsque l'on cherche la position relative de deux caméras, ou la position d'une caméra dans une scène tridimensionnelle, une possibilité est de commencer à établir des correspondances entre certains points. Dans le premier cas, on dispose de deux images et il faut mettre en correspondance des points d'intérêt extraits automatiquement dans les images par des algorithmes comme SIFT [Low04], qui permettront ensuite de résoudre le problème de calibrage des caméras par l'intermédiaire de la matrice fondamentale. Dans le second cas, on dispose d'une image et d'un modèle non structuré de la scène (par exemple un nuage de points 3D) et il faut mettre en correspondance des points d'intérêt de l'image avec des points 3D pour estimer position et orientation de la caméra dans son environnement, par résolution du problème dit *Perspective-n-Points* [HZ00, LF05]. Les résultats ont pour l'essentiel été obtenus dans le cadre des thèses de Nicolas Noury [Nou11] (soutenue en 2011), Srikrishna Bhat [Bha13] (soutenue en 2013), et Pierre Rolin (thèse commencée en septembre 2013). Ces trois thèses ont été co-encadrées avec Marie-Odile Berger.

Le **chapitre 2** porte sur l'appariement de points d'intérêt entre deux images par un algorithme basé sur la théorie de la détection a-contrario. Les méthodes classiques d'appariement consistent à pré-apparier les points d'intérêt selon une mesure de dissimilarité entre descripteurs photométriques associés à ces points, puis à éliminer les fausses correspondances inévitables à ce stade en imposant une contrainte géométrique par un algorithme d'échantillonnage itératif comme RANSAC [FB81]. La contribution développée dans [14, 36] est un modèle a-contrario permettant de détecter des groupes d'appariement vérifiant simultanément une contrainte photométrique et une contrainte géométrique. Cette approche présente comme avantage une meilleure précision de l'estimation de la matrice fondamentale et une plus grande robustesse à la présence de motifs répétés dans les images, qui sont classiquement mal appariés. L'intégration d'un tel modèle dans ASIFT [MY09, YM11] (qui permet l'appariement entre images présentant de forts changements de points de vue) est ensuite discutée dans [32]. Des modèles a-contrario de détection de groupes d'appariements sous diverses contraintes géométriques, basés sur des points d'intérêt dont la localisation n'est connue qu'avec une précision limitée, font l'objet de [33] et [35].

Les principales publications relatives à ce chapitre sont :

[14] F. Sur, N. Noury, and M.-O. Berger. An a contrario model for matching interest points under geometric and photometric constraints. *SIAM Journal on Imaging Sciences*, 6(4):1956–1978, 2013.

[32] N. Noury, F. Sur, and M.-O. Berger. How to overcome perceptual aliasing in ASIFT? In *Proceedings of the International Symposium on Visual Computing (ISVC), part I*, volume LNCS 6453, pages 231–242, Las Vegas, NV, USA, 2010.

[33] F. Sur. Robust matching in an uncertain world. In *Proceedings of the International Conference on Pattern Recognition (ICPR)*, pages 2350–2353, Istanbul, Turkey, 2010.

[35] F. Sur, N. Noury, and M.-O. Berger. Computing the uncertainty of the 8 point algorithm for fundamental matrix estimation. In *Proceedings of the British Machine Vision Conference (BMVC)*, volume 2, pages 965–974, Leeds, UK, 2008.

[36] N. Noury, F. Sur, and M.-O. Berger. Fundamental matrix estimation without prior match. In *Proceedings of the IEEE International Conference on Image Processing (ICIP)*, volume 1, pages 513–516, San Antonio (TX), USA, 2007.

L'estimation de la pose d'une caméra, lorsqu'un modèle non-structuré de la scène tridimensionnelle est disponible, fait l'objet du **chapitre 3**. Ce problème doit par exemple être résolu dans des applications de réalité augmentée (de manière à insérer de manière réaliste un objet synthétique dans une vue réelle) ou de navigation robotique visuelle (pour positionner un robot dans un environnement préalablement modélisé). Le but est alors de disposer d'appariements entre points d'intérêt de l'image test et points 3D de la scène. Nos contributions sont de deux ordres. D'abord, nous avons exploré l'utilisation d'un vocabulaire visuel pour fournir des appariements à l'algorithme de *Structure from Motion* (SfM) permettant d'obtenir le modèle par ajustement de faisceaux, puis pour fournir des appariements entre l'image test et le modèle [31]. Ensuite, nous avons proposé l'enrichissement du modèle par l'intégration de descripteurs issus de points de vue simulés [19, 22, 25]. Estimer la pose d'une image-test présentant un fort changement de point de vue par rapport aux images disponibles de la scène est en effet très difficile de par l'invariance limitée des descripteurs photométriques usuels. L'enrichissement par simulation permet de faciliter l'appariement entre points d'intérêt de la vue test et du modèle, et donc l'estimation de la pose.

Les principales publications relatives à ce chapitre sont :

[19] P. Rolin, M.-O. Berger, and F. Sur. Simulation de point de vue pour la mise en correspondance et la localisation. *Traitement du signal*, 32(2-3):169–194, 2015.

[22] P. Rolin, M.-O. Berger, and F. Sur. Enhancing pose estimation through efficient patch synthesis. In *Proceedings of the British Machine Vision Conference (BMVC)*, York, UK, 2016.

[25] P. Rolin, M.-O. Berger, and F. Sur. Viewpoint simulation for camera pose estimation from an unstructured scene model. In *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, pages 6320–6327, Seattle (WA), USA, 2015.

[31] S. Bhat, M.-O. Berger, and F. Sur. Visual words for 3D reconstruction and pose computation. In *Proceedings of the Joint 3DIM/3DPVT Conference (3DIMPVT)*, pages 326–333, Hangzhou, China, 2011.

La **deuxième partie** du mémoire est dédiée au bruit dans les images, que ce soit le bruit blanc intrinsèque à tout capteur numérique ou un bruit périodique ou quasi-périodique rencontré avec certains modes d'acquisition des images (principalement en télédétection satellitaire).

Le **chapitre 4** présente des contributions à l'estimation des principaux paramètres d'un capteur imageur numérique, à savoir le gain, le décalage (*offset*), et le niveau du bruit de lecture (*readout noise*). À partir d'une rafale d'images prise caméra immobile, il serait tentant

d'estimer le bruit en supposant qu'il est seul responsable des fluctuations aléatoires de l'intensité lumineuse mesurée en chaque pixel. Or, cette hypothèse est en pratique prise en défaut : la caméra n'est jamais rigoureusement immobile et l'intensité de la source lumineuse fluctue toujours, même faiblement. Dans [9, 27], nous avons proposé un modèle de formation d'une série d'images acquises sous une illumination scintillante et soumises à des micro-vibrations, ce modèle s'appuyant sur un processus stochastique de Cox-Gauss. Dans ce modèle, la variance temporelle du niveau de gris dépend de manière quadratique du niveau de gris moyen, au lieu d'une dépendance affine dans le cas classique d'une illumination d'intensité constante. Nous avons proposé un algorithme permettant, dans ce cadre a priori peu favorable, d'estimer par simple régression l'ensemble des paramètres du capteur. Nous avons également présenté dans [11] un algorithme d'estimation des paramètres gouvernant le niveau de bruit, basé sur une série d'images d'une grille quasi-périodique en présence de micro-vibrations, mais sous éclairage constant. Cet algorithme est utile pour caractériser rapidement ces paramètres dans une expérience issue de la mécanique des solides expérimentale.

Les principales publications relatives à ce chapitre sont :

[9] F. Sur and M. Grédiac. Measuring the noise of digital imaging sensors by stacking raw images affected by vibrations and illumination flickering. *SIAM Journal on Imaging Sciences*, 8(1):611–643, 2015.

[11] F. Sur and M. Grédiac. Sensor noise modeling by stacking pseudo-periodic grid images affected by vibrations. *IEEE Signal Processing Letters*, 21(4):432–436, 2014.

[27] F. Sur and M. Grédiac. Sensor noise measurement in the presence of a flickering illumination. In *Proceedings of the International Conference on Image Processing (ICIP)*, pages 1763–1767, Paris, France, 2014.

Le **chapitre 5** traite du bruit (quasi-)périodique, qui peut affecter certaines images sous la forme d'une texture périodique relativement haute-fréquence couvrant l'intégralité de l'image. Si la contribution de ce type de bruit est habituellement assez bien concentrée dans le spectre de l'image, toute la difficulté est de détecter cette contribution de manière automatique, en discriminant les composantes spectrales du bruit périodique de celles dues aux structures et textures présentes dans les images considérées. Nous avons proposé deux types de détection des composantes spectrales du bruit périodique, l'une basée sur un écart aux propriétés statistiques attendues du spectre de puissance moyen [8], l'autre basée sur une détection a-contrario [23]. Nous discutons également un algorithme d'interpolation des coefficients spectraux permettant de remplacer les composantes spectrales du bruit, qui s'appuie sur une minimisation sous contrainte de la variation totale. Nous avons passé en revue dans [7] l'essentiel des algorithmes de la littérature permettant de retirer le bruit quasi-périodique, dans le cadre du calcul de champs de déformation en mécanique des solides expérimentale, ces champs particuliers ne satisfaisant pas les hypothèses statistiques de [8, 23], plutôt adaptées aux images naturelles.

Les principales publications relatives à ce chapitre sont :

[7] M. Grédiac, F. Sur, and B. Blaysat. Removing quasi-periodic noise in strain maps by filtering in the Fourier domain. *Experimental Techniques*, 40(3):959–971, 2016.

[8] F. Sur and M. Grédiac. Automated removal of quasiperiodic noise using frequency domain statistics. *IS&T / SPIE Journal of Electronic Imaging*, 24(1):013003/1–19, 2015.

[23] F. Sur. An a-contrario approach to quasi-periodic noise removal. In *Proceedings of the IEEE International Conference on Image Processing (ICIP)*, pages 3841–3845, Québec City, Canada, 2015.

La **troisième partie** du mémoire expose des travaux relatifs à l'analyse et au traitement d'images issues d'expériences menées en mécanique des solides expérimentale, visant à caractériser les champs de déplacement et de déformation à la surface d'un matériau soumis à un effort de traction ou de compression. Ces champs de déplacement et de déformation sont obtenus par des méthodes optiques dites « sans contact ». Pour permettre la diffusion de ces méthodes dans l'industrie, il faut certifier les mesures en quantifiant des grandeurs métrologiques comme la résolution de mesure, la résolution spatiale, ou le biais de mesure. Le **chapitre 6** est un résumé des contributions dans ce domaine. Il s'agit de résultats issus d'une collaboration interdisciplinaire avec des collègues de l'Institut Pascal à Clermont-Ferrand spécialistes de mécanique, Michel Grédiac et Benoît Blaysat. L'essentiel de nos contributions vise à établir des formules prédictives pour les performances métrologiques, que l'on cherche à améliorer par restauration des champs calculés. Ces performances sont limitées en particulier par le bruit du capteur numérique employé. Le schéma numérique d'interpolation, indispensable dans certains cas de par la nature discrète de l'image numérique, peut également jouer un rôle sur les performances métrologiques. Le cadre de nos études est celui dit des petites déformations, se traduisant par des déplacements sous-pixelles sur les images du matériau déformé.

Deux techniques sans contact de mesure des champs de déplacement et déformation ont émergé depuis la fin des années 1990. Elles s'appuient toutes deux sur (au moins) deux images de la surface à analyser, l'une acquise avant et l'autre après déformation. L'originalité des problèmes soulevés tient au fait que les déformations ont bien souvent une amplitude très faible, l'évolution des images étant à peine perceptible à l'œil nu. La première technique est basée sur la détermination des champs de déplacement par maximisation de la corrélation entre deux vues de la surface, sur laquelle a été projeté un « mouchetis » aléatoire au cas où la surface ne présente pas une texture suffisamment contrastée. Nous avons contribué à l'analyse des performances métrologiques de ces méthodes basées sur la corrélation d'image (*digital image correlation*, DIC [SOS09]) en caractérisant la résolution de cette méthode. Il s'agit de prendre en compte l'interpolation nécessaire à l'établissement d'un déplacement sous-pixelle dans l'algorithme de Gauss-Newton modifié maximisant la corrélation [5]. Une validation expérimentale, prenant en compte un modèle réaliste du bruit d'un capteur numérique et les conditions expérimentales rencontrées, est proposée dans [6]. La seconde technique est basée sur l'analyse des déformations d'une fine grille périodique transférée sur la surface du matériau déformé, par la méthode dite de la grille [BGM09]. L'avantage de cette technique est qu'elle ne nécessite pas de schéma numérique d'optimisation (contrairement à la DIC), ce qui facilite son étude et permet d'établir des formules prédictives des performances métrologiques dans le cadre de l'analyse de Fourier. Nous avons caractérisé mathématiquement le biais de mesure et le bruit affectant les mesures (qui prend la forme d'un bruit spatialement corrélé) dans [13], puis avons vérifié expérimentalement dans [12] les formules prédictives obtenues. Il s'agit à notre connaissance de la première caractérisation des performances métrologiques de la méthode de la grille. Le biais de mesure se traduisant par une équation de convolution, nous avons proposé une première étude de la restauration des champs de déformation et déplacement dans [15]. Nous avons aussi étudié dans [10] le débruitage par moyennage d'une rafale d'images perturbée par les inévitables micro-vibrations transmises par la dalle du bâtiment dans lequel l'expérience a lieu. L'analyse des images de grilles se faisant par analyse de Fourier à fenêtre, nous avons étudié l'influence du choix de la fenêtre dans les performances métrologiques de la méthode [3]. Enfin, l'étude de l'effet du repliement spectral (*aliasing*) dans les méthodes de grilles est menée dans [2], où nous proposons une méthode simple pour

l'éviter en pratique. Une revue systématique de la littérature et une analyse complète de la méthode de la grille sont présentées dans [4].

Les principales publications relatives à ce chapitre sont :

[2] F. Sur, B. Blaysat, and M. Grédiac. Determining displacement and strain maps immune from aliasing effect with the grid method. *Optics and Lasers in Engineering*, 86:317–328, 2016.

[3] F. Sur and M. Grédiac. Influence of the analysis window on the metrological performance of the grid method. *Journal of Mathematical Imaging and Vision*, 56(3):472–498, 2016.

[4] M. Grédiac, F. Sur, and B. Blaysat. The grid method for in-plane displacement and strain measurement : a review and analysis. *Strain*, 52(3):205–243, 2016.

[5] B. Blaysat, M. Grédiac, and F. Sur. Effect of interpolation on noise propagation from images to DIC displacement maps. *International Journal for Numerical Methods in Engineering*, 108(3):213–232, 2016.

[6] B. Blaysat, M. Grédiac, and F. Sur. On the propagation of camera sensor noise to displacement maps obtained by DIC - an experimental study, *Experimental Mechanics*, 56(6):919–944, 2016.

[10] F. Sur and M. Grédiac. On noise reduction in strain maps obtained with the grid method by averaging images affected by vibrations. *Optics and Lasers in Engineering*, 66:210–222, 2015.

[12] M. Grédiac and F. Sur. Effect of sensor noise on the resolution and spatial resolution of displacement and strain maps estimated with the grid method. *Strain*, 50(1):1–27, 2014.

[13] F. Sur and M. Grédiac. Towards deconvolution to enhance the grid method for in-plane strain measurement. *AIMS Inverse Problems and Imaging*, 8(1):259–291, 2014.

[15] M. Grédiac, F. Sur, C. Badulescu, and J.-D. Mathias. Using deconvolution to improve the metrological performance of the grid method. *Optics and Lasers in Engineering*, 51(6):716–734, 2013.

Enfin, la conclusion générale et des perspectives constituant mes principales directions de recherche futures sont discutées dans le **chapitre 7**.

1.2.3 Encadrement scientifique

Les thèses et stages de Master recherche, tous co-encadrés avec Marie-Odile Berger, sont les suivants :

1. Depuis octobre 2013 : thèse de Pierre Rolin. *Synthèse de vues pour l'initialisation de pose*. Soutenance prévue début 2017.
2. Décembre 2008 à janvier 2013 : thèse de Srikrishna Bhat. *Vocabulaires visuels pour les problèmes d'estimation de pose*. Thèse soutenue le 22 janvier 2013. S. Bhat effectuée depuis un séjour postdoctoral chez J. Heikkilä (Université d'Oulu, Finlande).
3. Septembre 2006 à octobre 2011 : thèse de Nicolas Noury. *Mise en correspondance a-contrario de points d'intérêt sous contraintes géométrique et photométrique*. Thèse soutenue le 13 octobre 2011. N. Noury a été recruté chez Archos après sa soutenance.
4. Février à juillet 2013 : stage de Master recherche de P. Rolin. *Calcul de pose par simulation de points de vue*.
5. Mars à juillet 2006 : stage de Master recherche de N. Noury. *Élaboration de techniques fiables et dynamiques de mise en correspondance pour l'identification des structures et du mouvement dans une séquence vidéo*.

J'ai également pu co-encadrer des stages longs d'élèves-ingénieurs à temps plein au LORIA dans le cadre de leur scolarité :

1. Juin à août 2013 : N. Oishi (École des Mines de Nancy, trois mois entre la deuxième et troisième année). *Utilisation de méthodes d'égalisation d'histogrammes pour la mise en correspondance et le calcul de pose dans le cadre d'applications de réalité augmentée.*
2. Juin-juillet 2013 : G. Vandeneeckhoutte (ENSEIRB, six semaines entre la première et deuxième année). *Utilisation de descripteurs colorimétriques pour la mise en correspondance dans le cadre d'applications de réalité augmentée.*
3. Juillet à septembre 2012 : C.-E. Mabrier (École des Mines de Nancy, trois mois entre la deuxième et troisième année). *Construction de modèles texturés et calcul de pose en conditions d'appariement difficiles.*

Enfin, je propose régulièrement dans le cadre de mon service d'enseignement des projets scientifiques en deuxième ou troisième année à l'École des Mines de Nancy. Les sujets encadrés ces dernières années ont pour intitulés : *Minimisation de la variation totale sous contrainte et applications en traitement d'images et en finance, Optimisation convexe par algorithmes proximaux et application au débruitage par décomposition parcimonieuse, Moyennes non-locales pour le débruitage d'images, Déconvolution aveugle pour la restauration des images, Théorème d'échantillonnage de Shannon et étude du phénomène d'aliasing, Analyse par ondelettes et application aux séries chronologiques financières, Reconnaissance d'objets par une machine à vecteurs support multiclassées, Étude statistique des descripteurs SIFT, Analyse de données et analyse en composantes indépendantes, Analyse de données et classification non supervisée, Modèles ARCH/GARCH et séries financières, Modèles Vectoriels Auto-Régressifs (VAR).*

1.2.4 Réalisations logicielles

Lorsque la contribution s'y prête, je m'efforce de diffuser les logiciels implantant les algorithmes décrits dans les articles et permettant de reproduire les expériences numériques les illustrant.

1.2.4.1 Bibliothèques logicielles

1. The gridmethod toolbox : Analyse des images de grilles en mécanique des solides expérimentale (avec M. Grédiac et B. Blaysat). Bibliothèque Matlab implantant les algorithmes discutés en particulier dans :
 [4] M. Grédiac, F. Sur, B. Blaysat. The grid method for in-plane displacement and strain measurement : a review and analysis. *Strain*, 52(3):205–243, 2016.
 Pour une meilleure visibilité, la bibliothèque est disponible sur la plate-forme :
<http://www.thegridmethod.net>
 L'objectif est de faciliter l'utilisation de la méthode de la grille par la mise à disposition de la communauté des chercheurs en mécanique des solides d'un ensemble de fonctions Matlab efficaces et bien documentées.
2. Curve/matching dans **MegaWave2** : Appariement a-contrario de courbes géométriques normalisées dans les images (avec J.L. Lisani et P. Musé).
 Bibliothèque de codes en langage C (une dizaine de fonctions) intégrée dans le logiciel libre MegaWave2 implantant essentiellement les algorithmes décrits dans :
 [17] P. Musé, F. Sur, F. Cao, Y. Gousseau, J.-M. Morel. An a contrario decision method for

shape element recognition. *International Journal of Computer Vision*, vol. 69, no. 3, p. 295–315, 2006.

Megawave2 était un logiciel utilisé dans plusieurs universités, pour la recherche comme pour l’enseignement en traitement et analyse d’images. Page web :

<http://megawave.cmla.ens-cachan.fr/>

1.2.4.2 Codes Matlab autonomes

Ils sont disponibles sur ma page web à l’adresse suivante :

<https://members.loria.fr/FSur/software/>

1. VerifWV : Verification of the Wigner-Ville transform of rectangular, triangular, and Gaussian functions.

Code Matlab permettant la vérification numérique des formules établies dans :

[3] F. Sur, M. Grédiac. Influence of the analysis window on the metrological performance of the grid method. *Journal of Mathematical Imaging and Vision*, 56(3):472–498, 2016.

À notre connaissance, cet article présente le premier calcul (quelque peu technique et fastidieux) de la transformée de Wigner-Ville des fonctions triangles. Il est donc important d’en proposer une vérification numérique.

2. NESIF : Noise Estimation by Stacking Images affected by illumination Flickering.

Code Matlab implantant l’algorithme décrit dans :

[9] F. Sur, M. Grédiac. Measuring the noise of digital imaging sensors by stacking raw images affected by vibrations and illumination flickering. *SIAM Journal on Imaging Sciences*, vol. 8, no. 1, p. 611–643. 2015.

Ce code permet de reproduire les expériences numériques de notre article et facilite la réutilisation de l’algorithme d’estimation des paramètres d’un capteur numérique à partir d’une série d’images affectées par des vibrations et un scintillement de l’illumination.

3. ARPENOS : Automated Removal of quasi-PERiodic NOise using frequency domain Statistics.

Code Matlab implantant l’algorithme décrit dans :

[8] F. Sur, M. Grédiac. Automated removal of quasi-periodic noise using frequency domain statistics. *IS&T/SPIE Journal of Electronic Imaging*, vol. 24, no. 1, p. 013003/1–19, 2015.

Ce code permet de reproduire les expériences de l’article ainsi que des expériences additionnelles.

4. AC-ARPENOS : A-Contrario Automated Removal of quasi-PERiodic NOise using frequency domain Statistics.

Code Matlab implantant l’algorithme décrit dans :

[23] F. Sur. An a-contrario approach to quasi-periodic noise removal. *Proceedings of the IEEE International Conference on Image Processing (ICIP)*, p. 3841–3845, 2015.

Ce code permet de reproduire les expériences de l’article (et d’autres). La page web dédiée propose des images réelles affectées de bruit quasi-périodique. AC-ARPENOS comme ARPENOS implantent l’interpolation des coefficients de Fourier par minimisation contrainte de la variation totale.

5. Codes relatifs à mes enseignements : pour permettre aux étudiants de faire des expériences numériques sur des algorithmes récents de traitement des images, j'ai programmé plusieurs algorithmes classiques de la littérature, comme la génération de microtextures à phase aléatoire, le débruitage TV-L2, l'estimation du bruit par la méthode de Ponomarenko et al., un simulateur de compression JPEG, ou l'édition d'images par équation de Poisson.

Les codes Matlab correspondants sont disponibles ici :

<https://members.loria.fr/FSur/enseignement/photo/>

1.3 Guide de lecture

Les trois parties de ce mémoire sont largement indépendantes. Rappelons que la première partie traite de l'appariement de points d'intérêt entre deux images ou entre une image et un modèle de scène non structuré, avec comme objectif l'estimation de la pose d'une caméra. La deuxième partie résume les travaux relatifs au bruit quasi-périodique et au bruit causé par le capteur numérique. La troisième partie dresse un panorama de contributions à l'analyse d'images issues d'applications en mécanique des solides expérimentale, et en particulier à la quantification des performances métrologiques de techniques de mesure de champs de déformation et de déplacement.

De manière à faciliter la lecture, tous les chapitres sont organisés de la même manière : une introduction présentant de manière synthétique le contexte et les principales contributions, suivie d'un résumé détaillé de ces contributions, pour terminer par des perspectives de recherche ouvertes.

Les notations employées dans le mémoire peuvent différer de celles utilisées dans les articles originaux de manière à homogénéiser la présentation.

Première partie

Contributions à l'appariement de points
d'intérêt entre deux images ou entre une
image et un modèle de scène non-structuré

Appariement de points d'intérêt entre images

2.1	Introduction	21
2.2	Appariement de points d'intérêt sous contraintes photométriques et géométriques . . .	24
2.3	Comment rendre ASIFT davantage robuste à l' <i>aliasing</i> perceptuel ?	30
2.4	Prise en compte de l'incertitude de localisation des points d'intérêt	35
2.5	Conclusion et perspectives	39

2.1 Introduction

2.1.1 Contexte

La recherche de correspondances entre points d'intérêt appartenant à deux vues (ou davantage) de la même scène est bien souvent la première étape de nombreuses applications de la vision par ordinateur. Cette opération est généralement basée sur l'appariement de descripteurs photométriques locaux, de la manière suivante :

1. dans chaque vue, extraction des points d'intérêt et des descripteurs associés ;
2. appariement à l'aide d'une mesure de (dis-)similarité entre descripteurs ;
3. sélection de l'ensemble de correspondances le plus cohérent avec une contrainte géométrique imposée par un mouvement réaliste de caméra.

L'extraction des points d'intérêt et des descripteurs photométriques locaux dans l'étape 1 a été le sujet de nombreux travaux (voir par exemple les revues [ADP12, MS04a, MTS⁺06, MP07]). Les points d'intérêt peuvent être détectés comme extrema de différentes fonctions indicatrices dans une description espace-échelle (*scale space*) de l'image : par exemple *corner-ness* de Harris multi-échelle [MS04a], déterminant de la Hessienne [BETG08], approximation du Laplacien [Low04]. Les descripteurs photométriques rassemblent en général des statistiques sur la distribution de la direction du gradient dans une région autour du point d'intérêt considéré. La direction du gradient présente en effet l'avantage d'être invariante aux changements de contraste et est donc robuste aux changements d'illumination de la scène. Les régions à partir desquelles les descripteurs sont construits doivent être covariantes avec une transformation géométrique (et donc les descripteurs eux même doivent être invariants à ces transformations), de manière à ce que l'appariement soit, dans la mesure du possible, robuste à des changements

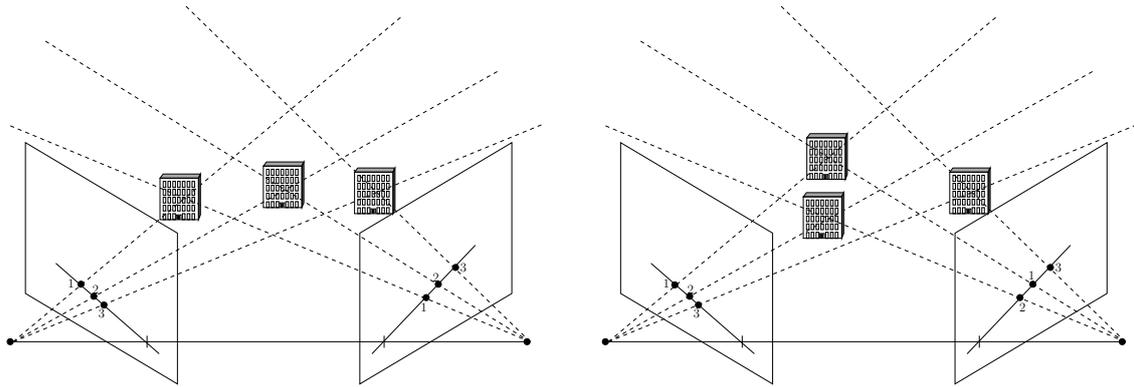


FIGURE 2.1 – Illusion des « clous multiples ». On considère deux vues d'une scène 3D présentant des motifs répétés. Même avec des caméras calibrées (la contrainte épipolaire est donc connue), il est impossible de décider de certains appariements en utilisant seulement l'information contenue dans des descripteurs invariants. Ici, les six points noirs ont des descripteurs identiques (car ils correspondent à des motifs répétés : il s'agit d'une situation d'aliasing perceptuel [WB91]), et puisqu'ils sont sur le même plan épipolaire, les trois points de l'image de gauche peuvent être associés aux trois points de l'image de droite dans chacune des six possibilités d'appariement (seules deux sont montrées ici). Bien sûr, seulement l'une des possibilités a un sens physique. Néanmoins, elle est indiscernable des cinq autres en utilisant seulement l'information disponible.

de point de vue de la caméra. La plupart du temps, les descripteurs sont invariants à des transformations affines ou à des similitudes comme dans l'algorithme SIFT de D. Lowe [Low04]. Dans l'étape 2, un point d'intérêt x de la première image est classiquement apparié à un point d'intérêt y dans la seconde image si le descripteur de y est le plus proche voisin du descripteur de x parmi tous les descripteurs extraits dans la seconde image. Pour éliminer les cas ambigus, un seuil supérieur peut être imposé sur le rapport entre la distance au plus proche voisin et au second plus proche voisin [Low04]. À ce stade, aucune contrainte géométrique n'est imposée et un certain nombre de fausses correspondances peuvent être constatées à l'issue de cette étape. L'étape 3 est souvent effectuée par l'intermédiaire de RANSAC [FB81] ou l'une des très nombreuses variantes [CMK03, CM05, MS04b, TM05, TZ00, ZK06]. L'objectif est de retenir un sous-ensemble satisfaisant une contrainte géométrique parmi les appariements issus de l'étape 2. Cette contrainte est par exemple une contrainte épipolaire dans le modèle de caméra sténopé, ou bien une contrainte homographique si on cherche à apparier des points sur un plan [HZ00]. La contrainte géométrique s'exprime à l'aide d'une transformation paramétrique qui lie la position d'un point d'intérêt et celle du point qui lui est apparié. L'idée de RANSAC est de générer aléatoirement ces transformations paramétriques à partir d'un ensemble de points appariés jusqu'à obtenir un sous-ensemble « assez gros ». Tout le problème dans RANSAC est de fixer le seuil au dessous duquel la contrainte géométrique est satisfaite.

Dans ce chapitre, nous étudions un certain nombre d'images présentant des motifs répétés, comme des fenêtres sur une façade ou une texture répétitive. Dans une telle situation, de nombreuses fausses correspondances peuvent être obtenues entre descripteurs semblables dont les points d'intérêt tombent « par hasard » sur les mêmes lignes épipolaires. Les psychologues de la perception connaissent cette situation sous le nom d'illusion des deux clous (*double-nail illusion*) depuis un article de Krol et van de Grind dans les années 1980 [KvdG80]. L'application de l'illusion au cas de l'appariement entre points de deux vues est illustrée par

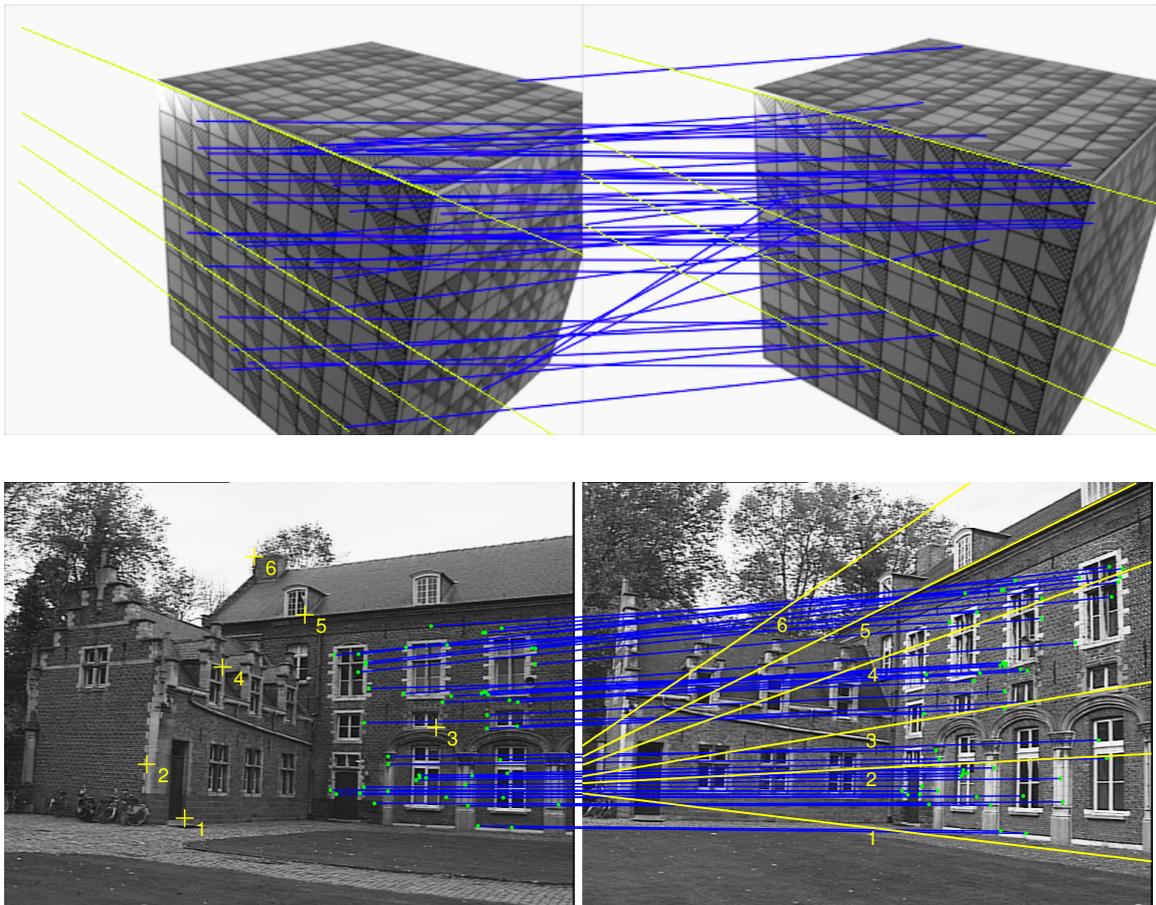


FIGURE 2.2 – Les correspondances satisfont ici à la fois la similarité des descripteurs et la contrainte épipolaire. Néanmoins, de nombreuses fausses correspondances peuvent être remarquées (par exemple sur la paire du bas, des correspondances peuvent être vues avec la quatrième fenêtre de l'image de droite qui n'est tout simplement pas visible dans la première image), et les droites épipolaires (en jaune) correspondent aux lignes de fuite. Ceci fournit bien sûr une estimation de la pose de la caméra incorrecte.

la figure 2.1. Dans certaines situations, cette illusion peut même faire échouer l'algorithme d'appariement et piéger RANSAC dans une fausse géométrie épipolaire, dont les droites épipolaires seraient en fait des lignes de fuite. Une illustration de ce phénomène est présentée dans la figure 2.2. Dans ces exemples, tout ensemble de points sur des droites 3D parallèles satisfait une contrainte épipolaire dégénérée dans laquelle les épipoles sont les points de fuite, quelque soit la manière dont les points sont appariés sur la ligne épipolaire. Si les points sur cette droite épipolaire ont des descripteurs semblables (ce qui peut être le cas en présence de motifs répétés), alors ces faux appariements peuvent générer le plus grand ensemble de consensus dans RANSAC et l'emporter sur des ensembles de correspondances plus petits mais corrects.

2.1.2 Contributions

Nos contributions à l'appariement de points d'intérêt portent sur plusieurs aspects. L'algorithme basé sur l'appariement au plus proche voisin suivi de RANSAC fait que les appariements au-delà du plus proche voisin ne sont jamais considérés. Comme on l'a vu, il s'agit d'un problème particulièrement ennuyeux en présence de motifs répétés indistinguables au niveau des descripteurs, cas dans lequel on se retrouve dans une situation d'*aliasing* perceptuel [WB91]. Ce phénomène peut rendre l'ensemble de correspondances cohérent avec une géométrie qui n'est pas celle induite par le véritable déplacement de la caméra. De plus, cet algorithme nécessite de fixer plusieurs seuils d'appariement. Dans [14, 36], nous avons proposé un algorithme a-contrario basé sur un critère de cohérence photométrique et géométrique permettant de regrouper les étapes 2 et 3 évoquées au début de l'introduction de ce chapitre. Cet algorithme est décrit dans la section 2.2. Les descripteurs utilisés ne sont pas invariants à tout déplacement de caméra, en particulier dès que le changement de point de vue dépasse un certain angle (30° d'après [MP07]). L'algorithme ASIFT [MY09, YM11], basé sur le principe de la simulation de point de vue, permet d'apparier des images en présence de changements de point de vue extrêmes. Nous présentons dans la section 2.3 les résultats de [32] : il s'agit d'une amélioration de ASIFT utilisant notre algorithme a-contrario, de manière à rendre ASIFT davantage robuste aux motifs répétés. Enfin, la section 2.4 discute brièvement une adaptation de notre modèle a-contrario de manière à tenir compte de l'incertitude de localisation des points d'intérêt [33, 35].

Les principales publications relatives à ce chapitre sont :

[14] F. Sur, N. Noury, and M.-O. Berger. An a contrario model for matching interest points under geometric and photometric constraints. *SIAM Journal on Imaging Sciences*, 6(4):1956–1978, 2013.

[32] N. Noury, F. Sur, and M.-O. Berger. How to overcome perceptual aliasing in ASIFT? In *Proceedings of the International Symposium on Visual Computing (ISVC), part I*, volume LNCS 6453, pages 231–242, Las Vegas (NV), USA, 2010.

[33] F. Sur. Robust matching in an uncertain world. In *Proceedings of the International Conference on Pattern Recognition (ICPR)*, pages 2350–2353, Istanbul, Turkey, 2010.

[35] F. Sur, N. Noury, and M.-O. Berger. Computing the uncertainty of the 8 point algorithm for fundamental matrix estimation. In *Proceedings of the British Machine Vision Conference (BMVC)*, volume 2, pages 965–974, Leeds, UK, 2008.

[36] N. Noury, F. Sur, and M.-O. Berger. Fundamental matrix estimation without prior match. In *Proceedings of the IEEE International Conference on Image Processing (ICIP)*, volume 1, pages 513–516, San Antonio (TX), USA, 2007.

Bon nombre des résultats présentés dans ce chapitre ont été obtenus par Nicolas Noury (thèse soutenue en octobre 2011 [Nou11]), doctorant que j'ai co-encadré avec Marie-Odile Berger.

2.2 Appariement de points d'intérêt sous contraintes photométriques et géométriques

Une possibilité pour éviter les problèmes d'appariements en présence de motifs répétés pourrait être d'augmenter la taille des régions d'intérêt [VS06] ou de vérifier séquentiellement les correspondances par cosegmentation [CMP10]. Incorporer l'information d'échelle de

la région dans l'appariement comme dans [TB10] permet aussi d'améliorer la distinguabilité des descripteurs. Rendre les correspondances moins ambiguës est aussi possible en prenant en compte la position relatives des points d'intérêt dans chaque image [DMSD06, AFE⁺09]. Néanmoins, lorsque trop de fausses correspondances sont présentes parmi les plus proches voisins, il peut être intéressant de permettre d'obtenir des correspondances au delà du plus proche voisin. Il s'agit de l'approche choisie par les auteurs de l'algorithme « RANSAC généralisé » [ZK06] qui utilise des informations géométriques et photométriques pour guider la recherche d'homographies.

La contribution proposée dans cet esprit est un modèle a-contrario définissant une métrique, le nombre de fausses alarmes, intégrant simultanément similarité photométrique et contrainte géométrique. L'algorithme d'appariement que nous déduisons permet d'associer des points d'intérêt au delà du plus proche voisin. Ce modèle a-contrario a été d'abord suggéré dans un article de L. Moisan et B. Stival [MS04b] (sous le terme *colored rigidity*), nous le complétons en nous basant sur le modèle a-contrario décrit par J. Rabin, J. Delon et Y. Gousseau dans [RDG08b, RDG08a, RDG09], qui permet d'apparier des descripteurs sans contrainte géométrique. Le modèle a-contrario de [MS04b] ne traite que le cas de la contrainte épipolaire; nous l'étendons aussi au cas des homographies. Remarquons que les auteurs de MLESAC guidé [TM05] évoquent eux aussi la possibilité d'incorporer une information a priori de nature photométrique (la corrélation croisée d'images dans leur cas) dans un modèle bayésien mesurant la probabilité d'un ensemble d'appariements.

2.2.1 Un modèle a-contrario pour les correspondances entre points d'intérêt sous contraintes photométriques et géométriques

Introduisons quelques notations. On suppose disposer de deux vues (images \mathcal{I}_1 et \mathcal{I}_2) de la même scène tridimensionnelle. Notons $(x_i, D(x_i))_{1 \leq i \leq N_1}$ (resp. $(y_j, D(y_j))_{1 \leq j \leq N_2}$) les N_1 (resp. N_2) couples formés des coordonnées des points d'intérêt et des descripteurs photométriques associés venant de \mathcal{I}_1 (resp. \mathcal{I}_2). Ils sont classiquement le résultat de l'algorithme SIFT [Low04].

En identifiant la caméra à un sténopé, si x_i et y_j sont les projections dans \mathcal{I}_1 et \mathcal{I}_2 du même point de la scène, alors y_j est situé sur la ligne épipolaire Fx_i associée à x_i , et x_i est situé sur la ligne épipolaire $F^T y_j$ associée à y_j , où F est la matrice fondamentale de \mathcal{I}_1 à \mathcal{I}_2 . Notons que si le mouvement de la caméra est une rotation autour de son centre optique ou si les points d'intérêt sont tous sur un plan commun, la matrice fondamentale n'est pas définie et il existe une transformation projective 2D (une homographie) H telle que $y_j = H(x_i)$ et $x_i = H^{-1}(y_j)$.

Le problème est donc de trouver un sous-ensemble \mathcal{S} de $\{1, \dots, N_1\} \times \{1, \dots, N_2\}$ et une matrice fondamentale F ou une homographie H de \mathcal{I}_1 à \mathcal{I}_2 de manière à ce que :

1. La distance entre des descripteurs en correspondance est en dessous d'un seuil δ_D

$$\forall (i, j) \in \mathcal{S}, d_D(D(x_i), D(y_j)) \leq \delta_D \quad (2.1)$$

2. La distance entre un point et la ligne épipolaire associée est en dessous d'un seuil δ_G , ce qui assure que la contrainte épipolaire soit satisfaite :

$$\forall (i, j) \in \mathcal{S}, d_G(x_i, y_j, F) := \max\{d_G(y_j, Fx_i), d_G(x_i, F^T y_j)\} \leq \delta_G \quad (2.2)$$

De manière alternative, pour la contrainte homographique :

$$\forall (i, j) \in \mathcal{S}, d_G(x_i, y_j, H) := \max\{d_G(y_j, H(x_i)), d_G(x_i, H^{-1}(y_j))\} \leq \delta_G \quad (2.3)$$

Comme on va le voir, le modèle statistique a-contrario permet de sélectionner automatiquement des seuils δ_D et δ_G .

2.2.2 Modèle a-contrario

La théorie a-contrario de la détection statistique a été introduite pour un problème de détection d'alignements dans des images [DMM00]. Elle s'est avérée adaptée à de nombreuses tâches de l'analyse d'images comme le groupement de Gestalts [DMM08, LvGRM14], la détection d'objets en mouvement dans des vidéos [DKP09], la détection de changement en imagerie satellitaire [RMLHM10] ou en imagerie par résonance magnétique [RFH⁺07], l'identification de formes géométriques [1], de segments [vGJMR10] ou d'arcs elliptiques [PGvG12], l'identification de regroupements de points [TMAM11], la détection d'objets multiples [RDGM10], la détection de « taches » dans des textures [GM09], ou la détection de structures convergentes pathologiques dans des tomosynthèse mammaires [PBM14], pour citer quelques articles très récents. Cette théorie de la détection est basée sur l'idée selon laquelle les caractéristiques intéressantes (appelé événements *significatifs*) ne peuvent être observé dans un *modèle de fond* (sous une hypothèse \mathcal{H}_0) qu'avec une probabilité très faible. Contrairement à la théorie classique de la décision statistique basée sur les p -valeurs, décider si un événement est significatif ou pas est basé sur le *nombre de fausses alarmes* (NFA) qui correspond à l'espérance d'occurrence d'un tel événement dans le modèle de fond, d'où l'expression « fausse alarme ». De manière plus précise, dans le cas d'une statistique réelle X qui n'est pas susceptible de prendre une petite valeur x sous \mathcal{H}_0 , et supposant que l'on cherche à détecter parmi N observations, alors le NFA de x est :

$$\text{NFA}(x) = N \Pr(X \leq x) \quad (2.4)$$

où $\Pr(X \leq x)$ est la valeur de la fonction de répartition de X en x sous l'hypothèse \mathcal{H}_0 . Dans les modèles a-contrario, la statistique X est en fait une fonction de variables aléatoires indépendantes (notons que des articles récents explorent aussi la non-indépendance [Des16, MGL13, FH11]), $\Pr(X \leq x)$ est donc calculé à partir des lois marginales, qui sont soit paramétriques (comme dans [DMM08]) soit estimées empiriquement (comme dans [1]). Une fois que le NFA a été défini, les caractéristiques significatives sont généralement celles pour lesquelles $\text{NFA} \leq 1$, ce qui signifie qu'au plus une telle observation est attendue sous \mathcal{H}_0 parmi les N testées. En ce sens, la détection a-contrario n'a pas de paramètre. Des expériences psychophysiques suggèrent que la détection a-contrario est en adéquation avec la perception humaine [BCM⁺14, DMM03]. Notons enfin que la détection par seuillage du NFA généralise la correction de Bonferroni [Bon36] sur les seuils de p -valeurs utilisés classiquement en statistique. Dans la correction de Bonferroni, le risque global est divisé par le nombre de tests effectués (ou de manière similaire à (2.10), la probabilité de la statistique de test est multipliée par le nombre de tests); dans la détection a-contrario consistant à seuiller le NFA, le nombre de tests correspond au nombre d'observations N et peut dépendre implicitement ou explicitement de la dimensionnalité de X .

Dans le cas qui nous intéresse, nous supposons disposer d'un ensemble \mathcal{S} de correspondances, ainsi que d'une transformation A (matrice fondamentale ou homographie) et de deux seuils δ_D et δ_G comme dans (2.1) et (2.2)-(2.3). La probabilité à estimer est :

$$p(\mathcal{S}, A, \delta_G, \delta_D) := \Pr\left(\forall (i, j) \in \mathcal{S}, d_G(x_i, y_j, A) \leq \delta_G \text{ et } d_D(D(x_i), D(y_j)) \leq \delta_D \mid \mathcal{H}_0\right) \quad (2.5)$$

Supposons également que la transformation A est estimée à partir d'un sous-ensemble minimal s de \mathcal{S} de cardinal m comme dans RANSAC ($m = 7$ si $A = F$ et $m = 5$ si $A = H$).

Définition 2.1 *Considérant $(x_i, D(x_i))$ et $(y_j, D(y_j))$ comme des variables aléatoires, l'hypothèse \mathcal{H}_0 est définie comme :*

1. $(d_D(D(x_i), D(y_j)))_{(i,j) \in \mathcal{S}}$, et $(d_G(x_i, y_j, A))_{(i,j) \in \mathcal{S} \setminus s}$ sont des variables aléatoires mutuellement indépendantes.
2. $(d_G(x_i, y_j, A))_{(i,j) \in \mathcal{S} \setminus s}$ sont identiquement distribués et leur fonction de répartition est notée f_G .
3. $(d_D(D(x_i), D(y_j)))_{(i,j) \in \mathcal{S}}$ sont identiquement distribués et leur fonction de répartition est notée f_D .

Comme la loi jointe de variables aléatoires indépendantes est leur loi produit, l'équation (2.5) s'exprime de la manière suivante.

Proposition 2.1

$$p(\mathcal{S}, A, \delta_G, \delta_D) = f_D(\delta_D)^k f_G(\delta_G)^{k-m} \quad (2.6)$$

où k est le cardinal de \mathcal{S} , et m celui de s .

Dans un test d'hypothèse classique, la détection des groupes significatifs se ferait dès que cette probabilité est en dessous d'un seuil de significativité (typiquement 5%), ce qui aurait pour effet de favoriser les grands groupes. En effet, toutes choses égales par ailleurs, la probabilité devient très faible lorsque k grandit. Dans le cadre d'un modèle a-contrario, on définit le NFA en multipliant la probabilité de fausse alarme $p(\mathcal{S}, A, \delta_G, \delta_D)$ par le nombre d'événements possibles. Ici, il y a $\min\{N_1, N_2\} - m$ choix pour $k \geq m$, $\binom{N_1}{k}$ choix pour les points d'intérêt dans l'image 1, $\binom{N_2}{k}$ choix pour les points d'intérêt dans l'image 2, $k!$ choix pour les correspondances, $\binom{k}{m}$ choix pour les ensembles minimaux permettant d'estimer la transformation A . Chaque ensemble minimal s peut donner $Q = 3$ matrices fondamentales (avec l'algorithme des « sept points ») ou $Q = 1$ homographie.

Par conséquent (cf *colored rigidity* dans [MS04b]),

Définition 2.2 *Un ensemble \mathcal{S} de correspondances est ε -significatif s'il existe*

1. deux seuils δ_G et δ_D tels que :

$$\forall (i, j) \in \mathcal{S}, d_G(x_i, y_j, A) \leq \delta_G, \quad (2.7)$$

$$\forall (i, j) \in \mathcal{S}, d_D(D(x_i), D(y_j)) \leq \delta_D, \quad (2.8)$$

2. une transformation A évaluée à partir de m correspondances appartenant à \mathcal{S} , tels que :

$$\text{NFA}(\mathcal{S}, A, \delta_G, \delta_D) := Q (\min\{N_1, N_2\} - m) k! \binom{N_1}{k} \binom{N_2}{k} \binom{k}{m} f_D(\delta_D)^k f_G(\delta_G)^{k-m} \leq \varepsilon \quad (2.9)$$

où k est le cardinal de \mathcal{S} , $m = 4$ et $Q = 1$ si A est une homographie, $m = 7$ et $Q = 3$ si A est une matrice fondamentale.

Comme f_D et f_G sont croissantes, on peut obtenir la proposition suivante.

Proposition 2.2 *Un ensemble \mathcal{S} de correspondances est ε -significative s'il existe une transformation A estimée à partir de m correspondances appartenant à \mathcal{S} tel que :*

$$NFA(\mathcal{S}, A) := Q(\min\{N_1, N_2\} - m) k! \binom{N_1}{k} \binom{N_2}{k} \binom{k}{m} f_D(\delta_D)^k f_G(\delta_G)^{k-m} \leq \varepsilon \quad (2.10)$$

où k est le cardinal de \mathcal{S} ,

$$\delta_G = \max_{(i,j) \in \mathcal{S}} \max\{d_G(y_j, Fx_i), d_G(x_i, F^T y_j)\}, \quad (2.11)$$

et

$$\delta_D = \max_{(i,j) \in \mathcal{S}} (d_D(D(x_i), D(y_j))) \quad (2.12)$$

Les sections 2.2.3 et 2.2.4 précisent le choix des fonctions de répartition f_G et f_D et des seuils δ_G et δ_D .

2.2.3 Modélisation de la contrainte géométrique

Dans le cas $A = F$, L. Moisan et B. Stival [MS04b] proposent de définir $d_G(y, Fx)$ comme la distance euclidienne entre y et la droite épipolaire Fx . Ils montrent, sous hypothèse de répartition uniforme dans l'image, que la probabilité qu'un point aléatoire (sous hypothèse de distribution uniforme) y soit à une distance inférieure à δ_G de la droite Fx est majorée par $2D\delta_G/S$ où D et S sont respectivement le diamètre et l'aire de chaque image. Sous hypothèse d'indépendance dans la répartition des points entre les deux images, l'équation (2.2) devient

$$\Pr(d_G(x, y, F) \leq \delta_G) = \left(\frac{2D}{S}\delta_G\right)^2 \quad (2.13)$$

Nous décidons de paramétrer f_G par un paramètre α :

$$f_G(\delta_G) = \left(\frac{2D}{S}\delta_G\right)^{2\alpha} \quad (2.14)$$

L'influence du paramètre α est discutée dans [14].

Dans le cas $A = H$ case, on adapte la définition de $f_G(\delta_G)$ comme dans [BLMT09, RDGM10] en :

$$f_G(\delta_G) = \left(\frac{\pi\delta_G^2}{S}\right)^{2\alpha} \quad (2.15)$$

En effet, $\pi\delta_G^2/S$ est la probabilité pour un point aléatoire dans une image d'aire S de tomber à une distance inférieure à δ_G d'un point fixé.

2.2.4 Modélisation de la contrainte photométrique

Comme l'espace des descripteurs n'est ni isotrope ni homogène, toute « bonne » métrique sur les descripteurs doit tenir compte du voisinage de $D(x)$. Nous inspirant des travaux de J. Rabin, J. Delon, et Y. Gousseau [RDG09], et nous basant sur [17], nous définissons :

$$d_D(D(x), D(y)) = \phi_{D(x)}(\text{dist}(D(x), D(y))) \quad (2.16)$$

où dist est une distance sur l'espace des descripteurs, et $\phi_{D(x)}$ est la fonction de répartition de $\text{dist}(D(x), D(\cdot))$ lorsque $D(\cdot)$ parcourt l'ensemble des descripteurs de l'image \mathcal{I}_2 .

Rappelons que, sous certaines hypothèses raisonnables de régularité de $\phi_{D(x)}$, $d_D(D(x), D(y))$ est distribué uniformément sur l'intervalle $[0, 1]$. Cette distance s'adapte donc automatiquement à l'hétérogénéité de l'ensemble des descripteurs comme une mesure de dissimilarité contextuelle [JSHV10]. De plus, on peut simplifier l'expression de f_D par $f_D(t) = t$ pour tout $t \geq 0$.

Les descripteurs SIFT sont faits de $N = 16$ histogrammes D^i ($i \in \{1, \dots, N\}$), chacun d'entre eux étant composé de $m = 8$ baquets discrétisant l'intervalle circulaire $[0, 2\pi[$ d'orientation des gradients. La distance suivante entre descripteurs est donc introduite dans [RDG09] :

$$\text{dist}(D(x), D(y)) = \sum_{i=1}^N \text{CEMD}(D^i(x), D^i(y)) \quad (2.17)$$

où CEMD est la distance du terrassier circulaire (*circular earth mover's distance*), bien adaptée à la comparaison d'histogrammes circulaires.

Notons, pour tout $i \in [1, N]$, $\varphi_{D^i(x)}$ la fonction de répartition de $\text{CEMD}(D^i(x), D^i(\cdot))$. D'après la discussion dans [RDG09] et sous hypothèse d'indépendance, $\phi_{D(x)}$ est défini comme :

$$\phi_{D(x)}(\delta) = \int_0^\delta \bigotimes_{i=1}^N \varphi_{D^i(x)}(t) dt \quad (2.18)$$

où \otimes est le produit de convolution. En effet, $\text{CEMD}(D(x), D(y))$ est la somme de N variables aléatoires dont les distributions de probabilité sont le produit de convolution des N distributions marginales. En pratique, la distribution $\varphi_{D^i(x)}$ est estimée empiriquement à partir de l'ensemble des descripteurs $D^i(y)$ où y parcourt l'ensemble des points d'intérêt de l'image \mathcal{I}_2 .

2.2.5 Algorithme

L'algorithme permettant de détecter les groupes d'appariements les plus significatifs par échantillonnage aléatoire est décrit sur la figure 2.1. L'heuristique \mathcal{H} consiste à classer les correspondances provisoires selon le produit des deux probabilités apparaissant dans (2.10). Il s'agit de la même approche que celle proposée dans [MS04b] pour la recherche de groupes d'appariements significatifs sous contrainte géométrique uniquement. Néanmoins, alors que dans [MS04b] cela assure de trouver le groupe le plus significatif à F ou H fixé, ce n'est pas le cas dans notre algorithme du fait de la compétition entre contraintes géométrique et photométrique. Les paramètres et choix d'implantation sont discutés dans [14].

2.2.6 Expérience illustrative

Les résultats et expériences de [14] montrent que chercher les correspondances au-delà du plus proche voisin, tout en adaptant automatiquement les seuils d'appariement, permet de trouver des ensembles de correspondances corrects dans des cas où l'algorithme basé sur le plus proche voisin échoue.

La figure 2.3 illustre le comportement de l'algorithme dans le cas d'une contrainte homographique, dans le cas d'une scène présentant principalement une texture répétitive. Comme on peut le voir, l'algorithme basé sur l'appariement au plus proche voisin suivi d'un seuil sur le rapport entre distance au plus proche et distance au second plus proche ne fournit tout

Algorithme 2.1 Détection a-contrario de groupes d'appariements significatifs.

Entrée : deux vues de la même scène 3D.

1. À l'aide de l'algorithme SIFT, extraire les points d'intérêt et les descripteurs dans chaque vue : $(x_i, D(x_i))_{i \in \{1, \dots, N_1\}}$ et $(y_j, D(y_j))_{j \in \{1, \dots, N_2\}}$.
2. Pour chaque $i \in \{1, \dots, N_1\}$,
 - (a) construire les distances empiriques d_D (cf. section 2.2.4),
 - (b) définir un ensemble \mathcal{E}_p de correspondances provisoires (appariement au sens de [RDG09]).
3. Itérer :
 - (a) choisir aléatoirement sept (*resp. quatre*) points x_i et sélectionner sept (*resp. quatre*) points $y_{j(i)}$ en correspondance,
 - (b) calculer les trois matrices fondamentales F possibles à partir de ces sept correspondances et aller en (c) pour chacune des matrices (*resp. calculer l'homographie H à partir de ces quatre correspondances et aller en (c)*),
 - (c) sélectionner le sous-ensemble de \mathcal{E}_p le plus significatif (heuristique \mathcal{H}).

Sortie : le groupe le plus significatif rencontré.

simplement aucun appariement correct. L'algorithme RANSAC n'a donc aucune chance de fournir un groupe d'appariements cohérent avec une homographie. Au contraire, l'appariement a-contrario trouve des correspondances correctes : sur les 312 appariements retrouvés, 98 sont des plus proches voisins, et 67 sont de rang plus grand que 10. Cette expérience illustre l'intérêt d'aller au delà du plus proche voisin.

2.3 Comment rendre ASIFT davantage robuste à l'*aliasing* perceptuel ?

Comme expliqué dans l'introduction, la plupart des algorithmes d'appariement de points d'intérêt ne fonctionnent plus dès que le changement de point de vue sur la scène dépasse 30° [MP07]. La raison est l'invariance limitée de ces descripteurs à une gamme restreinte de transformations affines. L'algorithme ASIFT [MY09, YM11] de J.-M. Morel et G. Yu permet de dépasser cette limitation par des simulations de points de vue. Dans cette section, nous expliquons notre contribution à une amélioration de ASIFT rendant cet algorithme plus robuste à la présence de motifs répétés. Nous nous baserons essentiellement sur l'algorithme tel que décrit dans [MY09], qui diffère légèrement de la version disponible en ligne [YM11].

2.3.1 ASIFT

La décomposition en valeurs singulières d'une transformation affine A (de déterminant positif) s'écrit de la manière suivante [LF06, MY09].

$$A = \lambda R_\psi \begin{pmatrix} t & 0 \\ 0 & 1 \end{pmatrix} R_\phi \quad (2.19)$$

avec $\lambda > 0$, R_ψ et R_ϕ des matrices de rotation, $\phi \in [0, 180^\circ)$, et $t \geq 1$.

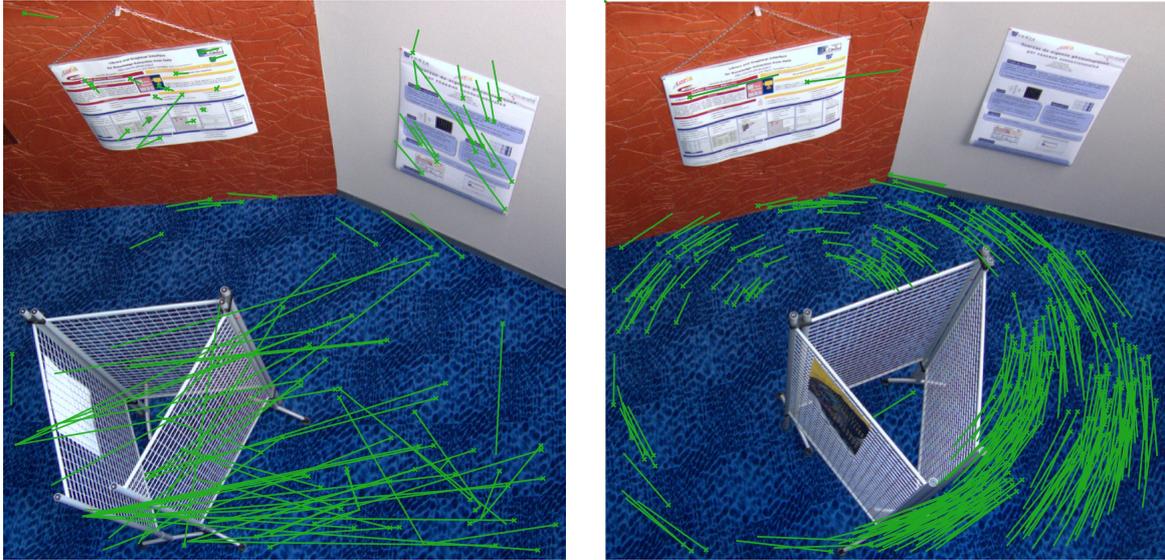


FIGURE 2.3 – Les deux images dans lesquelles on cherche des correspondances (sous contrainte homographique). On représente sur l'image de gauche (de droite, respectivement) le mouvement apparent des points d'intérêt entre leur localisation dans l'image de gauche (de droite, respectivement), marquée par une croix, et leur localisation dans l'image de droite (de gauche, respectivement), représentée par l'autre extrémité du segment. Il s'agit d'un problème difficile d'appariement, du fait de la texture répétitive. À gauche, correspondances établies par l'appariement classique au plus proche voisin des descripteurs. Très peu de bonnes correspondances sont identifiées. À droite, l'algorithme a-contrario proposé fournit un grand ensemble de bonnes correspondances. Dans ce cas, parmi les 312 appariements, 98 sont des plus proches voisins, et 67 sont de rang supérieure à 10.

Comme SIFT est invariant aux rotations et changements d'échelle induits par la transformation λR_ψ , une collection de descripteurs d'une image I invariants par transformation affine est obtenue dans l'algorithme ASIFT en extrayant des descripteurs SIFT des images simulées $I_{t,\phi}$ selon :

$$I_{t,\phi} = \begin{pmatrix} t & 0 \\ 0 & 1 \end{pmatrix} R_\phi(I) \quad (2.20)$$

En effet, la localisation des points d'intérêt SIFT est supposée covariante avec toute rotation et changement d'échelle λR_ψ appliquée à $I_{t,\phi}$, et les descripteurs associés y sont invariants.

D'après [MY09], il est suffisant de discrétiser t et ϕ comme suit : $t \in \{1, \sqrt{2}, 2, 2\sqrt{2}, 4\}$ et $\phi = \{0, b/t, \dots, kb/t\}$ avec $b = 72^\circ$ et $k = \lfloor t/b \cdot 180^\circ \rfloor$, $\lfloor x \rfloor$ désignant le plus grand entier inférieur au réel x .

L'étape suivante est d'apparier les descripteurs SIFT provenant des simulations sur les deux images I et I' . Ceci est réalisé par l'algorithme standard de [Low04] (plus proche voisin puis rapport des distances euclidiennes entre plus proche et second plus proche descripteur). Les correspondances obtenues sont ensuite reprojétées dans I et I' en commençant par les paires d'images simulées qui ont fourni le plus grand nombre de correspondances, à la condition qu'aucun point d'intérêt n'ait été préalablement placé à une distance plus petite que $\sqrt{3}$. Cette stratégie permet d'éviter des accumulations artificielles de points d'intérêt. Une dernière étape permet d'éliminer les fausses correspondances par RANSAC en imposant une contrainte épipolaire.

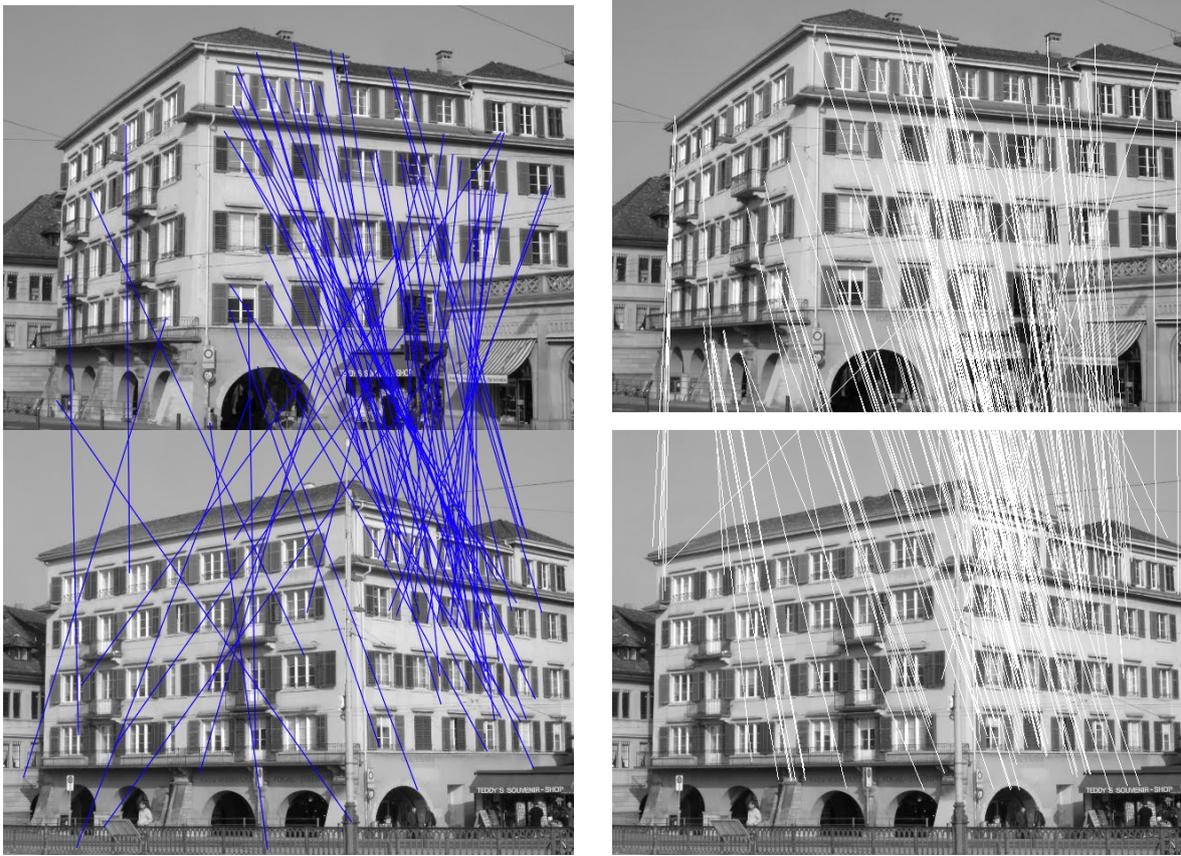


FIGURE 2.4 – Recherche de correspondances sous contrainte épipolaire. À gauche, recherche au plus proche voisin et contrainte épipolaire imposée par RANSAC. Un grand nombre de fausses correspondances peut être repéré; elles sont causées par l'illusion des doubles clous discutée précédemment. Il s'agit d'un problème d'autant plus difficile que l'aspect de la façade de gauche change fortement entre les deux prises de vue. À droite : ASIFT, qui permet de fournir davantage de correspondances que SIFT. Beaucoup de ces correspondances s'avèrent correctes, néanmoins, l'effet de l'illusion des doubles clous reste sensible. Ces images sont extraites de la Zürich Building Database [SSG03].

2.3.2 Modification de ASIFT

La figure 2.4 montre un exemple d'appariement de points d'intérêt en présence de motifs répétés et de fort changement de points de vue. On voit qu'ASIFT semble davantage robuste à la présence de motifs répétés que l'appariement SIFT classique. La raison est qu'en cherchant les paires d'images simulées fournissant le plus grand nombre d'appariements, on favorise les ensembles d'appariement entre des parties planaires de l'image quasiment recalées. Ce recalage implicite fournit davantage de bonnes correspondances. Nous proposons d'imposer explicitement ce recalage, en imposant une contrainte homographique entre appariements d'images simulées. D'autre part, le NFA introduit précédemment nous fournit un critère de *qualité* plus fin que le simple décompte du nombre de correspondances, et permet en outre de sélectionner des correspondances au-delà du plus proche voisin. ASIFT ainsi que les modifications que nous proposons sont décrits de manière succincte dans l'algorithme 2.2. Ces modifications sont discutées ci-dessous.

Algorithme 2.2 Improved ASIFT (**I-ASIFT**) et Standard ASIFT (*ASIFT*).**Entrée** : deux images I et I' .

1. Pour les deux images, générer les deux collections d'images $I_{t,\phi}$ et $I'_{t',\phi'}$ (eq. (2.20)), et extraire les descripteurs SIFT de chaque image générée.
2. Appariement des descripteurs SIFT entre chaque paire d'images générées :
 - I-ASIFT** - pour chaque paire $(I_{t,\phi}, I'_{t',\phi'})$ extraire le groupe de correspondances avec le plus faible NFA sous contrainte homographique (eq. (2.10)).
 - ASIFT* - pour chaque paire, appairer chaque descripteur de $I_{t,\phi}$ à son plus proche voisin dans $I'_{t',\phi'}$, à condition que le rapport entre la distance au plus proche et au second plus proche voisin est inférieur à 0,6.
3. Reprojection des points SIFT des $I_{t,\phi}$'s et $I'_{t',\phi'}$'s à I et I' :
 - I-ASIFT** - garder les groupes tels que $\log(\text{NFA}) < -50$, et les trier selon les NFA croissants. En partant du groupe le plus significatif, reprojeter une paire de descripteurs appariés seulement si le point d'intérêt ne tombe pas dans le voisinage d'un point déjà reprojété. Le voisinage est défini comme la reprojection dans I (resp. I') du cercle autour du descripteur extrait de l'image simulée dont le rayon est égale à l'échelle SIFT (au minimum 2 pixels).
 - ASIFT* - reprojeter les points appariés seulement s'il n'y a pas de point déjà placé à une distance inférieure à $\sqrt{3}$ pixels.
4. Élimination des fausses correspondances :
 - I-ASIFT** - utilisation de RANSAC *a contrario* [MS04b] pour imposer la contrainte épipolaire ou homographique, au choix de l'utilisateur.
 - ASIFT* - utilisation de RANSAC *a contrario* pour imposer la contrainte épipolaire.

Sortie : un ensemble de points en correspondance entre les images I et I' .

En premier lieu, nous remplaçons dans l'étape 2 l'appariement au plus proche voisin par notre algorithme d'appariement *a contrario* sous contrainte homographique. La raison pour utiliser une contrainte homographique est que lorsqu'on simule des transformations affine, on attend à ce que certaines d'entre elles approchent correctement des homographies relatives aux mêmes parties planaires de la scène. Ensuite, chaque groupe de correspondances entre images simulées est supposé appartenir à un morceau de plan, et donc doit satisfaire une contrainte homographique. Dans l'algorithme ASIFT standard tel que décrit dans [MY09], le nombre de groupes est limité a priori à cinq. Au contraire, nous gardons tous les groupes tels que $\log_{10}(\text{NFA}) < -50$. L'avantage du NFA est qu'il mesure la significativité du groupement : cela a du sens d'intégrer tous les groupes très significatifs, sans limite a priori sur leur nombre. En pratique, cela revient généralement à garder un nombre de groupes compris entre 5 pour une scène totalement plane et de l'ordre de 70 pour une scène multi-planaire. La limite du NFA à -50 peut sembler arbitraire, néanmoins nous avons constaté que des groupes avec un NFA plus grand était très souvent redondants avec des groupes possédant un NFA bien inférieur.

Pour améliorer la reprojection de l'étape 3, nous proposons d'utiliser le NFA comme critère d'adéquation. Comme le remarquent les auteurs de [HCH10], l'appariement par simulation de points de vue donne toujours un grand nombre de correspondances, bon nombre d'entre elles étant concentrées dans de petites zones. Le critère du NFA réalise un compromis entre la

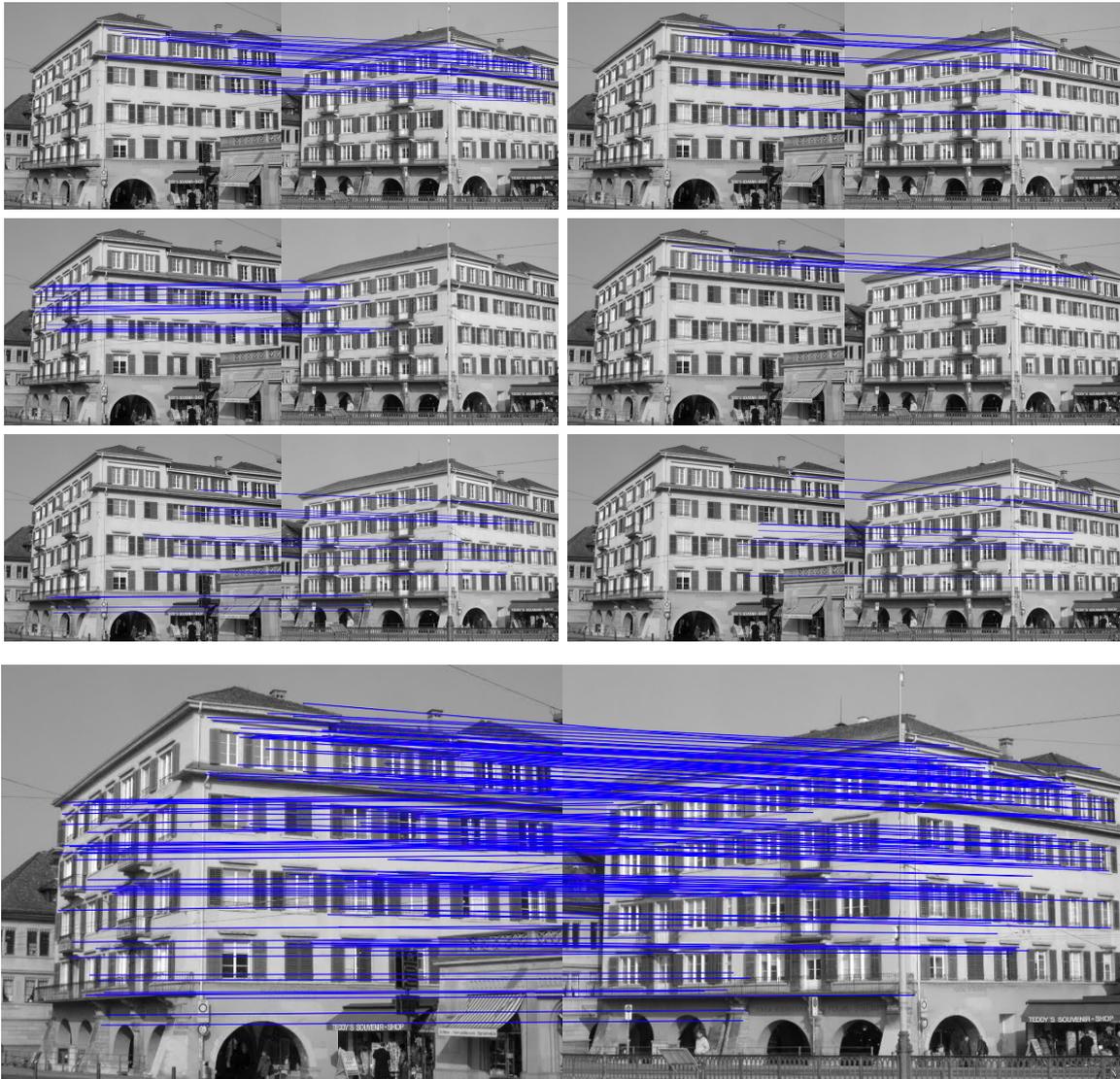


FIGURE 2.5 – Improved ASIFT : Zürich Building. I-ASIFT donne 151 correspondances, parmi lesquelles seulement 89 sont des plus proches voisins. Les six groupes avec les plus petits NFA sont montrés en haut : ils correspondent bien à des appariements entre morceaux localisés de plans.

taille du groupe et sa précision : les petits groupes peuvent avoir un NFA bien plus faible que des plus grands groupes si les descripteurs photométriques s'accordent quasi-parfaitement ou si la contrainte géométrique est très bien satisfaite. Il nous semble que favoriser les groupes avec un NFA le plus faible a plus de sens que favoriser systématiquement les grands groupes. Enfin, lors de la reprojection, on sélectionne les correspondances à partir de leur échelle afin d'éviter l'accumulation dans de petites zones.

L'application de cet algorithme à l'exemple de la figure 2.4 est illustrée par la figure 2.5. Les correspondances sont cette fois correctes.

2.3.3 Expériences illustratives

Les figures 2.6 et 2.7 montrent les résultats de notre algorithme sur un autre exemple illustratif. L'algorithme I-ASIFT permet d'obtenir 210 correspondances, toutes correctes. Par comparaison, SIFT permet d'obtenir 97 correspondances, un certain nombre d'entre elles étant incorrectes, et ASIFT 153 correspondances, avec les mêmes correspondances incorrectes. L'illusion des doubles clous explique les appariements incorrects de la figure 2.7. La contrainte homographique supplémentaire sur chaque groupe imposée dans I-ASIFT contribue à les éliminer.

D'autres expériences peuvent être trouvées dans [54].

La figure 2.8 montre deux vues distantes dans la séquence *Leuven Castle* [PVGV⁺04]. ASIFT fournit 94 appariements, principalement entre points de la façade principale. Remarquons que de nombreux appariements incorrects peuvent être observés. I-ASIFT fournit 118 appariements (24% seulement sont des plus proches voisins), distribués sur tout le bâtiment. Tous ces appariements (à l'exception de deux) sont corrects. La matrice fondamentale est estimée sur ces ensembles de correspondances, et les lignes épipolaires (en jaune) correspondant à quelques points manuellement sélectionnés dans l'image de gauche prouvent que I-ASIFT permet d'estimer correctement la géométrie épipolaire. Au contraire, le mouvement de la caméra ne peut pas être retrouvé à partir des appariements ASIFT. En comparaison, MSER [MCUP04] ne fournit que 5 appariements, et Harris/Hessian affine [MS04a] 20 à 30 appariements, principalement entre des motifs répétés improprement associés.

2.4 Prise en compte de l'incertitude de localisation des points d'intérêt

Les contributions de [33] et [35] concernent la prise en compte de l'incertitude de localisation des points d'intérêt dans l'appariement selon un modèle a-contrario. Tenir compte de l'incertitude de localisation des points à appairer est le sujet de plusieurs travaux, du recalage 3D [PT97] à l'appariement de points sous transformation projective [Bra06, OB06]. L'incertitude est d'une certaine manière incorporée dans MLESAC [TZ00], car la localisation des *inliers* y est supposée connue avec une erreur gaussienne, d'écart-type fixé a priori. Les auteurs de [TC05] adaptent MLESAC pour y incorporer l'incertitude sur les transformations candidates. Les auteurs de [RFP09] proposent d'accélérer RANSAC en limitant la recherche des correspondances aux régions correspondant à une certaine p -valeur du risque. Dans tous ces articles, l'erreur de localisation des points d'intérêt est supposée isotrope. Dans ce cadre, les auteurs de [KK01] donnent une réponse mitigée à la question « est-il réellement nécessaire de prendre en compte l'incertitude de localisation ? ». Néanmoins, la définition de l'incertitude est assez simple dans ce dernier article, en particulier elle est indépendante de l'algorithme d'extraction des points d'intérêt. Les auteurs de [ZGS⁺09] analysent eux précisément l'algorithme SIFT, et dérivent une incertitude de localisation anisotrope basée sur la matrice des dérivées secondes de l'image, qu'ils valident expérimentalement.

Nous avons proposé dans [33, 35] un algorithme RANSAC a-contrario prenant en compte l'incertitude de localisation des points à appairer, pas nécessairement isotrope. Nous donnons ici ses grandes lignes.

Supposons qu'une liste de correspondances provisoires \mathcal{S} faite de N paires de points d -dimensionnels $(x_i, y_i)_{1 \leq i \leq N}$ soit disponible (après un appariement exploitant une similarité

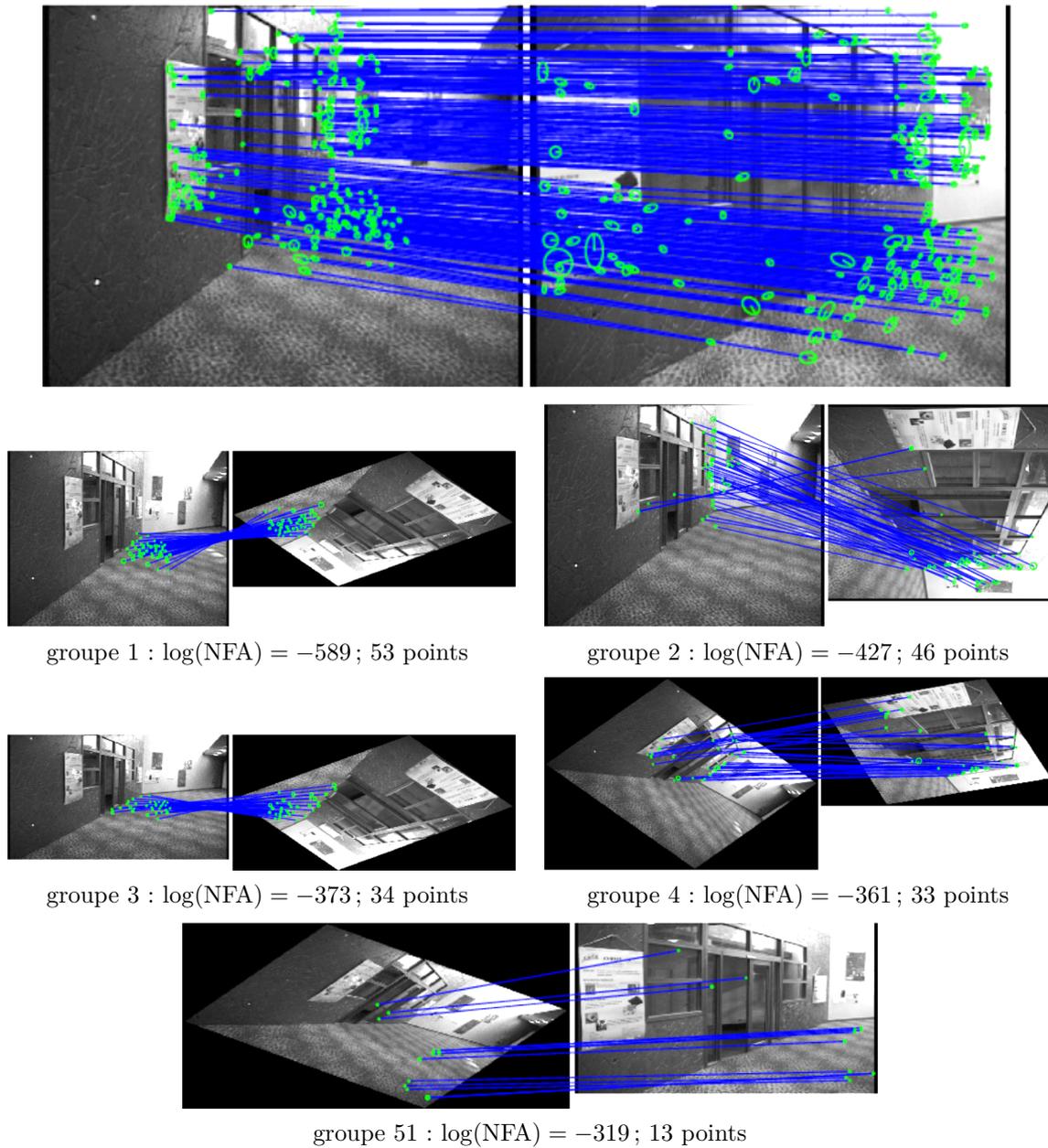


FIGURE 2.6 – En haut : 210 correspondances trouvées par I-ASIFT. Chaque ellipse verte est la reprojection d'un cercle avec un rayon égal à l'échelle SIFT dans l'image simulée par transformation affine. En dessous (groupes 1 à 4) : correspondances des quatre paires $(I_{t,\phi}, I'_{t,\phi'})$ correspondant aux groupes avec les NFA les plus faibles. On peut voir que ces groupes concernent en fait des points sur des morceaux de plans relativement concentrés. Dans cette expérience, 65 groupes sont gardés. Les 10 derniers groupes ne fournissent en fait que 14 correspondances. Par comparaison, les quatre groupes montrés en fournissent 116. Remarquons que les points sur le mur ou sur la moquette sont dispersés dans plusieurs groupes. En effet, l'homographie induite étant assez marquée, elle nécessite d'être approximée localement par plusieurs transformations affines. Dans le groupe 51, on remarque que l'appariement est piégé par l'aliasing perceptuel : les descripteurs sont semblables mais les correspondances sont cohérentes avec une homographie par hasard. Dix points de ce groupe sont effectivement reprojétés, mais tous sont éliminés par le RANSAC final qui impose la cohérence avec la géométrie épipolaire.

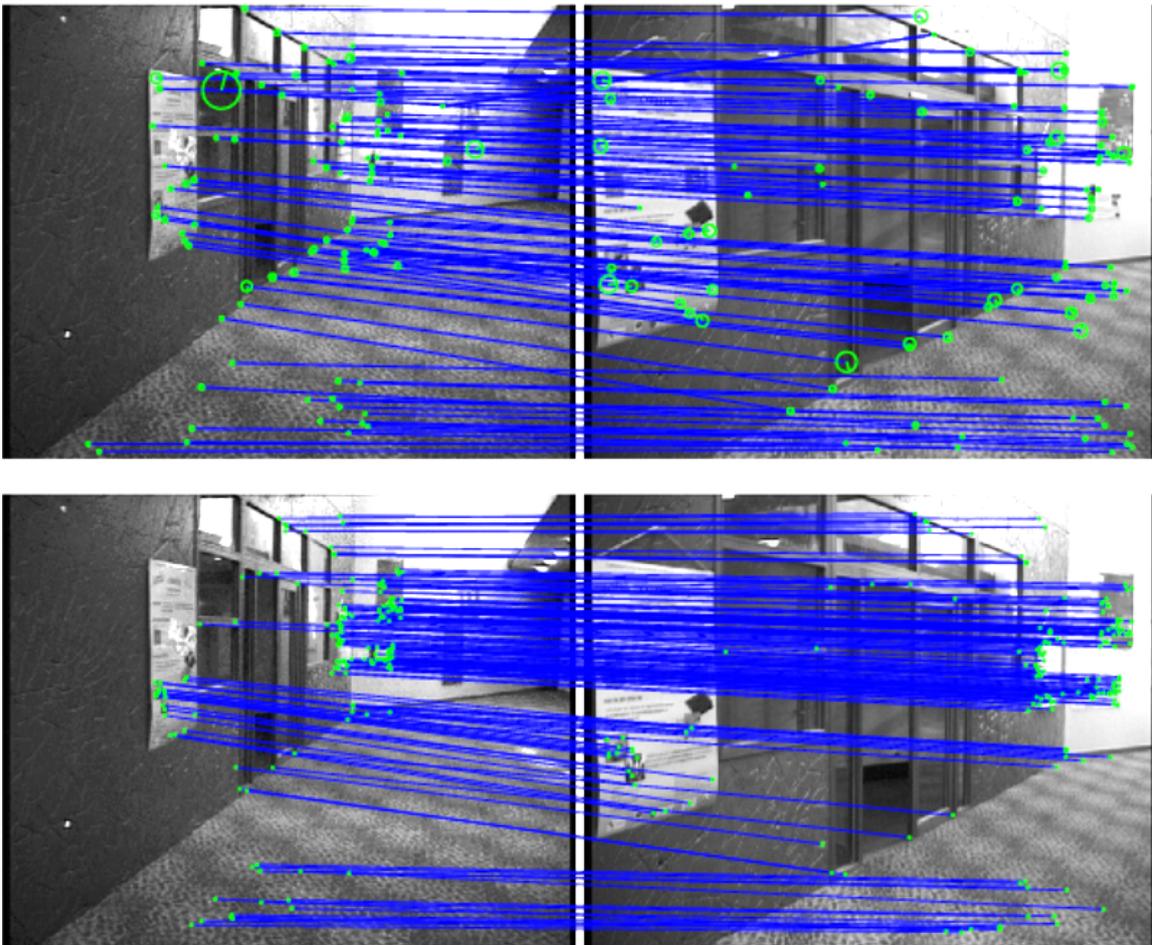


FIGURE 2.7 – En haut : appariement par SIFT (plus proche voisin, et seuil sur le rapport à 0,8, puis même RANSAC que dans ASIFT). 97 appariements sont trouvés. En bas : appariement par ASIFT. 153 appariements sont trouvés. Quelques points sur la moquette ne sont pas correctement appariés, et tout un groupe d'appariement est incorrect au premier plan, car les motifs répétés tombent par hasard sur la ligne épipolaire associée.

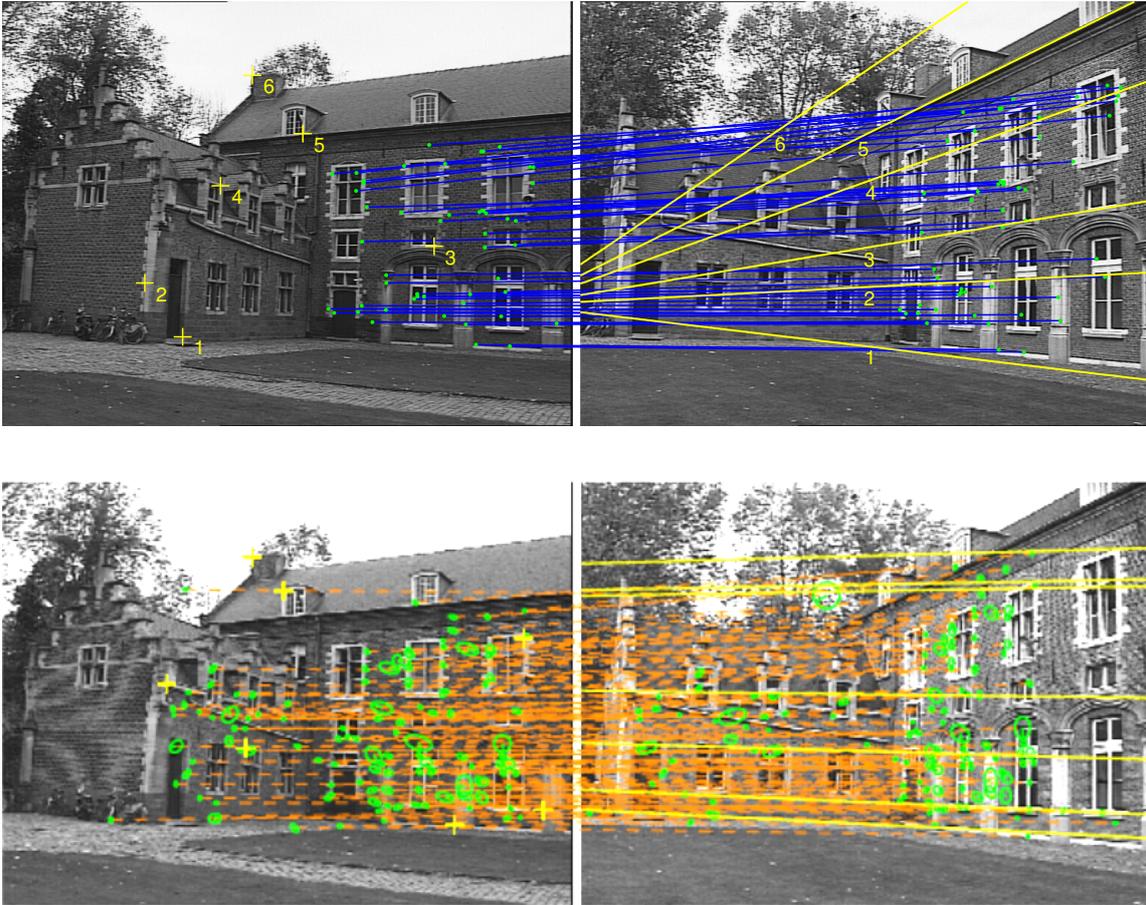


FIGURE 2.8 – En haut : résultat de ASIFT. En bas : résultat de I-ASIFT.

entre descripteurs par exemple). Les x_i (resp. y_i) sont modélisés comme des variables aléatoires gaussiennes de moyenne \bar{x}_i et de matrice de covariance Σ_{x_i} (resp. \bar{y}_i et Σ_{y_i}). Nous notons $x_i \sim \mathcal{N}(\bar{x}_i, \Sigma_{x_i})$.

À chaque étape de RANSAC, une transformation paramétrique \mathcal{T} est estimée, que l'on suppose basée sur m correspondances. Par exemple, si $d = 2$, l'estimation d'une homographie nécessite $m = 4$ correspondances, et si $d = 3$, $m = 5$. La localisation des points d'intérêt étant incertaine, l'estimation de \mathcal{T} sur la base de ces points l'est également.

Nous utilisons la propriété classique suivante :

Proposition 2.3 (*propagation de l'incertitude*) Si v est un vecteur aléatoire dans \mathbb{R}^d de loi gaussienne $\mathcal{N}(\bar{v}, \Sigma)$, et $f : \mathbb{R}^d \rightarrow \mathbb{R}^d$ est une transformation affine telle que $f(v) = f(\bar{v}) + A(v - \bar{v})$. Alors $f(v)$ suit la loi gaussienne $\mathcal{N}(f(\bar{v}), A\Sigma A^T)$.

Lorsque f n'est pas affine, il est classique d'utiliser une approximation de Taylor au premier ordre en v .

Dans [33, 35], nous montrons que l'incertitude de \mathcal{T} (représentée par la matrice de covariance $\Sigma_{\mathcal{T}}$) s'estime à partir de l'incertitude des $2m$ points ayant servi à son estimation et de la matrice Jacobienne de la transformation qui aux $2m$ points associe \mathcal{T} . Lorsque la mé-

thode de transformation linéaire directe (*Direct Linear Transform*) [HZ00] est utilisée, \mathcal{T} est obtenue par décomposition en valeurs singulières (SVD), et le « Jacobien de la SVD » [PL00] intervient alors dans le calcul. Dans [35], nous avons obtenu une expression analytique du Jacobien de la SVD dans le cas de l'estimation de la matrice fondamentale par l'algorithme dit « des 8 points ». Dans [33], nous utilisons plus simplement une estimation par différences finies.

La métrique utilisée dans notre adaptation de RANSAC est le carré de la distance de Mahalanobis entre couples de points d'intérêt, qui, sous certaines hypothèses d'indépendance des points d'intérêt, suit une loi du χ^2 à $2d$ degrés de liberté.

Plus précisément, si on note $d_M^s(x_i, y_i)$ le carré de la distance de Mahalanobis entre couples de points (x_i, y_i) appariés selon une transformation \mathcal{T} issue d'un ensemble minimal de points s , nous définissons les mesures suivantes de cohérence d'un ensemble S de correspondances potentielles :

$$d_{\max}(\mathcal{T}, S, s) = \max_{(x_i, y_i) \in S} d_M^s(x_i, y_i) \quad (2.21)$$

et :

$$d_{\text{sum}}(\mathcal{T}, S, s) = \sum_{(x_i, y_i) \in S} d_M^s(x_i, y_i) \quad (2.22)$$

Nous définissons alors l'hypothèse nulle \mathcal{H}_0 :

Définition 2.3 $\mathcal{H}_0 : (x_i, y_i)_{i \in S \setminus s}$ sont des variables aléatoires gaussiennes indépendantes, x_i de loi $\mathcal{N}(\bar{x}_i, \Sigma_{x_i})$, y_i de loi $\mathcal{N}(\bar{y}_i, \Sigma_{y_i})$, également indépendantes de la transformation \mathcal{T} de loi $\mathcal{N}(\bar{\mathcal{T}}, \Sigma_{\mathcal{T}})$ telles que $\forall i, \bar{y}_i = \bar{\mathcal{T}}x_i$.

Si δ est un réel positif, on obtient :

$$\Pr(d_{\max}(\mathcal{T}, S, s) \leq \delta \mid \mathcal{H}_0) = \left(f_{\chi_{2d}^2}(\delta) \right)^{k-m} \quad (2.23)$$

et :

$$\Pr(d_{\text{sum}}(\mathcal{T}, S, s) \leq \delta \mid \mathcal{H}_0) = f_{\chi_{2(k-m)d}^2}(\delta) \quad (2.24)$$

où $f_{\chi_{\alpha}^2}$ est la fonction de répartition du χ^2 à α degrés de liberté.

Un nombre de fausses alarmes est ensuite défini comme dans l'équation (2.10). Des résultats expérimentaux peuvent être trouvés dans [33, 35]. La difficulté est la non-linéarité des transformations étudiées (matrice fondamentale et homographie) ; l'approximation par des développements de Taylor locaux à l'ordre 1 semble assez grossière dans bon nombre de cas.

2.5 Conclusion et perspectives

Dans ce chapitre, nous avons rappelé nos résultats concernant l'appariement de points d'intérêt sous contraintes géométriques et photométriques, dans le cadre d'un modèle a-contrario. Nous avons également vu comment il était possible de rendre ASIFT davantage robuste à la présence de motifs répétés.

Un point resté en suspens à l'issue de ces travaux mériterait d'être examiné. Le NFA « photométrique » (au sens de [RDG09], voir la section 2.2.4) fournissant une mesure de la qualité des groupes d'appariements provisoires, c'est-à-dire avant d'imposer une contrainte géométrique, il est tentant de s'en servir pour guider un algorithme d'échantillonnage progressif. Les auteurs de PROSAC [CM05] utilisent le rapport entre distances euclidiennes des plus

proches et seconds plus proches descripteurs comme indicateur de qualité de l'appariement, une faible valeur de ce rapport étant censée indiquer une bonne confiance dans l'appariement. Ils adaptent alors RANSAC en générant des hypothèses successives par tirage préférentiel en favorisant les petites valeurs de ce rapport. Il serait intéressant de revisiter PROSAC à la lumière d'une métrique comme celle fournie par le NFA. D'autre part, une optimisation d'ASIFT a été récemment proposée [MMP15] : elle est basée sur différentes discrétisations de l'ensemble des transformations affines (t et ϕ dans l'équation (2.20)) et différents extracteurs de points d'intérêt. Il serait intéressant de voir si un échantillonnage progressif des paramètres t et ϕ et la recherche d'appariements provisoires dont la confiance serait mesurée par le NFA ne pourraient pas guider la recherche des t et ϕ optimaux, de manière à éviter un échantillonnage systématique.

Détermination de la pose d'une caméra

3.1	Introduction	41
3.2	Construction de mots visuels pour le calcul de pose	42
3.3	Simulation de points de vue pour l'appariement entre une image test et un modèle de la scène	43
3.4	Conclusion et perspectives	58

3.1 Introduction

3.1.1 Contexte

Ce chapitre traite de l'estimation de la *pose* d'une caméra, c'est-à-dire sa localisation et son orientation dans une scène tridimensionnelle. Il s'agit d'un des problèmes fondamentaux de la vision par ordinateur (voir la revue [LF05]), utilisé par exemple pour l'initialisation de la manipulation d'objets par un bras robotique [CBSF09], la relocalisation dans une approche de type *Simultaneous Localization And Mapping* (SLAM) en cas de perte du suivi [WKR07], la navigation d'un robot [LSC⁺15, LSB⁺15], la reconnaissance d'objets [RLSP06], et de façon générale pour les applications de localisation [SBS07] ou de réalité augmentée [GL06, BCL15].

Plus précisément, le problème abordé est celui de la détermination de la pose d'une caméra à partir d'une image test et d'un modèle de la scène tridimensionnelle. Dans nos travaux, le modèle de la scène sera essentiellement un nuage de points 3D obtenu par un algorithme de type *Structure from Motion* (SfM) à partir d'une collection d'images préalablement acquises. En pratique nous avons utilisé les logiciels VisualSfM [Wu11, WACS11] ou Bundler [Sna10, SSS06, SSS08]. Des descripteurs (généralement SIFT [Low04]) sont associés aux points du modèle 3D, et sont utilisés pour trouver des appariements entre des points d'intérêt de l'image test et les points du modèle 3D. Enfin, la pose de la caméra test est estimée à partir de ces correspondances par résolution du problème *Perspective-n-Point* (PnP) [DD95, HR11, LMNF09]. Dans nos travaux, nous nous sommes limités à l'estimation précise de la pose de la caméra dans un environnement de taille relativement limitée (une pièce, ou les environs proches d'un bâtiment). Il s'agit d'un problème réaliste dans une application de suivi où le calcul de pose serait utilisé pour une réinitialisation du système, ou dans une application dans laquelle un positionnement grossier par GPS serait disponible. L'estimation de pose en environnement très vaste [IZFB09] (voire mondial [LSHF12]) commence

d'ailleurs généralement par une phase d'élagage du modèle pour en limiter l'étendue. Nous supposons également que les paramètres intrinsèques des caméras sont connus.

Dans ce contexte, toute la difficulté du calcul de pose est que le modèle de la scène est construit à partir d'images prises avec des distances à la scène ou orientations potentiellement très différentes des conditions d'acquisition de l'image test. C'est par exemple le cas dans des applications de conduite autonome d'un robot [CRC16], ou des applications de localisation d'un drone prenant des images aériennes d'une scène construite à partir d'images « à hauteur d'homme » [SWC⁺14]. Les descripteurs utilisés n'étant invariants qu'à des changements de point de vue limités (selon les auteurs de [MP07], la robustesse de l'appariement est limitée à des changements d'orientation d'au plus 30°), en plus des limites numériques relevées dans [RO15], l'appariement des descripteurs préalable à la résolution du problème PnP constitue un problème très difficile.

3.1.2 Contributions

Nos contributions concernent essentiellement l'appariement de points entre le modèle de la scène et l'image test. Dans la thèse de Srikrishna Bhat [Bha13],[31, 34], nous avons étudié la construction de *mots visuels* permettant de construire des correspondances pour l'étape de SfM et décrire les points du modèle. Ces mots visuels sont obtenus par classification non supervisée des descripteurs des images permettant de construire le modèle. Ces travaux sont décrits en section 3.2. Dans la thèse de Pierre Rolin [19, 25, 40], nous avons proposé un enrichissement des descripteurs du modèle par une technique de *simulation de points de vue*. L'idée est inspirée de l'algorithme d'appariement de paires d'images ASIFT [MY09, YM11] : en ajoutant des descripteurs synthétiques correspondant à des points de vue non présents dans les images initiales, on facilite l'appariement d'une image test prise avec un point de vue très éloigné de celui de ces images. Ces travaux sont décrits en section 3.3.

Les principales publications relatives à ce chapitre sont :

[19] P. Rolin, M.-O. Berger, and F. Sur. Simulation de point de vue pour la mise en correspondance et la localisation. *Traitement du signal*, 32(2-3):169–194, 2015.

[22] P. Rolin, M.-O. Berger, and F. Sur. Enhancing pose estimation through efficient patch synthesis. In *Proceedings of the British Machine Vision Conference (BMVC)*, York, UK, 2016.

[25] P. Rolin, M.-O. Berger, and F. Sur. Viewpoint simulation for camera pose estimation from an unstructured scene model. In *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, Seattle (WA), USA, 2015.

[31] S. Bhat, M.-O. Berger, and F. Sur. Visual words for 3D reconstruction and pose computation. In *Proceedings of the Joint 3DIM/3DPVT Conference (3DIMPVT)*, pages 326–333, Hangzhou, China, 2011.

Bon nombre des résultats présentés dans ce chapitre ont été obtenus par Srikrishna Bhat (thèse soutenue en janvier 2013 [Bha13]) et Pierre Rolin (soutenance prévue début 2017), doctorants que j'ai co-encadrés avec Marie-Odile Berger.

3.2 Construction de mots visuels pour le calcul de pose

Si les vocabulaires visuels sont des outils standards de la classification d'images ou d'objets (depuis les travaux décrits dans [NS06, SZ03]), ils semblent davantage émergents pour les problèmes de SfM et d'identification de correspondances pour le calcul de pose ; citons la

reconstruction 3D dans des applications multimédia [AFS⁺11], la détection de boucles dans une trajectoire [AFDM08], la localisation rapide en grand environnement [IZFB09], ou le calcul de pose [XCYQ08].

L'étape préalable à la construction d'un modèle de la scène par ajustement de faisceaux et algorithme SfM est généralement basée sur l'appariement de descripteurs SIFT [Low04] : cette approche est notamment utilisée par les auteurs de [GL06, Sna10, Wu11]. Néanmoins, obtenir un modèle de bonne précision par ajustement de faisceaux nécessite de construire des ensembles de correspondances entre images balayant des points de vue assez différents, afin de diminuer l'erreur de parallaxe. Cette construction se fait par agrégation successive des images à partir des correspondances obtenues d'une première paire d'images. De nombreux paramètres sont nécessaires (comme souligné dans [MMM12]), ce qui fait qu'un ensemble de correspondances peut finalement ne concerner que des changements limités de points de vue. C'est pourquoi nous avons proposé de grouper en mots visuels les descripteurs de l'ensemble des images servant à construire le modèle dans une approche globale, non agrégative, préalable à tout SfM. Si les images initiales présentent une certaine continuité dans les changements de points de vue, une classification non supervisée devrait permettre de grouper ensemble les descripteurs associés au même point 3D. Les mots visuels présentent aussi l'intérêt de représenter de manière compacte l'information photométrique associée à un point du modèle. Ces mots visuels sont ensuite utilisés pour permettre l'appariement entre les points d'intérêt d'une image test et la scène, ce qui permet ensuite de calculer la pose.

Plus précisément, nous avons testé différentes méthodes de construction de mots visuels : k -moyennes (à la suite de [NS06, SZ03]), *mean-shift* [Che95, CM02], fermeture transitive des appariements obtenus par seuillage de la distance euclidienne entre descripteurs (de manière similaire à [AFDM08]), ainsi que les groupes de descripteurs correspondant au même point 3D obtenu directement en sortie d'un algorithme SfM classique.

Concernant l'appariement des points pour construire le modèle par SfM, nos expériences montrent que les mots obtenus par fermeture transitive fournissent un modèle présentant plus de point, et plus précis car ces mots permettent de grouper ensemble des descripteurs associés à des images présentant une plus grande variabilité de points de vue, alors que les mots obtenus par k -moyennes souffrent de la dépendance forte au nombre k de groupes qui doit être fixé a priori. La construction des mots par fermeture transitive est 100 à 200 fois plus rapide que par *mean-shift*.

Une fois qu'un algorithme d'appariement des descripteurs de l'image test et des mots visuels est défini, il est possible de calculer la pose par PnP. Si nos expériences ne montrent pas que les différentes stratégies se différencient, elles permettent de conclure que le paramètre intervenant dans les mots visuels (le seuil sur les distances entre descripteurs) n'est pas un paramètre sensible. Si des poses satisfaisantes peuvent être obtenues lorsque l'image test ne présente pas une forte variabilité de point de vue par rapport aux images permettant de construire le modèle, ce n'est plus le cas dès que la variabilité s'accroît. Ce constat motive les travaux décrits dans la section suivante.

3.3 Simulation de points de vue pour l'appariement entre une image test et un modèle de la scène

Nous avons proposé d'enrichir la description des points 3D en générant par simulation des descripteurs additionnels qui correspondent à des points de vue éloignés de ceux des images

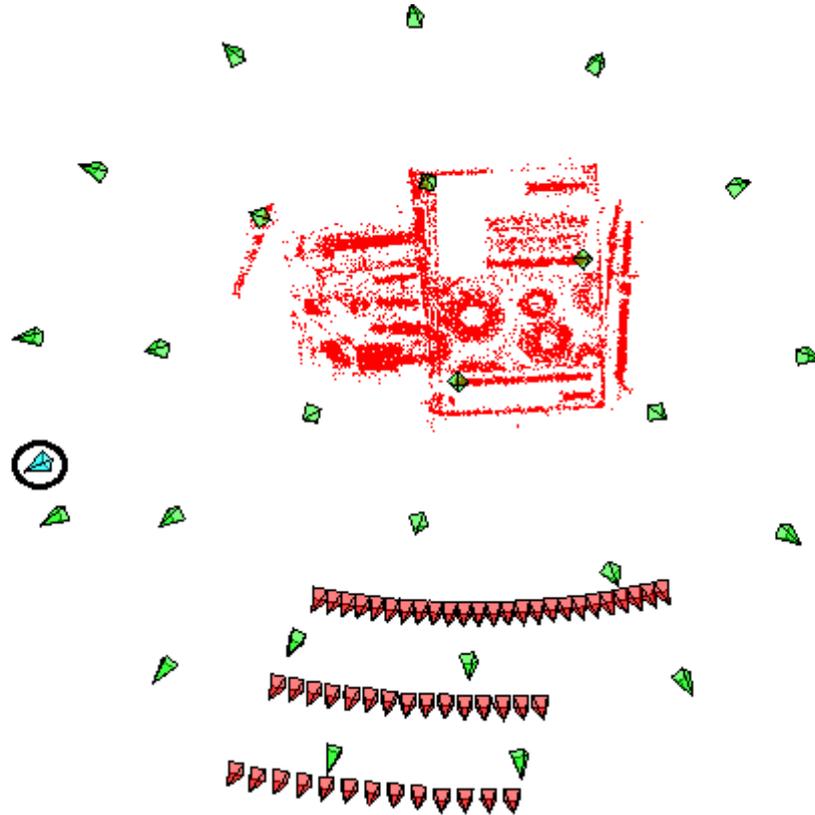


FIGURE 3.1 – Le modèle 3D de la scène (points rouge), les caméras ayant servi à le construire (en rouge pâle), une caméra éloignée dont on chercherait la pose (en cyan, position entourée), et les caméras virtuelles (en vert), réparties ici sur une demi-sphère centrée sur la scène. Les caméras virtuelles permettent de générer de nouveaux descripteurs pour chaque point du modèle.

ayant servi à la reconstruction initiale du modèle. Nous illustrons par plusieurs expériences qu'enrichir ainsi le modèle augmente le degré d'invariance de la description des points 3D. Cela facilite l'appariement, et donc le calcul de pose, lorsque la scène présente un fort changement d'aspect dans la nouvelle vue dont on cherche la pose. La simulation de points de vue a déjà montré son utilité dans le cadre de la mise en correspondance entre deux images présentant un changement de point de vue important, comme dans ASIFT [MY09] ou ses variantes [BZU10, MMP15][32]. Dans ASIFT, la simulation est faite en utilisant des transformations affines. Dans notre cas, en supposant la scène localement plane, toutes les vues d'une région autour d'un point 3D sont liées par des homographies avec le modèle sténopé ou des transformations affines avec le modèle de caméra orthographique. Nos descripteurs simulés seront donc générés à partir de vues synthétisées par un certain nombre de transformations de l'un de ces deux types, de manière à simuler un déplacement de la caméra dans des positions non représentées dans les images initiales. Ceci est illustré par les figures 3.1 et 3.2.

Notons qu'une méthode similaire est envisagée dans [KS14] et [WCL⁺08], mais les vues simulées sont uniquement fronto-parallèles. Dans [HCH10], la simulation est utilisée pour améliorer la reconnaissance d'objets, mais, contrairement à ce qui est proposé dans notre méthode, la géométrie de l'objet n'est pas prise en compte pour faire les simulations. Dans [IZFB09] la

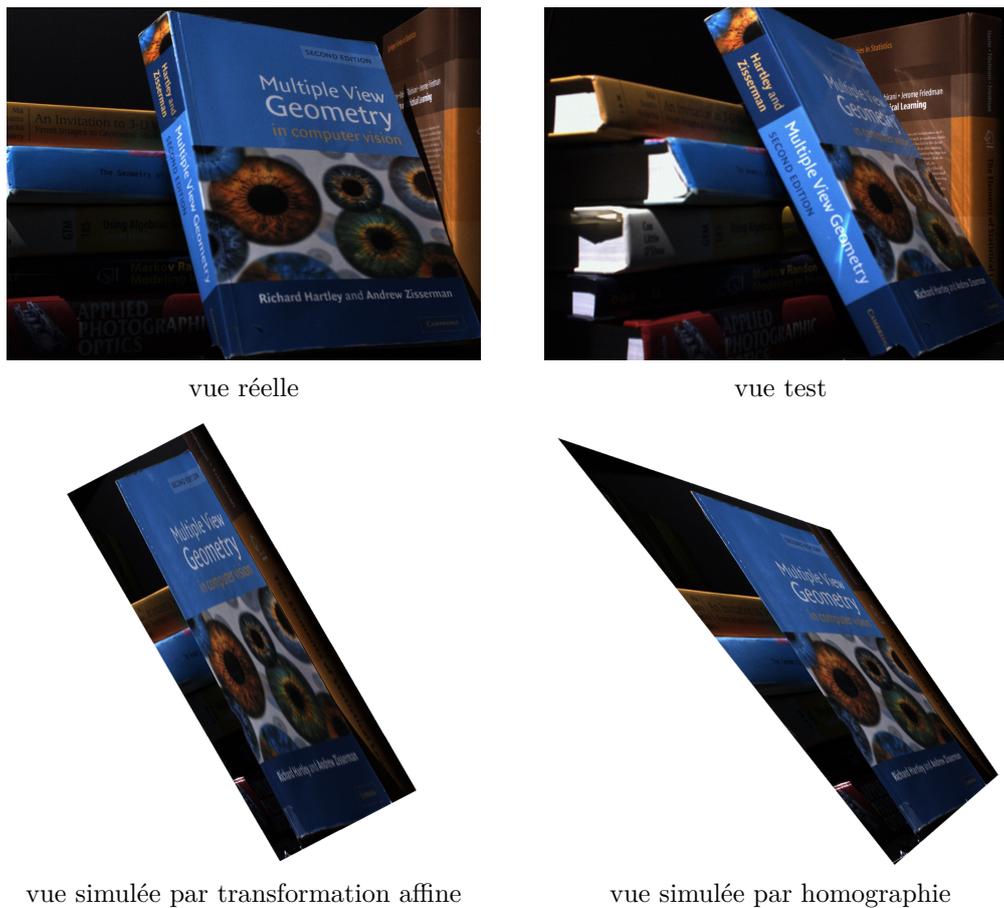


FIGURE 3.2 – Un exemple de simulation. Notons le fort changement de point de vue entre la vue test et la vue réelle la plus proche de celle-ci ayant servi à reconstruire le modèle. La vue test pourra difficilement être appariée au modèle à partir de descripteurs issus de vues comme la vue réelle. Les vues simulées de la couverture du livre depuis la position test sont obtenues à partir de la vue réelle. La simulation par homographie ressemble davantage à la vue test.

simulation est utilisée dans le cadre de la localisation d'une caméra dans un grand environnement, mais il semble que les changements de point de vue considérés soient relativement faibles. Dans le cadre de la navigation robotique, les auteurs de [CRC16] simulent des vues cohérentes avec la trajectoire estimée du robot.

La méthode que nous proposons permet de localiser une caméra par rapport à un modèle obtenu par SfM, même lorsque la pose cherchée est très éloignée des vues ayant été utilisées pour construire le modèle. Cette invariance aux changements de points de vue est obtenue en complétant la description de chaque point du modèle par des descripteurs issus d'images simulées prenant en compte la géométrie locale du point 3D. Nos premières expériences sur une scène centrée sur un objet montrent que l'ajout de vingt-cinq points de vue virtuels permet de calculer des poses là où des méthodes basées uniquement sur SIFT (comme celle décrite dans [GL06]) échouent. De manière plus générale, la simulation augmente le taux de correspondances correctes entre l'image et le modèle, ce qui permet à RANSAC de converger plus rapidement vers un ensemble de consensus fournissant la pose cherchée.

Nous supposons disposer d'un modèle d'une scène, constitué d'un nuage de points, et que chacun de ces points est associé à un ensemble de descripteurs SIFT provenant des vues réelles dans lesquelles il a été repéré par l'algorithme de SfM. Nous supposons également que la scène est localement plane autour des points 3D, et que l'on a associé à chaque point le vecteur normal du plan sur lequel il se trouve (ce vecteur étant calculé par analyse en composante principale locale sur le modèle 3D initial, voir [HDD⁺92]). Étant donnée une vue réelle d'une zone plane autour d'un point 3D, comment synthétiser une vue de cette zone à partir d'une nouvelle position de caméra, afin d'en extraire un nouveau descripteur SIFT ? Si on modélise les caméras comme des sténopés, deux vues d'un même plan sont liées par une homographie. Dans le modèle de caméras affines (lorsque la profondeur de la scène est faible devant la focale), les deux vues sont liées par une transformation affine, ce qui n'est pas très réaliste lorsque la différence de points de vue est assez grand et la focale relativement courte.

3.3.1 Cas des homographies

Soient deux caméras représentées par leurs matrices de projection $P_1 = K_1[R_1|T_1]$ et $P_2 = K_2[R_2|T_2]$, où K_i est la matrice des paramètres intrinsèques pour un capteur à pixels carrés (supposée connue), et R_i, T_i déterminent la pose dans un repère commun, $i \in \{1, 2\}$. Considérons un plan de l'espace d'équation $n^T X + d = 0$ (où n est un vecteur normal au plan, d un paramètre réel, et X des coordonnées d'un point de l'espace). La transformation induite par le plan entre les deux caméras est alors l'homographie H donnée par l'équation homogène [HZ00] :

$$H = K_2(R - Tn^T/d)K_1^{-1} \quad (3.1)$$

où $R = R_2R_1^T$ et $T = -R_2(C_2 - C_1)$ et le centre optique C_i vérifie $C_i = -R_i^T T_i$, $i \in \{1, 2\}$.

Remarquons que dans le cas où les deux caméras partagent le même axe optique et que celui-ci porte le vecteur n , cette homographie se réduit à une similitude.

Si P_1 est la matrice de projection d'une caméra réelle, P_2 celle d'une caméra virtuelle, et I_1 et I_2 les images du plan dans ces deux caméras, alors $HI_1 = I_2$, soit :

$$K_2R_2(R_1^T + (C_2 - C_1)n^T/d)K_1^{-1}I_1 = I_2. \quad (3.2)$$

Rappelons que la matrice R_2 s'écrit $R_2 = R_Z(\kappa)R_Y(\phi)R_X(\omega)$ où (X, Y, Z) est un repère orthonormé tel que Z est l'axe optique de la caméra et (κ, ϕ, ω) sont les angles d'Euler associés. Les descripteurs SIFT étant supposés invariants par similitude (plane), on voit que toute rotation autour de l'axe optique ou tout changement de focale de la caméra 2 fournira les mêmes descripteurs. Donc la pose de la caméra virtuelle n'a besoin d'être fixée qu'à une rotation selon l'axe optique près, et la focale est arbitraire. Comme il l'a été souligné dans [MY11], ce raisonnement sur des images idéales continues reste valable pour des images discrètes sous réserve de respect de la condition de Shannon-Nyquist. Néanmoins la position de la caméra est ici importante, T_2 intervenant dans (3.2).

La donnée du plan, d'une pose de caméra réelle, et de la pose de la caméra virtuelle (à une rotation selon l'axe optique près) permet de simuler avec l'équation (3.2) une vue de laquelle nous allons extraire un descripteur SIFT.

3.3.2 Cas des transformations affines

Dans le cas de deux caméras affines, notons $(\lambda_i, \psi_i, t_i, \phi_i)$ les éléments caractéristique de la caméra $i \in \{1, 2\}$ dans un repère associé à un plan repéré par son vecteur normal n ,

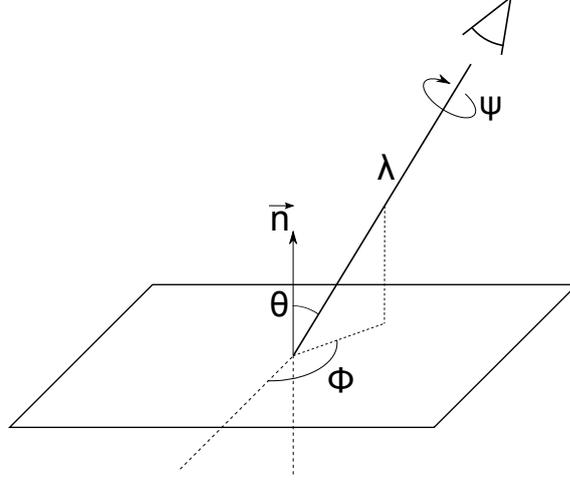


FIGURE 3.3 – Position d’une caméra affine par rapport à la normale d’un morceau de plan, avec les notations de l’équation (3.3) où $t = 1/\cos(\theta)$.

comme illustré par la figure 3.3. Les angles ϕ_i et θ_i sont respectivement la longitude et la latitude de l’axe optique de la caméra. Le paramètre $t_i = 1/\cos(\theta_i)$ est le *tilt* de la caméra. Le paramètre ψ_i correspond à la rotation de la caméra autour de son axe optique et λ_i au zoom. La transformation induite par le plan entre une vue fronto-parallèle de ce plan et la vue i est donnée par la transformation affine suivante [MY09, OCLF10] :

$$A_i = \lambda_i \begin{pmatrix} \cos(\psi_i) & -\sin(\psi_i) \\ \sin(\psi_i) & \cos(\psi_i) \end{pmatrix} \begin{pmatrix} t_i & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \cos(\phi_i) & -\sin(\phi_i) \\ \sin(\phi_i) & \cos(\phi_i) \end{pmatrix}. \quad (3.3)$$

Par composition, la transformation affine induite par le plan entre les deux caméras est :

$$A = A_2 A_1^{-1}. \quad (3.4)$$

Avec les mêmes notations que dans le cas des homographies, $A I_1 = I_2$ soit $A_1^{-1} I_1 = A_2^{-1} I_2$. L’invariance aux similitudes des descripteurs SIFT nous permet d’écrire que toutes les valeurs de $\psi_1, \psi_2, \lambda_1, \lambda_2$ fournissent les mêmes descripteurs SIFT, que l’on choisit donc arbitrairement à $\psi_1 = \psi_2 = 0, \lambda_1 = \lambda_2 = 1$.

Ainsi la donnée des positions relatives (t_i, ϕ_i) des caméras réelles et virtuelles par rapport à la normale à une partie plane de la scène permet de simuler une vue avec l’équation (3.4) de laquelle on extraira un descripteur SIFT.

3.3.3 Résumé de la méthode proposée

3.3.3.1 Génération des descripteurs synthétiques

Pour chaque point du modèle 3D associé à une direction normale, et pour chaque position de caméra virtuelle, on génère une vue (selon une transformation homographique ou affine selon la méthode choisie), puis on extrait un descripteur SIFT dans cette vue que l’on associe au point 3D. Un exemple de simulation est visible sur la figure 3.2.

3.3.3.2 Position des caméras virtuelles

La position des caméras virtuelles est choisie de manière à compléter les points de vue des caméras ayant permis de construire le modèle. Comme le justifie la discussion de la section 3.3.2, le cas affine ne nécessite que de positionner les caméras sur une demi-sphère orientée par la normale considérée, alors que le cas homographique nécessiterait de préciser leur distance par rapport à la scène.

Dans cette première étude nous plaçons les caméras virtuelles dans les mêmes positions dans les deux cas : il s'agit de vingt-cinq positions régulièrement réparties sur une demi-sphère s'appuyant sur un plan moyen de la scène, de rayon égal à la distance de la plus proche caméra à la scène, comme dans la figure 3.1 ; les caméras sont dirigées vers le barycentre de la scène. Nous simulons donc un grand nombre de directions d'observation de la scène, mais pas de variations de la distance de la caméra à la scène. Néanmoins, les expériences montrent que ces simulations sont suffisantes pour calculer des poses relativement éloignées des vues de reconstruction et des vues virtuelles.

Cet échantillonnage est arbitraire pour le moment, mais devra à terme être défini en fonction de la géométrie de la scène et des points de vue utilisés pour construire le modèle.

3.3.3.3 Choix de la vue utilisée pour la simulation et extraction d'un descripteur SIFT

Étant donné un point du modèle 3D (associé à des descripteurs venant de plusieurs vues réelles) et un point de vue à simuler, il faut également choisir à partir de quelle vue réelle réaliser la simulation. Parmi les vues dans lequel le point 3D est visible, la vue à partir de laquelle la simulation est réalisée est la plus proche angulairement du point de vue qu'on veut simuler, ce qui est un choix classique pour limiter l'influence des spécularités et des artefacts de discrétisation.

La simulation produit une image de taille 100×100 pixels centrée sur un point du modèle, qui correspond à l'apparence de ce point observé à partir d'une caméra virtuelle. L'algorithme SIFT permet alors d'extraire des couples de points d'intérêt et descripteurs dans cette image. On ajoute alors à la liste des descripteurs de ce point 3D le descripteur extrait de l'image dont le point d'intérêt est le plus proche de la position théorique de la projection du point 3D, si cette distance est inférieure à 10 pixels. Ce seuil correspond à une distance de reprojction typique des points du modèle obtenu par SfM.

3.3.3.4 Correspondances image/modèle

On commence par extraire les descripteurs SIFT de la nouvelle vue. La méthode de mise en correspondance utilisée est celle proposée dans [GL06]. Pour apparier un point d'intérêt p_1 de la nouvelle vue à un point 3D, on considère les distances d_1 et d_2 du descripteur SIFT de p_1 aux deux plus proches classes de descripteurs. Si d_1/d_2 est inférieur à un seuil λ on retient la correspondance. La recherche des plus proches voisins est accélérée comme dans [GL06] par une recherche approchée [MA10, ML14].

3.3.3.5 Perspective-n-Points

Le calcul de pose se fait par une estimation robuste de type RANSAC [FB81] basée sur l'algorithme PnP proposé dans [HR11]. Bien entendu, plus la proportion de correspondances

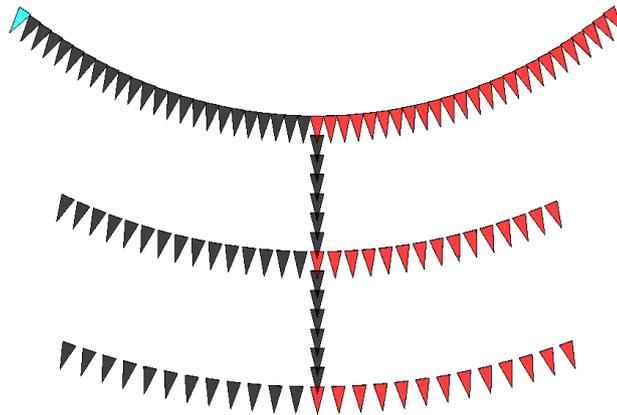


FIGURE 3.4 – Les positions des 119 caméras de la base Robot Data Set (vues de dessus). Elles sont réparties sur des cercles centrés sur la scène, non-représentée. En rouge les caméras servant à la reconstruction par SfM, en cyan la caméra de test (en haut à gauche).

correctes dans l'étape précédente est grande, plus le nombre d'itérations requises dans RANSAC peut être diminué.

3.3.4 Étude expérimentale

Les expériences suivantes montrent qu'en présence de fortes variations de direction de vue ou de profondeur la simulation de point de vue améliore considérablement l'estimation de la pose. La pose peut être calculée dans des situations où une approche basée uniquement sur SIFT, telle que celle de [GL06], échoue. Plus généralement, pour un nombre fixé d'itérations de RANSAC, la pose est calculée avec plus de précision en utilisant la simulation. À la fin de cette section nous discutons les problèmes de temps de calcul et les améliorations envisageables.

3.3.4.1 Protocole expérimental

La méthode proposée est évaluée sur quatre séquences d'images : la séquence numéro 2 de la base *Robot Data Set* [ADP12] avec la première illumination proposée (la reconstruction de la scène est présentée dans la figure 3.1 et les positions des caméras utilisées dans la figure 3.4) et trois séquences spécialement acquises, illustrées par la figure 3.5. Ces séquences sont composées d'images de taille 1600×1200 pixels et les scènes associées sont globalement planes par morceaux et centrées sur un objet.

Toutes les expériences utilisent le même protocole. Un modèle 3D de la scène est construit avec VisualSfM. La pose d'une vue test est calculée dans trois scénarios : **S** où le modèle est la reconstruction obtenue par SfM sans simulation, **A** où le modèle de **S** est enrichi par des descripteurs issus de simulations affines, et **H** où le modèle de **S** est enrichi par des descripteurs issus de simulations homographiques.

Pour comparer les trois scénarios, cent poses sont calculées pour la même vue test dans chaque cas en utilisant le même nombre d'itérations de RANSAC. La variabilité de ces cent poses est évaluée visuellement. Lorsque ces poses sont superposées, nous calculons également



FIGURE 3.5 – Images représentatives des quatre séquences. Livre vient de la base Robot Data Set [ADP12]. Les autres séquences sont acquises spécialement pour l'expérience.

l'écart type (reporté dans les figures). L'échelle étant un paramètre libre de toute reconstruction SfM, les écart types sont exprimés en pourcentage de la distance à la scène. De plus, pour chaque expérience, des contours des objets de la scène sont reprojétés dans la vue test en utilisant les poses calculées.

Comme les taux d'*inliers* dans les correspondances image/modèle sont très variables d'une séquence à une autre (par exemple de 4 % à 23 % pour le scénario **S**), nous utilisons un nombre d'itérations de RANSAC différent pour chaque séquence. Cependant, pour rendre possible la comparaison de la variabilité, le même nombre d'itérations est utilisé pour les trois scénarios.

3.3.4.2 Robustesse du calcul de pose aux changements de direction de vue

Nous montrons ici que la simulation de point de vue améliore significativement la précision des poses calculées lorsque la vue test est éloignée des vues réelles et a donc un aspect très différent.

Nous présentons d'abord les résultats sur la séquence Livre (figure 3.1) pour laquelle la pose réelle de la vue test est connue. Il est donc possible de déterminer si une correspondance 2D/3D est correcte ou non, en reprojétant le point 3D en utilisant la pose de la vérité terrain.

Si la distance de reprojection est inférieure à 20 pixels la correspondance est considérée comme correcte (ce seuil correspond à $\mu + 3\sigma$ avec μ et σ respectivement la moyenne et l'écart type de l'erreur de reprojection de l'étape SfM ; les images sont de taille 1600×1200 pixels). Dans cette expérience, la proportion de correspondances correctes est de 23 % dans le scénario **S**, 30 % dans le scénario **A** et 37 % dans le scénario **H**.

La figure 3.6 montre la répartition des correspondances 2D/3D parmi les vues réelles et simulées dans le scénario **H**. Le point de vue qui contribue le plus au calcul de pose est virtuel et proche de la caméra test. Globalement, les points de vue simulés produisent 85 % de l'ensemble de consensus de RANSAC. Ces graphes illustrent la pertinence de la méthode proposée et l'augmentation du taux d'*inliers* obtenue grâce aux simulations.

Les résultats du calcul de pose sont illustrés dans la figure 3.7 ($N = 1000$). Les poses estimées sont visuellement plus précises dans les scénarios **A** et **H** que dans le scénario **S**. Avec 500 itérations dans RANSAC, le calcul de la pose échoue dans **S**, alors que les résultats sont corrects dans **H**. En augmentant le nombre d'itérations à 1 000, la variabilité de la pose n'est que légèrement réduite dans **S** alors que dans **H** toutes les poses calculées sont superposées.

Un phénomène remarquable se produit dans **A** (et dans une moindre mesure dans **S**). Dans cette expérience, les poses calculées se répartissent en trois catégories : la plupart des poses sont proches du point de vue attendu, quelques unes sont totalement fausses et un groupe de poses erronées se trouve face à la couverture du livre. Cet ensemble d'erreurs est provoqué par un motif répété de la scène, à savoir l'œil de la couverture qui apparaît également sur la tranche du livre. La reprojection des bords de la couverture dans la figure 3.7 illustre bien le phénomène. Dans ce cas, les simulations homographiques produisent plus de correspondances en dehors de ce motif répété, ce qui permet d'obtenir des poses correctes dans **H**. L'influence des motifs répétés est discutée par exemple dans [32, 14][RSSH11].

Ces expériences ont été reproduites sur les séquences **Poster** et **Bureau** avec des résultats similaires, comme illustré par les figures 3.8 et 3.9. Dans tous les cas présentés, la simulation améliore la précision de l'estimation de la pose, ce qui est illustré par la meilleure superposition des positions de caméra estimées ou des quadrilatères correspondant à la projection de contours 3D de la scène par les caméras estimées.

Le tableau 3.1 donne pour chaque séquence la taille des modèles utilisés et les temps de calcul des étapes de mise en correspondance et de calcul de pose. Le code est exécuté sous Matlab sur un processeur Intel Core i7 sans optimisation. Les temps de calcul sont raisonnables pour un prototype.

La figure 3.10 (base de donnée décrite dans [CZSP12]) présente une illustration dans un environnement plus grand. Si l'estimation de pose est possible sans simulation, l'utilisation de la simulation améliore grandement la vitesse de convergence de RANSAC par diminution de la proportion de fausses correspondances (*outliers*).

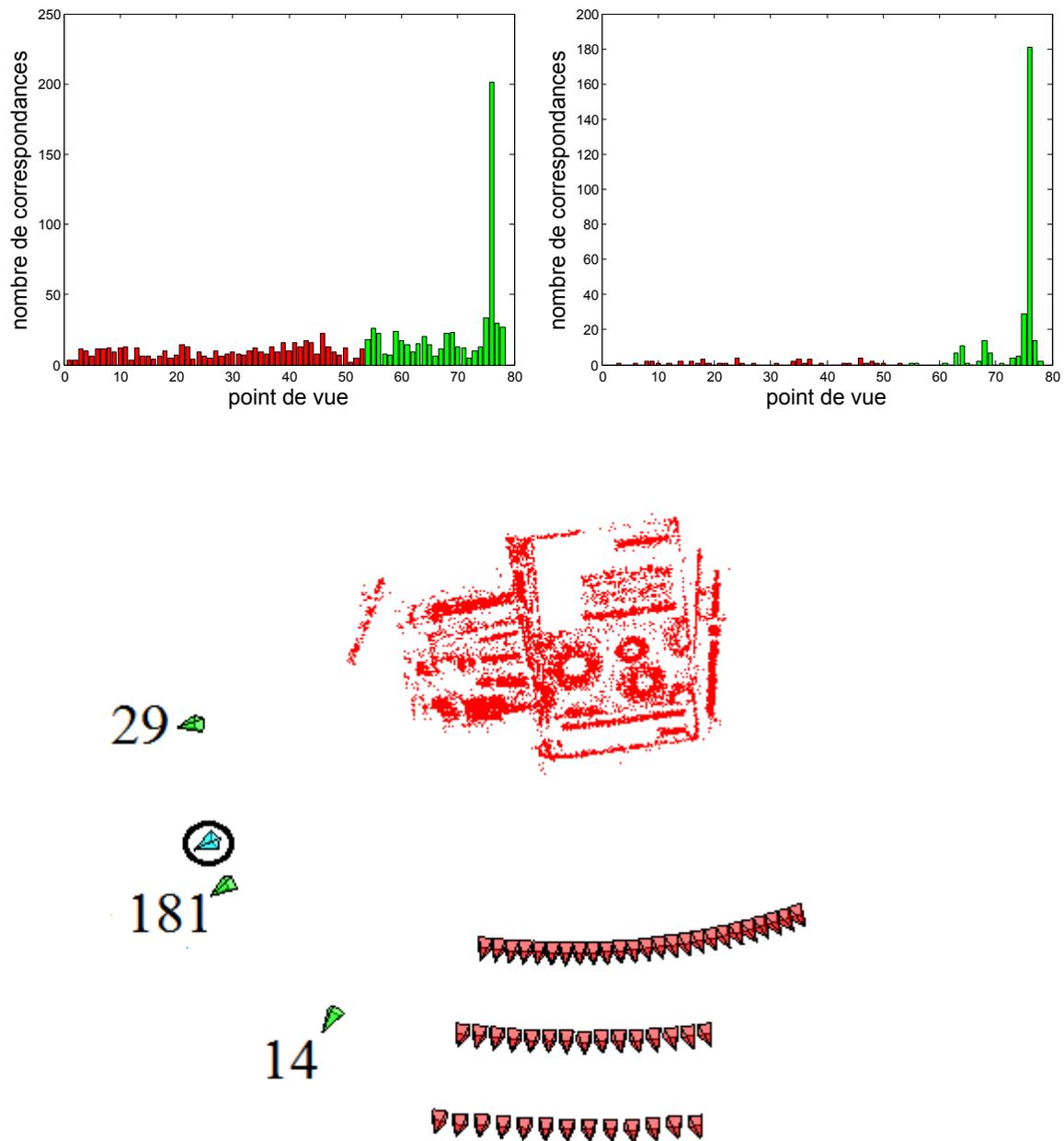


FIGURE 3.6 – Séquence Livre : nombre de correspondances associées à chaque point de vue (réel en rouge, virtuel en vert), pour l'ensemble des correspondances image/modèle (en haut à gauche) et dans l'ensemble de consensus trouvé par RANSAC (en haut à droite). Les points de vue contribuant le plus restent les mêmes, et sont proches de la pose cherchée. Les trois points de vue contribuant le plus et le nombre de correspondances associées sont montrés en bas.

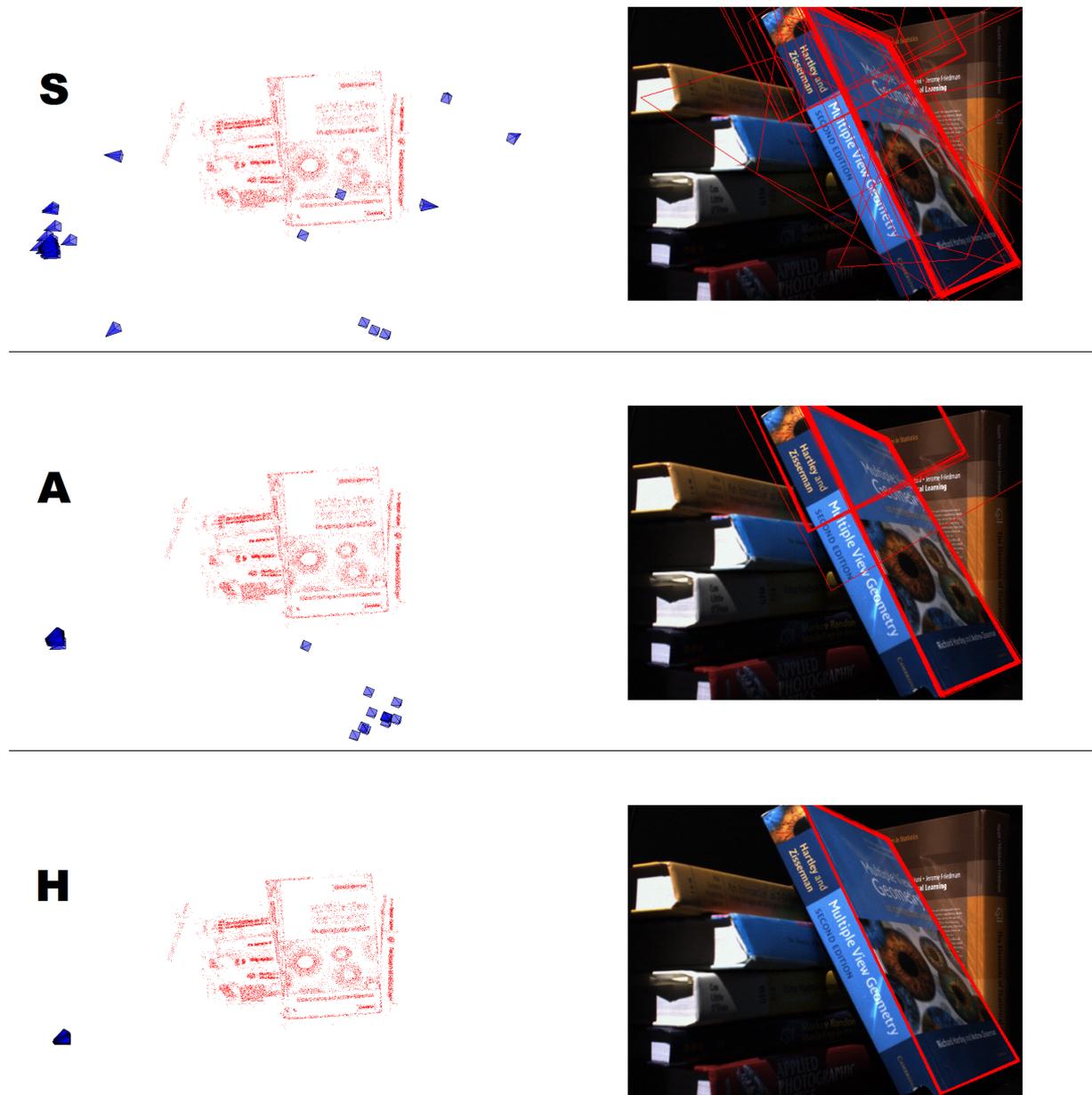


FIGURE 3.7 – Séquence Livre : 100 poses calculées avec $N = 1\ 000$ itérations de RANSAC, et la reprojection des bords de la couverture en utilisant ces 100 poses. Dans le scénario **H** l'écart type de la position de la caméra est de 0,29 % de la distance à la scène.

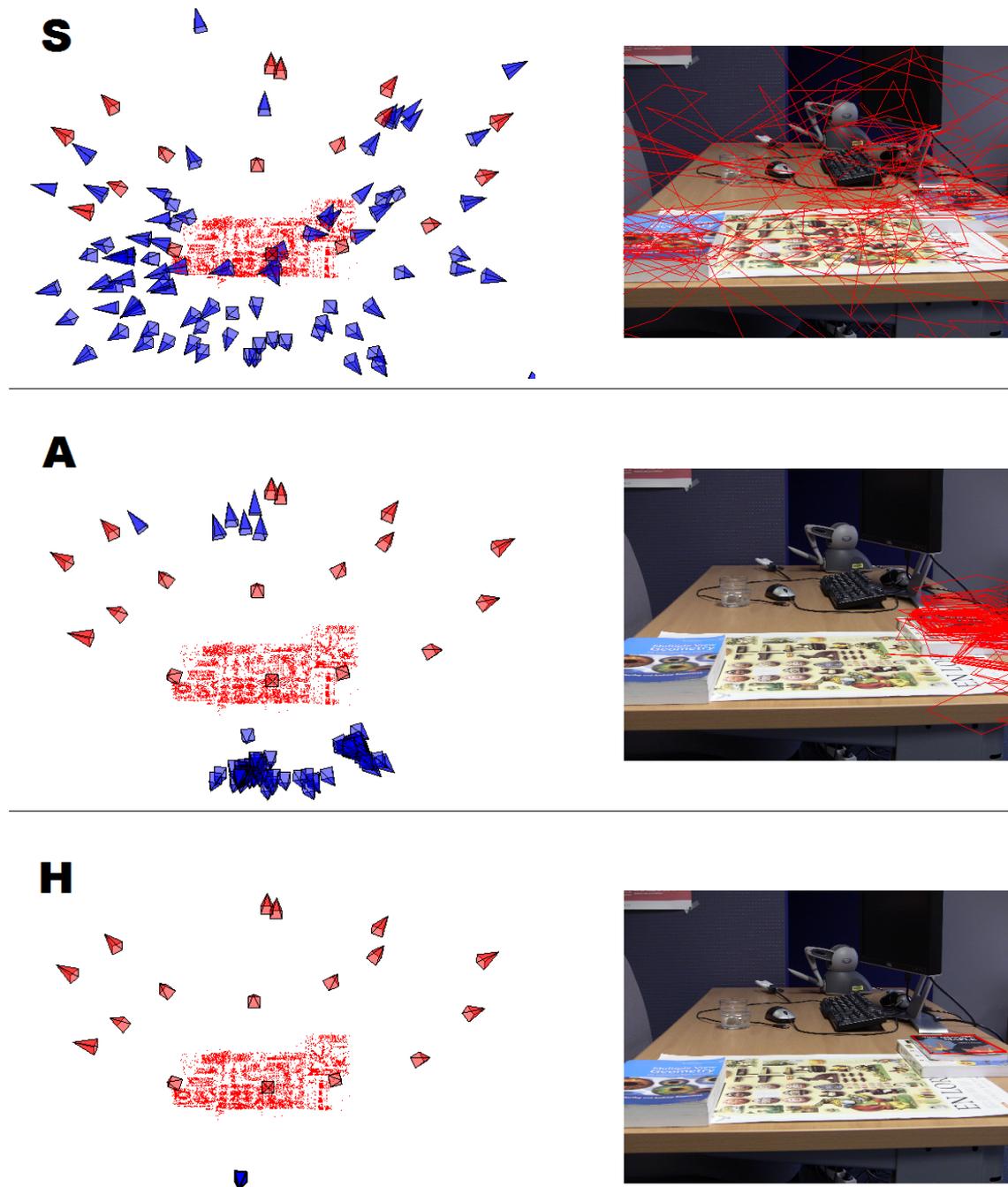


FIGURE 3.8 – Séquence Poster : 100 poses calculées avec $N = 1\,000$ itérations de RANSAC, et la reprojection des bords du livre en utilisant ces 100 poses. Dans le scénario **H** l'écart type de la position de la caméra est de 0,07 % de la distance à la scène.

	Livre	Poster
(1) nb de vues réelles	53	17
(2) nb de points 3D	15 269	7 552
(3) nb de descripteurs	225 207 / 403 662 / 386 970	47 643 / 161 596 / 224 923
(4) temps de calcul (s)	76,7 / 82,4 / 81,4	70,2 / 99,5 / 120,8
(5) nb de correspondances	1 272 / 809 / 1 097	1 144 / 1 293 / 1 092
	Bureau	Mur
(1) nb de vues réelles	17	6
(2) nb de points 3D	3 525	2 527
(3) nb de descripteurs	15 109 / 33 396 / 45 393	10 765 / 59 325 / 61 690
(4) temps de calcul (s)	11,0 / 16,9 / 22,3	3,0 / 10,2 / 10,2
(5) nb de correspondances	892 / 779 / 657	322 / 338 / 266

TABLE 3.1 – Nombre de vues réelles (1), nombre de points 3D dans le modèle SfM (2), nombre de descripteurs dans les scénarios **S/A/H** (3), temps de calcul en secondes pour la mise en correspondance image/modèle (4), nombre de correspondances (5).

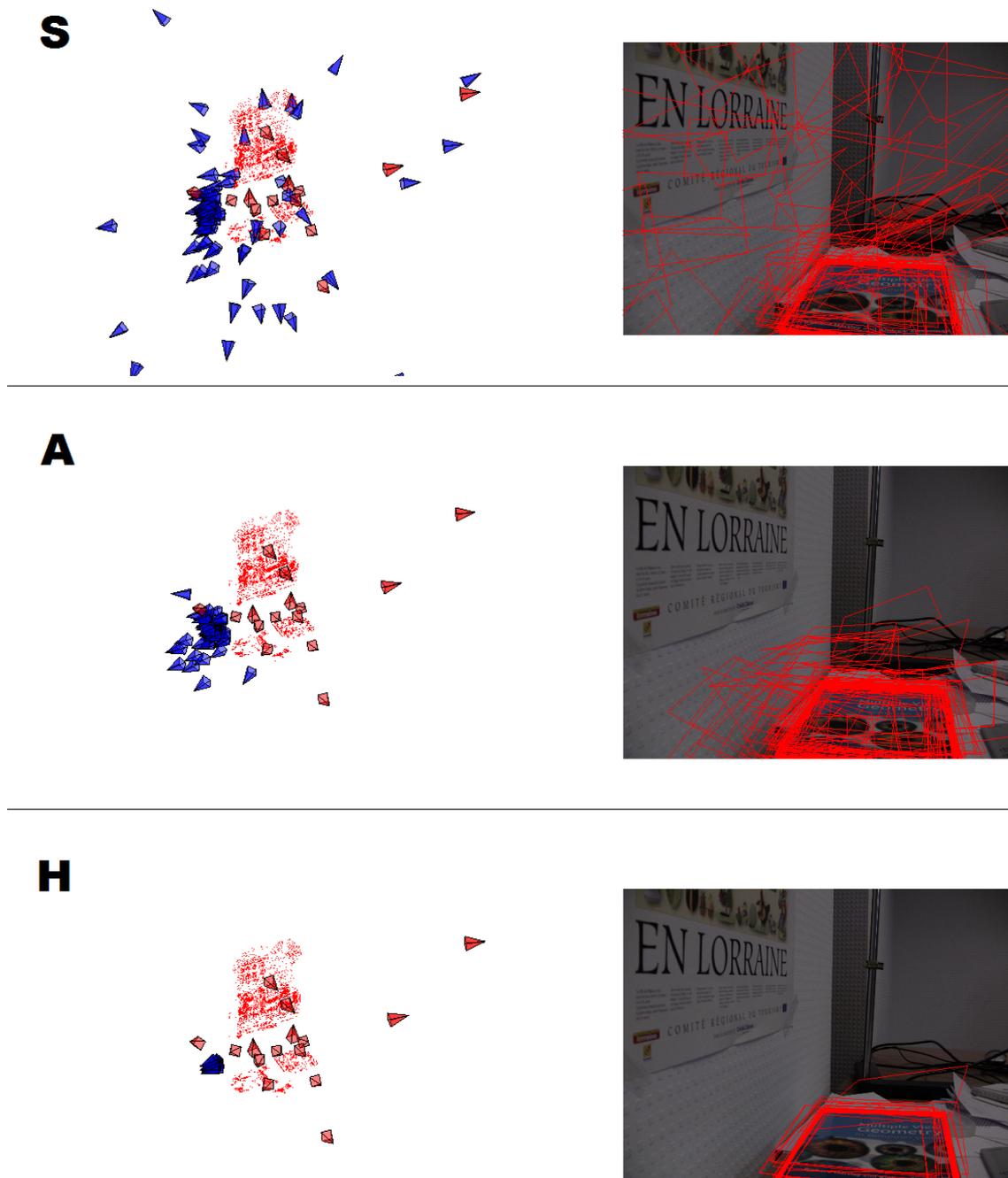


FIGURE 3.9 – Séquence Bureau : 100 poses calculées avec $N = 5\,000$ itérations de RANSAC, et la reprojection des bords du livre de droite en utilisant ces 100 poses. Dans le scénario **H** l'écart type de la position de la caméra est de 3,04 % de la distance à la scène.

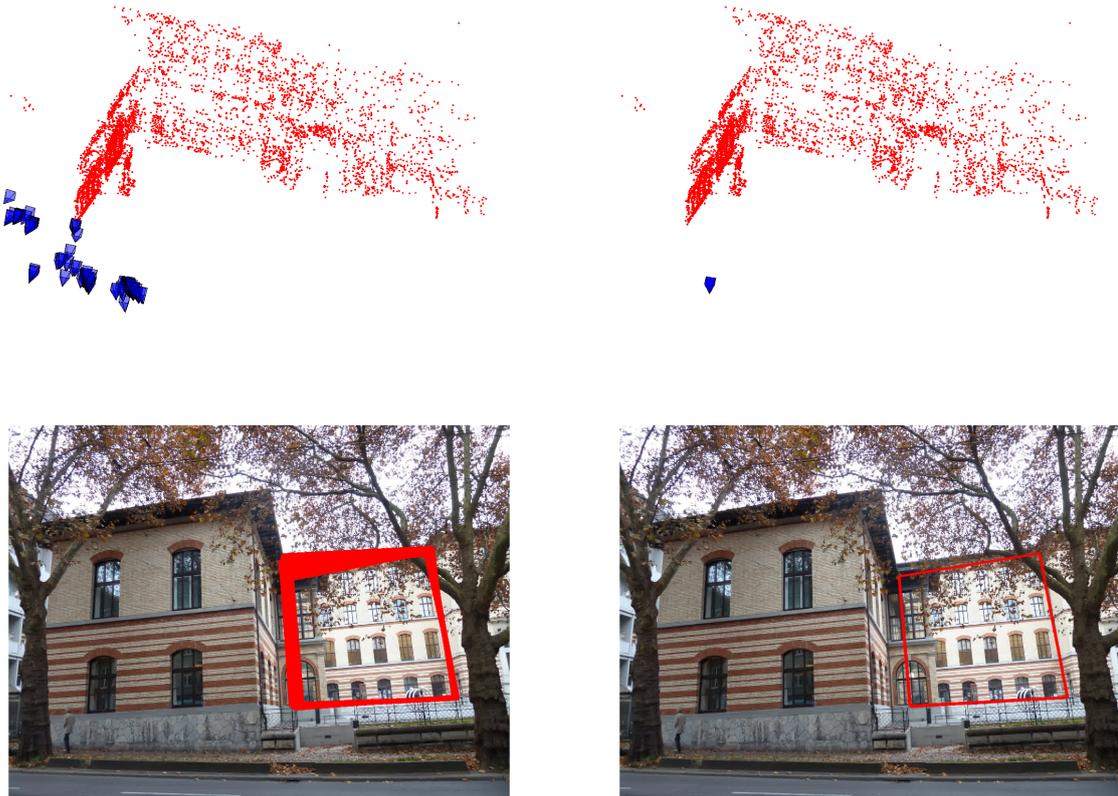


FIGURE 3.10 – 100 poses calculées sur la séquence venant de [CZSP12] à l'aide de RANSAC sans simulation (à gauche) et par notre approche avec simulation (à droite). Les poses calculées figurent sur la première ligne. Les contours de la façade en arrière-plan obtenus dans chaque cas sont superposés à l'image test dans les illustrations de la seconde ligne. On observe une réduction significative de la variabilité de la pose en utilisant notre méthode (les 100 poses calculées sont quasiment toutes superposées), même si dans tous les cas l'erreur de reprojection des points 3D appariés dans l'image test est assez faible.

3.4 Conclusion et perspectives

Dans ce chapitre nous avons résumé nos résultats concernant l'estimation de la pose d'une caméra à partir d'un modèle non structuré de l'environnement. Comme on l'a vu, enrichir le modèle par des descripteurs simulés permet d'améliorer le calcul de pose dans le cas des forts changements de point de vue.

Les travaux en cours de finalisation dans la thèse de Pierre Rolin concernent essentiellement deux axes.

En premier lieu, le fort taux de correspondances erronées parmi les correspondances provisoires (approximativement 90 à 50%, d'autant plus grand dans le cas où la simulation est utilisée) fait que l'algorithme RANSAC nécessite un grand nombre d'itérations. Nous travaillons sur une accélération de RANSAC basé sur un échantillonnage progressif des correspondances provisoires, de manière à commencer par celles ayant de meilleures chances d'être correctes. L'accélération de RANSAC est principalement possible par deux voies. Une possibilité est d'éviter de tester les hypothèses générées sur l'ensemble des correspondances provisoires par une approche « préemptive » comme dans [MC04, Nis05]. Néanmoins, ces approches ne semblent pas adaptées à des situations à faible taux de correspondances correctes. Une autre possibilité est de générer les hypothèses prioritairement parmi des sous-ensembles qui contiennent un plus fort taux d'*inliers*. Ceci est possible si une métrique est disponible pour classer les correspondances des plus vraisemblables au moins vraisemblables. Il s'agit de l'idée de base de PROSAC [CM05] : dans le contexte de l'appariement de points entre deux images, il est proposé d'utiliser le rapport des distances entre la plus proche correspondance et la seconde plus proche correspondance comme critère de correction. Une autre approche est discutée dans [LSHF12], à savoir la co-occurrence de descripteurs dans les images. Nos premiers résultats montrent qu'un algorithme similaire à PROSAC permet une accélération significative de la recherche d'un ensemble de consensus permettant une estimation précise de la pose.

D'autre part, le placement des caméras virtuelles décrit dans la figure 3.1 n'est possible que lorsque la scène est de taille relativement limitée (à quelques objets par exemple) : une telle approche n'est plus possible dans un scénario où la scène est un bâtiment tout entier par exemple. Nous étudions le placement des caméras virtuelles sur des hémisphères centrés sur des morceaux de plans de la scène. D'autre part, la segmentation du modèle en parties planes, associée à des contraintes de visibilité, permet d'accélérer fortement la simulation en procédant à une simulation des morceaux de plans visibles dans leur ensemble, au lieu de simuler une image pour chaque position de caméra virtuelle et chaque point 3D comme décrit dans la section 3.3.3.1. Il s'agit de la voie explorée dans [22].

Deuxième partie

Contributions à l'étude de deux types de bruit dans les images numériques

Estimation des paramètres d'un capteur numérique

4.1	Introduction	61
4.2	Modélisation du bruit en présence d'une illumination scintillante et de micro-vibrations	65
4.3	Estimation des paramètres de la caméra	67
4.4	Résultats expérimentaux	69
4.5	Estimation du bruit d'un capteur par analyse d'une série d'images de grilles pseudo-périodiques affectée par des vibrations	74
4.6	Conclusion et perspectives	79

4.1 Introduction

4.1.1 Contexte

Le bruit est un phénomène inhérent à tout capteur d'image numérique. S'il est crucial de bien le caractériser pour régler les paramètres d'algorithmes de débruitage (comme expliqué par exemple dans les revues [LCBM12] ou [Mil13]), il est également important de caractériser le bruit pour quantifier les performances métrologiques de mesures sans contact. Le bruit se traduit en effet comme une incertitude sur le niveau de gris, qui se propage en une incertitude sur la quantité mesurée à partir de l'image acquise. Comme nous le verrons plus précisément au chapitre 6 dans le cadre de la mécanique des solides expérimentale, il est important d'établir la résolution dans les cartes de déplacement ou de déformation obtenues par mesure de champ sans contact, et cette résolution est essentiellement limitée par le bruit dans les images (voir par exemple [12], [HR12, MN13]).

4.1.1.1 Modèle Poisson-Gauss

Dans le cadre de la sortie « brute » (*raw image*) d'un capteur CCD ou CMOS (c'est-à-dire sans démosaïçage pour former une image couleur, correction γ , ou compression avec perte d'information comme JPEG), les auteurs de [HK94] ont identifié quatre grandes sources de bruit, à savoir :

- le courant d'obscurité résiduel (*dark current*), qui dépend de la température ambiante et de la durée de l'exposition ;
- le bruit dit « de grenaille » (*shot noise*), qui modélise l'incertitude quantique du décompte des électrons générés par les photons incidents en un photosite ;
- le bruit de lecture (*read-out noise*) généré par l'amplification du signal électrique ;
- le bruit de quantification induit par la conversion du signal analogique en un signal numérique.

Un modèle stochastique couramment utilisé pour le niveau de gris mesuré en un pixel est le suivant [ADGM13, BKB⁺10, FM06, FTKE08, HK94] :

$$u(x, y) = g \eta_{p(x,y)+d(x,y)}(x, y) + \delta(x, y) \quad (4.1)$$

où :

- $u(x, y)$ est le niveau de gris mesuré au photosite correspondant au pixel (x, y) dans l'image brute en sortie ;
- $g > 0$ est le gain du système électronique ;
- le nombre d'électrons $\eta_{p(x,y)+d(x,y)}$ générés en (x, y) est une variable aléatoire de Poisson, de moyenne $p(x, y) + d(x, y)$, $p(x, y)$ dénotant le nombre d'électrons générés par les photons incidents (en proportion égale à l'efficacité quantique), et $d(x, y)$ est le nombre d'électrons dus au courant d'obscurité ;
- $\delta(x, y)$ est une variable aléatoire gaussienne de moyenne μ (un décalage fixe imposé par le fabricant du capteur) et de variance σ^2 , cette variable modélisant les différentes sources de bruit annexes comme les bruits de lecture et de quantification ;
- les variables aléatoires $\eta_{p(x,y)+d(x,y)}(x, y)$ et $\delta(x, y)$ sont statistiquement indépendantes entre elles, et spatialement indépendantes.

Le bruit affectant le niveau de gris est donc hétéroscédastique (sa variance n'est *a priori* pas constante) et dépend du signal sous-jacent.

Dans cette formulation, l'amplitude maximale, due à la capacité limitée de chaque photosite, et la nature discrète du niveau de gris $u(x, y)$ sont ignorées.

Dans les conditions expérimentales « standards » (température inférieure à 20°C et durée de la pose égale à une fraction de seconde), le courant d'obscurité est négligeable [ADGM13, HK94], raison pour laquelle dans tout ce chapitre on considérera $d = 0$.

Dans le modèle de (4.1), il est facile de calculer l'espérance et la variance du niveau de gris mesuré [BKB⁺10, DLR⁺08, FM06, FTKE08, RVU08] :

$$\begin{cases} E(u(x, y)) = gp(x, y) + \mu \\ \text{Var}(u(x, y)) = g^2p(x, y) + \sigma^2 \end{cases} \quad (4.2)$$

où E et Var dénotent respectivement espérance et variance. Il s'ensuit que la variance du bruit est affinement liée à l'espérance du niveau de gris par la relation :

$$\text{Var}(u(x, y)) = gE(u(x, y)) + \sigma^2 - g\mu \quad (4.3)$$

4.1.1.2 Estimation des paramètres du capteur

Estimer l'ensemble des paramètres du modèle relatif au capteur (g, μ, σ) peut être effectué par l'intermédiaire de la méthode dite de *transfer of photons* [emv10, Hyt06]. Néanmoins,

cette méthode requiert un protocole expérimental rigoureux et plusieurs acquisitions de données, ici des images de mires uniformes éclairées à l'aide d'une source lumineuse à intensité réglable.

L'estimation de la pente g et de l'ordonnée à l'origine $\sigma^2 - g\mu$ dans l'équation (4.3) est néanmoins suffisante dans de nombreuses applications, par exemple pour la stabilisation de la variance du bruit par transformée d'Anscombe généralisée [MF13, MSB95] (après [Ans48]) ou l'identification du modèle de caméra [TCR14] pour ne citer que deux exemples.

Des méthodes d'estimation de g et de $\sigma^2 - g\mu$ simples à mettre en œuvre sont décrites dans divers articles. L'idée sous-jacente est la construction de différents couples de valeurs $(E(u), \text{Var}(u))$ suivie d'une estimation des paramètres de la relation affine par régression. Les couples espérance / variance peuvent être estimés sur des régions d'images homogène [DLR⁺08, FTKE08, RVU08] résultant d'une segmentation, ou bien par la sélection de petits images (*patches*) homogènes [BKB⁺10, LTO14]. Bien sûr, ces approches ne peuvent fonctionner lorsque l'image considérée est très texturée et présente peu de régions homogènes. Dans ce cas, les auteurs de [UVLC13] modélisent la texture comme un Brownien fractionnaire bidimensionnel et proposent un estimateur des paramètres du capteur. La méthode développée dans [FTKE08] a été également rendue davantage robuste aux textures de haute-fréquence dans [AF14a]. D'autre part, une estimation jointe des paramètres du bruit et de la segmentation par l'intermédiaire d'un algorithme EM est proposée dans [JPTC14]. Les auteurs de [AF14b] modélisent des *patches* arbitraires comme des mélanges de gaussiennes et se passent de la sélection de *patches* homogènes. Une idée récente permet de se passer de l'estimation par régression, en estimant les paramètres qui donneront les meilleures performances à une transformation de stabilisation de la variance du bruit, sous hypothèse d'un bruit de Poisson dans [JXH14] ou d'un bruit de Poisson-Gauss dans [MF14, PH14]. Lorsque l'image disponible a subi par exemple une compression JPEG ou une correction γ , le modèle de (4.3) n'est plus valable ; on peut alors estimer une fonction du niveau de bruit non paramétrique, comme les auteurs de [LSK⁺08] ou [CBM14]).

Les travaux de ce chapitre sont effectués sous hypothèse du modèle de Poisson-Gauss. Nous nous sommes intéressés à l'estimation des paramètres du capteur lorsque l'on dispose d'une séquence d'images prise avec des réglages de caméra identiques (même durée d'exposition, ouverture, et sensibilité). Une manière simple d'estimer l'espérance et la variance du niveau de gris serait d'utiliser des estimateurs temporels. Il s'avère néanmoins que les fluctuations aléatoires du niveau de gris $u(x, y, t)$ au cours du temps ne sont pas uniquement causées par le bruit du capteur. En premier lieu, des micro-vibrations sont difficilement évitables : la dalle en béton formant le sol dans un bâtiment est soumise à des vibrations d'amplitude non négligeable [ET84], et le mouvement du miroir dans un appareil photographique de type *reflex* ou le mécanisme de refroidissement d'une caméra professionnelle induisent également des vibrations. La turbulence atmosphérique est également une autre source potentielle d'instabilité de la scène imagée. La seconde source de fluctuation du niveau de gris est le phénomène de *flickering* de la source lumineuse, que nous traduirons par *scintillement*. Les lumières artificielles sont particulièrement sujettes à ce phénomène [CCM⁺09], de par la nature alternative du courant domestique, en particulier les tubes à néon [Loh33] ou les lampes à diodes électroluminescentes [jee10]. Dans le cas d'un appareil photographique *reflex*, une légère variabilité de la durée d'obturation cause un effet équivalent. Ces phénomènes sont illustrés par la figure 4.1. Les micro-vibrations mécaniques comme le scintillement de la source lumineuse empêchent donc d'estimer simplement les paramètres de la relation affine (4.3) par le simple calcul de la moyenne et variance temporelles en chaque pixel.

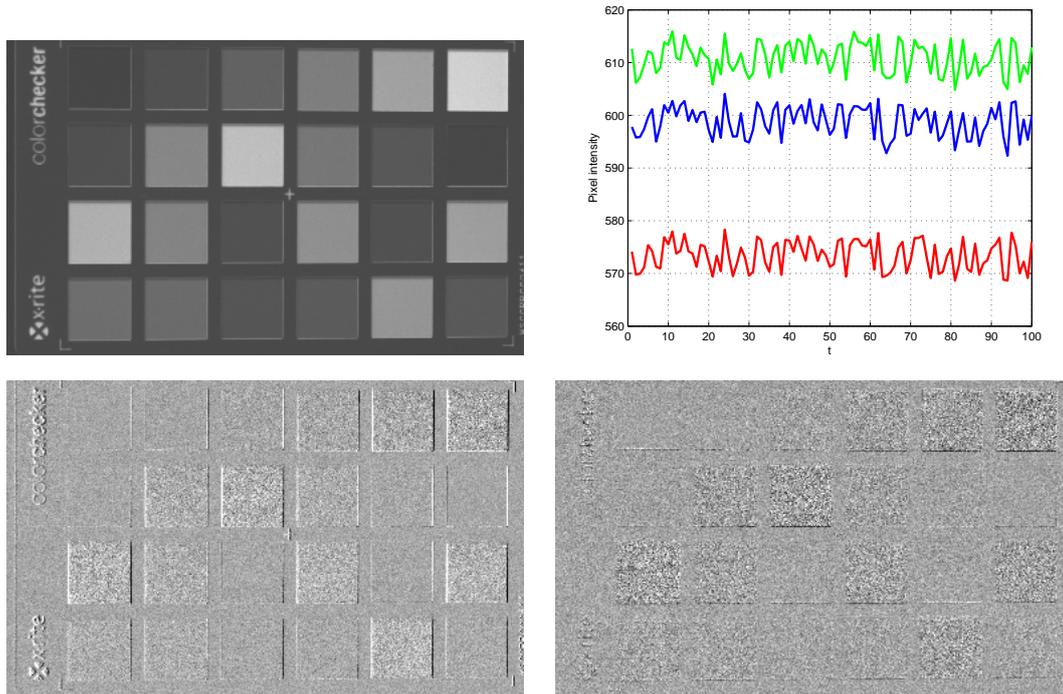


FIGURE 4.1 – Une mire de calibration X-rite Colorchecker, éclairée par un tube au néon, est photographiée par un appareil photographique Canon EOS 350D (sensibilité ISO 100), muni d'un objectif de 50mm et fixé à un trépied. Une rafale de 100 images est prise. En haut à gauche : une des images brutes de la série (canal « rouge »). En haut à droite : un niveau de gris moyen est calculé sur trois carrés de taille 10×10 appartenant à différents patches uniformes de la grille, la moyenne ayant pour effet d'atténuer les fluctuations aléatoires dues au bruit. Le graphique montre l'évolution de l'intensité moyenne de ces carrés. Les fluctuations en phase entre les différents patches mettent en évidence le scintillement de l'illumination. En bas : deux images correspondant à la différence entre deux paires de photographies de la mire. En l'absence de vibrations, ces deux différences devraient être semblables et n'être constituées que d'un bruit de variance proportionnelle au niveau de gris du pixel. Ici, on peut voir les contours des patches, dans une direction puis dans l'autre : ils sont causés par les micro-vibrations qui se traduisent par un mouvement apparent de la mire.

4.1.2 Contributions

Ce chapitre traite de l'influence des micro-vibrations mécaniques et du scintillement de la source lumineuse sur l'estimation des paramètres du capteur. Un modèle théorique de la formation d'une rafale d'images sous ces hypothèses est présenté en section 4.2. Un algorithme d'estimation de chacun des paramètres g, μ, σ , ainsi que de l'amplitude des vibrations, est expliqué en section 4.3. L'algorithme est uniquement basé sur des estimateurs statistiques et ne requiert pas de recalage permettant de compenser les vibrations ni de segmentation préalable. Une sélection de résultats expérimentaux est présentée en section 4.4. Ces résultats sont discutés en détail dans [9] ; des expériences additionnelles sont présentées dans un rapport de recherche compagnon [50], et une présentation simplifiée est disponible dans [27]. Enfin, la section 4.2 présente une estimation des paramètres de la relation affine (4.3) robuste aux vibrations, dans le cas de séries d'images d'une grille pseudo-périodique. La motivation de cette dernière étude était la détermination des paramètres de la caméra sans avoir à modifier

le protocole expérimental dans le cadre d'une application à la mécanique des solides, qui fera l'objet du chapitre 6. Les détails de cette dernière application peuvent être trouvés dans [11].

Les principales publications relatives à ce chapitre sont :

[9] F. Sur and M. Grédiac. Measuring the noise of digital imaging sensors by stacking raw images affected by vibrations and illumination flickering. *SIAM Journal on Imaging Sciences*, 8(1):611–643, 2015.

[11] F. Sur and M. Grédiac. Sensor noise modeling by stacking pseudo-periodic grid images affected by vibrations. *IEEE Signal Processing Letters*, 21(4):432–436, 2014.

[27] F. Sur and M. Grédiac. Sensor noise measurement in the presence of a flickering illumination. In *Proceedings of the International Conference on Image Processing (ICIP)*, pages 1763–1767, Paris, France, 2014.

Le code Matlab NESIF (Noise Estimation by Stacking Images affected by illumination Flickering) disponible à l'URL :

<https://members.loria.fr/FSur/software/NESIF/>
implémente les algorithmes décrits dans [9] et [27].

4.2 Modélisation du bruit en présence d'une illumination scintillante et de micro-vibrations

Le but est de modéliser une série de T images d'une scène statique, éclairée par une illumination scintillante, la caméra étant soumise à des micro-vibrations.

4.2.1 Modèle cox-gaussien

Le nombre d'électrons générés par les photons étant proportionnel à l'éclairement, le nombre moyen de ces électrons générés en un photosite (x, y) lors de l'acquisition de la t -ème image peut s'exprimer comme :

$$(1 + \gamma_t)p(x + \alpha_t, y + \beta_t) \quad (4.4)$$

où :

- (α_t, β_t) est le vecteur de translation modélisant les déplacements sur le plan du capteur causés par les vibrations entre une image de référence et la t -ème image ;
- $1 + \gamma_t$ est la fluctuation relative de l'intensité lumineuse.

Remarquons que nous ne modélisons pas le flou de bougé, qui nécessiterait d'intégrer p au cours du temps.

Les hypothèses de la section 4.1.1.1 permettent d'exprimer le niveau de gris $u(x, y, t)$ du pixel (x, y) de la t -ème image comme :

$$u(x, y, t) = g \eta_{(1+\gamma_t)p(x+\alpha_t, y+\beta_t)}(x, y, t) + \delta(x, y, t) \quad (4.5)$$

Les deux processus stochastiques γ_t et (α_t, β_t) sont considérés indépendants et identiquement distribués, de moyenne nulle. La variance de γ est notée σ_γ^2 et la matrice de covariance (de taille 2×2) de (α_t, β_t) est notée $\text{Cov}(\alpha, \beta)$.

Dans un tel modèle, le paramètre de la variable de Poisson est lui-même aléatoire. Le processus $\eta_{(1+\gamma_t)}$ est appelé *processus de Poisson doublement stochastique*, ou processus de

Cox [Cox55, DVJ03]. Le niveau de gris $u(x, y, t)$ est donc la réalisation d'un processus cox-gaussien.

Les formules d'espérance et de variance totale permettent alors de calculer l'espérance et la variance de u en tout pixel :

$$\begin{cases} E(u(x, y, t)) &= gE(p(x + \alpha_t, y + \beta_t)) + \mu \\ \text{Var}(u(x, y, t)) &= g^2E(p(x + \alpha_t, y + \beta_t)) + g^2(1 + \sigma_\gamma^2)E(p^2(x + \alpha_t, y + \beta_t)) \\ &\quad - g^2(E(p(\cdot + \alpha_t, \cdot + \beta_t)))^2 + \sigma^2 \end{cases} \quad (4.6)$$

On peut voir que le terme $E(p^2(x + \alpha_t, y + \beta_t))$ ne nous permet pas d'écrire simplement $\text{Var}(u)$ en fonction de $E(u)$. Néanmoins, un développement en série de Taylor donne :

$$\begin{aligned} p(x + \alpha_t, y + \beta_t) &= p(x, y) + (\alpha_t, \beta_t)\nabla p(x, y) \\ &\quad + \frac{1}{2}(\alpha_t, \beta_t)H_p(x, y)(\alpha_t, \beta_t)^T + R_p(x, y, \alpha_t, \beta_t) \end{aligned} \quad (4.7)$$

où ∇ est le gradient, $H_p(x, y)$ la matrice hessienne de p en (x, y) , et le reste R_p est un polynôme multivarié en (α_t, β_t) avec des monômes de degré total supérieur ou égal à 3.

Après calculs, on obtient :

$$\begin{aligned} \text{Var}(u) &= \sigma_\gamma^2 E(u)^2 + (g - 2\mu\sigma_\gamma^2)E(u) + \sigma^2 - g\mu + \sigma_\gamma^2\mu^2 \\ &\quad + (1 + \sigma_\gamma^2)\nabla E(u)^T \text{Cov}(\alpha, \beta)\nabla E(u) + M_p(x, y, \alpha_t, \beta_t) \end{aligned} \quad (4.8)$$

avec un terme M_p négligeable devant le terme en $\text{Cov}(\alpha, \beta)$.

Ainsi, la relation affine (4.3) devient, à cause du scintillement de l'illumination, une relation quadratique, et les micro-vibrations ajoutent un terme additionnel, bien sûr positif.

Lorsque le gradient est négligeable, la relation se simplifie en :

$$\text{Var}(u) = \sigma_\gamma^2 E(u)^2 + (g - 2\mu\sigma_\gamma^2)E(u) + \sigma^2 - g\mu + \sigma_\gamma^2\mu^2 \quad (4.9)$$

Bien sûr, $\sigma_\gamma = 0$ simplifie la relation quadratique (4.9) en la relation affine classique décrite par l'équation (4.3).

Même dans le cas où l'illumination ne scintille pas, la relation (4.8) montre que, en présence de vibrations, la variance est toujours plus élevée que ce qui est attendu.

L'espérance et la variance empiriques estimées par moyennage temporel en tout pixel (x, y) sont des estimateurs non biaisés et consistants de $E(u)$ et de $\text{Var}(u)$.

4.2.2 Limites du modèle

Le développement de Taylor est valide sous hypothèse de mouvements faibles, typiquement des mouvements apparents sous-pixelliques sur le capteur. On peut remarquer que la relation (4.8) est valable localement : en d'autres termes le processus (α_t, β_t) peut dépendre de (x, y) . Si la partie non statique de la scène ne couvre qu'une partie limitée de l'image (par exemple un arbre se balançant sous le vent ou un passant traversant la scène) l'estimation de la majeure partie des couples variance / espérance empiriques ne sera pas perturbée. Néanmoins, en toute généralité les scènes non-statiques ne rentrent pas dans le cadre du modèle présenté ici.

La modélisation décrite dans la section 4.1.1.1 procède à un certain nombre de simplifications, en particulier concernant la non-uniformité des paramètres du capteur. En pratique,

le gain, l'efficacité quantique, le bruit de lecture varient de photosite à photosite, ce qui entraîne les phénomènes de DSNU (*dark signal non-uniformity*) et de PRNU (*photo response non-uniformity*) [emv10, HK94, Hyt06]. Ceci a pour effet d'introduire une variabilité supplémentaire (en plus de celle due à l'estimation sur un échantillon de taille finie) dans les nuages de points représentant les couples variance / espérance du niveau de gris.

4.3 Estimation des paramètres de la caméra

Dans l'article [9], nous avons proposé un algorithme permettant de tirer profit du scintillement de l'illumination pour estimer tous les paramètres du capteur, c'est-à-dire σ , g , et μ et pas simplement la pente g et l'ordonnée à l'origine $\sigma^2 - g\mu$ de la relation affine classique (4.3).

On suppose disposer d'une série d'images $u(x, y, t)$, affectée par un scintillement temporel de l'illumination, les images se formant selon le modèle stochastique décrit par (4.5). Les grandes étapes de l'algorithme sont les suivantes :

1. Élimination des pixels (x, y, t) pour lequel le gradient moyen est trop grand pour avoir été causé seulement par le bruit, de manière à éliminer les pixels dont l'estimation de la variance est biaisée par les vibrations. Ceci est fait par l'intermédiaire d'un test du χ^2 . Les pixels défectueux ou présentant une saturation du niveau de gris sont également éliminés. La variance théorique des pixels restant est donc supposée satisfaire l'équation (4.9).
2. Identification du paramètre de scintillement γ_t , et élimination de l'effet du scintillement dans la variance estimée en chaque pixel, de manière à se ramener à la relation affine (4.3). Cette étape centrale est détaillée dans la section 4.3.1.
3. Estimation des paramètres de bruit à l'aide d'une régression linéaire.

4.3.1 Estimation du scintillement relatif et élimination de son effet dans la variance

Soit $v(x, y, t)$ le rapport, à un pixel (x, y) donné, entre le niveau de gris $u(x, y, t)$ et le niveau de gris moyen $\bar{u}(x, y)$:

$$v(x, y, t) = \frac{u(x, y, t)}{\bar{u}(x, y)} \quad (4.10)$$

On peut alors simplifier les espérances conditionnelles :

$$\begin{cases} E(v | \gamma_t) = \frac{1 + \gamma_t + \mu/(gp)}{1 + \mu/(gp)} + \mathcal{O}(1/T) \\ \text{Var}(v | \gamma_t) = \frac{g(1 + \gamma_t)E(u) + \sigma^2 - g(1 + \gamma_t)\mu}{E(u)^2} + \mathcal{O}(1/T) \end{cases} \quad (4.11)$$

On peut voir de cette expression que $E(v | \gamma_t)$ fournit une estimation de $1 + \gamma_t/(1 + \mu/(gp))$ avec une variance minimale pour les grandes valeurs de niveau de gris moyen $E(u) = gp + \mu$. De plus, une estimation de $E(v(x, y, t) | \gamma_t)$ peut être obtenue en moyennant, pour chaque t , $v(x, y, t)$ sur un ensemble de pixels pour lesquels $\bar{u}(x, y)$ est constant.

Ceci motive la définition de l'ensemble

$$\mathcal{S}^* = \{(x, y), u^* - 3\sigma_{u^*} \leq \bar{u}(x, y) \leq u^* + 3\sigma_{u^*}\} \quad (4.12)$$

où u^* désigne le quantile à 99% dans la distribution empirique des niveaux de gris moyens $\bar{u} = gp + \mu + \mathcal{O}(1/T)$, et σ_{u^*} est l'écart-type associé. Une estimation de $E(v(x, y, t) | \gamma_t)$ est calculée comme une moyenne empirique sur l'ensemble \mathcal{S}^* .

Une fois que $v^*(t)$ est calculée, la quantité définie pour tout (x, y, t) par $\tilde{u}(x, y, t) = u(x, y, t) - v^*(t)\bar{u}(x, y)$ satisfait :

$$\begin{cases} E(\tilde{u}(x, y, t) | \gamma_t) = -\gamma_t \frac{\mu}{u^*} (u^* - \mu - gp(x, y)) \\ \text{Var}(\tilde{u}(x, y, t) | \gamma_t) = g^2(1 + \gamma_t)p(x, y) + \sigma^2 \end{cases} \quad (4.13)$$

Soit $K = E(\tilde{u}(x, y, t) | \gamma_t) / (E(v^*(t)) - 1)$. Après calcul, on établit l'expression suivante :

$$K = -\mu \left(1 - \frac{gp(x, y)}{u^* - \mu} \right) \quad (4.14)$$

On voit que K ne dépend pas de t . Pour $p(x, y)$ fixé, il y a une relation de proportionnalité entre $E(\tilde{u}(x, y, t) | \gamma_t)$ et $E(v^*(t) - 1)$ pour tout t . Le coefficient de proportionnalité K peut être estimé par régression linéaire, de la manière suivante. De la même manière que pour l'estimation de $E(v^*(t))$, on remarque de (4.13) que la variance $\text{Var}(\tilde{u}(x, y, t) | \gamma_t)$ est minimale pour les petites valeurs $p(x, y)$. On définit donc l'ensemble :

$$\mathcal{S}_* = \{(x, y), u_* - 3\sigma_{u_*} \leq \bar{u}(x, y) \leq u_* + 3\sigma_{u_*}\} \quad (4.15)$$

où u_* est le quantile à 1% de la distribution des niveaux de gris moyen $\bar{u}(x, y)$, et σ_{u_*} est l'écart-type associé. Une estimation de $E(\tilde{u}(x, y, t) | \gamma_t)$ pour les pixels (x, y) tels que $gp(x, y) = u_* - \mu$ est donc donnée par :

$$v_*(t) = \frac{1}{\#\mathcal{S}_*} \left(\sum_{(x, y) \in \mathcal{S}_*} \tilde{u}(x, y, t) \right) \quad (4.16)$$

Une régression linéaire sur le nuage des points $(v^*(t) - 1, v_*(t))$ donne alors une estimation de la pente K définie par (4.14).

Un simple calcul montre alors la relation :

$$\mu = \frac{Ku^*}{K - u^* + u_*} \quad (4.17)$$

Il est donc possible de déduire de l'estimation de K une estimation de μ .

On montre alors que

$$\tilde{\gamma}(t) = (v^*(t) - 1) \left(1 + \frac{\mu}{u^* - \mu} \right) \quad (4.18)$$

est une estimation de γ_t pour chaque t .

Par conséquent, on définit :

$$\hat{u}(x, y, t) = u(x, y, t) - (1 + \tilde{\gamma}_t)(\bar{u}(x, y) - \mu) \quad (4.19)$$

qui satisfait :

$$\begin{cases} E(\hat{u}) = \mu \\ \text{Var}(\hat{u}) = gE(u) + \sigma^2 - g\mu + \mathcal{O}(1/T) \end{cases} \quad (4.20)$$

Cette expression de la variance est la même que dans le modèle affine standard (4.3). Il est donc possible d'estimer g et $\sigma^2 - g\mu$ par régression linéaire, puis σ^2 connaissant μ .

4.3.2 Estimation de l'amplitude des vibrations

Les pixels affectés par les vibrations sont éliminés de l'estimation précédente. Si on les réintègre, on remarque alors qu'une régression linéaire multiple entre la variance empirique \bar{u}^2 , \bar{u} , $(\partial\bar{u}/\partial x)^2$, $(\partial\bar{u}/\partial y)^2$, and $\partial\bar{u}/\partial x \cdot \partial\bar{u}/\partial y$ dans l'équation (4.8) fournit une estimation de (α_t, β_t) pour tout t .

4.3.3 L'effet « obturateur déroulant » (*rolling shutter*)

L'obturateur d'un appareil photographique reflex est constitué de deux rideaux parcourant le capteur, l'espace entre les rideaux permettant d'éclairer une partie du capteur seulement, la distance entre les rideaux permettant de contrôler la durée d'obturation. Lorsque la durée d'exposition est inférieure au temps de *synchro-X*, le deuxième rideau commence sa course avant que le premier rideau l'ait terminée, et tous les photosites du capteur ne sont alors pas forcément éclairés simultanément. Ceci provoque l'effet *rolling shutter*. Notons que les caméras basées sur un capteur CMOS peuvent être sujettes à cet effet même en l'absence d'obturateur mécanique, l'image étant acquise ligne à ligne. Si la source lumineuse scintille à une fréquence suffisamment élevée, le paramètre γ_t ne peut plus être considéré être uniforme spatialement.

L'effet *rolling shutter* ne change pas la relation quadratique de (4.9) qui reste valable en tout point, mais elle rend impossible l'estimation de γ_t proposée dans la section 4.3.1, car dans cette estimation on suppose que γ_t est uniforme dans les ensembles \mathcal{S}^* et \mathcal{S}_* . Nous proposons dans [9] une localisation $\gamma_t(y)$ de cette estimation, en nous basant sur une bande de pixels entourant la ligne y parallèle à la fente inter-rideaux. Notons que fixer précisément la largeur de la bande nécessiterait de connaître la vitesse des rideaux, le temps d'exposition, ainsi que la taille du capteur.

4.4 Résultats expérimentaux

Nous proposons dans cette section une sélection de quelques résultats expérimentaux parmi les expériences sur des séries d'images synthétiques ou réelles montrés dans [9]. Des expériences additionnelles sont également disponibles dans le rapport de recherche [50].

Dans tous les cas, on suppose disposer d'une série de T images d'une scène quasi-statique (aux micro-vibrations près), acquise à paramètres d'ouverture de diaphragme, de durée d'exposition, et de réglage de sensibilité constants. L'algorithme proposé permet d'estimer la matrice de covariance du processus (α_t, β_t) représentant les vibrations, de γ_t modélisant la fluctuation relative de l'illumination, et les paramètres du capteur g (gain), μ (décalage), σ (bruit de lecture). Une comparaison avec le logiciel de A. Foi¹ est également présentée. Ce logiciel implante l'algorithme décrit dans [FTKE08] et fournit une estimation de g et de $\sigma^2 - g\mu$. Il est basé sur l'analyse d'une image unique. Il n'est donc pas affecté par le scintillement de la source et les vibrations. Afin de fournir des estimations sur la même quantité de données, ce logiciel est lancé avec ses paramètres par défaut sur une image composite où les T images de la série forment une mosaïque de taille $\sqrt{T} \times \sqrt{T}$.

1. v. 2.31, disponible à l'URL <http://www.cs.tut.fi/~foi/sensornoise.html>

4.4.1 Exemple illustratif

La figure 4.2 montre un exemple illustratif de fonctionnement de notre algorithme. $T = 100$ images d'une mire X-Rite Colorchecker illuminée par un tube à néon sont prises avec un appareil photo reflex Canon EOS 6D. Il s'agit de l'expérience correspondant à la ligne "ISO 6400" dans la table 4.1, donc une vitesse d'obturation rapide de $1/320$ sec., inférieure à la vitesse de synchro-X et donc sujette à l'effet *rolling shutter*. Le nuage de points formés par les couples $(v^*(t)-1, v_*(t))$ pour $1 \leq t \leq T$ montre la relation de proportionnalité de pente $K < 0$. La relation de proportionnalité est quasiment parfaite en dépit de l'effet *rolling shutter*.

Le scintillement relatif $1 + \gamma_t(y)$ est estimé à l'aide de la méthode décrite dans la section 4.3.3 et on estime $\sigma_\gamma = 0.24$. La dépendance de $\gamma_t(y)$ à y peut être remarquée. Le graphe de la variance empirique de $u(x, y, t)$ contre la moyenne empirique montre la tendance quadratique prédite par (4.9). Le graphe de $\tilde{u}(x, y, t)$ n'est pas linéaire lorsque γ_t est supposé constant sur toute l'image : les pixels correspondants à certains *patches* de même couleur sur la mire ne sont pas corrigés correctement, ce qui explique les groupes de points en dehors de la tendance linéaire dans la partie supérieure du graphe. Dans ce cas, la droite de régression ne donne pas une pente et une ordonnée à l'origine cohérentes avec les gains, décalages, et niveaux de bruit de lecture attendus.

Avec une estimation effectuée selon le correctif décrit dans la section 4.3.3, la tendance devient bien linéaire et une estimation correcte des paramètres de la caméra est possible, comme on peut le constater dans les expériences de la section 4.4.2. Dans ces nuages de points, les points en vert sont des données aberrantes éliminés par la première étape de l'algorithme (pixels affectés par les vibrations), et l'estimation des paramètres de la caméra est basée sur les points bleus.

4.4.2 Estimation des paramètres de deux appareils photographiques

Deux appareils photographiques reflex de marque Canon (EOS 350D avec un capteur APS-C produisant des images 12 bits et EOS 6D avec un capteur 24×36 produisant des images 14 bits) prennent des séries d'images de deux scènes différentes illuminées par un tube à néon. Ces deux scènes correspondent respectivement à la mire X-Rite Colorchecker (présentant donc de larges régions homogènes) et à des objets sur un bureau (donc assez difficiles à segmenter), comme illustré sur la figure 4.3. Plusieurs sensibilités ISO sont testées ; les durées d'exposition correspondantes, qui dépendent de l'ouverture du diaphragme, sont également mentionnées.

La série de $T = 100$ images au format RAW est traitée linéairement avec le logiciel `dcraw`² et le canal de Bayer correspondant au Rouge est extrait. Les autres canaux donnent des résultats semblables, cf. [50].

Dans la table 4.1, on donne :

1. la valeur obtenue par la méthode de transfert de photon³ (le décalage μ est estimé indépendamment par de courtes expositions effectuées bouchon d'objectif fermé).
2. les estimations g et $\sigma^2 - g\mu$ du logiciel de Foi et al. ;
3. les résultats de la méthode proposée ;

2. available at <http://www.cybercom.net/~dcoffin/dcraw/>
Options : `dcraw -c -D -4 -r 1 1 1 1`

3. EOS 350D : from www.astrosurf.com/comolli/strum41.htm
EOS 6D : from www.astrosurf.com/comolli/strum54.htm

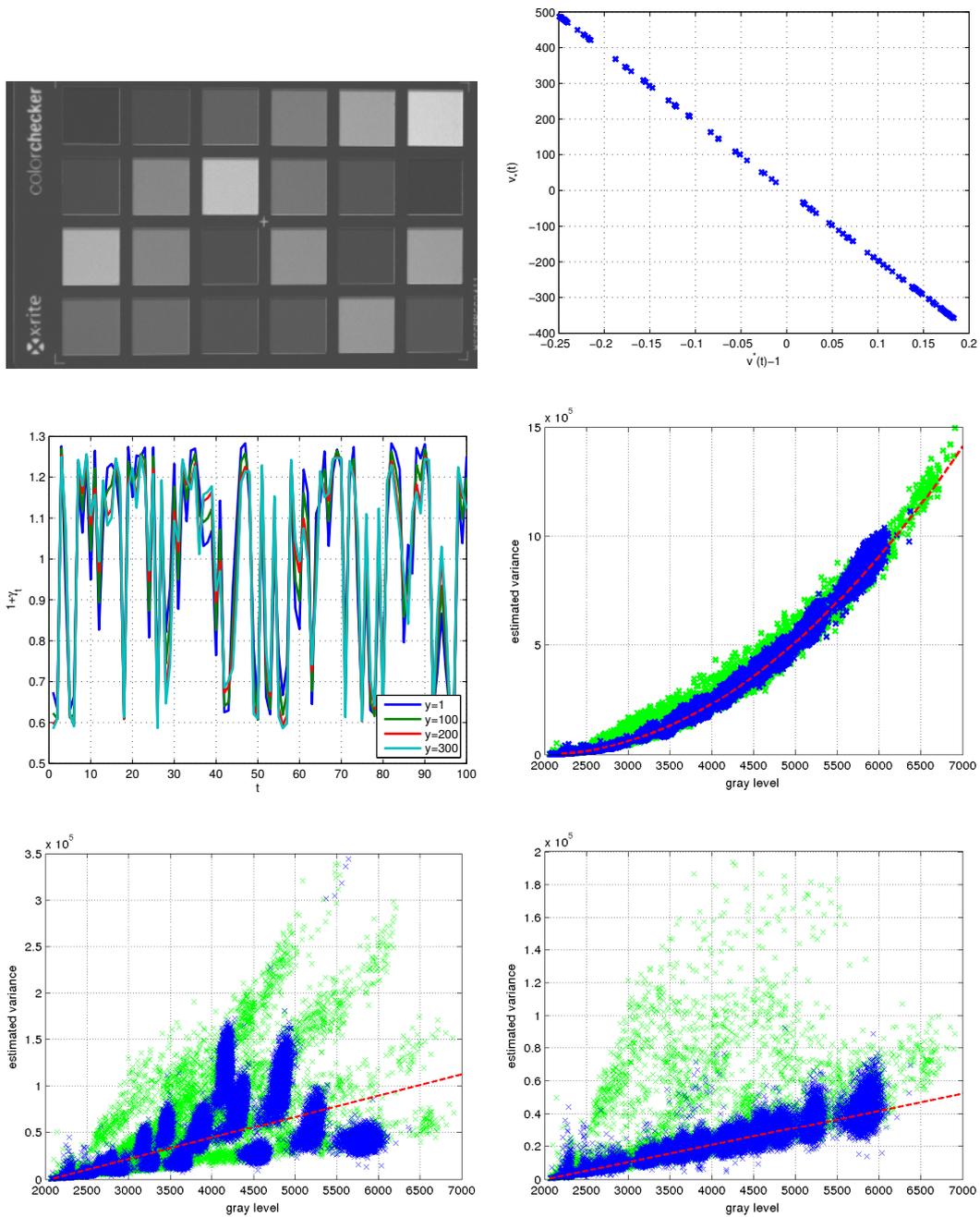


FIGURE 4.2 – Exemple illustratif. En haut : une image de la série, et le graphe de $v_*(t)$ contre $v^*(t)-1$. Au milieu : estimation de la fluctuation relative de l'illumination $1 + \gamma_t(y)$ pour différentes valeurs de y , et graphique de la variance empirique de $u(x, y, t)$ contre la moyenne empirique de $u(x, y, t)$. En bas : graphique de la variance empirique de $\hat{u}(x, y, t)$ contre la moyenne empirique de $u(x, y, t)$ (estimation de γ_t uniforme sur toute l'image, à gauche), et les mêmes statistiques où l'effet rolling shutter est pris en compte (à droite).



FIGURE 4.3 – À gauche : scène « de fouillis » pour l'expérience avec le Canon EOS 350D ; à droite : scène « de fouillis » pour l'expérience avec le Canon EOS 6D.

4. l'estimation de la pente g et de l'ordonnée à l'origine $\sigma^2 - g\mu$ d'une droite de régression sur $(E(u), \text{Var}(u))$, sans tenir compte de la tendance quadratique.

Les sensibilités ISO marquées d'une astérisque sont traitées avec la méthode décrit dans la section 4.3.3 pour tenir compte de l'effet *rolling shutter*.

Les résultats de la méthode de [FTKE08] suggère que les paramètres par défaut du logiciel ne sont pas toujours adéquats. La surestimation (attendue) de g par la quatrième méthode montre qu'il est important de corriger le scintillement, spécialement pour les grandes valeurs d'ISO donnant des vitesses d'obturations très rapides. On peut voir que la méthode proposée permet d'identifier correctement g , mais aussi μ et la plupart du temps également σ .

La table 4.2 montre l'estimation de la matrice de covariance du processus modélisant les vibrations, ainsi que de la variation relative d'illumination, pour l'appareil Canon EOS 350D avec la scène « fouillis ».

En ce qui concerne la durée d'exposition, comme le tube à néon est alimenté par le courant à 50 Hz, son scintillement ne devrait pas être notable pour les faibles valeurs d'ISO car l'intégration des photons se fait sur une durée supérieure à quelques dixièmes de seconde. Néanmoins, σ_γ est estimé à moins de 0.01 for $t \simeq 1/10s$, et $\sigma_\gamma \simeq 0.15$ for $t \simeq 1/200s$. Ceci confirme que le scintillement a un effet notable principalement pour les durées d'exposition courte. Ceci montre également qu'une petite variabilité dans l'exposition effective peut être mesurée. Elle est potentiellement causée par des imperfections mécaniques ou électroniques de l'obturateur. On peut noter comme attendu que des temps d'exposition similaires donnent des variabilités relatives d'illumination semblables. À l'exception de quelques cas, on peut voir aussi que l'intensité du scintillement semble être plus petite dans le canal vert que dans les canaux rouge ou bleu. Ceci pourrait s'expliquer par la sensibilité de ces canaux au spectre de la lumière du tube à néon. Enfin, la petite valeur de σ_γ explique l'estimation de moins bonne qualité de μ pour les faibles ISO, et donc de σ .

Les résultats de la table 4.2 montrent aussi que l'estimation proposée des paramètres de vibration est assez fiable, car ils sont du même ordre de grandeur pour les quatre canaux de la même série, alors qu'ils sont traités indépendamment. En comparaison avec la table 4.1, on

ISO	t (sec.)	1. Photon transfer method				2. Foi et al.				3. Proposed method				4. Affine estimation				
		g	σ	μ	$\sigma^2 - g\mu$	g	σ	μ	$\sigma^2 - g\mu$	g	σ	μ	$\sigma^2 - g\mu$	g	σ	μ	$\sigma^2 - g\mu$	
EOS 350D Colorch.	100	1/15	0.096	2.21	256	-19.92	0.101	-22.93	0.087	2.27	260.0	-17.49	0.112	2.27	260.0	-17.49	0.112	-25.18
	200	1/30	0.191	2.38	256	-44.57	0.188	-44.33	0.181	2.45	259.1	-40.84	0.209	2.45	259.1	-40.84	0.209	-49.49
	400	1/30	0.384	2.88	256	-91.92	0.374	-89.61	0.358	2.53	256.8	-85.55	0.406	2.53	256.8	-85.55	0.406	-100.7
	800 *	1/60	0.762	4.07	256	-182.1	0.749	-177.7	0.726	3.03	254.7	-175.7	1.079	3.03	254.7	-175.7	1.079	-269.8
1600 *	1/125	1.531	6.36	256	-355.3	1.492	-342.2	1.455	4.62	254.8	-349.4	3.741	4.62	254.8	-349.4	3.741	-963.7	
EOS 6D Colorch.	100	1/3	0.176	4.85	2047	-338.6	0.724	-171.8	0.209	5.38	2158.4	-422.2	0.269	5.38	2158.4	-422.2	0.269	-573.4
	200	1/8	0.361	4.93	2047	-735.6	0.629	-1381.8	0.360	3.01	2054.7	-729.9	0.431	3.01	2054.7	-729.9	0.431	-896.2
	400	1/15	0.705	5.31	2047	-1442	1.042	-2356.9	0.697	3.16	2053.7	-1421.2	0.876	3.16	2053.7	-1421.2	0.876	-1880.1
	800 *	1/40	1.372	6.10	2047	-2053.4	1.546	-3230.6	1.29	3.85	2040.2	-2618.1	1.52	3.85	2040.2	-2618.1	1.52	-3145.8
	1600 *	1/80	2.69	7.91	2047	-5534	2.961	-6125.0	2.57	6.85	2047.5	-5222.5	6.22	6.85	2047.5	-5222.5	6.22	-13846
	3200 *	1/160	5.12	11.77	2047	-10355	5.475	-11143	5.11	10.30	2045.4	-10355	22.8	10.30	2045.4	-10355	22.8	-49609
6400 *	1/320	10.52	19.03	2047	-21131	10.77	-21631	10.49	16.25	2043.7	-21170	122.0	16.25	2043.7	-21170	122.0	-281840	
EOS 350D Clutter	100	1/13	0.096	2.21	256	-19.92	0.114	-24.42	0.093	1.99	247.3	-19.07	0.132	1.99	247.3	-19.07	0.132	-29.89
	200	1/30	0.191	2.38	256	-44.57	0.212	-49.5	0.182	2.25	255.2	-41.27	0.213	2.25	255.2	-41.27	0.213	-50.43
	400	1/60	0.384	2.88	256	-91.92	0.391	-92.6	0.375	3.51	258.9	-84.81	0.730	3.51	258.9	-84.81	0.730	-186.4
	800 *	1/125	0.762	4.07	256	-182.1	0.775	-182.1	0.720	3.88	258.3	-170.9	1.135	3.88	258.3	-170.9	1.135	-345.5
1600 *	1/250	1.531	6.36	256	-355.3	1.570	-365.5	1.503	6.60	258.7	-345.2	4.64	6.60	258.7	-345.2	4.64	-1207.1	
EOS 6D Clutter	100	1/2	0.176	4.85	2047	-338.6	0.244	-489.6	0.202	5.36	2115.4	-399.0	0.197	5.36	2115.4	-399.0	0.197	-397.2
	200	1/5	0.361	4.93	2047	-735.6	0.428	-874.0	0.334	5.32	2079	-666.8	0.321	5.32	2079	-666.8	0.321	-637.8
	400	1/10	0.705	5.31	2047	-1442	0.775	-1580.8	0.660	7.28	2094.3	-1330	0.643	7.28	2094.3	-1330	0.643	-1294.7
	800	1/20	1.372	6.10	2047	-2053.4	1.437	-2929.1	1.270	6.71	2056.2	-2566.2	1.256	6.71	2056.2	-2566.2	1.256	-2542
	1600 *	1/40	2.69	7.91	2047	-5534	2.72	-5528.4	2.51	9.77	2058.7	-5072.8	4.28	9.77	2058.7	-5072.8	4.28	-8959.1
	3200 *	1/80	5.12	11.77	2047	-10355	5.29	-10689	4.96	14.74	2059.7	-10000	8.427	14.74	2059.7	-10000	8.427	-17639
6400 *	1/160	10.52	19.03	2047	-21131	10.82	-21782	10.31	21.85	2058.2	-20752	52.98	21.85	2058.2	-20752	52.98	-123430	

TABLE 4.1 – Estimation des paramètres de la caméra (canal Rouge).

ISO	channel	σ_α	σ_β	$\text{Cov}_{\alpha,\beta}$	σ_γ
100	R	0.047	0.19	-0.0086	0.0088
	G1	0.039	0.16	-0.0065	0.0088
	G2	0.039	0.16	-0.0064	0.0086
	B	0.047	0.19	-0.0086	0.0099
200	R	0.043	0.059	-0.0016	0.012
	G1	0.039	0.057	-0.0014	0.0092
	G2	0.038	0.056	-0.0013	0.0099
	B	0.037	0.059	-0.0014	0.012
400	R	0.037	0.049	-0.0014	0.057
	G1	0.028	0.045	-0.0011	0.042
	G2	0.025	0.045	-0.00099	0.042
	B	0.031	0.044	-0.0012	0.059
800	R	0.042	0.078	-0.0015	0.058
	G1	0.03	0.073	-0.001	0.042
	G2	0.026	0.074	-0.00083	0.041
	B	0.035	0.071	-0.0011	0.056
1600	R	0.066	0.043	-0.0021	0.17
	G1	0.027	0.032	-0.00035	0.12
	G2	0.027	0.03	-0.00013	0.12
	B	0.037	0.058	-0.00032	0.17

TABLE 4.2 – Estimation de l'amplitude des vibrations et du scintillement de l'illumination.

peut voir que l'effet des vibrations tend à s'atténuer quand la durée d'exposition augmente à cause du plus grand temps d'intégration pour former l'image.

4.5 Estimation du bruit d'un capteur par analyse d'une série d'images de grilles pseudo-périodiques affectée par des vibrations

Cette section traite de l'estimation des paramètres de bruit d'un capteur, mais à partir d'une série d'images d'une grille pseudo-périodique de pas réduit à quelques pixels. Il s'agit d'une présentation succincte des résultats de [11]. Les méthodes basées sur la segmentation d'une image unique (comme par exemple [FTKE08]) sont ici inopérantes, les images de grilles ne se prêtant pas à une segmentation. Le problème est motivé par un problème en mécanique des solides expérimentale qui sera détaillé dans le chapitre 6. Dans ce cadre et par contraste avec les résultats précédemment décrits, la lumière est supposée stable et ne pas produire de scintillement (un guide de lumière à LED sur alimentation stabilisée est utilisé ici), seules les micro-vibrations empêchent d'estimer la variance empirique du bruit par moyennage temporel. Les micro-vibrations sont d'autant plus problématiques que le pas de grille n'est que de quelques pixels.

4.5.1 Approche proposée

Plus précisément, on cherche un algorithme d'estimation de la variance du bruit dépourvue du biais causé par les micro-vibrations, en utilisant la pseudo-périodicité de la grille. La figure 4.4 (à gauche) montre un détail d'une grille déposée sur un échantillon déformé. Une

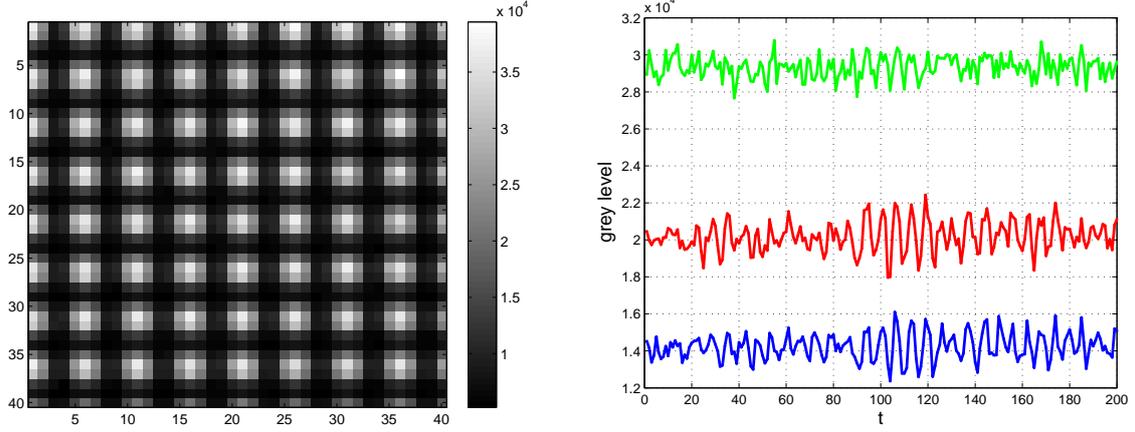


FIGURE 4.4 – À gauche : détail d'une grille. Dans cet exemple illustratif, les déformations de la grille sont à peine visibles à l'œil nu. À droite : évolution typique du niveau de gris de trois pixels distants (situés entre des lignes) au cours du temps. Les oscillations « en phase » sont causées par des micro-vibrations mécaniques et ne peuvent pas être expliquées par le bruit du capteur.

image de grille déformée est modélisée comme une fonction 2D pseudo-périodique [BGM09] :

$$s(x, y) = \frac{A}{2} \left(2 + \gamma \ell(2\pi f x + \phi_1(x, y)) + \gamma \ell(2\pi f y + \phi_2(x, y)) \right) \quad (4.21)$$

où $A > 0$ est l'illumination moyenne (supposée constante), $\gamma \in [0, 1]$ est le contraste des lignes de la grille, le profil des lignes ℓ est une fonction réelle 2π -périodique d'amplitude 1 et de valeur moyenne 0, f est la fréquence de la grille (par exemple, $f = 1/5$ pixel $^{-1}$ dans la figure 4.4), et $\phi_1(x, y)$ et $\phi_2(x, y)$ représentent les modulations de phase de la grille porteuse, selon les directions x et y respectivement.

En pratique, les dérivées de ϕ_1 et ϕ_2 sont très petites par rapport à $2\pi f$ pour la plupart des matériaux réels ($10^{-2} - 10^{-4}$ contre $2\pi/5$ pixel $^{-1}$). Il s'agit d'une propriété fondamentale que nous allons exploiter.

Nous montrons dans [11] par des développements de Taylor à l'ordre 1, justifiés par les ordres de grandeur des dérivées partielles de ϕ_1 et ϕ_2 , que l'algorithme suivant permet de mesurer les paramètres gouvernant le bruit de capteur à partir d'une série de T images affectées par des micro-vibrations. Chaque image de la série se modélise, avec les notations de la section 4.2, en :

$$u(x, y, t) = s(x + \alpha_t, y + \beta_t) + n_{s(x+\alpha_t, y+\beta_t)} \quad (4.22)$$

où n désigne le bruit, dont la variance dépend du signal sous-jacent.

Un développement de Taylor justifie d'écrire à l'ordre 1, pour chaque image t :

$$\frac{1}{N} \sum_{k,l} u(x + kp, y + lp, t) = s(x + \alpha_t, y + \beta_t) + \frac{1}{N} \sum_{k,l} n_{s(x+\alpha_t, y+\beta_t)}(x + kp, y + lp, t) \quad (4.23)$$

où (k, l) parcourt un ensemble de cardinal N . En pratique la somme est calculée avec des poids gaussiens en fonction de l'éloignement au pixel considéré (x, y) de manière à localiser l'estimation (ce qui permet de considérer que l'éclaircissement n'est constant que localement).

Si N est suffisamment grand, cette moyenne fournit une estimation de $s(x + \alpha_t, y + \beta_t)$, et retrancher cette quantité à (4.22) fournit la composante de bruit $n_{s(x+\alpha_t, y+\beta_t)}$ pour tout x, y, t .

Toujours à l'ordre 1, l'espérance empirique de $s(x + \alpha_t, y + \beta_t)$ est égale à

$$\frac{1}{T} \sum_{t=1}^T s(x + \alpha_t, y + \beta_t) = s(x, y) + (\bar{\alpha}, \bar{\beta}) \nabla s(x, y) \quad (4.24)$$

et l'espérance de la variance empirique de $n_{s(x+\alpha_t, y+\beta_t)}$ (processus d'espérance nulle) est :

$$E \left(\frac{1}{T} \sum_{t=1}^T n_{s(x+\alpha_t, y+\beta_t)}^2 \right) = \frac{1}{T} \sum_{t=1}^T \text{Var}(n_{s(x+\alpha_t, y+\beta_t)}) \quad (4.25)$$

$$= as(x, y) + a(\bar{\alpha}, \bar{\beta}) \nabla s(x, y) + b \quad (4.26)$$

qui converge vers $as(x, y) + b$ dès que T est assez grand.

L'algorithme proposé consiste finalement en les étapes suivantes.

1. Calcul des moyennes spatiales sur les pixels distant d'un multiple de p autour du pixel courant (x, y) à t donné, comme dans (4.23), pour obtenir la composante « débruitée » $s(x + \alpha_t, y + \beta_t)$, puis soustraction aux images de la série pour obtenir la composante de bruit $n_{s(x+\alpha_t, y+\beta_t)}$ pour tout (x, y, t) .
2. Estimer $s(x, y)$ grâce à (4.24) et $as(x, y) + b$ grâce à (4.26), puis estimer a et b par régression linéaire.

L'étape 1 est nommée dans la suite *Non-Random Signal Reduction* (NRSR).

4.5.2 Expériences

On considère une série de $T = 200$ images de grilles acquise par une caméra PCO Sensicam QE. Elle est affectée par des micro-vibrations, comme l'illustre la figure 4.4 (à droite).

La procédure décrite dans la section précédente est appliquée. Les résultats sont illustrés par la figure 4.5. Le graphique a est la composante « débruitée » $s(x + \alpha_t, y + \beta_t)$, estimée aux mêmes points que ceux suivis dans la figure 4.4 (à droite). Le graphique b est la partie aléatoire du signal, centrée à la valeur de la moyenne temporelle. Il s'agit du résultat de l'algorithme NRSR. Le signal résiduel est bien distribué aléatoirement, et son amplitude est plus grande pour la ligne rouge que pour les lignes bleue et verte, ce qui est cohérent avec la dépendance affine entre la variance du bruit et le niveau de gris. Les graphiques c et d montrent la distribution des variances empiriques comme fonction de la moyenne empirique sur les données brutes (comme sur la figure 4.4) puis après NRSR (comme sur le graphique b). Le premier graphique montre bien que les micro-vibrations induisent un fort biais sur l'estimation de la variance. Dans notre expérience, il s'avère que les vibrations sont surtout distribuées verticalement, et n'affectent donc pas les points de l'image de grille situés sur les lignes verticales (voir l'équation (4.8)). Ici, ces points correspondent à la partie basse du graphique. Le second graphique montre une tendance linéaire, et une régression linéaire donne, avec les notations de (4.3) : $a = g = 8.4547$ et $b = \sigma_d^2 - g\mu_\delta = -5378.1$. La droite de régression est superposée en rouge. Sans NRSR, les paramètres estimés à partir du graphique c) sont $a = 21$ et $b = -35332$. Ces valeurs peuvent être vérifiées sur les estimations du constructeur donnée dans les caractéristiques de la caméra PCO Sensicam QE [sen05]. Il indique $g = 0.519$ (mode *high gain*), $\mu_d = 45.46$ et $\sigma_d^2 = 1.33$ sur une échelle 12 bits. Pour obtenir les valeurs

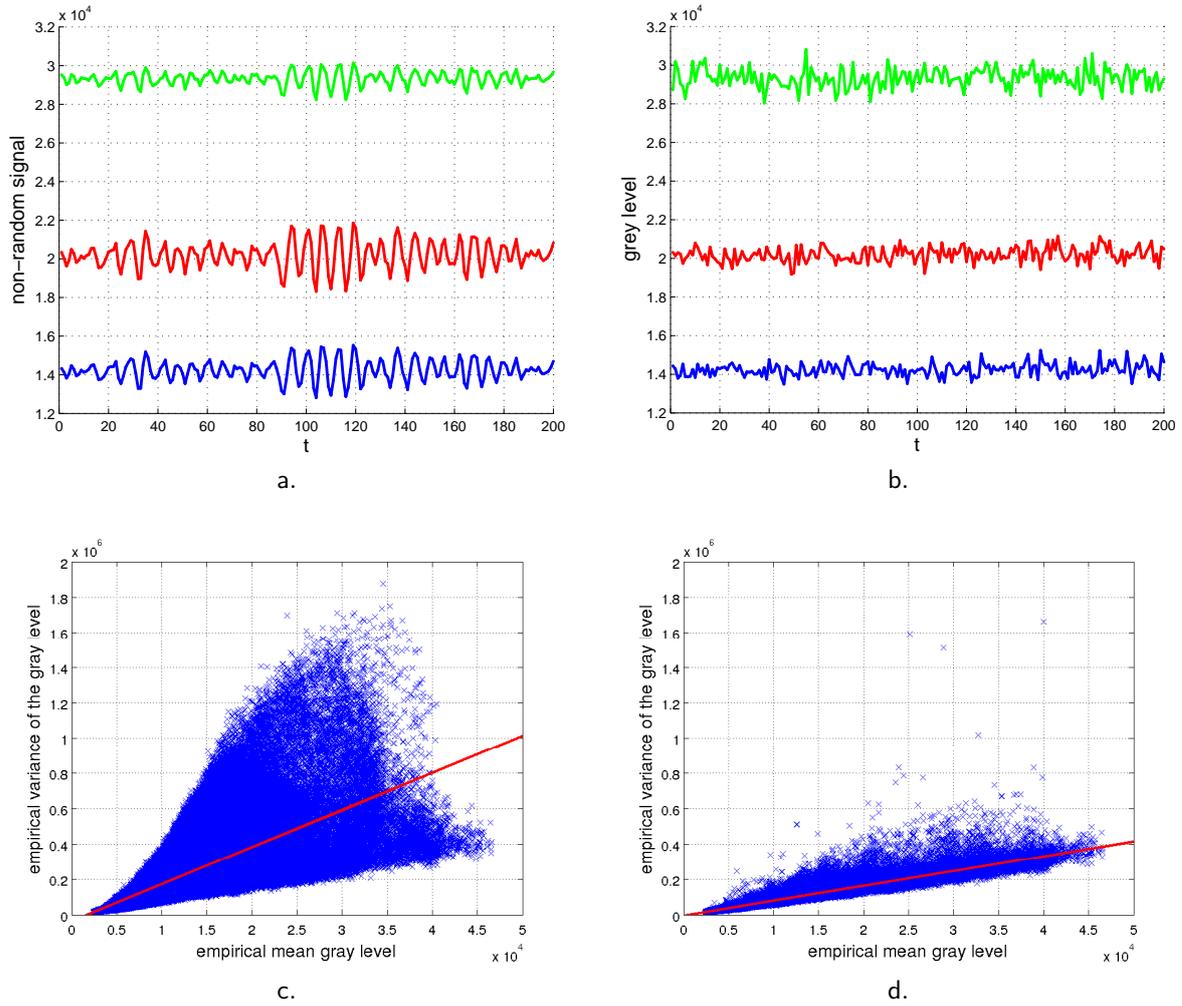


FIGURE 4.5 – a : oscillations non-aléatoires estimées. b : composante de bruit autour du niveau de gris moyen. c : variance empirique contre niveau de gris moyen sur les données brutes. d : variance empirique contre niveau de gris moyen après NRSR (même axes que dans c).

théoriques pour a et b (sur une échelle 16 bits), il faut multiplier g par 2^4 et $\sigma_d^2 - g\mu_\delta$ par 2^8 . Ceci donne $a = 2^4 \times 0.519 = 8.304$ et $b = (1.33 - 45.46 \times 0.519) \times 2^8 = -5699.5$. Ces données constructeur sont en relativement bonne adéquation avec les valeurs trouvées plus haut.

Comme application (motivée par [12] au chapitre 6), nous vérifions que les paramètres estimés permettent d'appliquer la transformée d'Anscombe généralisée (GAT, décrite dans [MSB95] d'après [Ans48]) qui stabilise à 1 la variance de tout processus de Poisson-Gauss comme celui de (4.1). Avec nos notations, la GAT s'écrit :

$$\text{GAT}(s(x, y)) = \frac{2}{g} \sqrt{gs(x, y) + \frac{3}{8}g^2 + \sigma_\delta^2 - g\mu_\delta} \quad (4.27)$$

La figure 4.6 a montre la variance empirique contre la moyenne empirique après NRSR et GAT. Le bruit est à présent homoscedastique, avec une variance approximativement constante et effectivement égale à 1. Le bénéfice de la NRSR dans l'estimation des paramètres du bruit

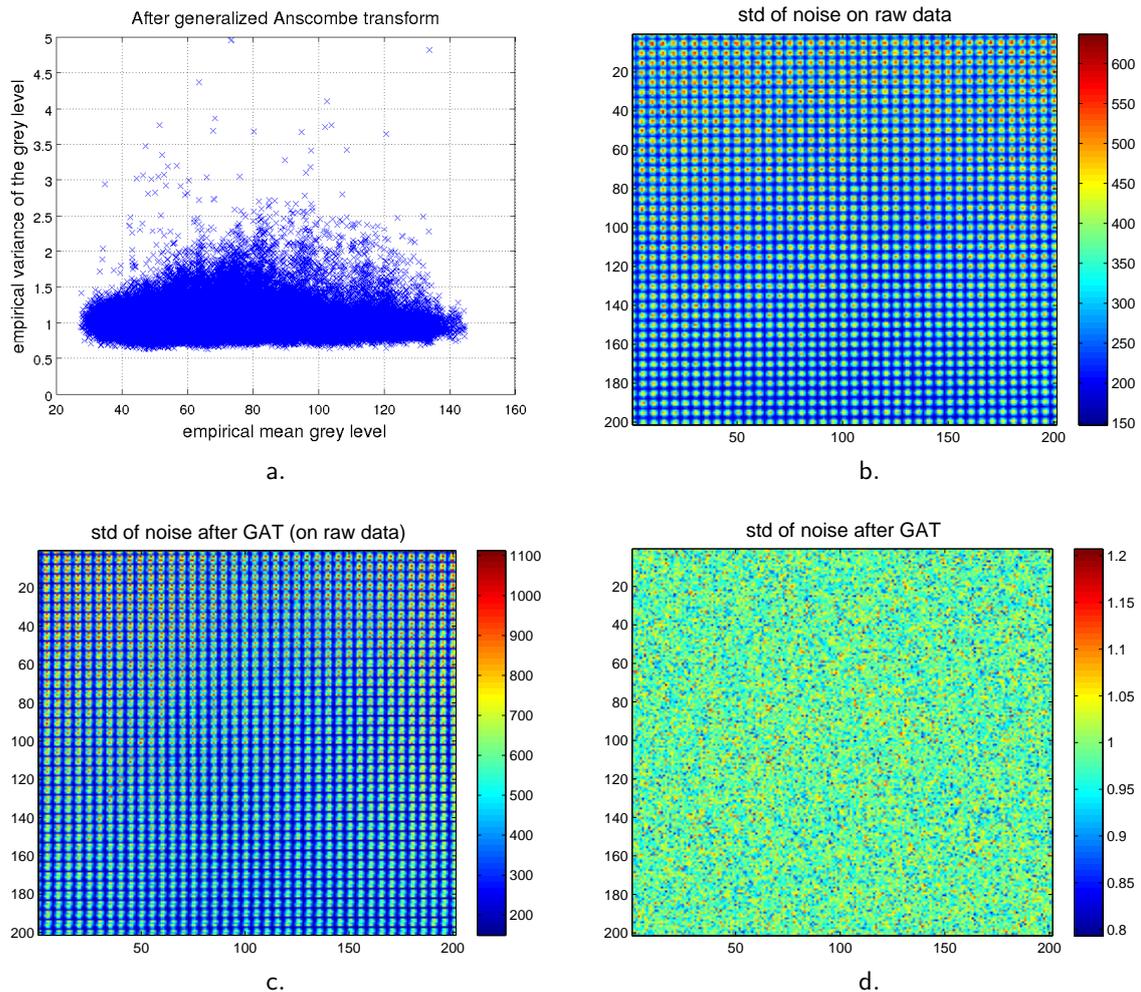


FIGURE 4.6 – a : variance empirique contre niveau de gris moyen empirique après transformée d'Anscombe généralisée avec les paramètres du bruit estimés par notre méthode. b : carte de l'écart-type du bruit sur les données brutes. c : carte après GAT mais sans NRSR sur les données brutes. d : carte après estimation des paramètres du bruit par NRSR puis GAT.

est ensuite vérifié graphiquement en comparant les distributions spatiales des écart-types empiriques. La carte b (estimée directement depuis les données brutes, sans NRSR ni GAT) montre clairement le motif de grille, dû à la relation (4.3) entre variance du bruit et niveau de gris moyen. Si la GAT est appliquée sur les données brutes (avec les paramètres estimés par régression linéaire sur ces données, c'est-à-dire la ligne rouge dans le graphique c de la figure 4.5), alors la carte c montre toujours le motif de grille. Ceci signifie que l'estimation par NRSR est requise pour éliminer le biais dû aux micro-vibrations dans l'estimation de la variance. La carte d, estimée après GAT avec les données traitées par NRSR, ne présente en effet aucun motif visible. Le bruit a maintenant une variance normalisée à 1, ce qui illustre la robustesse de notre procédure d'estimation de a et b .

4.6 Conclusion et perspectives

Les résultats résumés dans ce chapitre traitent de l'estimation des paramètres gouvernant le niveau de bruit à partir d'une série d'images d'une scène quasi-statique sous les mêmes paramètres de prise de vue. Utiliser une série d'images plutôt qu'une image unique permet de s'affranchir des problèmes de segmentation préalable, mais rend le procédé d'estimation sensible à des phénomènes extrinsèques comme les micro-vibrations ou un scintillement de la source lumineuse (ou, de manière équivalente, une variabilité de l'exposition). Nous avons proposé un modèle stochastique de Cox-Gauss de formation de l'image étendant le modèle classique de Poisson-Gauss. Un algorithme pratique, robuste à l'effet *rolling shutter*, a été également proposé. Il permet l'estimation de tous les paramètres du capteur (gain, décalage, et niveau du bruit de lecture) avec une précision comparable à la méthode de transfert de photon, spécialement lorsqu'un fort scintillement de la source est observé.

De manière intéressante, un lien peut être établi avec les ICCD dans lesquels un amplificateur multiplie les photons incidents avant acquisition par le CCD. Une relation quadratique entre variance du bruit et niveau de gris moyen est établie dans [Bod99] (citée par [RKBW13, RKWB12]). La différence avec le modèle standard décrit dans (4.1) est que le gain g est modélisé dans un ICCD comme une variable aléatoire, donnant une relation légèrement différente de celle que nous établissons. Il nous semble que considérer l'amplification comme une source de scintillement justifierait d'envisager l'ICCD comme un cas d'application de notre modèle.

D'autre part, le caractère non-constant de l'illumination dans une rafale d'images a été noté il y a quelques années par les auteurs de [BLMT09]. Dans une application de débruitage dit « multi-images », basé sur une rafale, ce phénomène rend difficile l'estimation du niveau de bruit pour un débruitage efficace. Les auteurs de [BLMT09] dépassent ce problème en égalisant les histogrammes des images de la rafale par *midway equalization* [Del04], ce qui leur permet un débruitage efficace. Cette solution ne permet néanmoins pas de garantir l'estimation des paramètres de bruit, et il nous semble qu'il serait intéressant d'étendre notre estimation à une rafale d'images présentant un mouvement de caméras, avec des applications potentielles au débruitage multi-images tel qu'envisagé par exemple dans [BLMT09] ou plus récemment dans [LYT⁺14].

Par ailleurs, on peut remarquer que l'hypothèse d'illumination constante est mise à mal dans certaines applications comme l'imagerie par fluorescence dans laquelle un processus de photo-destruction (*photobleaching*) peut être observé. Les auteurs de [JCP⁺14] proposent alors une modélisation paramétrique de ce phénomène (la photo-destruction induisant une décroissance exponentielle du flux de photons) et une estimation par algorithme EM. Explorer de telles séquences d'images dans notre formalisme nous semble une perspective intéressante.

Enfin, dans le cas des grilles pseudo-périodiques utilisées en mécanique des solides expérimentale, sujettes à micro-vibrations, nous avons proposé un autre algorithme simple d'estimation des paramètres gouvernant le niveau de bruit. Cet algorithme est pratique car il est simple à mettre en œuvre une fois que la manipulation a été montée, et ne nécessite donc pas d'acquérir au préalable des images de mire uniforme sous illumination contrôlée.

Élimination du bruit périodique ou quasi-périodique

5.1	Introduction	81
5.2	Détection du bruit quasi-périodique comme déviation statistique des coefficients spectraux	83
5.3	Détection a-contrario du bruit quasi-périodique	90
5.4	Interpolation des coefficients spectraux par minimisation sous contrainte de la variation totale	95
5.5	Conclusion et perspectives	98

5.1 Introduction

5.1.1 Contexte

Dans certains cas, les images numériques peuvent être intégralement couvertes d'un motif répétitif parasite. Dans la littérature, on parle de « bruit périodique » ou de « bruit quasi-périodique » car le motif n'est souvent pas rigoureusement périodique. Le phénomène peut apparaître à cause d'interférences électriques perturbant l'acquisition de l'image, d'un capteur mal calibré ou de données manquantes, et est la cause, par exemple, de lignes manquantes, de bandes parasites (phénomènes de *striping* ou *banding*), ou d'un bruit de fond plus complexe. Les applications de télédétection comme l'imagerie satellitaire sont particulièrement sensibles au phénomène [Sch06]. Certaines techniques de reproduction d'images peuvent aussi présenter une forme de bruit périodique, comme l'impression en similitravure (*halftone printing*) ou l'affichage sur moniteur à tube cathodique qui laisse apercevoir les lignes de balayage. Certaines photographies à haute résolution de peintures laissent voir la trame de la toile, qui se traduit par un motif périodique [CDCS12, CYG⁺16, HMdC87]. Enfin, une source de bruit quasi-périodique peut aussi être l'*aliasing*. Le repliement de spectre lors de la numérisation de signaux périodiques dont la fréquence ne satisfait pas la condition de Nyquist est susceptible de générer des franges périodiques sur certaines textures. Néanmoins, l'*aliasing* se traduit par un bruit périodique localisé à la zone couverte par la texture haute-fréquence.

Alors que l'estimation et l'élimination du bruit blanc est le sujet d'une littérature abondante (voir par exemple les revues récentes [LCBM12, Mil13]), il s'avère que le bruit quasi-périodique n'a été abordé que dans relativement peu d'études. La majorité des méthodes

existantes suivent les grandes lignes de ce qui a été proposé à l'aube du traitement des images provenant des sondes astronautiques [RDF⁺71]. Un bruit quasi-périodique se traduisant par des pics assez bien localisés dans le spectre, il convient de détecter ces pics puis de les éliminer par un filtre coupe-bande. Les approches élémentaires sont décrites dans [GW06], et des problèmes pratiques discutés dans [WL10, WWN⁺12, ICST15], [7]. La difficulté est d'automatiser la détection des pics spectraux dus au bruit, ou, de manière équivalente, la construction des filtres coupe-bandes. Il est souvent difficile de distinguer dans le spectre ces pics de ceux engendrés par une texture localisée ou une structure répétée, très courante dans un environnement construit par l'homme. Certains auteurs [AB02, AB08, HT05] suggèrent de détecter les pics dans le spectre à l'aide d'un filtre médian local sur le spectre d'amplitude. La propriété de localisation d'un bruit quasi-périodique dans le spectre permet également de considérer le problème de séparation entre l'image sous-jacente et le bruit comme une séparation de source aveugle [ZP01]. Par exemple, ces deux composantes peuvent être supposées morphologiquement distinctes dans le sens de [FSBM10], c'est-à-dire qu'elles peuvent avoir une représentation parcimonieuse dans deux bases dites incohérentes. Les auteurs de [SED05] présentent quelques exemples de décomposition en composantes morphologiquement indépendantes (décomposition MCA), utilisant une représentation de l'image sur une base de *curvelets* et de la composante périodique sur une base DCT (*discrete cosine transform*). Néanmoins, une texture localisée haute-fréquence est susceptible d'être aussi représentée dans la composante périodique, et par conséquent d'être supprimée avec le bruit quasi-périodique. Des exemples de décomposition MCA sont montrés dans [8] et [CYG⁺16]. Les décompositions entre structure et texture par approche variationnelle, comme celles discutées dans [AGCO06], mériteraient d'être étudiées de manière approfondie dans ce cas, mais présentent *a priori* les mêmes limites. Notons enfin qu'un algorithme d'élimination de bandes parasites est décrit dans [FWL12] dans le cadre des bruits blancs convolués. Les bruits quasi-périodiques considérés dans ce chapitre ne sont néanmoins pas bien modélisés dans ce cadre.

5.1.2 Contributions

Dans ce chapitre, nous discutons deux méthodes de détection du bruit quasi-périodique dans les images. Toutes les deux sont basées sur la remarque que le bruit quasi-périodique est la seule structure périodique présente dans un ensemble d'images (*patches*) couvrant l'intégralité de l'image, contrairement à une structure périodique localisée. Deux méthodes de détection des coefficients spectraux du bruit sont discutées : l'une comme détection de déviations aux propriétés statistiques attendues dans le spectre d'une image naturelle (section 5.2), l'autre dans le cadre de la détection a-contrario (section 5.3). Comme l'application d'un filtre coupe-bande revient à mettre à zéros certains coefficients spectraux et est donc susceptible de générer du *ringing* dans l'image débruitée, nous discutons également d'une méthode d'interpolation des coefficients spectraux couverts par le bruit à l'aide d'une minimisation sous contrainte de la variation totale (section 5.4).

Les principales publications relatives à ce chapitre sont :

[7] M. Grédiac, F. Sur, and B. Blaysat. Removing quasi-periodic noise in strain maps by filtering in the Fourier domain. *SEM Experimental Techniques*, 40(3):959–971, 2016.

[8] F. Sur and M. Grédiac. Automated removal of quasiperiodic noise using frequency domain statistics. *IS&T / SPIE Journal of Electronic Imaging*, 24(1):013003/1 – 013003/19, 2015.

[23] F. Sur. An a-contrario approach to quasi-periodic noise removal. In *Proceedings of the IEEE International Conference on Image Processing (ICIP)*, pages 3841–3845, Québec City, Canada, 2015.

Le code Matlab ARPENOS (Automated Removal of quasi-Periodic NOise using frequency-domain Statistics) disponible à l'URL :

<https://members.loria.fr/FSur/software/ARPENOS/>
implémente l'algorithme décrit dans [8].

Le code Matlab AC-ARPENOS (A-Contrario Automated Removal of quasi-Periodic NOise using frequency-domain Statistics) disponible à l'URL :

<https://members.loria.fr/FSur/software/ACARPENOS/>
implémente l'algorithme décrit dans [23].

5.2 Détection du bruit quasi-périodique comme déviation statistique des coefficients spectraux

En exploitant la remarque selon laquelle le bruit quasi-périodique est la seule composante périodique présente dans un ensemble d'images couvrant l'intégralité de l'image, il est tentant de considérer le *spectre d'amplitude* (ou de puissance) *moyen* correspondant à la moyenne des spectres d'amplitudes (ou de puissance) de ces images. Dans ce spectre moyen, les composantes spectrales dues aux textures localisées ou aux motifs répétés sont atténuées, et les composantes dues au bruit quasi-périodique sont réhaussées. Il s'avère que l'idée n'est pas nouvelle et a déjà été suggérée dans les années 1980 dans un court article [CLP83], avec une application au *destriping* d'images satellitaire [SCW88]. Néanmoins, ces auteurs présentent le spectre moyen comme un outil permettant de faciliter le repérage des coefficients spectraux du bruit et la construction « à la main » du filtre coupe-bande adéquat. Nous proposons de systématiser cette méthode et de détecter automatiquement ces coefficients comme des déviations aux propriétés statistiques du spectre moyen, qui, en tout cas pour des images naturelles, est censé décroître comme l'inverse d'une puissance positive de la fréquence, comme discuté dans [Fie87, OT01, TO03, vdSvH96, HHH09]. Notons que des similarités existent avec la contribution développée dans [CDCS12]. Les auteurs de cet article proposent en effet de construire un filtre de Wiener pour débruiter l'image, ce qui nécessite d'estimer les fonctions d'autocorrélation du bruit et de l'image non-bruitée. Une moyenne de *patches* est proposée dans ce but. Les approches proposées dans [CDCS12] et [8] présentent donc une similitude dans leur estimation par *patches*, le spectre de puissance et la fonction d'autocorrélation étant liés par le théorème de Wiener-Khinchine.

5.2.1 Modèle d'imagette

Notre méthode est basée sur la modélisation stochastique d'une imagette. On suppose qu'une imagette aléatoire p est la somme de la composante non bruitée p_0 , d'un bruit périodique n , et d'un bruit blanc gaussien w , c'est-à-dire :

$$p(x, y) = p_0(x, y) + n(x, y) + w(x, y) \quad (5.1)$$

où

— p_0 est une imagette aléatoire ;

- $n = \tau_{t_x, t_y}(n_0)$ est la translation périodique d'un motif n_0 (à de petites variations près, ce qui motive la modélisation de n_0 comme une imagerie aléatoire d'espérance $E(n_0)$), par un déplacement aléatoire (t_x, t_y) uniformément distribué dans $[1, X] \times [1, Y]$. Comme n_0 représente le bruit quasi-périodique, $E(n_0)$ est la transformée de Fourier inverse d'un spectre réduit à quelques composantes non nulles. Notons que si nous présentons ici le modèle original de [8], on peut également simplement supposer n stationnaire.
- w est un bruit blanc gaussien de variance $\text{Var}(w)$;
- les variables aléatoires précédentes sont mutuellement indépendantes.

Si on note en capitales d'imprimerie les transformées de Fourier discrètes, on conclut après calcul que :

$$E(|P|^2) = E(|P_0|^2) + E(|N_0|^2) + XY\text{Var}(w) \quad (5.2)$$

L'espérance du spectre de puissance de l'imagerie p est donc fait de trois composantes, à savoir :

1. l'espérance du spectre de puissance de l'imagerie non bruitée p_0 . De nombreuses études s'attachent à la modélisation du spectre de puissance des images naturelles (voir, par exemple, [Fie87, vdSvH96]), comme une loi de puissance en l'inverse de la fréquence. Il existe donc, des réels positifs α et A tels que :

$$E(|P_0(f)|^2) = \frac{A}{f^\alpha} \quad (5.3)$$

Dans la plupart des images naturelles, $\alpha \simeq 2$. Des modèles plus sophistiqués existent : dans [OT01, TO03] par exemple, il est montré que A et α dépendent en fait aussi de la direction associée aux coefficients spectraux. Nous nous contentons d'utiliser le modèle isotrope de (5.3) ;

2. l'espérance du spectre de puissance de la composante quasi-périodique n_0 , qui est principalement constituée de quelques pics ou régions bien isolées dans le spectre ;
3. la contribution du bruit blanc w , constante sur l'ensemble du spectre. Elle est dominée par les deux premières composantes, spécialement pour les fréquences basses et moyennes. Elle peut néanmoins être visible dans les hautes fréquences à cause de la décroissance en loi de puissance de $E(|P_0(f)|^2)$ et de l'aspect parcimonieux de $E(|N_0|^2)$.

Si on néglige le bruit blanc, on voit que le spectre de puissance moyen des imageries est réduit aux pics dus au bruit quasi-périodique auxquels s'ajoutent la distribution en loi de puissance.

5.2.2 Algorithme

Étant donnée une image affectée d'un bruit quasi-périodique, l'algorithme 5.1 détecte les composantes du bruit périodique et les élimine.

L'estimation se fait par régression linéaire robuste sur le nuage de points des $(\log(f), I(f))$ pour f entre f_0 et f_1 cycles par pixels, de la même manière que dans [OT01] par exemple. La régression robuste donne des estimations de A et α non influencées par les pics dus au bruit quasi-périodique. L'estimation au moindre carré donne aussi l'écart-type σ des résidus. On restreint la détection des coefficients aberrants aux fréquences $f > f_2$, car les basses fréquences ne peuvent correspondre à des motifs répétés.

Les choix des paramètres sont discutés soigneusement dans [8].

Algorithme 5.1 Élimination de bruit quasi-périodique par détection de déviations statistiques à une loi de puissance.

Entrée : image i (taille $X \times Y$, et transformée de Fourier I), taille d'imagette L .

1. Extraction des imagettes non-recouvrantes $L \times L$ distribués uniformément sur i (ce qui fournit P imagettes) et calcul du spectre de puissance moyen $|\widetilde{P}|^2$.
2. Estimation par régression linéaire des paramètres A et α de la loi de puissance du spectre de puissance :

$$\log\left(|\widetilde{P}|^2(f)\right) = A - \alpha \log(f) \quad (5.4)$$

3. Détection des déviations à la loi de puissance (ξ, η) telles que

$$\frac{\log\left(|\widetilde{P}|^2(\xi, \eta)\right) - (A - \alpha \log(f))}{\sigma} > 3 \quad (5.5)$$

où σ est l'écart-type des résidus de régression.

4. Construction de la carte des pics M_o^P dus au bruit quasi-périodique dans le spectre de taille $L \times L$ tel que $M_o^P(\xi, \eta) = 1$ si le coefficient spectral en (ξ, η) est aberrant, et 0 sinon.
5. Interpolation de M_o^P de la taille $L \times L$ à $X \times Y$, ce qui donne une carte M_o des pics dans le spectre de l'image originale.

Sortie : \widehat{n} , estimation du bruit quasi-périodique comme la transformée de Fourier inverse de $M_o I$, et \widehat{i} , estimation de l'image débruitée, comme $i - \widehat{n}$ (i.e., transformée de Fourier inverse de $(1 - M_o)I$).

5.2.3 Exemples illustratifs

La figure 5.1 a. montre une image i de taille 600×581 pixels obtenue par la sonde spatiale Mariner 4. Elle est affectée d'un bruit quasi-périodique. Son spectre de puissance est montré dans la figure 5.1 b. Les pics correspondant au bruit quasi-périodique sont à peine visibles et difficilement séparables des autres composantes, ce qui illustre la difficulté d'une détection automatique. Ceci est confirmé par la distribution des coefficients de ce spectre de puissance en fonction de la fréquence, qui montre une tendance décroissante mais des coefficients très éparpillés (cf. figure 5.1 c). La figure 5.1 d. montre le spectre de puissance moyen sur un ensemble de 870 imagettes de taille $L \times L = 128 \times 128$, uniformément distribuées sur l'ensemble de l'image, à une distance de $L/8 = 16$ pixels l'une de l'autre (les imagettes se recouvrent légèrement car elles sont en fait multipliées comme dans [OT01] par une fenêtre de Hann pour éviter le biais dans le spectre moyen induit par les discontinuités aux extrémités de l'image). La régression linéaire robuste permet d'estimer les paramètres $A = 11.27$ and $\alpha = 1.47$, voir la figure 5.1 e. Ici, le segment rouge correspond à la tendance linéaire (entre f_0 et f_1 cycles par pixel), et le segment vert correspond à la limite à 3σ (au delà de f_2). Les coefficients aberrants sont ceux au-dessus de ce dernier segment. On peut voir qu'un simple seuil constant sur toutes les fréquences ne peut convenir.

La figure 5.2 a. montre la localisation des coefficients aberrants dans le spectre de puissance moyen et la figure 5.2 b. montre le spectre de l'image originale multiplié par le filtre coupe-bande $1 - M_o$.

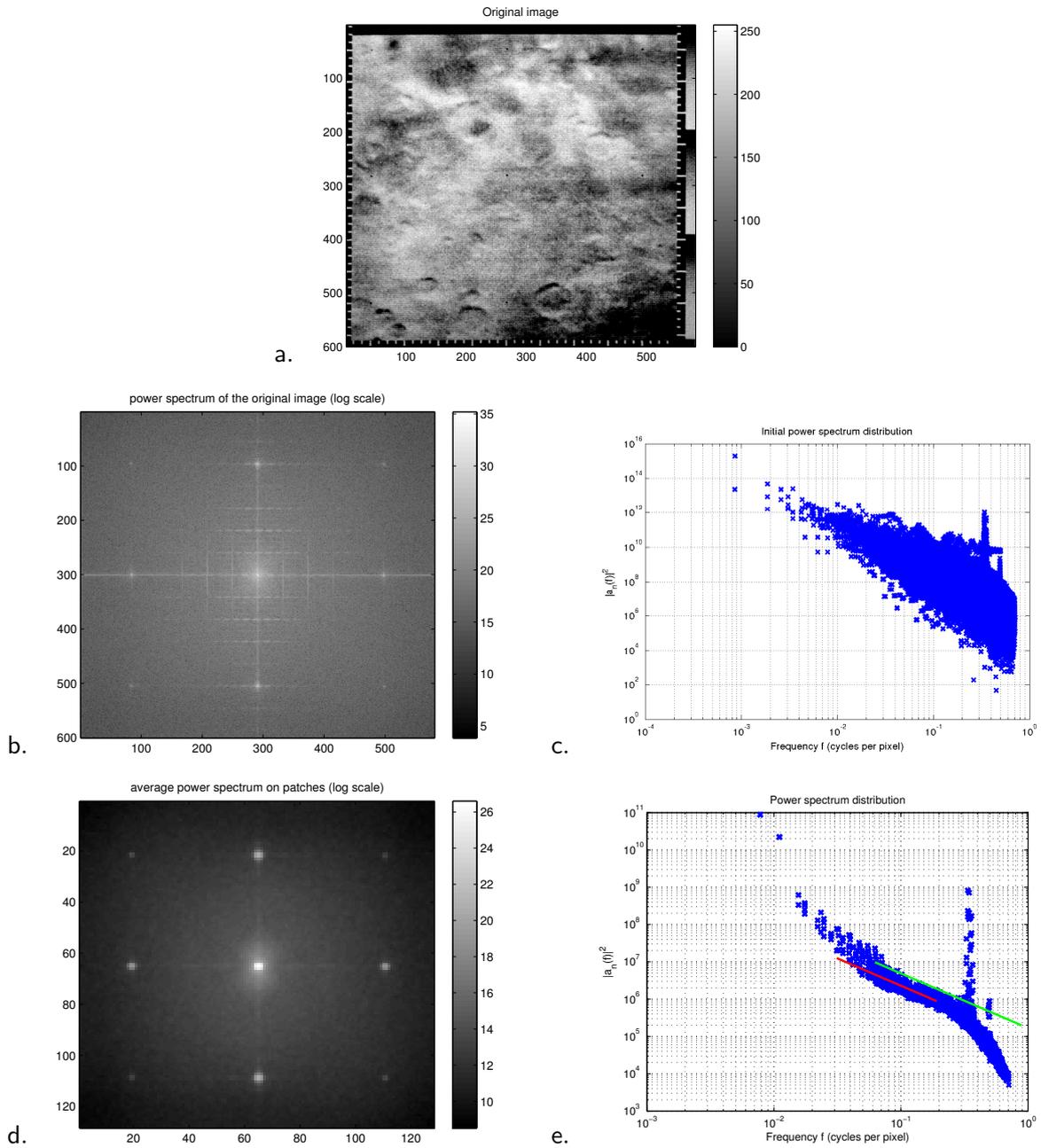


FIGURE 5.1 – *Expérience Mariner 4 (1)*. a : image originale. b : spectre de puissance (échelle logarithmique). c : distribution des coefficients du spectre de puissance en fonction de la fréquence. d : Spectre de puissance moyen. e : distribution des coefficients du spectre de puissance moyen en fonction de la fréquence.

La sortie de notre algorithme est l'image débruitée \hat{i} et le bruit quasi-périodique \hat{n} sont illustrés par la figure 5.3 (a et b). Comme dans le *bruit de méthode* défini dans [BCM05], cette dernière composante ne doit idéalement pas contenir de structures provenant de l'image originale. On peut voir que le bruit quasi-périodique a été effectivement retiré de l'image \hat{i} . La composante pseudo-périodique \hat{n} est en fait plus complexe qu'une simple sinusoïde. Un détail des images originales et débruitée est montré sur la figure 5.3 (c et d).

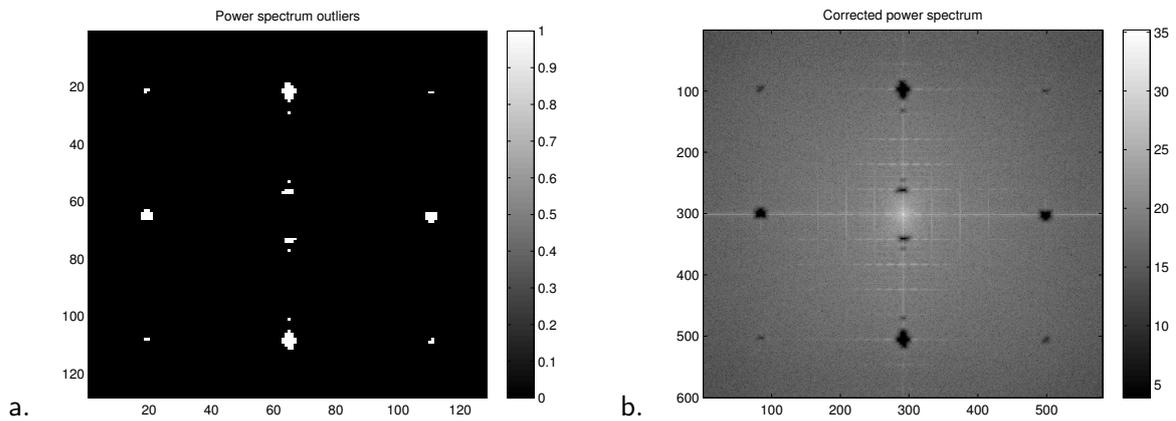


FIGURE 5.2 – *Expérience Mariner 4 (2)*. a : coefficients aberrants du spectre. b : Spectre de puissance corrigé.

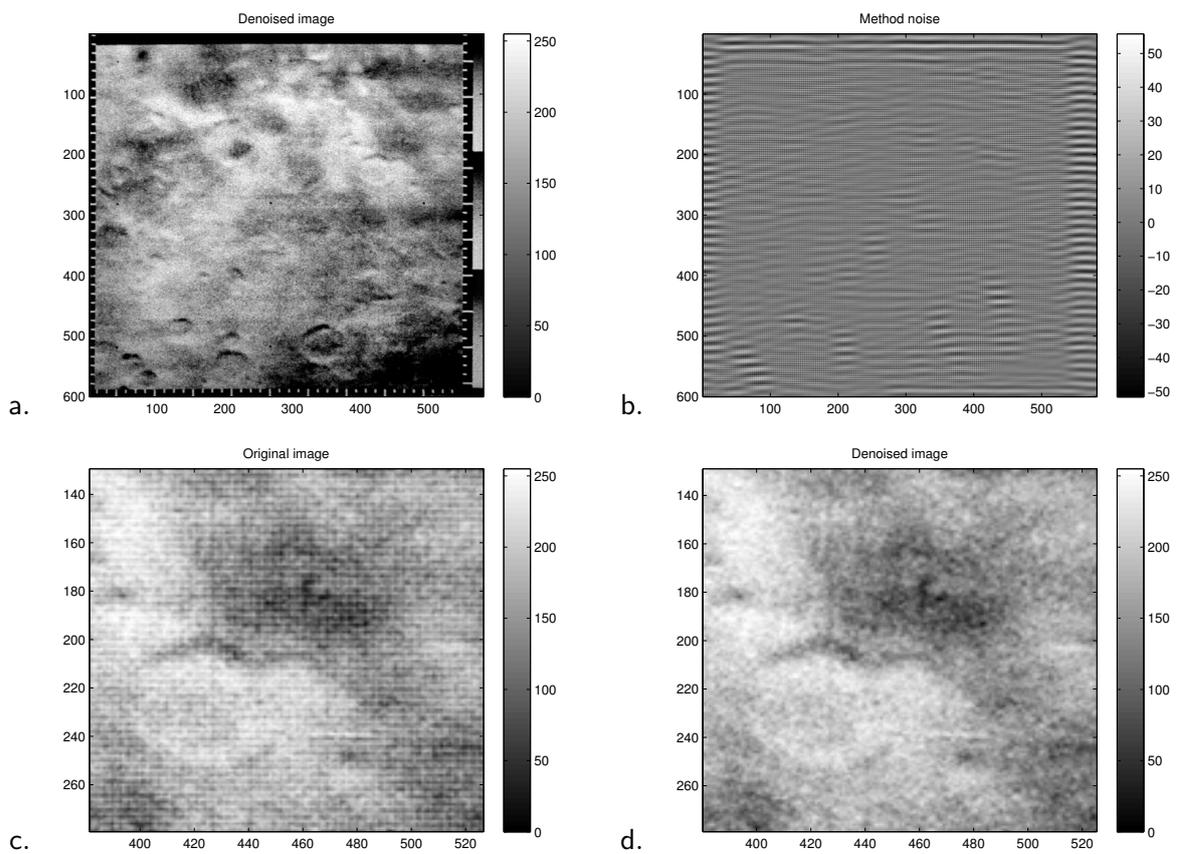


FIGURE 5.3 – *Expérience Mariner 4 (3)*. a : image débruitée \hat{i} . b : composante de bruit \hat{n} . c : détail de l'image bruitée. d : détail de l'image débruitée.

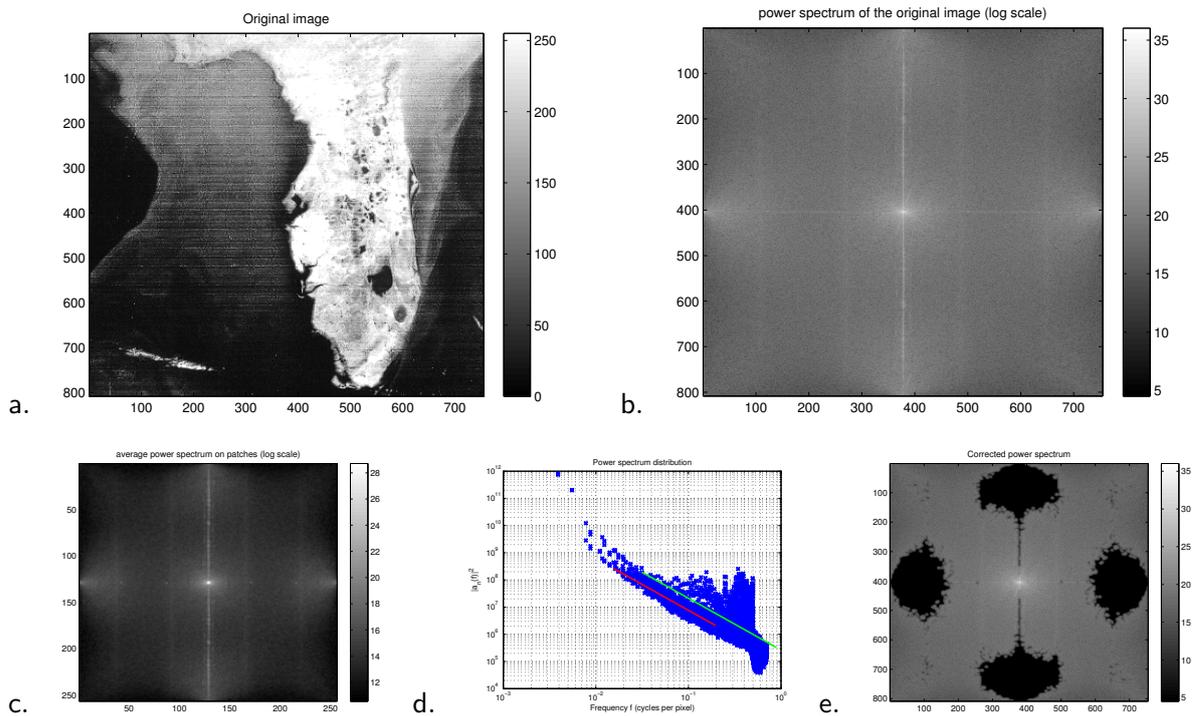


FIGURE 5.4 – *Expérience Floride (1)*. a : image bruitée. b : spectre de puissance. c : spectre de puissance moyen. d : distribution des coefficients du spectre de puissance moyen. e : Spectre de puissance corrigé.

Une autre expérience est montrée dans les figures 5.4 et 5.5. Sur cette vue satellitaire de la Floride, un fort phénomène de *striping* est visible dans la mer. Ceci explique les lignes verticales visibles dans le spectre. De plus, des “blobs” peuvent être vus au milieu des bords gauche/droite et haut/bas du spectre. Toutes ces composantes sont détectées comme des coefficients aberrants. Il se trouve que l’image originale est fortement aliasée, et que dans cet exemple l’*aliasing* a un effet visible sur l’intégralité de l’image, comme on le voit dans l’aspect pixellisé des côtes dans le détail présenté dans la figure 5.5). Naturellement, notre algorithme ne permet pas d’éliminer le bruit blanc affectant également cette image. Enfin, remarquons dans la figure 5.5 g des franges parasites qui, même si elles ne sont pas visibles à première vue, affectent également l’image débruitée complémentaire f. Il s’agit de *ringing* causé par l’annulation de nombreux coefficients du spectre. Nous reviendrons sur ce phénomène en section 5.4.

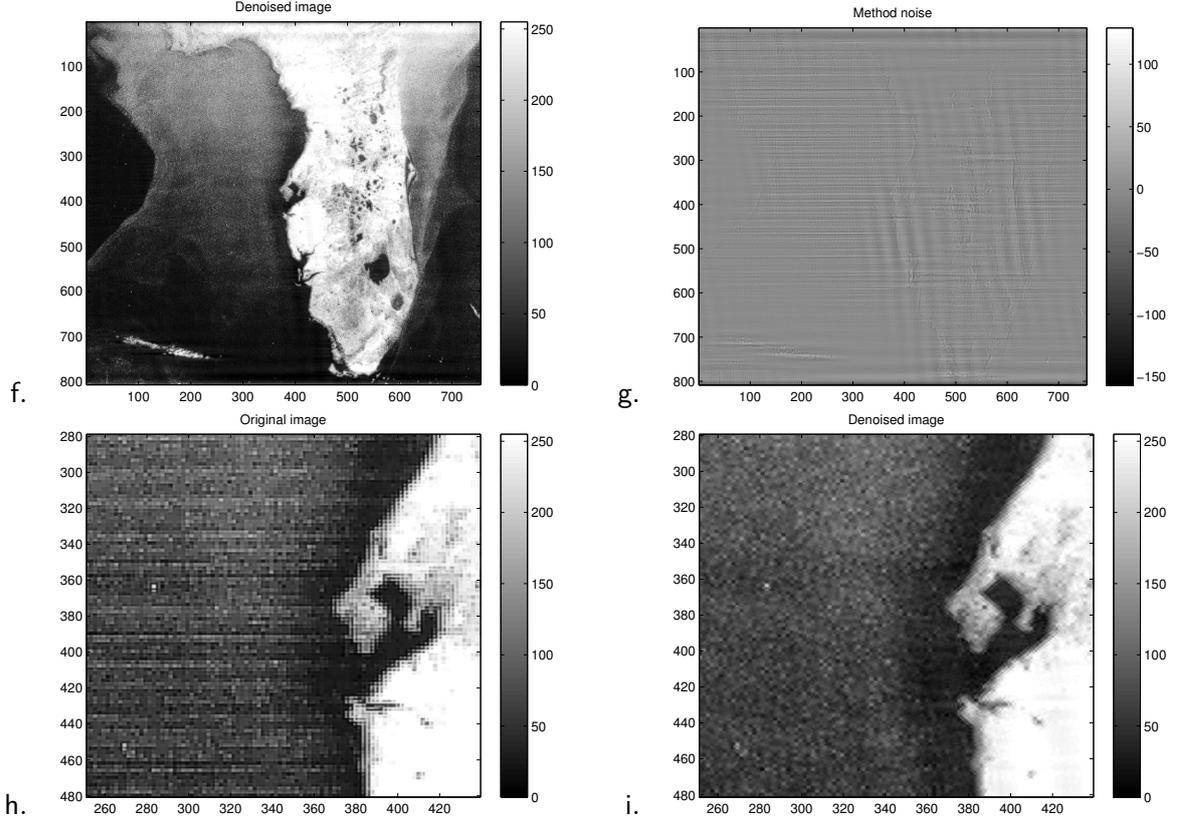


FIGURE 5.5 – Expérience Floride (2). f : image débruitée. g : estimation du bruit. h : détail de l'image bruitée. i : détail de l'image débruitée.

5.3 Détection a-contrario du bruit quasi-périodique

Nous envisageons dans cette section une détection a-contrario des « pics » du spectre causés par le bruit quasi-périodique à la place de la détection par la « règle des 3σ ». Nous ne rappelons pas ici le principe de la détection a-contrario, qui a été détaillée dans le chapitre 2.

Plutôt que considérer le spectre de puissance *moyen*, nous considérons le spectre de puissance *minimal* défini comme suit. Notons le coefficient de Fourier $c_{n,m}^p$ de fréquences (n, m) cycles par image de la p -ème imagerie. Nous souhaitons caractériser les couples (n, m) pour lesquels le minimum des coefficients du spectre d'amplitude $c_{n,m} = \min_{1 \leq p \leq P} |c_{n,m}^p|$ ait une valeur trop grande pour provenir d'imageries d'une image non-bruitée. Dans ce cas, c'est qu'il y a un pic significatif en (n, m) dans le spectre de *toutes* les imageries.

On définit le processus de fond pour (n, m) comme $C_{n,m} = \min_{1 \leq p \leq P} |C_{n,m}^p|$ avec les $C_{n,m}^p$ désignant des variables aléatoires indépendantes et identiquement distribuées, leur fonction de répartition complémentaire étant notée $F_{n,m}$ (il s'agit du complément à 1 de la fonction de répartition). Grâce à l'hypothèse d'indépendance, on calcule :

$$\Pr(C_{n,m} \geq x) = \Pr\left(\min_{1 \leq p \leq P} |C_{n,m}^p| \geq x\right) = \Pr(\forall p, |C_{n,m}^p| \geq x) = (F_{n,m}(x))^P \quad (5.6)$$

Le problème est de modéliser la fonction de répartition des $C_{n,m}^p$ selon le modèle de fond,

c'est-à-dire dans une image non bruitée. La décroissance attendue en fonction de la fréquence nous amène à couvrir le spectre par $R = L/8$ anneaux concentriques (centrés sur les basses fréquences), comme dans la figure 5.6 d, la fonction de répartition étant la même pour les coefficients appartenant au même anneau. Chaque fonction de répartition complémentaire est alors estimée empiriquement à partir des coefficients de Fourier des P imagelettes appartenant à l'ensemble considéré. En d'autres termes, pour tout x ,

$$F_{\mathcal{R}(n,m)}(x) = \frac{1}{P\#\mathcal{R}(n,m)} \#\left\{ (n', m', p) \text{ t.q. } |c_{n',m'}^p| \geq x, (n', m') \in \mathcal{R}(n,m), 1 \leq p \leq P \right\} \quad (5.7)$$

où $\#\cdot$ dénote le cardinal de tout ensemble fini, et $\mathcal{R}(n,m)$ désigne l'anneau auquel (n,m) appartient.

Par hypothèse, dans un anneau donné, seulement une petite minorité des coefficients peuvent appartenir à un pic causé par le bruit quasi-périodique. Il est donc valide de considérer que $F_{\mathcal{R}(n,m)}(x)$ ainsi estimé est la probabilité pour que $|C_{n,m}| \geq x$ dans une image non-bruitée, en accord avec le processus de fond. Le nombre attendu de fausses alarmes sur un anneau donné \mathcal{R} (où la fonction de répartition complémentaire est $F_{\mathcal{R}}$) est donc $\#\mathcal{R} \cdot F_{\mathcal{R}}(x)$. Comme il y a R tels anneaux,

$$\text{NFA}(|c_{n,m}|) = R\#\mathcal{R}(n,m) \left(F_{\mathcal{R}(n,m)}(|c_{n,m}|) \right)^P \quad (5.8)$$

est une définition valide du nombre de fausses alarmes (NFA) (voir la Définition 4 et la Proposition 2 dans [GM09]), dans le sens où il satisfait :

Proposition 5.1 *L'espérance du nombre de coefficients spectraux suivant le modèle de fond ayant un NFA inférieur à ε est plus petit que ε , c'est-à-dire :*

$$E(\#\{(n,m) \text{ t.q. } \text{NFA}(|C_{n,m}|) \leq \varepsilon\}) \leq \varepsilon \quad (5.9)$$

Preuve. La preuve est classique et basée sur le lemme suivant donnant une majoration des p -valeurs d'un estimateur statistique.

Lemme 5.1 *Pour toute variable aléatoire X et tous réels s et t , si $F(t) = \Pr(X \geq t)$, alors $\Pr(F(X) \leq s) \leq s$.*

Soit $\mathcal{F}_{n,m} = F_{\mathcal{R}(n,m)}^P$. D'après l'équation (5.8), $\text{NFA}(C_{n,m}) = R\#\mathcal{R}(n,m)\mathcal{F}_{n,m}(C_{n,m})$. On calcule alors successivement :

$$\begin{aligned} E(\#\{(n,m) \text{ t.q. } \text{NFA}(C_{n,m}) \leq \varepsilon\}) &= E\left(\sum_{n,m} \mathbb{1}_{\mathcal{F}_{n,m}(C_{n,m}) \leq \frac{\varepsilon}{R\#\mathcal{R}(n,m)}}(C_{n,m})\right) \\ &= \sum_{n,m} \Pr\left(\mathcal{F}_{n,m}(C_{n,m}) \leq \frac{\varepsilon}{R\#\mathcal{R}(n,m)}\right) \\ &\leq \sum_{n,m} \frac{\varepsilon}{R\#\mathcal{R}(n,m)} \\ &\leq \varepsilon \sum_{i=1}^R \sum_{(n,m) \in \mathcal{R}_i} \frac{1}{R\#\mathcal{R}(n,m)} \leq \varepsilon \end{aligned}$$

En effet, $\sum_{(n,m) \in \mathcal{R}_i} \frac{1}{\#\mathcal{R}(n,m)} = 1$; ce qui prouve la proposition 5.1. \square

Algorithme 5.2 Élimination de bruit quasi-périodique par méthode a-contrario.

Entrée : image i (taille $X \times Y$, et transformée de Fourier I), taille d'imagette L .

1. Extraction des imagettes non-recouvrantes (pour se conformer à l'hypothèse de non-indépendance dans (5.6)) de taille $L \times L$ distribués uniformément sur i (ce qui fournit P imagettes) et calcul de $F_{\mathcal{R}}$ par l'intermédiaire de (5.7).
2. Calcul du spectre de puissance minimum : pour tout (n, m) , $|c_{n,m}| = \min_{p=1\dots P} |c_{n,m}^p|$.
3. Pour tout (n, m) , calcul de $\text{NFA}(|c_{n,m}|)$ avec (5.8).
4. Construction de la carte des pics $M_o^{\mathcal{P}}$ dus au bruit quasi-périodique dans le spectre de taille $L \times L$ tel que $M_o^{\mathcal{P}}(n, m) = 1$ si $\text{NFA}(|c_{n,m}|) \leq 1$, et 0 sinon.
5. Interpolation de $M_o^{\mathcal{P}}$ de la taille $L \times L$ à $X \times Y$, ce qui donne une carte M_o des pics dans le spectre de l'image originale.

Sortie : \hat{n} , estimation du bruit quasi-périodique comme la transformée de Fourier inverse de $M_o I$, et \hat{i} , estimation de l'image débruitée, comme $i - \hat{n}$ (i.e., transformée de Fourier inverse de $(1 - M_o)I$).

Intuitivement, les fréquences (n, m) telles que $\min_p c_{n,m}^p$ a un NFA inférieur à 1 ne viennent probablement pas d'une image non-bruitée, et peuvent être considérées comme venant du bruit quasi-périodique.

Remarquons au passage que la détection a-contrario de structures spectrales est utilisé dans [CM10] pour détecter un *aliasing* et dans [XLF07] pour caractériser le flou de bougé, sans, néanmoins, proposer de modélisation statistique du spectre en général ou du spectre minimum en particulier.

Une fois les pics détectés, la même stratégie que celle décrite dans la section 5.2 est employée pour construire un filtre coupe-bande. L'algorithme 5.2 décrit la méthode d'élimination du bruit quasi-périodique résultante.

5.3.1 Exemple illustratif

La figure 5.6 montre un exemple illustratif et les principales étapes de l'algorithme. **a** est une image obtenue par une sonde Mariner 4, affectée d'un bruit quasi-périodique, et **b** montre son spectre de puissance. **c** montre le spectre de puissance minimum (étape 2, ici $L = 128$ pixels). Quatre pics sont clairement visibles, en plus de la composante basse-fréquence centrale. **d** montre la distribution des $R = L/8 = 16$ anneaux concentriques sur lesquels l'amplitude des coefficients de Fourier est supposée identiquement distribuée. Le disque central (en bleu foncé) n'est pas pris en compte dans l'algorithme de détection car les basses fréquences ne correspondent pas à des motifs répétés, et de plus l'hypothèse de distribution identique est douteuse dans ce disque. **e** est la carte des logarithmes du NFA (avec une valeur arbitraire dans le rond central) : un grande majorité des coefficients donne un NFA supérieur à 1 ($\log(\text{NFA}) > 0$), exceptés quatre pics avec un NFA très faible ($\simeq 10^{-180}$). **f** est le spectre corrigé $(1 - M_o)I$ (étape 5) : on peut voir qu'en plus des autres pics, quelques coefficients ont été détectés comme significatifs (avec un NFA $\simeq 10^{-1}$, donc proche du seuil de 1). **g** est l'image débruitée \hat{i} , et **h** est l'estimation de la composante de bruit \hat{n} (étape 6). Des détails sont montrés en **i**.

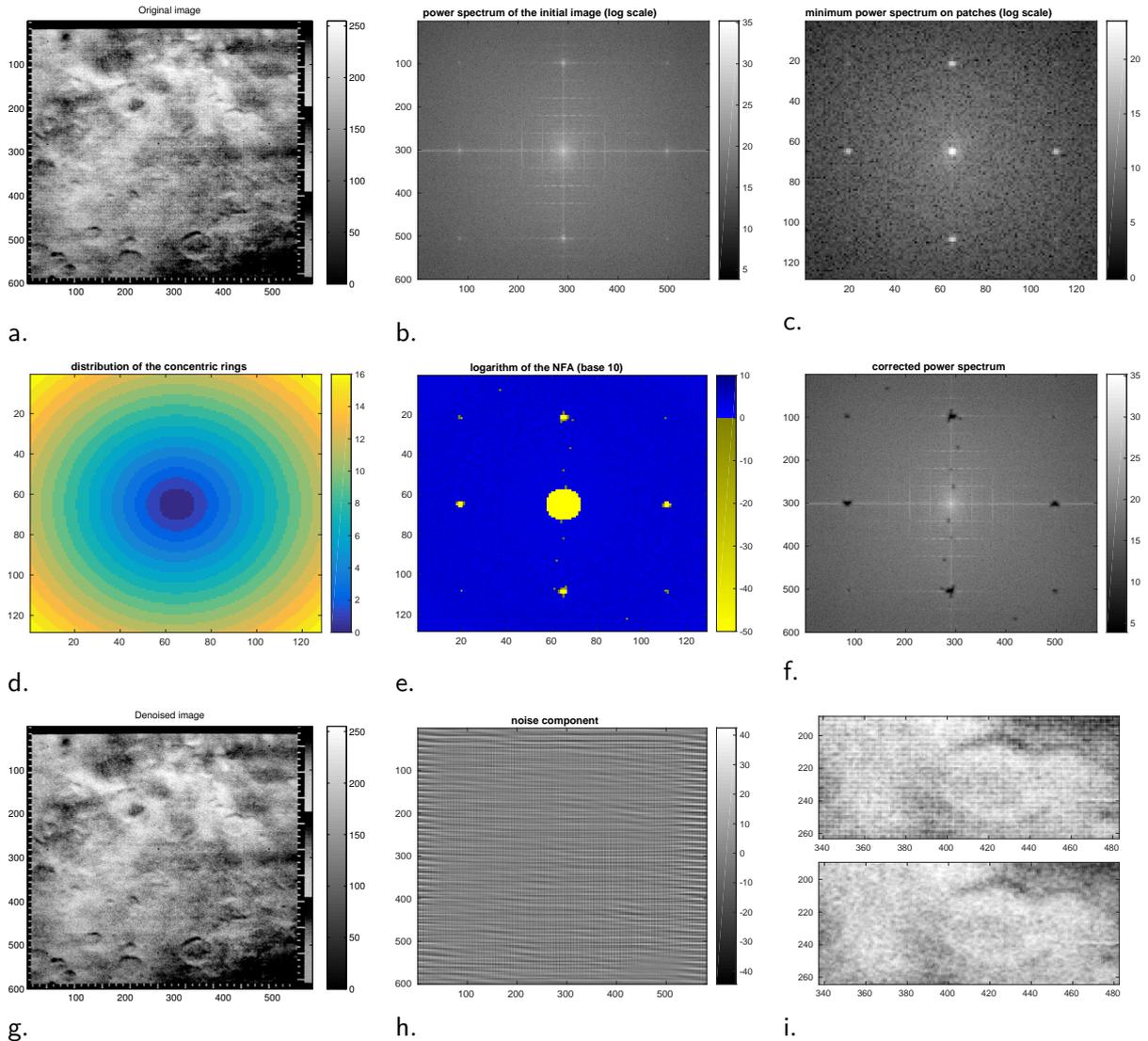


FIGURE 5.6 – a : image originale. b : spectre de puissance. c : spectre de puissance minimum sur les imquettes de taille 128×128 pixels. d : distribution des 16 anneaux concentriques. e : carte du NFA. f : spectre de puissance où les composantes spectrale du bruit ont été filtrées. g : image débruitée. h : composante de bruit. i : détail des images originales et débruitées.

5.3.2 Quelques expériences complémentaires

D'autres expériences sont présentées sur la figure 5.7. D'abord, un bruit périodique synthétique a été ajouté à l'image Mandrill. L'image sous-jacente et le bruit sont visuellement bien séparées par notre algorithme. Cela montre que la détection a-contrario est capable de détecter les composantes spectrales du bruit même en présence d'une texture haute-fréquence (la fourrure) dans la grande majorité des imquettes. Les deux autres images montrent des expériences réelles.

L'image Lunar Orbiter est affectée de *striping*, qui est correctement enlevé. L'image Mariner 6 est affecté par des motifs périodiques dont le spectre est bien plus complexe que quelques

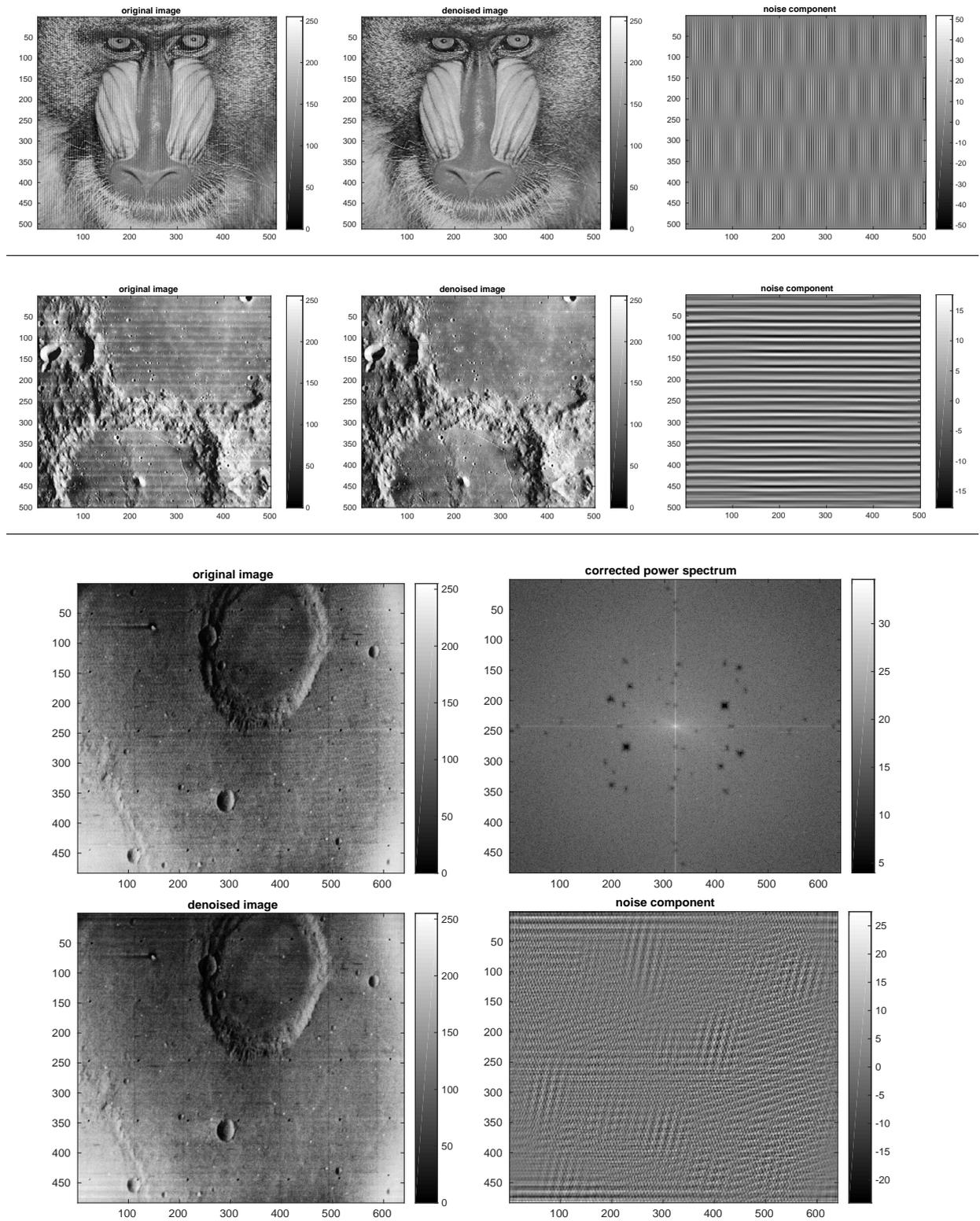


FIGURE 5.7 – De haut en bas : expériences Mandrill, Lunar Orbiter, et Mariner 6.



FIGURE 5.8 – À gauche : un détail de Carry on, de Norman Rockwell. La trame de la toile est bien visible. Au centre : bruit quasi-périodique éliminé. À droite : le bruit quasi-périodique.

pics bien séparés. L'algorithme parvient encore à séparer les composantes du bruit de l'image.

La figure 5.8 présente les résultats de la séparation entre bruit quasi-périodique et image sous-jacente sur une image couleur, en opérant sur chaque canal rouge, vert et bleu indépendamment. Il s'agit d'un exemple d'élimination de la trame de la toile visible sur les images à haute résolution de peinture. Ces problèmes sont discutés dès les années 1980 dans [HMdC87], où des modèles auto-régressifs de textures sont utilisés pour le filtrage, et dans des articles récents [CDCS12, CYG⁺16], comme précédemment mentionné.

5.4 Interpolation des coefficients spectraux par minimisation sous contrainte de la variation totale

Dans certaines situations, appliquer un filtre coupe-bande à l'image induit un effet de *ringing*. Pour éviter cet effet, il est nécessaire d'interpoler les coefficients de Fourier annulés par le filtre. Des méthodes basiques d'interpolation sont décrites dans [7] et [AB08]. En suivant [GM98] (voir aussi [Moi01, ZF05], nous proposons dans cette section une interpolation par minimisation sous contrainte de la variation totale.

Soit u l'image affectée par un bruit quasi-périodique, et Ω l'ensemble des fréquences où la contribution du bruit a été localisée par l'un des algorithmes décrits dans les sections précédentes. Le problème est de trouver \tilde{u} de manière à ce que les spectres de \tilde{u} et u coïncident sur le complémentaire de Ω (noté Ω^C) et à ce que le spectre de \tilde{u} restreint à Ω permette de minimiser la variation totale de \tilde{u} .

Nous considérons la variation totale régularisée d'une image v , définie par :

$$\text{TV}_\varepsilon(v) = \int \|\nabla v\|_\varepsilon \quad (5.10)$$

où $\|\nabla v\|_\varepsilon = \sqrt{\varepsilon^2 + \|\nabla v\|^2}$, $\|\cdot\|$ étant la norme euclidienne dans \mathbb{R}^2 et ∇ le gradient. Avec

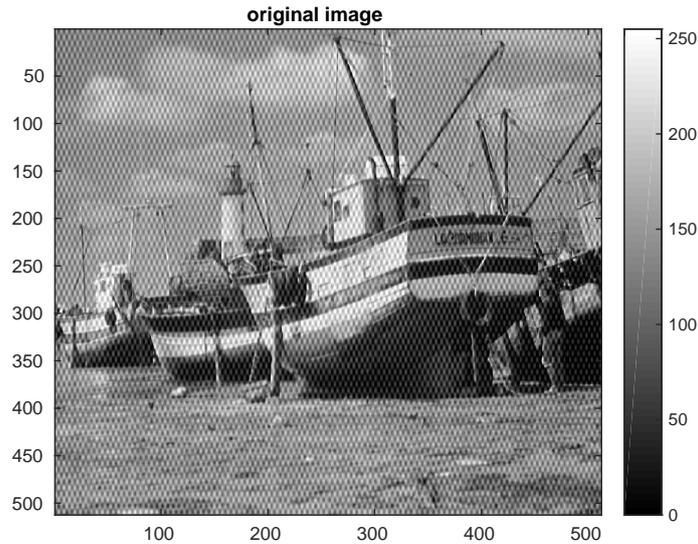


FIGURE 5.9 – Une image affectée par un bruit quasi-périodique synthétique.

une petite valeur de ε , $\text{TV}_\varepsilon(v)$ est différentiable aux points où le gradient de v s'annule, sans trop diffuser de la notion standard de variation totale.

Le problème est de trouver \tilde{u} , solution du problème suivant de minimisation sous contrainte :

$$\begin{cases} \tilde{u} = \operatorname{argmin}_v \text{TV}_\varepsilon(v) \\ \text{sous contrainte : } \mathcal{F}(v)|_{\Omega^C} = \mathcal{F}(u)|_{\Omega^C} \end{cases} \quad (5.11)$$

où $\mathcal{F}(\cdot)$ désigne la transformée de Fourier d'une image.

Comme dans [GM98, Moi01], nous utilisons un algorithme de descente de gradient projeté.

La figure 5.9 montre une image affectée d'un bruit quasi-périodique synthétique. On voit dans la figure 5.10 l'effet de *ringing* provoqué par le filtre coupe-bande : la raison en est ici la superposition de la contribution du bruit avec la contribution de segments de l'image (des mats obliques sont parallèles à une direction d'ondulation du bruit périodique). Comme noté dans [Moi01], le *ringing* est moins visible dans les régions texturées que dans les régions uniformes. La minimisation TV sous contrainte permet alors d'interpoler les coefficients spectraux correspondant à cette contribution de manière à éliminer l'effet de ringing.

La figure 5.11 compare un détail avant et après minimisation TV. On voit que l'interpolation des coefficients spectraux « reconstruit » correctement le segment causé par les mats obliques visible dans le spectre.

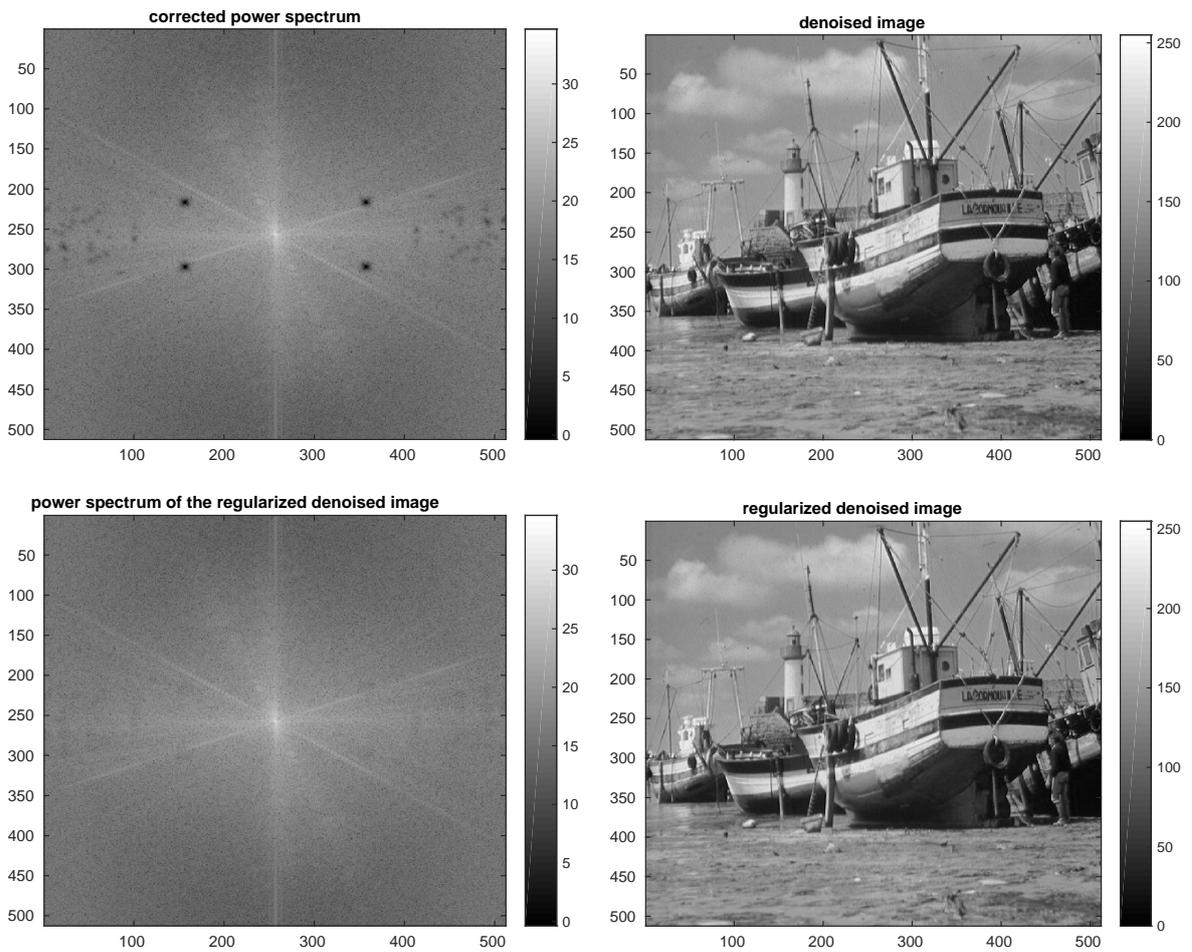


FIGURE 5.10 – En haut : spectre de puissance après application du filtre coupe-bande, et image correspondante débruitée. En bas : spectre de puissance de l'image reconstruite par minimisation TV sous contrainte, et image correspondante.

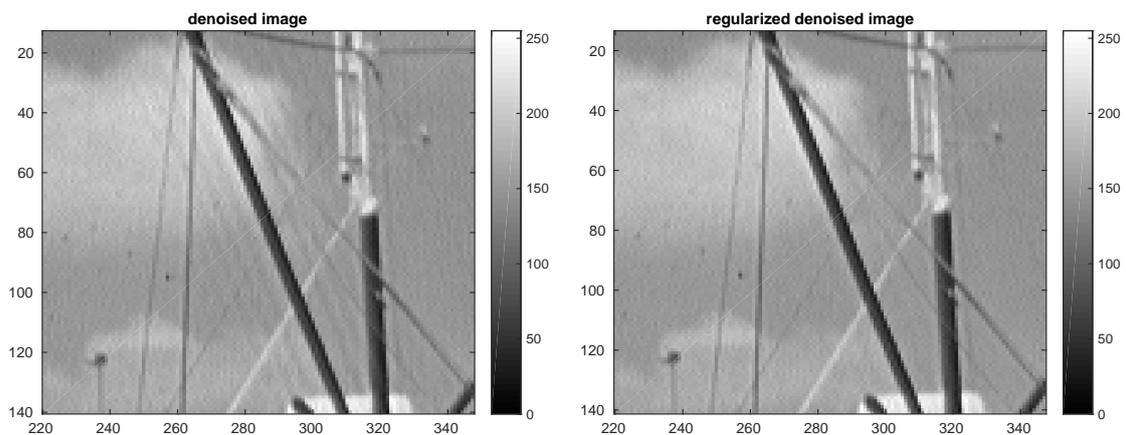


FIGURE 5.11 – Détail du résultat du filtre passe-bande (à gauche) et de la minimisation TV sous contrainte (à droite). L'effet de ringing cernant le mat oblique a quasiment disparu.

5.5 Conclusion et perspectives

Dans ce chapitre, nous avons présenté deux algorithmes de détection d'un bruit quasi-périodique. Le premier est basé sur le spectre de puissance moyen dans lequel les composantes du bruit sont détectées comme valeurs aberrantes dans la distribution attendue des coefficients spectraux en $1/f^\alpha$, le second est basé sur le spectre de puissance minimum et les composantes du bruits sont détectées comme événements significatifs dans un modèle a-contrario.

Remarquons que les deux approches s'appliquent à des images pour lesquels les imagerie présentent des propriétés statistiques semblables, donc essentiellement des images naturelles. Nous avons pu observer par exemple que nos algorithmes ne permettent pas la détection automatique du bruit quasi-périodique dans des cartes de déformation issues de la mécanique des solides expérimentale, ce bruit pouvant être causé par des défauts de grille ou un problème d'*aliasing* (voir la discussion de la section 6.2.2 dans le chapitre 6). Dans ce cas, la détection des composantes spectrales du bruit par un expert reste nécessaire, voir [7]. Les spectres de ces images présentant la particularité d'être fortement anisotropes, on pourrait analyser une détection a-contrario dans laquelle les fonctions de répartition empirique ne seraient plus estimées sur des régions en anneau (comme sur la figure 5.6 d) mais en imposant une subdivision angulaire supplémentaire de ces anneaux.

Il est possible d'étendre ce travail en envisageant un modèle a-contrario pour le spectre de puissance moyen. Avec les notations de (5.6), la distribution de $\tilde{C}_{n,m} = \frac{1}{P} \sum_{p=1}^P |C_{n,m}^p|$ est une convolution normalisée itérée P fois de la distribution de $|\tilde{C}_{n,m}|$ par elle-même. Un NFA peut alors être défini de manière similaire à l'équation (5.6), dans le même esprit que ce qui est fait dans [RDG09] pour l'appariement de descripteurs SIFT vus comme des histogrammes circulaires. D'après le théorème limite central, la distribution des $\tilde{C}_{n,m}$ peut même être approchée, pour de grandes valeurs de P , par une distribution normale de valeur moyenne approximativement égale à $1/f^\alpha$. De la même manière, on pourrait envisager de remplacer la répartition empirique $F_{n,m}^P$ dans (5.6) par une fonction paramétrique issue des distributions de Gumbel, Fréchet ou Weibull, selon la théorie des valeurs extrêmes [Col01]. La détection des composantes du bruit peut alors se faire par détection de grandes déviations ou par détection a-contrario.

On peut remarquer que lorsque le bruit quasi-périodique cherché est spatialement localisé (ce qui est le cas en présence de saturation locale de l'image, ou pour l'*aliasing*), ses composantes sont éliminées du spectre de puissance minimum dès qu'il n'est pas visible dans une des imagerie. Il sera par contre visible dans le spectre de puissance moyen. Formaliser une détection de zones de l'image présentant des propriétés spectrales communes serait intéressant. Des perspectives de recherche relatives à ce sujet sont discutées dans le chapitre 7.

Enfin, même si l'algorithme de descente de gradient projeté de la section 5.4 donne ici des résultats satisfaisants (l'explication étant qu'il y a peu de coefficients spectraux à interpoler), il semble pertinent d'envisager les méthodes récentes d'interpolation ou extrapolation de spectre décrites dans [AM16, CJ14, Jal12], basées sur des algorithmes primal-dual.

Troisième partie

Contributions à la mécanique des solides expérimentale

Quantification des performances métrologiques en mécanique des solides expérimentale

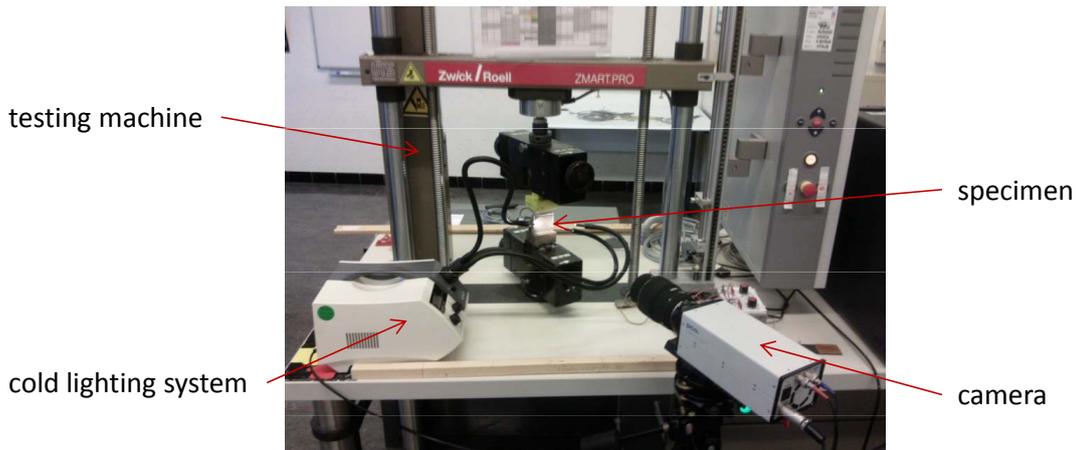
6.1 Introduction	101
6.2 Quantification des performances métrologiques de la méthode de la grille	104
6.3 Restauration des champs de déplacement et déformation	110
6.4 Quantification des performances métrologiques de la méthode basée sur la corrélation d'images (DIC)	113
6.5 Conclusion et perspectives	116

6.1 Introduction

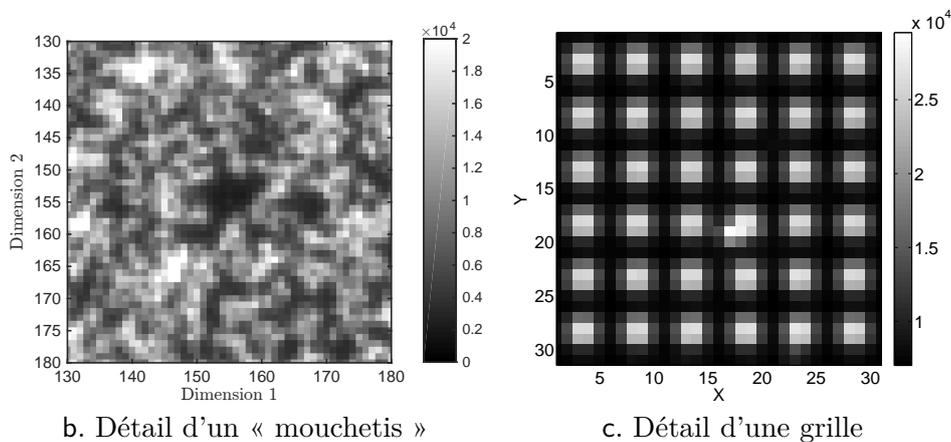
6.1.1 Contexte

Un des problèmes de la mécanique des solides expérimentale est la mesure des déformations subies par un matériau soumis à des sollicitations mécaniques. Les jauges de déformation électriques sont utilisées classiquement, mais elles ne fournissent qu'une information ponctuelle. Des techniques optiques de mesure sans contact ont émergé ces dernières années. Elles fournissent des champs entiers de mesures, ce qui est particulièrement intéressant lorsque le champ de déformation présente une hétérogénéité. Ces techniques sont basées sur la comparaison d'images de la surface du matériau considéré, avant et après sollicitation mécanique et donc déformation. Plus précisément, un motif contrasté est déposé sur la surface du matériau, et l'évolution de l'image du motif permet de déduire un champ de déplacement, puis de déformation. Deux principales techniques ont été développées ; elles se distinguent par la nature du motif contrasté. L'une est la méthode de grille [BGM09, BGMR09], qui est basée sur l'analyse spectrale d'une grille régulière transférée sur la surface du matériau à analyser. L'autre est la corrélation d'images (*digital image correlation*, DIC) basée sur un « mouchetis » (*speckle*) déposé à la surface du matériau [SOS09]. Notons que la surface nue du matériau peut suffire dans certains cas à établir un champ de déplacement par DIC.

Le problème est caractérisé par les très faibles déformations généralement mesurées (souvent inférieures au pourcent), qui font que les déformations se traduisent essentiellement par



a. Un essai mécanique typique



b. Détail d'un « mouchetis »

c. Détail d'une grille

FIGURE 6.1 – Illustration du protocole expérimental. a : une éprouvette est fixée aux mors d'une machine d'essai qui permet d'imposer une compression ou une traction. Une caméra permet de prendre une vue fronto-parallèle de la surface de l'éprouvette. Des images de la surface de l'éprouvette sont réalisées avant et après sollicitation mécanique. b : exemple d'image de « mouchetis » déposé à la surface de l'échantillon. c : exemple d'image de grille.

une faible variation de l'intensité lumineuse mesurée en chaque pixel, difficilement visible à l'œil nu. Un problème majeur est de quantifier les performances métrologiques de ces méthodes afin qu'elles puissent être utilisées de manière industrielle. Si la résolution des mesures ponctuelles obtenues par les jauges de déformation électriques peut être caractérisée, ce n'est pas encore le cas des mesures de champs sans contact, qui ne donnent souvent qu'une valeur numérique non certifiée.

Un environnement expérimental typique est décrit sur la figure 6.1. Pour donner un ordre de grandeur, un pixel représente une surface de côté 40 micromètres sur l'échantillon. Dans le cas de l'image de grille, le pas est de 5 pixels. La figure 6.2 montre un exemple de champ de déformation obtenu par la méthode de grille.

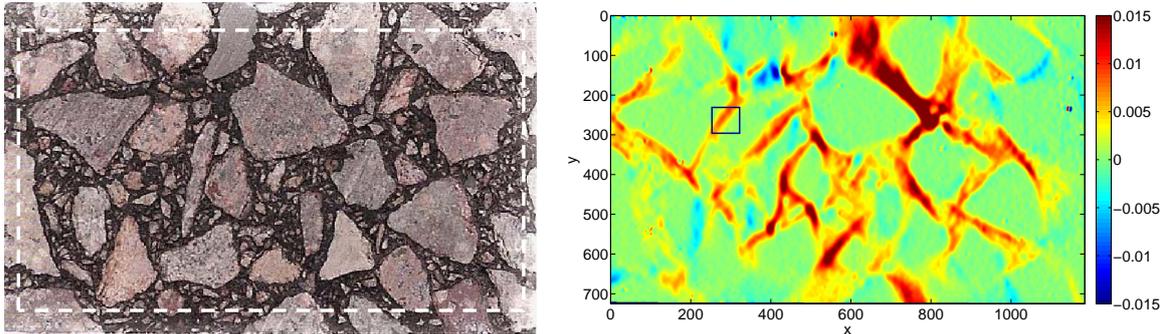


FIGURE 6.2 – Test de compression d'un échantillon d'asphalte (à gauche), et champ de déformation résultant obtenu par la méthode de grille (à droite). Notons qu'un asphalte est formé de granulats enrobés dans du bitume, et que les granulats ne se déforment pas. Cette illustration est tirée de [GT13].

6.1.2 Contributions

Ce chapitre présente nos contributions à l'analyse et au traitement des images issues de la mécanique des solides expérimentale. Les résultats obtenus sont le fruit d'une collaboration avec Michel Grédiac et Benoît Blaysat, respectivement professeur et maître de conférences à l'Institut Pascal à Clermont-Ferrand, dont le domaine de recherche est précisément la mécanique expérimentale. Nous présentons nos résultats concernant la caractérisation des performances métrologiques de la méthode de grille dans la section 6.2. Ces performances sont essentiellement limitées par la méthode numérique d'estimation ainsi que par le bruit inhérent au capteur numérique utilisé pour l'acquisition des images. La section 6.3 propose des améliorations de ces performances par différentes techniques de restauration d'images. Enfin, dans la section 6.4 nous proposons une nouvelle formule prédictive de la résolution du déplacement lorsqu'il est mesuré par corrélation d'images. Pour les deux méthodes, nous nous attachons à vérifier expérimentalement les formules prédictives théoriques obtenues, en particulier en tenant compte de la nature hétéroscédastique du bruit réel.

Les principales publications relatives à ce chapitre sont :

[2] F. Sur, B. Blaysat, and M. Grédiac. Determining displacement and strain maps immune from aliasing effect with the grid method. *Optics and Lasers in Engineering*, 86:317–328, 2016.

[3] F. Sur and M. Grédiac. Influence of the analysis window on the metrological performance of the grid method. *Journal of Mathematical Imaging and Vision*, 56(3):472–498, 2016.

[4] M. Grédiac, F. Sur, and B. Blaysat. The grid method for in-plane displacement and strain measurement : a review and analysis. *Strain*, 52(3):205–243, 2016.

[5] B. Blaysat, M. Grédiac, and F. Sur. Effect of interpolation on noise propagation from images to DIC displacement maps. *International Journal for Numerical Methods in Engineering*, 108(3):213–232, 2016.

[6] B. Blaysat, M. Grédiac, and F. Sur. On the propagation of camera sensor noise to displacement maps obtained by DIC - an experimental study. *Experimental Mechanics*, 56(6):919–944, 2016.

[10] F. Sur and M. Grédiac. On noise reduction in strain maps obtained with the grid method by averaging images affected by vibrations. *Optics and Lasers in Engineering*, 66:210–222, 2015.

[12] M. Grédiac and F. Sur. Effect of sensor noise on the resolution and spatial resolution of displacement and strain maps estimated with the grid method. *Strain*, 50(1):1–27, 2014.

[13] F. Sur and M. Grédiac. Towards deconvolution to enhance the grid method for in-plane strain measurement. *AIMS Inverse Problems and Imaging*, 8(1):259–291, 2014.

[15] M. Grédiac, F. Sur, C. Badulescu, and J.-D. Mathias. Using deconvolution to improve the metrological performance of the grid method. *Optics and Lasers in Engineering*, 51(6):716–734, 2013.

Le code Matlab VerifWV (Verification of the Wigner-Ville transform) disponible à l’URL : <https://members.loria.fr/FSur/software/VerifWV/> implémente une vérification numérique des formules établies dans [3].

Le code Matlab The GridMethod toolbox disponible sur le site web dédié : <http://www.thegridmethod.net> implémente l’ensemble de nos codes Matlab permettant l’analyse des images de grilles et l’estimation des champs de déplacement et déformation.

6.2 Quantification des performances métrologiques de la méthode de la grille

Dans nos premiers travaux, nous nous sommes focalisés sur la méthode de grille afin d’en quantifier les performances métrologiques et de les améliorer par différentes techniques. Nous avons d’abord étudié un modèle mathématique des images de grille, sous hypothèse de bruit blanc gaussien dans les images. Nous avons ensuite montré que le champ de déformation obtenu par la transformée de Fourier à fenêtre est en fait au premier ordre la convolution de la “vraie” déformation et de la fenêtre d’analyse. Nous avons aussi caractérisé le transfert du bruit numérique dans les champs de déformation sous forme d’un bruit auto-corrélé. Ces résultats théoriques reposent sur l’hypothèse des petites déformations qui justifie les approximations au premier ordre. Nous en avons ensuite déduit une caractérisation métrologique de la méthode de la grille par l’intermédiaire des biais de mesure, résolution de mesure, et résolution spatiale.

6.2.1 Modèle d’image de grille et estimation des champs de déplacement et déformation

Une grille, c’est-à-dire deux réseaux orthogonaux de lignes parallèles légèrement déformées par les contraintes mécaniques, peut être modélisée de la manière suivante lorsque les lignes sont alignées sur les axes de l’image (voir [BGM09, Sur00]).

$$s(x, y) = \frac{A(x, y)}{2} \left(2 + \gamma \cdot \ell(2\pi f x + \phi_x(x, y)) + \gamma \cdot \ell(2\pi f y + \phi_y(x, y)) \right) \quad (6.1)$$

où :

- $A(x, y) > 0$ représente l’illumination globale, qui peut varier légèrement sur la surface du champ imagé ;
- $\gamma \in [0, 1]$ est le contraste des lignes ;
- le profil d’une ligne ℓ est une fonction réelle 2π -périodique d’amplitude 1 et de valeur moyenne 0 sur une période. Nous notons d_k les coefficients de Fourier de ℓ ;
- f est la fréquence nominale de la grille (supposée identique dans les deux dimensions), le pas de grille étant $p = 1/f$.
- $\phi_x(x, y)$ et $\phi_y(x, y)$ représentent les phases selon les axes x et y respectivement, modulées spatialement.

Notons que ce modèle est valide pour les images de grilles avant et après déformation, les phases avant déformation modélisant les défauts de fabrication des grilles employées. Dans tous les cas, les dérivées des phases sont très petites par rapport à $2\pi f$. D'autre part, si un profil sinusoïdal serait idéal [12], il n'est pas possible de l'obtenir en pratique. C'est pourquoi un certain nombre d'harmoniques d_k , $|k| > 1$, sont observées. Il est possible d'aligner les grilles sur l'image car des réglages fins de la position de la caméra sont disponibles et on utilise un objectif macrophotographique présentant de très faibles distorsions géométriques sur le champ du capteur de la caméra.

Les champs de déplacement et déformation étant liés aux phases, il convient d'abord de les estimer. La différence par rapport aux méthodes générales d'analyse de franges en interférométrie [Kem07, Kem15] est que la fréquence spatiale de la grille est connue et n'a pas à être estimée au préalable. Une méthode utilisant la *Geometric Phase Analysis* (GPA) a récemment été proposée [DXW⁺14a, DXW14b]. Elle est inspirée de procédures classiques pour le traitement de franges en interférométrie, par exemple pour la profilométrie [Tak90, TM83] ou pour l'observation microscopique de réseaux atomiques [HSK98]. Il s'agit essentiellement de filtrer les « lobes » dans le spectre causés par les modulations de phase et centrés sur la fréquence nominale de la grille dans chaque direction, puis de récupérer les phases par transformée de Fourier inverse. La quantification des performances métrologiques de cette méthode semble difficile : cela nécessite en particulier de quantifier le « bon » filtrage des lobes dans le spectre initial. Une autre approche, à la suite de [Sur00], a été proposée pour l'estimation des phases, basée elle uniquement sur la fréquence nominale de la grille [BGM09, BGMR09] et d'une complexité algorithmique considérablement réduite. Dans ce dernier cas, nous avons établi des expressions quantifiant les différents paramètres métrologiques de la méthode. Dans la suite du chapitre, nous focaliserons l'analyse sur cette méthode. Notons au passage que les auteurs de [DXW⁺14a, DXW14b] semblent appliquer un filtre sur le spectre des images de grille de manière à ne garder que les composantes des lobes qui émergent du bruit. Selon la valeur du seuil, il se peut très bien que la GPA soit en fait basée uniquement sur la composante de fréquence égale à celle de la grille, à l'instar de la méthode proposée dans [BGM09, BGMR09].

La méthode de [BGM09, BGMR09] consiste à calculer la transformée de Fourier à fenêtre de l'image de grille, puis à estimer les phases comme les arguments de la valeur de cette transformée aux harmoniques principales de la grille (fréquences $(f, 0)$ et $(0, f)$). Autrement dit, si on note

$$\mathcal{F}_w(f)(x, y, u, v) = \iint_{\mathbb{R}^2} s(\xi, \eta) w(\xi - x, \eta - y) e^{-2i\pi(\xi u + \eta v)} d\xi d\eta \quad (6.2)$$

où w est la fenêtre d'analyse, les phases (à une constante additive près) sont estimées par :

$$\begin{cases} \phi_x(x, y) \simeq \arg(\mathcal{F}_w(f)(x, y, f, 0)) \\ \phi_y(x, y) \simeq \arg(\mathcal{F}_w(f)(x, y, 0, f)) \end{cases} \quad (6.3)$$

où \arg désigne l'argument de tout nombre complexe non nul. Les phases étant déterminées modulo 2π , l'utilisation d'un algorithme de dépliement est nécessaire [GP98, HBLG02]. On verra dans la section suivante une expression plus fine de cette approximation. Notons que $\mathcal{F}_w(f)(x, y, f, 0)$ et $\mathcal{F}_w(f)(x, y, 0, f)$ s'expriment comme des produits de convolution entre le signal complexe $se^{-2i\pi f \cdot}$ et la fenêtre symétrique w . Ceci permet en pratique de calculer ces quantités efficacement à l'aide d'une transformée de Fourier rapide 2D [4].

Notons $\phi^1 = (\phi_x^1 \ \phi_y^1)^T$ et $\phi^2 = (\phi_x^2 \ \phi_y^2)^T$ les vecteurs des phases en tout point, respectivement avant et après déformation, et $\mathbf{u} = (u_x, u_y)$ le champ de déplacement.

Le champ de déplacement à la surface de l'échantillon vérifie [4] :

$$\mathbf{u}(\mathbf{x}) = -\frac{p}{2\pi} \left(\phi^2(\mathbf{x} + \mathbf{u}(\mathbf{x})) - \phi^1(\mathbf{x}) \right) \quad (6.4)$$

On peut voir que le champ cherché \mathbf{u} est solution d'une équation de point fixe. Il s'avère que les itérations gouvernées par

$$\mathbf{u}^{n+1}(\mathbf{x}) = -\frac{p}{2\pi} \left(\phi^2(\mathbf{x} + \mathbf{u}^n(\mathbf{x})) - \phi^1(\mathbf{x}) \right) \quad (6.5)$$

pour n entier naturel, convergent vers le déplacement cherché \mathbf{u} . La fonctionnelle est en effet contractante : son coefficient de Lipschitz est très petit car les dérivées de phases le sont, et quelques itérations (voire une seule) suffisent en pratique à estimer le champ \mathbf{u} . Cette présentation et l'algorithme itératif résultant n'étaient à notre connaissance pas employés en mécanique expérimentale. Une estimation par simple différence de phases $\mathbf{u}^0 = -\frac{p}{2\pi} (\phi^2 - \phi^1)$ était utilisée dans la littérature, jusque l'introduction dans [BGM09] d'une estimation revenant à considérer $\mathbf{u}^1 = -\frac{p}{2\pi} (\phi^2(\cdot + \mathbf{u}^0) - \phi^1)$. Des résultats préliminaires suggèrent que l'algorithme itératif de (6.5) permet d'améliorer l'estimation de champs de déformation d'amplitude assez grande (10%), pour lesquels des artefacts numériques subsistent avec une seule itération.

Une fois que les composantes du champ de déplacement \mathbf{u} sont connues, la quantité intéressant les mécaniciens est le tenseur de déformations linéarisé, dont les composantes sont :

$$\epsilon_{ij} = \frac{1}{2} \left(\frac{\partial u_i}{\partial j} + \frac{\partial u_j}{\partial i} \right) \quad (6.6)$$

où $i, j \in \{x, y\}$.

Une revue de la littérature, une analyse complète des méthodes basées sur les grilles en mécanique expérimentale, ainsi qu'une discussion de leurs limites pratiques, sont disponibles dans [4].

6.2.2 Discussion du phénomène de repliement spectral

La transformée de Fourier d'une image de grille s (modélisée par l'équation (6.1)) s'écrit de la manière suivante [2] :

$$\begin{aligned} \mathcal{F}(s)(u, v) = \mathcal{F}(A)(u, v) + \frac{\gamma}{2} \sum_{k \in \mathbb{Z}} d_k \mathcal{F}(A) * \mathcal{F}\left(e^{ik\phi_x}\right)(u - kf, v) \\ + \frac{\gamma}{2} \sum_{k \in \mathbb{Z}} d_k \mathcal{F}(A) * \mathcal{F}\left(e^{ik\phi_y}\right)(u, v - kf) \end{aligned} \quad (6.7)$$

Si on suppose que les dérivées partielles de ϕ_l et ϕ_y sont bornées, on peut démontrer par la méthode de la phase stationnaire, que pour tout k , $\mathcal{F}\left(e^{ik\phi_l}\right)$ a un support essentiellement inclus dans

$$\left[-\frac{|k|}{p} M_l^x, \frac{|k|}{p} M_l^x \right] \times \left[-\frac{|k|}{p} M_l^y, \frac{|k|}{p} M_l^y \right] \quad (6.8)$$

où $M_l^x = (p/2\pi) \max(|\partial\phi_l/\partial x|)$ et $M_l^y = (p/2\pi) \max(|\partial\phi_l/\partial y|)$.

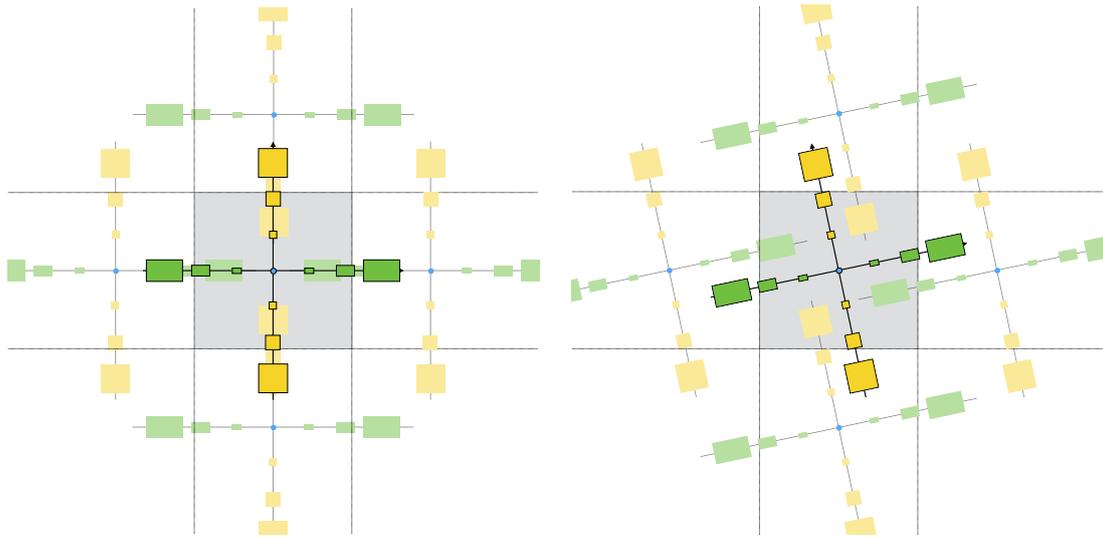


FIGURE 6.3 – Une illustration du phénomène d’aliasing sur le spectre d’une image de grille (à gauche, pour une grille alignée sur les pixels; à droite, pour une grille tournée). Les copies de $\mathcal{F}(s)$ translattées de $\{(0, 1), (0, -1), (1, 0), (-1, 0)\}$ (unité : nombre de cycles par pixel) sont montrées en couleur atténuée, superposée à $\mathcal{F}(s)$ non translatté centré sur le carré gris central. Dans l’image de gauche, les harmoniques principales en $(f, 0)$ et $(0, f)$ sont couvertes par un lobe venant des copies, « replié » sur la partie centrale grise. Dans l’image de droite, le simple fait de tourner la grille imagée fait que ces composantes repliées ne couvrent pas les harmoniques principales.

Ceci montre que le spectre d’une image de grille est essentiellement fait d’une composante centrale causée par le terme $\mathcal{F}(A)$ et de « lobes » centrés aux multiples de la fréquence f le long des axes verticaux et horizontaux, causés par les termes $\mathcal{F}(e^{ik\phi_x})$ and $\mathcal{F}(e^{ik\phi_y})$ et les harmoniques d_k du profil des lignes de la grille.

Ainsi, dès que $kf \leq 0.5$ cycle par pixel pour une k -ème harmonique non négligeable, l’échantillonnage provoque un repliement spectral (*aliasing*). Dans le cas typique où $f = 1/5$ pixel $^{-1}$, la troisième harmonique se replie. L’estimation des phases par (6.3), et par conséquence des champs de déplacement et de déformation, est donc susceptible d’être perturbée par le phénomène d’*aliasing* dès lors qu’un lobe associé à une harmonique se replie sur le lobe de l’harmonique principale. La transformée de Fourier étant covariante par rotation de l’image, on remarque qu’une simple rotation des grilles permet d’éviter que l’*aliasing* gêne l’estimation, comme illustré en figure 6.3. La figure 6.4 discute un exemple réel d’estimation de champ de déformation. Remarquons que cette astuce pratique permettant de limiter l’influence de l’*aliasing* est indépendante du procédé d’estimation des modulations de phase et s’applique à la GPA [DXW⁺14a, DXW14b, HSK98].

L’article [2] discute en détail l’influence de l’*aliasing* dans les méthodes de grille.

6.2.3 Biais de mesure, résolution de mesure, et résolution spatiale

Dans le cas d’images de grilles bruitées par un bruit blanc gaussien, nous montrons dans [13] qu’une estimation plus fine que celle de l’équation (6.3) est obtenue à l’ordre 1.

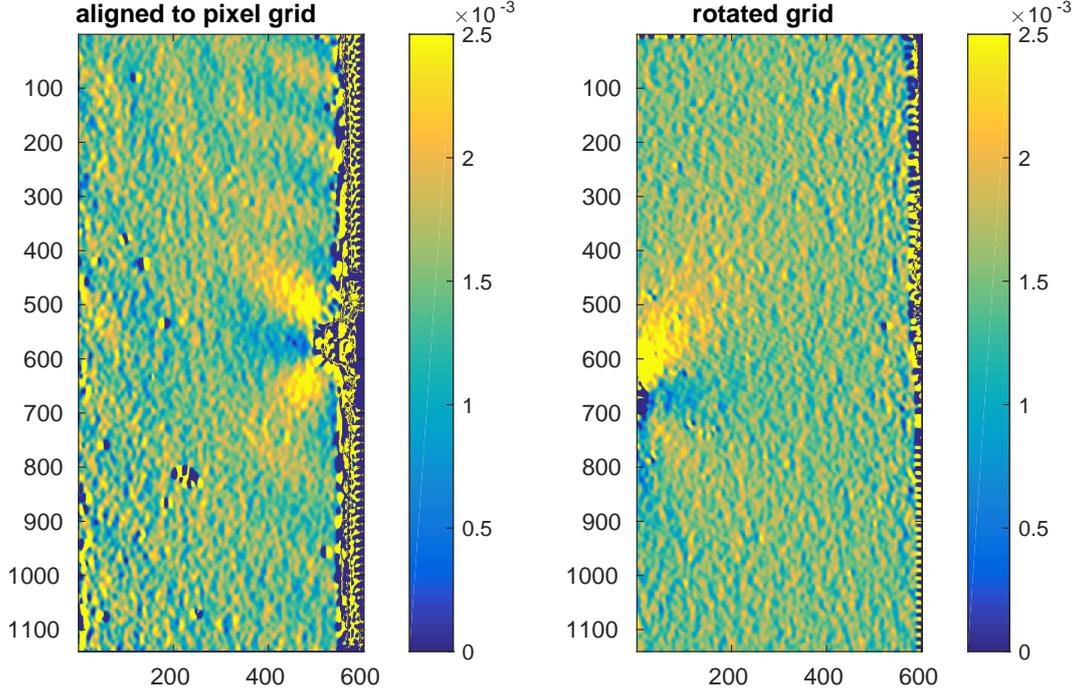


FIGURE 6.4 – Dans cette expérience, deux grilles sont collées de part et d'autre d'une éprouvette trouée subissant un essai de traction. La grille de gauche présente des lignes alignées sur les axes de l'image, la grille de droite est faite de lignes penchées d'une trentaine de degrés. Le trou est situé sur la grille alignée au niveau de la ligne 600 sur la droite, et sur la grille tournée au niveau de la ligne 700 sur la gauche. Les champs de déformation ϵ_x estimés devraient être en principe symétriques par rapport à un axe vertical. Les franges parasites visibles à gauche ne sont pas visibles à droite. Elles sont causées par le phénomène de repliement spectral : tourner la grille permet bien de l'éviter.

Elle s'exprime comme suit.

$$\begin{cases} \arg(\mathcal{F}_w(f)(x, y, f, 0)) = \arg d_1 + w_\sigma * \phi_x(x, y) + n_x(x, y) \\ \arg(\mathcal{F}_w(f)(x, y, 0, f)) = \arg d_1 + w_\sigma * \phi_y(x, y) + n_y(x, y) \end{cases} \quad (6.9)$$

où n_x et n_y désignent un bruit gaussien spatialement corrélé. Cette équation n'est valable que si la taille σ de la fenêtre est suffisamment grande pour estimer une composante fréquentielle de fréquence f , en accord avec le théorème d'incertitude [Mal99]. On doit imposer en pratique $\sigma f \geq 1$. Les composantes n_x et n_y ont été caractérisées dans [13] pour une fenêtre d'analyse gaussienne et dans [3] pour les principales fenêtres de la littérature. Dans le cas d'une fenêtre gaussienne, la matrice de covariance du bruit n affectant la phase s'exprime de la manière suivante :

$$\text{Covar}(n(\xi, \eta), n(\xi', \eta')) = \frac{v \Delta_x \Delta_y}{8\pi\sigma^2 P^2} e^{-\frac{(\xi-\xi')^2 + (\eta-\eta')^2}{4\sigma^2}} \quad (6.10)$$

où v est la variance du bruit blanc gaussien affectant l'image de la grille, $\Delta_x \Delta_y$ est l'aire d'un pixel, et $P = |d_1| \gamma A / 2$ est une constante dépendant uniquement de l'image de grille.

Cette relation entre phase estimée et phase réelle se propage ensuite sur les champs de déplacement, puis de déformation à l'aide des propriétés de dérivation d'un champ stationnaire

rappelées par exemple dans [Abr97]. Elle permet d'exprimer (toujours dans le cas d'une fenêtre gaussienne) la matrice de covariance du bruit n' affectant la dérivée de la phase selon la première composante par :

$$\text{Covar}(n'(\xi, \eta), n'(\xi', \eta')) = \frac{v\Delta_x\Delta_y}{16\pi\sigma^4 P^2} \left(1 - \frac{(\xi - \xi')^2}{2\sigma^2}\right) e^{-\frac{(\xi - \xi')^2 + (\eta - \eta')^2}{4\sigma^2}} \quad (6.11)$$

La relation symétrique est bien sûr valide pour la dérivée selon la seconde composante.

Ces équations montrent bien que la mesure des champs de déplacement et déformation par cette méthode est principalement limitée par :

- la taille σ de la fenêtre d'analyse à cause de la convolution, qui atténue l'amplitude des phases estimées (l'atténuation dépendant bien sûr de la fréquence des composantes spectrales des phases) et qui corrèle spatialement les estimations ;
- le bruit spatialement corrélé n_x et n_y .

Les performances métrologiques de la méthode sont alors quantifiées par :

- le biais de mesure relatif [4], mesurant la perte d'amplitude relative d'une sinusoïde de référence de fréquence donnée ;
- la résolution spatiale, définie ici à la suite de [WLLD15] comme la plus petite période d'une sinusoïde pouvant être estimée avec un biais de mesure relatif supérieur à un certain seuil (cette définition est bien adaptée car elle permet de comparer des fenêtres à support compact avec la fenêtre gaussienne par exemple) ;
- la résolution de mesure, définie comme l'écart-type du bruit dans le champ de mesure considéré (phase, déplacement, ou déformation) comme dans [CS12].

Nous avons déterminé dans [3] toutes ces quantités dans le cas d'images de grilles affectées d'un bruit gaussien, pour les principales fenêtres utilisées dans la littérature, à savoir les fenêtres birectangles [Kem07], triangle-rectangles [Sur00], bitriangles [AFV⁺04], ou gaussiennes [BGM09]. L'analyse générale de [3] est basée sur la transformée de Wigner-Ville de ces fenêtres [Coh95, HBB92, Mal99], certains résultats étant obtenus dans [13] de manière *ad hoc* dans le cas des fenêtres gaussiennes.

Plusieurs arguments sont en faveur d'une fenêtre d'analyse gaussienne. Les fenêtres à support compact doivent avoir une taille multiple de la période de la grille (ce qui peut être difficile à réaliser en pratique lorsque le pas n'est pas un multiple entier de pixels), contrairement aux fenêtres gaussiennes. Si les résolutions de mesure des déplacements sont comparables pour toutes les fenêtres, la résolution de mesure des déformations est bien meilleure pour une fenêtre gaussienne. Le caractère stationnaire du bruit corrélé est davantage valide dans le cas gaussien : c'est une conséquence de la décroissance rapide de la transformée de Wigner-Ville des gaussiennes [Fla98]. La fonction d'autocorrélation du bruit est également caractérisée dans [3, 13] : la fonction d'autocorrélation du bruit dans les champs de déplacement et déformation présente l'avantage d'être davantage régulière dans le cas des fenêtres gaussiennes, ce qui induit un bruit assez « régulier », plus agréable à l'œil et moins susceptible de cacher des phénomènes localisés pertinents dans ces champs.

Dans le cas plus réaliste d'un bruit numérique hétéroscédastique suivant le modèle Poisson-Gauss rappelé au chapitre 4, nous avons montré dans [12] en analysant une succession d'images dans une expérience réelle que la prédiction théorique de la résolution de mesure était expérimentalement très bien vérifiée une fois qu'une transformation stabilisant la variance (la transformée d'Anscombe généralisée [MSB95]) a été appliquée. Il convient également de réduire l'influence des inévitables micro-vibrations entre images, empêchant en pratique d'estimer des variances empiriques temporelles en chaque pixel des champs de déplacement et

déformation. À notre connaissance, cette étude constitue la première caractérisation de la résolution de mesure due au bruit de capteur des champs de déformation par la méthode de grille.

6.3 Restauration des champs de déplacement et déformation

L'expression établie par (6.9) laisse entrevoir une restauration des champs de déplacement et déformation par déconvolution non-aveugle (on connaît ici le noyau de convolution, qui est la fenêtre d'analyse), comme illustré par une expérience menée sur des images de grille synthétiques présentée dans la figure 6.5. Une restauration par déconvolution permet effectivement de réhausser certains détails.

Dans [15] nous avons étudié l'effet d'algorithmes de type Richardson-Lucy [Ric72, Luc74] (hypothèse d'un bruit de Poisson), régularisation de Tikhonov [TA77] (hypothèse d'un bruit blanc gaussien), filtrage de Wiener [Wie64] (hypothèse d'un bruit stationnaire de fonction de corrélation donnée par [3, 13]). Les résultats présentés dans cet article sont encourageants mais limités principalement par le fait que le bruit dans les images réelles n'est pas un processus gaussien, mais plutôt un processus de Poisson-Gauss comme discuté dans le chapitre 4. Un exemple de résultat obtenu sur un champ de déformation réel est présenté sur la figure 6.6 (la figure 6.7 présente un détail). Il s'agit d'un essai de traction selon l'axe vertical sur une éprouvette faite d'un alliage à mémoire de forme. Des détails deviennent visibles après déconvolution, en particulier les « aiguilles de martensites » apparaissant comme de fins segments rouges sur la figure 6.7. Ces aiguilles correspondent à un phénomène mécanique avéré car elles deviennent franchement visibles dans le champ non-déconvolué lorsqu'une force plus grande est appliquée à l'échantillon, cf. [DGBB12]. La restauration a donc pu révéler un phénomène qui n'était pas clairement visible sur les champs de déformation bruts.

De manière à améliorer la résolution de mesure ou de faciliter une déconvolution, il serait intéressant de *débruiter* les images de grille préalablement à l'estimation des champs de déplacement et déformation. Le but est ici de contrôler les performances métrologiques : il n'est donc pas question d'utiliser directement des algorithmes généraux de traitement d'images (voir les revues systématiques [LCBM12, Mil13]), comme les très performants BM3D [DFKE07] ou NL-means [BCM05]. Une approche plus simple permettant de contrôler la résolution de mesure serait de moyenniser une série de T images de grille. Néanmoins, le protocole expérimental nécessitant de forts agrandissements, on ne peut pas négliger les micro-vibrations qui affectent inévitablement la stabilité de la caméra par rapport à l'échantillon imagé. C'est l'objet de l'étude présentée dans [10]. Nous avons montré que, pour une image s quelconque, un développement limité donne l'expression de l'espérance de la moyenne empirique \bar{v} d'une série de T images bruitées :

$$E(\bar{v}) \simeq s + \frac{1}{2} \operatorname{div} (\operatorname{Var}(\alpha, \beta) \cdot \nabla s) \quad (6.12)$$

où s est l'image statique idéale (non-bruitée), div dénote la divergence d'un champ de vecteur, et $\operatorname{Var}(\alpha, \beta)$ est la matrice de covariance du processus (α_t, β_t) modélisant les vibrations comme au chapitre 4.

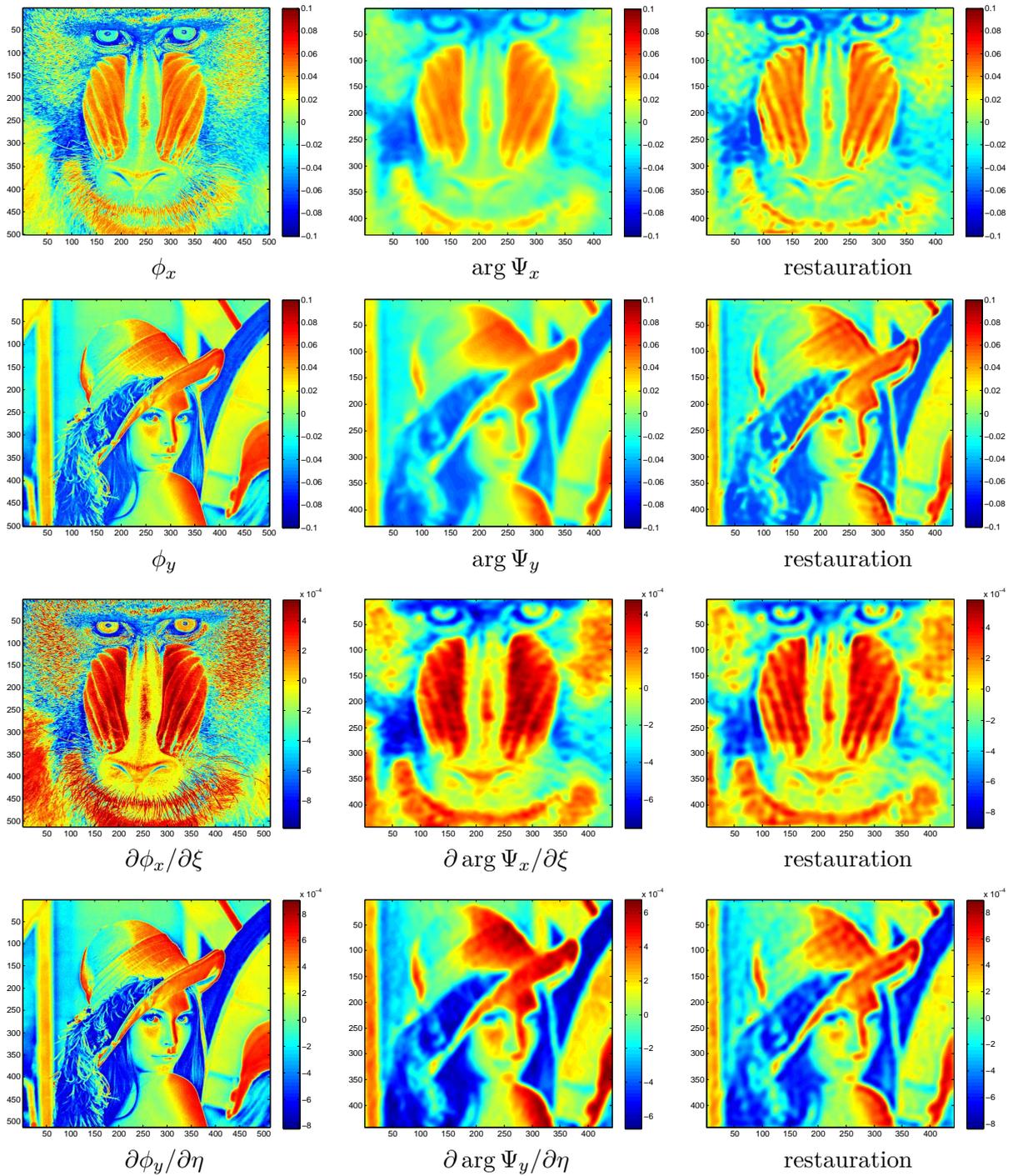


FIGURE 6.5 – Une expérience de restauration sur des images de grilles synthétiques générées selon l'équation (6.1) en imposant les phases ϕ_x et ϕ_y et leurs dérivées comme les images de la colonne de gauche. Un bruit blanc gaussien est également ajouté aux images. Les phases et dérivées de phases estimées comme argument de la transformée de Fourier à fenêtre (colonne du milieu) semblent effectivement suivre l'équation de convolution (6.9). La colonne de droite montre la restauration par filtre de Wiener tenant compte de l'autocorrélation du bruit caractérisée dans [13] et rappelée par les équations (6.10) et (6.11).

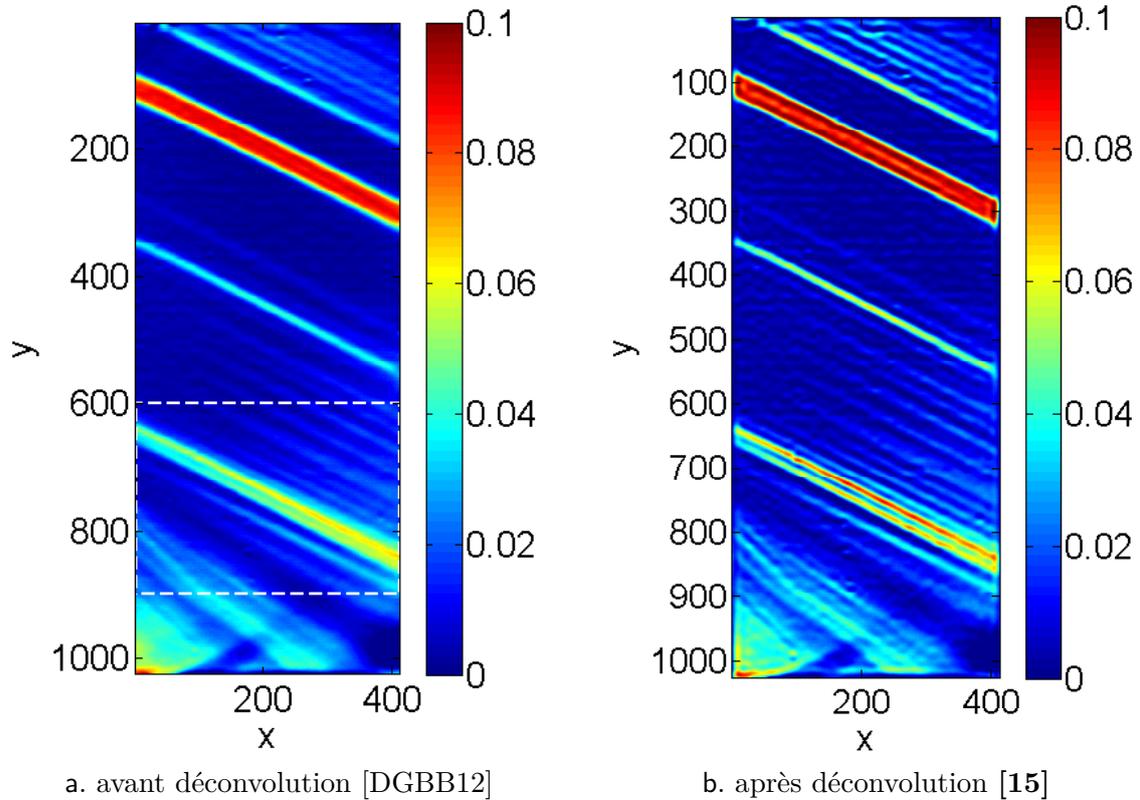


FIGURE 6.6 – Essai de traction sur un alliage à mémoire de forme. Champ de déformation ϵ_y , avant (à gauche) et après (à droite) déconvolution.

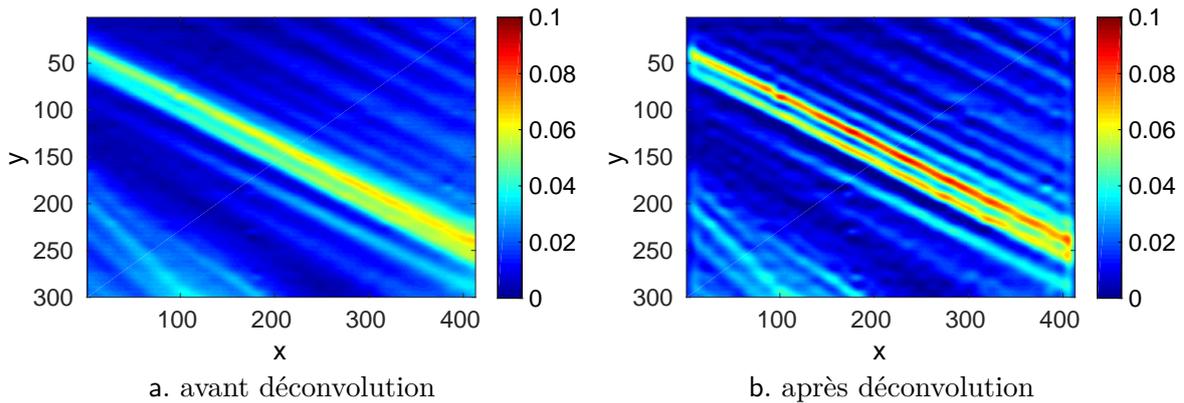


FIGURE 6.7 – Essai de traction sur un alliage à mémoire de forme. Détail de la figure 6.6 (rectangle blanc). Le cercle blanc a un diamètre de 6σ , où σ est l'écart-type de la fenêtre gaussienne d'analyse. L'écart entre les deux aiguilles de martensite étant inférieur à cette distance, elles ne sont pas clairement résolues dans le champ de déformation brut.

Sans surprise, la moyenne empirique est donc un estimateur biaisé de l'image de grille non bruitée inconnue. Nous montrons néanmoins que pour des images de grilles modélisées par l'équation (6.1), et sous l'approximation au premier ordre donnée par (6.9), l'influence de ce biais décroît tout de même en $\mathcal{O}(1/\sqrt{T})$ dans les champs de déplacement, et n'influence pas les champs de déformation. Il est donc légitime de moyenner une série d'images sans se soucier des micro-vibrations. À titre d'illustration, la figure 6.8 illustre la décroissance du niveau de bruit quand le nombre d'images T augmente.

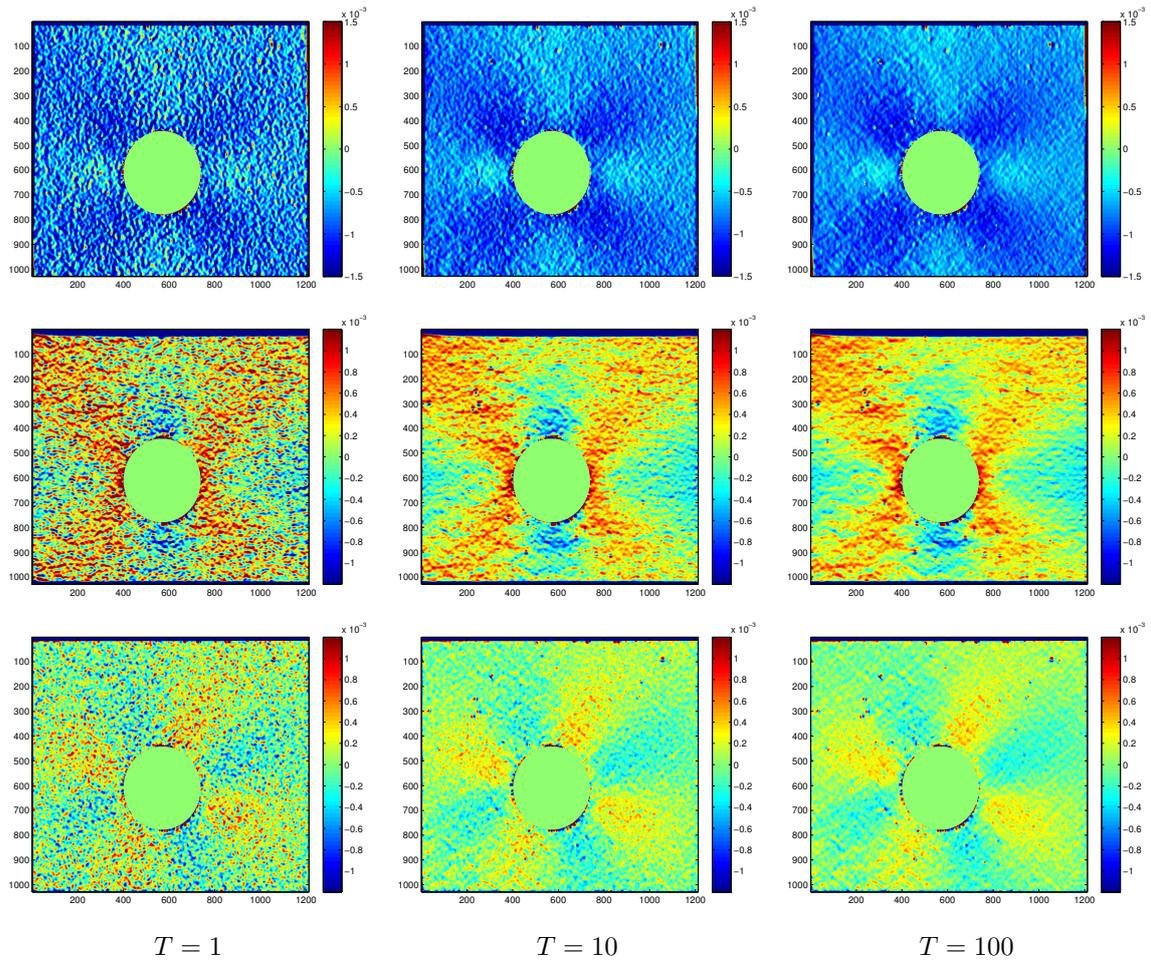


FIGURE 6.8 – Évolution des composantes de déformation ε_x (en haut), ε_y (au milieu), et ε_{xy} (en bas), estimée sur la grille moyennée obtenue à partir de $T \in \{1, 10, 100\}$ images (de gauche à droite ; $T = 1$ correspond à une image brute, non moyennée).

6.4 Quantification des performances métrologiques de la méthode basée sur la corrélation d'images (DIC)

Dans les articles [5] et [6] nous avons étudié la méthode d'estimation des champs de déplacement basée sur la corrélation d'images de « mouchetis » (DIC).

La contribution de [5] est une nouvelle formule de prédiction de la résolution de mesure du déplacement, limitée par le bruit affectant le capteur, qui tient en compte l'interpolation nécessaire dans l'estimation des déplacements sous-pixeliques.

En mécanique expérimentale, de nombreuses méthodes variantes de la méthode de maximisation de la corrélation entre imagerie sont utilisées et font l'objet d'une littérature abondante depuis les années 1980 [CRS85], comme décrit dans les livres et revues [SOS09, SH15]. Considérons deux images de la surface du matériau testé, l'une (notée f) acquise avant déformation, l'autre (notée g) après. Sous hypothèse de conservation de l'intensité lumineuse, le champ de déformation \mathbf{u} vérifie, au sein d'une région d'intérêt Ω , $f(\mathbf{x}) = g(\mathbf{x} + \mathbf{u}(\mathbf{x}))$. On peut voir qu'un déplacement sous-pixelique \mathbf{u} rend nécessaire l'interpolation de g .

Dans cette section, nous nous intéressons à la formulation dans laquelle l'objectif est de minimiser le résidu optique de manière à trouver le déplacement \mathbf{u}^* comme solution de :

$$\mathbf{u}^* = \arg \min_{\mathbf{u} \in \mathcal{U}} \sum_{\mathbf{x} \in \Omega} \left(f(\mathbf{x}) - g(\mathbf{x} + \mathbf{u}(\mathbf{x})) \right)^2 \quad (6.13)$$

Notons qu'il ne s'agit pas d'une corrélation au sens strict mais bien de la somme des carrés des résidus. En pratique, les déplacements \mathbf{u} ne sont pas déterminés en chaque point de la région d'intérêt mais sont cherchés dans l'espace vectoriel des cinématiques possibles, engendrés par N fonctions de forme $(\phi_i)_{1 \leq i \leq N}$ (classiquement, des fonctions polynomiales d'interpolation de degré 1 ou 2). Ceci permet de réduire fortement la dimensionnalité du problème tout en assurant la régularité du déplacement solution. Tout déplacement s'écrit donc $\mathbf{u} = \sum_{i=1}^N \lambda_i \phi_i$. Le problème de minimisation (6.13) est donc équivalent au problème suivant, portant à présent sur $\boldsymbol{\lambda} = (\lambda_i)_{1 \leq i \leq N}$:

$$\boldsymbol{\lambda}^* = \arg \min_{\boldsymbol{\lambda} \in \mathbb{R}^N} \sum_{\mathbf{x} \in \Omega} \left(f(\mathbf{x}) - g \left(\mathbf{x} + \sum_{i=1}^N \lambda_i \phi_i(\mathbf{x}) \right) \right)^2 = \arg \min_{\boldsymbol{\lambda} \in \mathbb{R}^N} \|\mathbf{r}(\boldsymbol{\lambda})\|^2 \quad (6.14)$$

où $\mathbf{r}(\boldsymbol{\lambda})$ est le vecteur des résidus et $\|\cdot\|$ désigne la norme euclidienne. Dans la suite, nous notons $J_{\mathbf{r}}(\boldsymbol{\lambda}^*)$ la matrice Jacobienne de \mathbf{r} en $\boldsymbol{\lambda}^*$. Elle vérifie, pour $1 \leq i \leq \#\Omega$ et $1 \leq j \leq N$,

$$(J_{\mathbf{r}}(\boldsymbol{\lambda}^*))_{i,j} = -(\nabla g(\mathbf{x}_i + \mathbf{u}_{\boldsymbol{\lambda}}(\mathbf{x}_i)))^T \phi_j(\mathbf{x}_i) \quad (6.15)$$

Par la simplification $\nabla g(\mathbf{x}_i + \mathbf{u}_{\boldsymbol{\lambda}}(\mathbf{x}_i)) \simeq \nabla f(\mathbf{x}_i)$, on remarque que $J_{\mathbf{r}}(\boldsymbol{\lambda}^*) = -L$ où

$$L_{i,j} = (\nabla f(\mathbf{x}_i))^T \phi_j(\mathbf{x}_i) \quad (6.16)$$

Nous noterons dans la suite $M = L^T L$.

Ce problème de minimisation est résolu par un algorithme de Gauss-Newton modifié. En propageant l'incertitude sur f et g (modélisée comme un bruit blanc gaussien de variance σ^2) par une technique similaire à celle de [35], nous avons montré dans [5] que l'incertitude sur la valeur de $\boldsymbol{\lambda}^*$ peut être modélisée par une variable aléatoire gaussienne $\boldsymbol{\delta}_{\boldsymbol{\lambda}}$ de moyenne et matrice de covariance :

$$\begin{cases} \langle \boldsymbol{\delta}_{\boldsymbol{\lambda}} \rangle = A(f + Pg) \\ \Sigma_{\boldsymbol{\delta}_{\boldsymbol{\lambda}}} = \sigma^2 A (I + PP^T) A^T \end{cases} \quad (6.17)$$

où f et g désignent ici, par léger abus de notation, les vecteurs-colonnes des $f(\mathbf{x}_i)$ et $g(\mathbf{x}_i)$ respectivement, où $A = M^{-1} J_{\mathbf{r}}^T$, et où la matrice P est formée de manière à ce que :

$$g(\cdot + \mathbf{u}_{\boldsymbol{\lambda}}(\cdot)) = Pg \quad (6.18)$$

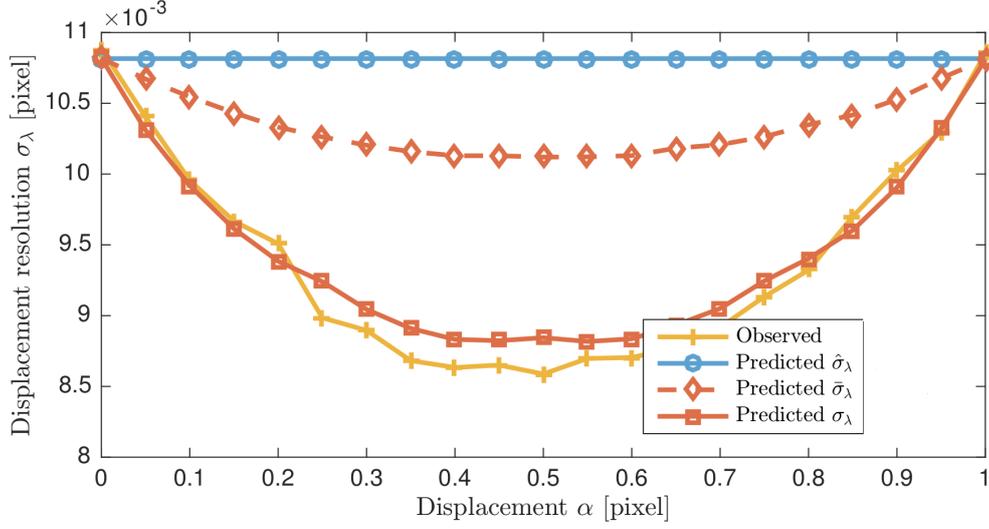


FIGURE 6.9 – Dans le cas d’un déplacement uniforme de α pixel (α entre 0 et 1), confrontation de la résolution moyenne sur les λ obtenues par estimation empirique (Observed), par la formule de la littérature ($\hat{\sigma}_\lambda$, cf. (6.20)) et par les deux formules proposées ($\bar{\sigma}_\lambda$ donnée par la simplification $J_r = -L$ et σ_λ , cf. (6.19)).

Les interpolations usuelles (ici bilinéaires ou bicubiques) permettent bien cette écriture, car la valeur interpolée en un point est combinaison linéaire des valeurs aux pixels entiers dans un voisinage. Notons qu’on peut approcher comme précédemment J_r par $-L$.

Le terme $\langle \delta_\lambda \rangle$ est connu comme étant le biais d’estimation causé par le schéma d’interpolation (voir par exemple [SBS00]).

L’écart-type de l’erreur sur le coefficient λ_i s’écrit donc :

$$\sigma_{\lambda_i} = \sigma \sqrt{[A(I + PP^T)A^T]_{ii}} \quad (6.19)$$

L’écart-type obtenu par la simplification $J_r = -L$ sera noté $\bar{\sigma}_\lambda$.

Il s’agit d’une généralisation de la formule précédemment disponible dans la littérature [Rét10, GLH⁺15] qui exprimait l’écart-type de l’erreur par :

$$\hat{\sigma}_{\lambda_i} = \sigma \sqrt{[2M^{-1}]_{ii}} \quad (6.20)$$

En effet, dans le cas où l’interpolation n’est pas prise en compte, on simplifie dans (6.19) $PP^T = I$ et $AA^T = M^{-1}LL^TM^{-T} = M^{-1}M^TM^{-T} = M^{-1}$, d’où (6.20).

Dans [5], nous propageons ensuite ce terme d’erreur au déplacement mesuré par DIC. Nous proposons une validation numérique sur des images simulées affectées d’un bruit blanc gaussien. Par exemple dans le cas d’un déplacement de translation uniforme, la figure 6.9 confronte différentes valeurs prédites pour la résolution à une valeur empirique obtenue par simulation de Monte-Carlo. Logiquement, la prédiction fournie par (6.20) ne dépend pas de la valeur du déplacement sous-pixellicque imposé car elle ne tient pas compte de l’interpolation. La formule donnée par (6.19) est en bonne adéquation avec la valeur empirique, et la formule simplifiée surestime l’erreur.

Dans [6], nous avons validé expérimentalement cette formule, c'est-à-dire sur des données issues du laboratoire, et pas des données synthétiques. Pour ce faire, nous confrontons les résolutions prédites aux résolutions empiriques obtenues par estimation sur une série d'images. Alors que les formules prédictives reposent sur l'hypothèse d'un bruit blanc gaussien, le bruit d'un capteur numérique est en fait hétéroscédastique et dépend du signal. Prendre en compte un tel modèle de bruit peut se faire de deux manières. Une première possibilité est de stabiliser la variance du bruit (une fois les paramètres du capteur estimés) par la transformée d'Anscombe généralisée [MSB95], comme dans [12] pour la méthode de grille. Cette transformation n'est pas linéaire et induit donc un biais dans l'estimation des déplacements. Néanmoins, ce biais semble négligeable en pratique. Une seconde possibilité est d'intégrer dans la formule prédictive la matrice de covariance de l'erreur sur les images f et g (respectivement Σ_f et Σ_g). Ce sont des matrices diagonales car le bruit est supposé indépendant de pixel à pixel, mais les valeurs diagonales ne sont pas identiques en chaque pixel. La formule (6.19) devient alors :

$$\sigma_{\lambda_i} = \sigma \sqrt{[A(\Sigma_f + P\Sigma_g P^T)A^T]_{ii}} \quad (6.21)$$

Dans les deux cas se pose le problème de l'estimation des variances empiriques à partir d'une série d'images. En effet, les micro-vibrations posent inévitablement des problèmes à ce niveau d'agrandissement des images ; les moyennes temporelles sur un même pixel n'intègrent pas uniquement la variabilité due au bruit et sont biaisées. Nous avons donc terminé la validation en utilisant l'estimation du déplacement par corrélation sur des images de grilles, sur lesquelles nous pouvons appliquer l'algorithme NRSR d'estimation de la variance débarrassée de l'influence des micro-vibrations (cf. section 4.5 au chapitre 4).

La figure 6.10 montre un de nos résultats de validation dans le cas d'une série de 100 images d'un essai de translation uniforme de l'échantillon. On constate que dans les deux possibilités évoquées, la résolution prédite est en bonne adéquation avec la résolution empirique et que les deux méthodes présentent des performances similaires. Une validation dans le cas d'un déplacement non-uniforme (déformation d'une éprouvette trouée soumise à un test de traction) est également discutée dans [6].

6.5 Conclusion et perspectives

Dans ce chapitre, nous avons présenté nos contributions à l'analyse et au traitement d'images servant à la détermination de champs de déplacement et de déformation en mécanique des solides, que ces images soient issues de la méthode de la grille ou d'algorithmes basés sur la corrélation d'images (DIC). Notre souci est d'établir des formules prédictives caractérisant les performances métrologiques des méthodes d'estimation, et de les valider en conditions expérimentales réelles.

Des prolongements naturels de ces travaux sont envisageables dans plusieurs directions.

Tout d'abord, il conviendrait d'aller au delà de l'étude préliminaire de [15] dans la restauration des champs de déplacement et de déformation. Cette étude tient compte de la nature corrélée du bruit affectant ces cartes, néanmoins l'hypothèse de bruit blanc gaussien dans les images de grille est bien trop grossière. Le fait que l'on puisse disposer facilement d'une série d'images affectées de micro-vibrations, donc légèrement décalées les unes par rapport aux autres, laisse entrevoir l'utilisation de techniques de super-résolution pour augmenter virtuellement la résolution des images acquises, qu'elles soient de grilles ou de mouchetis. Ces deux points sont repris dans les perspectives de recherches détaillées dans le chapitre 7.

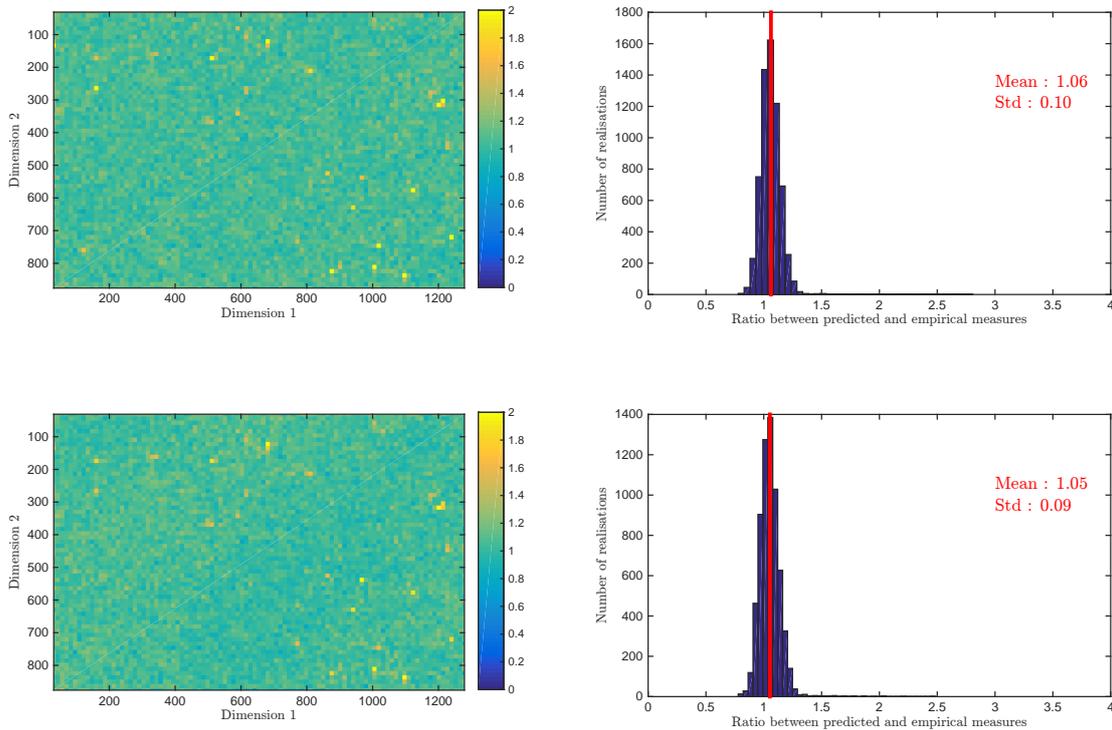


FIGURE 6.10 – Rapport entre résolution prédite et résolution observée dans un champ de déplacement. À gauche : valeur du rapport en chaque pixel. À droite : histogramme de répartition des rapports. En haut : après stabilisation de la variance par GAT. En bas : après intégration de la variance du bruit en chaque pixel dans la formule prédictive. On voit que les rapports sont proches de 1 avec une faible dispersion, ce qui permet de valider notre formule prédictive de la résolution.

Par ailleurs, il serait utile de définir une règle de décision permettant de discriminer les détails réels du bruit spatialement corrélé affectant les champs de déformation. On peut noter qu'un problème semblable existe en tomographie par émission de positons (PET) et en imagerie par résonance magnétique fonctionnelle (fMRI), deux techniques produisant des images perturbées par un bruit corrélé (voir par exemple [FCF⁺95]). Dans le cas d'un bruit blanc, la détection statistique peut se faire par une simple correction de Bonferroni [Bon36] pour contrôler le taux d'erreur. Néanmoins, cette approche n'est pas adaptée dans le cas d'un bruit corrélé. Dans le cas de la détection en imagerie PET ou fMRI, des techniques basées sur la théorie des champs aléatoires ont été développées [BPK03, Wor03]. Le développement de telles techniques pour l'analyse de champs de déformation reste à faire.

Enfin, il serait intéressant d'établir des formules prédictives pour la *geometric phase analysis* (GPA) dans le cadre de la mécanique expérimentale [DXW⁺14a, DXW14b], similaires à celles établies pour la méthode de grille, de manière à comparer les performances métrologiques respectives de ces méthodes sur des bases objectives. Dans le même esprit, une image de grille étant essentiellement un signal bidimensionnel modulé en amplitude et en fréquence (voir l'équation (6.1)), il serait intéressant d'examiner une méthode comme la décomposition modale empirique (*empirical mode decomposition*, EMD) proposée dans [HSL⁺98], sa généra-

lisation aux images dans le cadre d'une analyse par ondelettes [GTO14], ou, dans un premier temps, les différentes méthodes rappelées dans [Kem07, Kem15] pour l'analyse de franges d'interférométrie en optique. La difficulté est bien sûr d'établir des formules prédictives des performances métrologiques lorsque les cartes de déplacement et déformation sont estimées par un algorithme non-linéaire.

Quatrième partie

Quelques perspectives de recherche

Conclusion et perspectives

7.1 Conclusion générale	121
7.2 Quelques perspectives de recherche	121

7.1 Conclusion générale

Les travaux de recherche présentés dans ce mémoire se situent dans les domaines de l'analyse et du traitement des images numériques. Leurs applications vont de problèmes de la vision par ordinateur à la caractérisation des performances métrologiques de mesures de champ sans contact en mécanique des solides expérimentale. Une sélection de résultats récents a été présentée, concernant l'appariement de points d'intérêt, le bruit dans les images, ainsi que l'analyse des images et des méthodes d'estimation de la mécanique expérimentale.

À la fin de chaque chapitre, les développements naturels des résultats présentés ont été décrits. L'objet de la section suivante est de reprendre et de développer quelques perspectives constituant mes principales directions de recherche.

7.2 Quelques perspectives de recherche

L'analyse et le traitement des images sont intrinsèquement liés, dans la mesure où un traitement efficace des images ne peut s'affranchir d'une analyse du contenu sémantique. Par exemple, la restauration des images doit tenir compte de la présence de contours qui doivent être préservés (ce qui motive les diffusions anisotropes [Wei98] ou l'introduction de la variation totale [ROF92]), ou d'une redondance dans l'information contenue dans les images (ce qui est à la base des algorithmes non-locaux [BCM05]). Les thèses co-encadrées traitent de problèmes de vision par ordinateur, discipline qui enrichit l'analyse sémantique des images en incorporant une modélisation géométrique de la formation des images dans une caméra. Dans le futur, j'entends développer en particulier les deux directions de recherche suivantes, dans le domaine du traitement et de l'analyse des images.

7.2.1 Vers une analyse spectrale non-locale

Dans le chapitre 5, la recherche de méthodes de caractérisation du bruit quasi-périodique nous a amené à introduire le spectre d'amplitude moyen et le spectre d'amplitude minimum, définis respectivement comme la moyenne ou le minimum des modules des coefficients de Fourier issus d'imagettes couvrant l'ensemble d'une image. Les spectres d'amplitude moyen et minimum présentent l'avantage que leurs coefficients significativement élevés ont de bonne chance d'être provoqués par le bruit périodique qui est la seule structure répétitive couvrant l'intégralité de l'image. Des arguments statistiques permettent ensuite de détecter ces composantes spectrales communes à toutes les imagettes.

Au delà de la moyenne ou du minimum calculé sur l'ensemble des imagettes, il serait intéressant d'envisager une pondération des spectres d'amplitude des imagettes par un critère de similarité à une imagette donnée de manière à localiser la détection des composantes quasi-périodiques. Des applications seraient envisageables à diverses situations, comme la détection d'un bruit quasi-périodique qui n'affecterait pas l'intégralité de l'image (à cause de saturations par exemple), des textures haute-fréquences et de manière générale le problème de décomposition entre structure et texture [AGCO06, BLMV10], et sans doute certaines formes d'*aliasing*. Il est également envisageable de pondérer la moyenne par d'autres quantités qu'une mesure de similarité entre imagettes. Par exemple, un moyennage pondéré de spectres issus d'une rafale d'images a été récemment proposé [DS15] pour la restauration du flou de bougé. Notons qu'il ne s'agit pas de spectres d'amplitude dans ce travail, la phase est conservée. La pondération y est proportionnelle à l'amplitude des coefficients spectraux de manière à favoriser les composantes spectrales venant des images peu bruitées. Un tel moyennage pondéré des spectres ou des spectres d'amplitude constituerait une sorte de moyenne non-locale [BCM05] ou de filtrage bilatéral [TM98] adaptés aux composantes spectrales. On peut également remarquer que l'analyse de spectres issus d'imagettes similaires est à la base de l'algorithme de débruitage BM3D [DFKE07]. Les auteurs de [LXZG15] adaptent le seuillage permettant le débruitage par une modélisation statistique des coefficients d'une décomposition en cosinus discret (DCT). L'avantage de ces méthodes spectrales est leur faible complexité algorithmique, par comparaison à des méthodes variationnelles de restauration comme celles discutées dans [AGCO06]. Dans une optique de détection localisée, adapter la forme des imagettes dont on extrait les composantes spectrales semble crucial. On peut noter que c'est l'objet de la décomposition *shape adaptive discrete cosine transform* (SA-DCT) [Sik95], utilisée par les auteurs de [FKE07], sur la base de masques binaires locaux adaptés aux formes sous-jacentes décrite dans [KEA06]. Plutôt qu'un masquage binaire, il serait également pertinent d'envisager un filtrage bilatéral comme proposé dans l'algorithme *Dual-Domain Image Denoising* [KZ13] ou son extension *Dual-Domain Filtering* [KZ15]. Enfin, des applications à l'estimation des niveaux de bruit en fonction de l'intensité lumineuse ou des applications au débruitage d'images peuvent être envisagées, à l'instar des auteurs de [CB13, PLZ⁺07] qui estiment le bruit à partir de la transformée DCT d'imagettes, ou bien des auteurs de [YS11] qui débruitent les images par seuillage des coefficients de la DCT locale.

En résumé, nous pensons qu'il serait pertinent d'unifier ces différentes méthodes basées sur l'analyse de spectres d'imagettes, que ce soit des spectres de puissance issus de la transformée de Fourier discrète ou de coefficients de DCT. On retrouve des méthodes similaires dans des problèmes a priori très différents de l'analyse et du traitement des images. En ce qui concerne les problèmes de détection comme dans le cas du bruit périodique, une méthode adaptée doit aussi être développée, de manière à repérer les composantes spectrales saillantes. Une telle

méthode pourrait être développée dans le cadre de la décision a-contrario ou dans celui de la théorie des valeurs extrêmes [Col01] comme suggéré au chapitre 5.

7.2.2 Étude et amélioration des performances métrologiques en mécanique des solides

Nous envisageons des perspectives aux travaux décrits au chapitre 6 dans, principalement, deux directions.

Tout d'abord, nous souhaitons aller au delà de l'étude préliminaire de [15] dans la restauration des champs de déplacement et de déformation issus de la méthode de grille. En effet, si la déconvolution par filtre de Wiener étudiée dans [15] incorpore la corrélation spatiale du bruit dans les cartes de déplacement et déformation, l'hypothèse sous-jacente est celle d'un bruit blanc gaussien dans les images de grille. Une possibilité serait en premier lieu d'utiliser des techniques de déconvolution tenant compte de la nature hétéroscédastique du bruit sur les capteurs et en particulier du modèle Poisson-Gauss : le rapport signal-sur-bruit n'étant expérimentalement pas très favorable, il semble important de caractériser correctement le bruit. Il conviendrait également d'étudier l'influence de la résolution limitée de l'optique employée, régie par sa *point spread function* (PSF). D'autre part, nos résultats établissent, comme attendu, un compromis dans l'analyse de Fourier à fenêtre entre résolution fréquentielle (la largeur de la fenêtre d'analyse doit être suffisante par rapport au pas de grille pour accéder à l'information portée par la grille) et résolution spatiale (se traduisant par la relation de convolution entre phase estimée et phase réelle imposant le biais de mesure). Or, les techniques de réassignement introduites par F. Auger et P. Flandrin [AF95, FACM03] permettent d'améliorer la localisation de l'information dans les spectrogrammes. Il serait intéressant de voir comment ces techniques se traduisent dans l'analyse par transformée de Fourier à fenêtre employée dans la méthode de grille, qui a pour particularité de ne nécessiter le calcul de la transformée de Fourier qu'à la fréquence nominale de la grille, sans avoir à calculer l'intégralité du spectrogramme. Il n'est pas clair que le réassignement soit, dans ce cadre, équivalent à une déconvolution comme dans [15]. La question mérite d'être approfondie, d'un point de vue théorique comme pratique.

Nous entendons ensuite étudier l'apport des techniques de super-résolution pour les champs de déplacement et déformation, que ce soit par la méthode de grille ou par les approches de type DIC. Dans un cadre applicatif proche, les auteurs de [TEC09] imposent des mouvements contrôlés à la caméra pour reconstruire à partir d'une série d'images des champs thermographiques super-résolus, c'est-à-dire de résolution supérieure à la résolution de celui obtenu à partir d'une image seule, et avec un niveau de bruit inférieur. Dans notre protocole expérimental, il ne semble pas nécessaire d'imposer un mouvement contrôlé. De manière paradoxale, les micro-vibrations qui affectent l'acquisition d'une série d'images et se traduisent par des décalages sous-pixelles pourraient en effet être employées favorablement, à condition que l'on soit capable de les estimer précisément. Les techniques de super-résolution étant néanmoins susceptibles d'amplifier le bruit dans les hautes fréquences, elles présentent bien sûr des limites, discutées par exemple dans [LS04] ou [ADBS16]. Les auteurs de [LS04] mentionnent par exemple une limite d'augmentation de la résolution des images par un facteur de 2,5 à niveau de bruit constant. Des bornes de type « Cramér-Rao » sont établies dans [ADBS16] pour l'estimation des déplacements entre images, avec des modèles d'images déterministes ou stochastiques. De manière intéressante, le cadre déterministe semble bien adapté aux images de grilles, et le modèle stochastique aux images de *speckle*. D'autre part, le problème de la

super-résolution étant un problème inverse mal posé, il convient d'imposer une contrainte de régularité de la solution. Si des contraintes de type Tikhonov ou variation totale sont très utilisées en traitement d'images « classique » [FREM04], on peut se demander quelle serait la contrainte la plus adaptée aux champs de déplacement ou déformation. Pour l'estimation du décalage sous-pixelique constant entre images imposé par les micro-vibrations, les principales méthodes mentionnées dans [TH86] doivent être analysées. Par exemple, il serait intéressant de spécialiser les résultats de [FZB02] concernant le recalage sous-pixelique par corrélation de phase à nos images de grilles ou de mouchetis translattées par les micro-vibrations. Citons aussi [RMF15] quantifiant les performances de l'estimation de décalages en fonction du bruit et de l'amplitude de ces décalages. D'autre part, dans les expériences dynamiques de chargement, l'échantillon se déforme continûment et progressivement au cours du temps (du moins dans un premier temps) sous l'effet d'une traction ou d'une compression. Dans ce cas, il n'est plus possible de considérer que, dans une série d'images, les déplacements entre points physiques soient uniquement causés par des micro-vibrations uniformes sur tout le champ. Il est nécessaire de tenir compte de ces déformations continues microscopiques non-uniformes dans une restauration par super-résolution ou dans un débruitage de la série d'images, la difficulté étant que ce sont précisément ces quantités que l'on cherche à restaurer. On peut remarquer que ce problème semble présenter des similitudes avec celui rencontré en imagerie par résonance magnétique (IRM) volumique fœtale, dans lequel les acquisitions de basse résolution doivent être recalées, comme par exemple dans [TBH⁺15].

Ces perspectives de recherche ne constituent pas simplement une application de techniques classiques du traitement des images aux données de la mécanique des solides. D'une part, les problèmes de traitement d'images afférents sont toujours ouverts et font l'objet de travaux dans une communauté très active. D'autre part, l'originalité et la difficulté des travaux proposés résident dans le souci de quantifier la précision des champs de déplacement ou déformation obtenus, de manière à certifier les performances métrologiques des méthodes d'estimation. Contrairement à de nombreuses applications du traitement d'images « classique », il s'agit d'améliorer des mesures dérivées des images et pas les images elles-mêmes, ces mesures émergeant à peine du bruit. Il s'agit bien sûr de travaux pluridisciplinaires à mener de concert avec des collègues mécaniciens, en premier lieu Michel Grédiac et Benoît Blaysat.

Publications personnelles

Livre

- [1] F. Cao, J.L. Lisani, J.-M. Morel, P. Musé, and F. Sur. *A theory of shape identification*. Number 1948 in Lecture Notes in Mathematics. Springer, 2008. 264 pages.

Articles dans des journaux internationaux

- [2] F. Sur, B. Blaysat, and M. Grédiac. Determining displacement and strain maps immune from aliasing effect with the grid method. *Optics and Lasers in Engineering*, 86:317–328, 2016.
- [3] F. Sur and M. Grédiac. Influence of the analysis window on the metrological performance of the grid method. *Journal of Mathematical Imaging and Vision*, 56(3):472–498, 2016.
- [4] M. Grédiac, F. Sur, and B. Blaysat. The grid method for in-plane displacement and strain measurement: a review and analysis. *Strain*, 52(3):205–243, 2016.
- [5] B. Blaysat, M. Grédiac, and F. Sur. Effect of interpolation on noise propagation from images to DIC displacement maps. *International Journal for Numerical Methods in Engineering*, 108(3):213–232, 2016.
- [6] B. Blaysat, M. Grédiac, and F. Sur. On the propagation of camera sensor noise to displacement maps obtained by DIC - an experimental study. *Experimental Mechanics*, 56(6):919–944, 2016.
- [7] M. Grédiac, F. Sur, and B. Blaysat. Removing quasi-periodic noise in strain maps by filtering in the Fourier domain. *Experimental Techniques*, 40(3):959–971, 2016.
- [8] F. Sur and M. Grédiac. Automated removal of quasiperiodic noise using frequency domain statistics. *IS&T / SPIE Journal of Electronic Imaging*, 24(1):013003/1–19, 2015.
- [9] F. Sur and M. Grédiac. Measuring the noise of digital imaging sensors by stacking raw images affected by vibrations and illumination flickering. *SIAM Journal on Imaging Sciences*, 8(1):611–643, 2015.
- [10] F. Sur and M. Grédiac. On noise reduction in strain maps obtained with the grid method by averaging images affected by vibrations. *Optics and Lasers in Engineering*, 66:210–222, 2015.
- [11] F. Sur and M. Grédiac. Sensor noise modeling by stacking pseudo-periodic grid images affected by vibrations. *IEEE Signal Processing Letters*, 21(4):432–436, 2014.
- [12] M. Grédiac and F. Sur. Effect of sensor noise on the resolution and spatial resolution of displacement and strain maps estimated with the grid method. *Strain*, 50(1):1–27, 2014.

-
- [13] F. Sur and M. Grédiac. Towards deconvolution to enhance the grid method for in-plane strain measurement. *AIMS Inverse Problems and Imaging*, 8(1):259–291, 2014.
- [14] F. Sur, N. Noury, and M.-O. Berger. An a contrario model for matching interest points under geometric and photometric constraints. *SIAM Journal on Imaging Sciences*, 6(4):1956–1978, 2013.
- [15] M. Grédiac, F. Sur, C. Badulescu, and J.-D. Mathias. Using deconvolution to improve the metrological performance of the grid method. *Optics and Lasers in Engineering*, 51(6):716–734, 2013.
- [16] F. Cao, J. Delon, A. Desolneux, P. Musé, and F. Sur. A unified framework for detecting groups and application to shape recognition. *Journal of Mathematical Imaging and Vision*, 27(2):91–119, 2007.
- [17] P. Musé, F. Sur, F. Cao, Y. Gousseau, and J.-M. Morel. An a contrario decision method for shape element recognition. *International Journal of Computer Vision*, 69(3):295–315, 2006.
- [18] F. Cao, P. Musé, and F. Sur. Extracting meaningful curves from images. *Journal of Mathematical Imaging and Vision*, 22(2-3):159–181, 2005.
-

Articles dans des journaux nationaux

- [19] P. Rolin, M.-O. Berger, and F. Sur. Simulation de point de vue pour la mise en correspondance et la localisation. *Traitement du signal*, 32(2-3):169–194, 2015.
- [20] P. Musé, F. Sur, and J.-M. Morel. Sur les seuils de reconnaissance des formes. *Traitement du Signal*, 20(3):279–294, 2003.
-

Chapitre de livre

- [21] P. Musé, F. Sur, F. Cao, Y. Gousseau, and J.-M. Morel. Shape recognition based on an a contrario methodology. In H. Krim and A. Yezzi, editors, *Statistics and analysis of shapes*, chapter 5, pages 107–136. Birkhauser, 2006.
-

Articles dans des actes de conférences internationales

- [22] P. Rolin, M.-O. Berger, and F. Sur. Enhancing pose estimation through efficient patch synthesis. In *Proceedings of the British Machine Vision Conference (BMVC)*, York, UK, 2016.
- [23] F. Sur. An a-contrario approach to quasi-periodic noise removal. In *Proceedings of the IEEE International Conference on Image Processing (ICIP)*, pages 3841–3845, Québec City, Canada, 2015.
- [24] B. Blaysat, M. Grédiac, and F. Sur. On noise prediction in maps obtained with global DIC. In H. Jin, S. Yoshida, L. Lamberti, and M.-T. Lin, editors, *Advancement of Optical Methods in Experimental Mechanics, Volume 3*, Conference Proceedings of the Society for Experimental Mechanics Series (SEM 2015), pages 211–216. 2016.

-
- [25] P. Rolin, M.-O. Berger, and F. Sur. Viewpoint simulation for camera pose estimation from an unstructured scene model. In *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, pages 6320–6327, Seattle (WA), USA, 2015.
- [26] M. Grédiac and F. Sur. Stabilizing heteroscedastic noise with the generalized Anscombe transform. Application to accurate prediction of the resolution in displacement and strain maps obtained with the grid method. In H. Jin, C. Sciammarella, S. Yoshida, and L. Lamberti, editors, *Advancement of Optical Methods in Experimental Mechanics, Volume 3*, Proceedings of the Society for Experimental Mechanics (SEM 2014), pages 225–230. 2015.
- [27] F. Sur and M. Grédiac. Sensor noise measurement in the presence of a flickering illumination. In *Proceedings of the IEEE International Conference on Image Processing (ICIP)*, pages 1763–1767, Paris, France, 2014.
- [28] M. Grédiac, F. Sur, C. Badulescu, and J.-D. Mathias. Deconvolving strain maps obtained with the grid method. In H. Jin, C. Sciammarella, S. Yoshida, and L. Lamberti, editors, *Advancement of Optical Methods in Experimental Mechanics, Volume 3*, Conference Proceedings of the Society for Experimental Mechanics (SEM 2013), pages 21–26. 2014.
- [29] F. Sur and M. Grédiac. Enhancing with deconvolution the metrological performance of the grid method for in-plane strain measurement. In *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, pages 1563–1567, Vancouver (BC), Canada, 2013.
- [30] F. Sur. Illumination-invariant representation for natural colour images through SIFT matching. In *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, pages 1962–1966, Vancouver (BC), Canada, 2013.
- [31] S. Bhat, M.-O. Berger, and F. Sur. Visual words for 3D reconstruction and pose computation. In *Proceedings of the Joint 3DIM/3DPVT Conference (3DIMPVT)*, pages 326–333, Hangzhou, China, 2011.
- [32] N. Noury, F. Sur, and M.-O. Berger. How to overcome perceptual aliasing in ASIFT? In *Proceedings of the International Symposium on Visual Computing (ISVC), part I*, volume LNCS 6453, pages 231–242, Las Vegas (NV), USA, 2010.
- [33] F. Sur. Robust matching in an uncertain world. In *Proceedings of the International Conference on Pattern Recognition (ICPR)*, pages 2350–2353, Istanbul, Turkey, 2010.
- [34] S. Bhat, M.-O. Berger, G. Simon, and F. Sur. Transitive closure based visual words for point matching in video sequences. In *Proceedings of the International Conference on Pattern Recognition (ICPR)*, pages 3300–3303, Istanbul, Turkey, 2010.
- [35] F. Sur, N. Noury, and M.-O. Berger. Computing the uncertainty of the 8 point algorithm for fundamental matrix estimation. In *Proceedings of the British Machine Vision Conference (BMVC)*, volume 2, pages 965–974, Leeds, UK, 2008.
- [36] N. Noury, F. Sur, and M.-O. Berger. Fundamental matrix estimation without prior match. In *Proceedings of the IEEE International Conference on Image Processing (ICIP)*, volume 1, pages 513–516, San Antonio (TX), USA, 2007.
- [37] A. Cerri, D. Giorgi, P. Musé, F. Sur, and F. Tomassini. Shape recognition via an a contrario model for size functions. In *Proceedings of the International Conference on Image Analysis and Registration (ICIAR)*, volume LNCS 4142, pages 410–421, Povia de Varzim, Portugal, 2006.

-
- [38] Y. Guermeur, M. Maumy, and F. Sur. Model selection for multi-class SVMs. In *Proceedings of the International Symposium on Applied Stochastic Models and Data Analysis (ASMDA)*, pages 506–517, Brest, France, 2005.
- [39] P. Musé, F. Sur, F. Cao, and Y. Gousseau. Unsupervised thresholds for shape matching. In *Proceedings of the IEEE International Conference on Image Processing (ICIP)*, volume 2, pages 647–650, Barcelona, Spain, 2003.
-

Articles dans des actes de conférences nationales

- [40] P. Rolin, M.-O. Berger, and F. Sur. Simulation de point de vue pour la localisation d’une caméra à partir d’un modèle non structuré. In *Actes de la conférence Reconnaissance de Formes et Intelligence Artificielle (RFIA)*, Rouen, France, 2014.
- [41] N. Noury, F. Sur, and M.-O. Berger. Modèle a contrario pour la mise en correspondance robuste sous contraintes épipolaires et photométriques. In *Actes du congrès Reconnaissance des Formes et Intelligence Artificielle (RFIA)*, Caen, France, 2010.
- [42] N. Noury, F. Sur, and M.-O. Berger. Modèles statistiques pour l’estimation de la matrice fondamentale. In *Acte du congrès francophone des jeunes chercheurs en vision par ordinateur ORASIS*, Obernai, France, 2007.
-

Communications à des conférences internationales (ateliers, conférences à sélection sur résumé, tutoriels)

- [43] M. Grédiac, B. Blaysat, and F. Sur. Comparison of the metrological performance of grid method and DIC. In *SEM Annual Conference & Exposition on Experimental and Applied Mechanics*, Orlando (FL), USA, 2016.
- [44] B. Blaysat, M. Grédiac, and F. Sur. An experimental study of camera sensor noise propagation to displacement maps obtained by DIC. In *Photomechanics*, Delft, Netherlands, 2015.
- [45] M. Grédiac and F. Sur. How noise propagates from camera sensor noise to displacement and strain maps obtained with the grid method. In *International Conference on Experimental Mechanics (ICEM)*, Cambridge, UK, 2014.
- [46] M. Grédiac, F. Sur, C. Badulescu, and J.-D. Mathias. Improving the spatial resolution of the grid method with deconvolution. In *Photomechanics*, Montpellier, France, 2013.
- [47] V. Cortier, J. Detrey, P. Gaudry, F. Sur, E. Thomé, M. Turuani, and P. Zimmermann. Ballot stuffing in a postal voting system. In *Proceedings of the international workshop on requirements engineering for electronic voting systems REVOTE*, Trento, Italy, 2011.
- [48] F. Sur, M.-O. Berger, and N. Noury. Point correspondences and repeated patterns: beyond the curse of perceptual aliasing. In *Tutorial “Tools and Methods in Image Registration” at IEEE International Conference on Computer Vision and Pattern Recognition (CVPR)*, Colorado Springs (CO), USA, 2011.
- [49] F. Sur, J.-M. Morel, and G. Yu. Similarity and affine invariant point detectors and descriptors. In *Tutorial “Tools and Methods in Image Registration” at IEEE International Conference on Computer Vision and Pattern Recognition (CVPR)*, Colorado Springs (CO), USA, 2011.

Rapports de recherche

- [50] F. Sur and M. Grédiac. Measuring the noise of imaging sensors in the presence of vibrations and illumination flickering: modeling, algorithm, and experiments. Research report 8672, INRIA, 2015.
- [51] F. Sur and M. Grédiac. An automated approach to quasi-periodic noise removal in natural images. Research report 8660, INRIA, 2015.
- [52] M. Grédiac, F. Sur, C. Badulescu, and J.-D. Mathias. Using deconvolution to improve the metrological performance of the grid method. Research Report 8127, INRIA, 2012.
- [53] F. Sur and M. Grédiac. Towards deconvolution to enhance the grid method for in-plane strain measurement. Research report 8126, INRIA, 2012.
- [54] F. Sur, N. Noury, and M.-O. Berger. Image point correspondences and repeated patterns. Research report 7693, INRIA, 2011.
- [55] F. Sur. Robust matching in an uncertain world. Research Report 7374, INRIA, 2010.
- [56] N. Noury, F. Sur, and M.-O. Berger. Determining point correspondences between two views under geometric constraint and photometric consistency. Research report 7246, INRIA, 2010.
- [57] F. Sur. Invariant image descriptors and affine morphological scale-space. Research Report 6250, INRIA, 2007.
- [58] P. Musé, F. Sur, F. Cao, J.L. Lisani, and J.-M. Morel. A theory of shape identification. Research Report 5766, INRIA, 2005.
- [59] F. Cao, A. Desolneux, J. Delon, J.L. Lisani, P. Musé, and F. Sur. A unified framework for detecting groups and application to shape recognition. Research Report 5695, INRIA, 2005.
- [60] P. Musé, F. Sur, F. Cao, Y. Gousseau, and J.-M. Morel. An *a contrario* decision method for shape element recognition. Research Report 2004-16, CMLA, ENS Cachan, 2004.
- [61] F. Cao, J. Delon, A. Desolneux, P. Musé, and F. Sur. An *a contrario* approach to hierarchical clustering validity assessment. Research Report 2004-13, CMLA, ENS Cachan, 2004.
- [62] P. Musé, F. Sur, F. Cao, Y. Gousseau, and J.-M. Morel. Accurate estimates of false alarm number in shape recognition. Research Report 2004-01, CMLA, ENS Cachan, 2004.
- [63] F. Cao, P. Musé, and F. Sur. Extracting meaningful curves from images. Research Report 5067, INRIA, 2003.
- [64] P. Musé, F. Sur, and J.-M. Morel. Recherche dans les grandes bases de formes. Research Report 2002-02, CMLA, ENS Cachan, 2002.

Mémoires

- [65] F. Sur. *A contrario decision for shape recognition*. PhD thesis, Université Paris Dauphine, Paris (France), 2004.
- [66] F. Sur. *Reconstruction de surfaces en imagerie sismique 3D*. Master's thesis, Université Paris Dauphine, 1999. (Mémoire de DEA).

Documents pédagogiques (polycopiés)

- [67] C. Chambon, Y. Gueniffey, S. Herrmann, and F. Sur. *Analyse des chroniques*. École des Mines de Nancy, 2009-2016. Notes du cours du département GIMA « Modélisation et Prévision », 195 pages.
- [68] F. Sur. *Programmation dynamique, chaînes de Markov, files d'attente*. École des Mines de Nancy, 2010-2014. Notes du cours de tronc commun « Recherche Opérationnelle », 134 pages.
- [69] F. Sur. *Initiation au traitement du signal et applications*. École des Mines de Nancy, 2009-2012. Notes de cours électif, 111 pages.
- [70] F. Sur. *Langage C et aléa*. École des Mines de Nancy, 2005-2009. Notes d'un cours de semaine départementale, 27 pages.
- [71] F. Sur. *Quelques problèmes géométriques en vision par ordinateur*, 2009. Version longue d'un texte écrit pour l'épreuve des TIPE (concours d'entrée aux Grandes Écoles), 20 pages.

Bibliographie

- [AB02] I. Aizenberg and C. Butakoff. Frequency domain median-like filter for periodic and quasi-periodic noise removal. In *SPIE Proceedigns of Image Processing: Algorithms and Systems*, volume 4767, pages 181–191, 2002.
- [AB08] I. Aizenberg and C. Butakoff. A windowed Gaussian notch filter for quasi-periodic noise removal. *Image and Vision Computing*, 26(10):1347–1353, 2008.
- [Abr97] P. Abrahamsen. A review of Gaussian random fields and correlation functions. Technical report, Norwegian Computing Center, Oslo, 1997.
- [ADBS16] C. Aguerrebere, M. Delbracio, A. Bartesaghi, and G. Sapiro. Fundamental limits in multi-image alignment. *IEEE Transactions on Signal Processing*, 64(21):5707–5722, 2016.
- [ADGM13] C. Aguerrebere, J. Delon, Y. Gousseau, and P. Musé. Study of the digital camera acquisition process and statistical modeling of the sensor raw data. Technical report, HAL hal-00733538, 2013.
- [ADP12] H. Aanæs, A. Dahl, and K.S. Pedersen. Interesting interest points: A comparative study of interest point performance on a unique data set. *International Journal of Computer Vision*, 97(1):18–35, 2012.
- [AF95] F. Auger and P. Flandrin. Improving the readability of time-frequency and time-scale representations by the reassignment method. *IEEE Transactions on Signal Processing*, 43(5):1068–1089, 1995.
- [AF14a] L. Azzari and A. Foi. Gaussian-Cauchy mixture modeling for robust signal-dependent noise estimation. In *Proceedings of the International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 5357–5361, Florence, Italy, 2014.
- [AF14b] L. Azzari and A. Foi. Indirect estimation of signal-dependent noise with non-adaptive heterogeneous samples. *IEEE Transactions on Image Processing*, 23(8):3459–3467, 2014.
- [AFDM08] A. Angeli, D. Filliat, S. Doncieux, and J.-A. Meyer. Fast and incremental method for loop-closure detection using bags of visual words. *IEEE Transactions on Robotics*, 24(5):1027–1037, 2008.
- [AFE⁺09] W. Aguilar, Y. Frauel, F. Escolano, M. E. Martinez-Perez, A. Espinosa-Romero, and M. A. Lozano. A robust graph transformation matching for non-rigid registration. *Image and Vision Computing*, 27(7):897–910, 2009.
- [AFS⁺11] S. Agarwal, Y. Furukawa, N. Snavely, I. Simon, B. Curless, S. M. Seitz, and R. Szeliski. Building Rome in a day. *Communications of the ACM*, 54(10):105–112, 2011.
- [AFV⁺04] S. Avril, E. Ferrier, A. Vautrin, P. Hamelin, and Y. Sirel. A full-field optical method for the experimental analysis of reinforced concrete beams repaired with

- composites. *Composites Part A: Applied Science and Manufacturing*, 35(7–8):873–884, 2004.
- [AGCO06] J.-F. Aujol, G. Gilboa, T. Chan, and S. Osher. Structure-texture image decomposition – modeling, algorithms, and parameter selection. *International Journal of Computer Vision*, 67(1):111–136, 2006.
- [AM16] R. Abergel and L. Moisan. The Shannon total variation. Technical Report MAP5 2016-19, Université Paris Descartes, 2016.
- [Ans48] F.J. Anscombe. The transformation of Poisson binomial and negative-binomial data. *Biometrika*, 35(3-4):246–254, 1948.
- [Att54] F. Attneave. Some informational aspects of visual perception. *Psychological Review*, 61(3):183–193, 1954.
- [BCL15] M. Billinghurst, A. Clark, and G. Lee. A survey of augmented reality. *Foundations and Trends in Human-Computer Interactions*, 8(2-3):73–272, 2015.
- [BCM05] A. Buades, B. Coll, and J.-M. Morel. A review of image denoising algorithms, with a new one. *SIAM Journal on Multiscale Modeling and Simulation*, 4(2):490–530, 2005.
- [BCM⁺14] S. Blusseau, A. Carboni, A. Maiche, J.-M. Morel, and R. Grompone von Gioi. A psychophysical evaluation of the a contrario detection theory. In *Proceedings of the IEEE International Conference on Image Processing (ICIP)*, Paris, France, 2014.
- [BETG08] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool. Speeded-Up Robust Features (SURF). *Computer Vision and Image Understanding*, 110(3):346–359, 2008.
- [BGM09] C. Badulescu, M. Grédiac, and J.-D. Mathias. Investigation of the grid method for accurate in-plane strain measurement. *Measurement Science and Technology*, 20(9):095102, 2009.
- [BGMR09] C. Badulescu, M. Grédiac, J.-D. Mathias, and D. Roux. A procedure for accurate one-dimensional strain measurement using the grid method. *Experimental Mechanics*, 49(6):841–854, 2009.
- [Bha13] S. Bhat. *Visual words for pose computation*. PhD thesis, Université de Lorraine, Nancy (France), 2013.
- [BKB⁺10] J. Boulanger, C. Kervrann, P. Bouthemy, P. Elbau, J.-B. Sibarita, and J. Salamero. Patch-based nonlocal functional for denoising fluorescence microscopy image sequences. *IEEE Transaction on Medical Imaging*, 29(2):442–454, 2010.
- [BLMT09] T. Buades, Y. Lou, J.-M. Morel, and Z. Tang. A note on multi-image denoising. In *Proceedings of the International Workshop on Local and Non-Local Approximation in Image Processing*, pages 1–15, Tuusula, Finland, 2009.
- [BLMV10] A. Buades, T. M. Le, J. M. Morel, and L. A. Vese. Fast cartoon + texture image filters. *IEEE Transactions on Image Processing*, 19(8):1978–1986, 2010.
- [Bod99] F.R. Boddeke. *Quantitative fluorescence microscopy*. PhD thesis, Technische Universiteit Delft, Delft (Netherlands), 1999.
- [Bon36] C.E. Bonferroni. Teoria statistica delle classi e calcolo delle probabilità. *Pubblicazioni del R. Istituto Superiore di Scienze Economiche e Commerciali di Firenze*, 8:3–62, 1936.

-
- [BPK03] M. Brett, W. Penny, and S. Kiebel. Introduction to random field theory. In J. Ashburner, K. Friston, and W. Penny, editors, *Human brain function, Part II*. Elsevier, 2nd edition, 2003.
- [Bra06] S. Brandt. On the probabilistic epipolar geometry. *Image and Vision Computing*, 26(3):405–414, 2006.
- [BZU10] C. Le Brese, J.J. Zou, and B. Uy. An improved ASIFT algorithm for matching repeated patterns. In *Proceedings of the IEEE International Conference on Image Processing (ICIP)*, pages 2949–2952, Hong Kong, 2010.
- [CB13] M. Colom and A. Buades. Analysis and extension of the Ponomarenko et al. method, estimating a noise curve from a single image. *Image Processing On Line (I POL)*, 3:173–197, 2013.
- [CBM14] M. Colom, A. Buades, and J.-M. Morel. Nonparametric noise estimation method for raw images. *Journal of the Optical Society of America A*, 31(4):863–871, 2014.
- [CBSF09] A. Collet, D. Berenson, S.S. Srinivasa, and D. Ferguson. Object recognition and full pose registration from a single image for robotic manipulation. In *Proceedings of the International Conference on Robotics and Automation (ICRA)*, pages 48–55, Kobe, Japan, 2009.
- [CCM⁺09] R. Cai, J.F.G. Cobben, J.M.A. Myrzik, J.H. Blom, and W.L. Kling. Flicker responses of different lamp types. *IET Generation, Transmission & Distribution*, 3(9):816–824, 2009.
- [CDCS12] B. Cornelis, A. Doms, J. Cornelis, and P. Schelkens. Digital canvas removal in paintings. *Signal Processing*, 92(4):1166–1171, 2012.
- [Che95] Y. Cheng. Mean shift, mode seeking, and clustering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 17(8):790–799, 1995.
- [CJ14] A. Chambolle and K. Jalalzai. Adapted basis for nonlocal reconstruction of missing spectrum. *SIAM Journal on Imaging Sciences*, 7(3):1484–1502, 2014.
- [CLP83] M. Cannon, A. Lehar, and F. Preston. Background pattern removal by power spectral filtering. *Applied Optics*, 22(6):777–779, 1983.
- [CM02] D. Comaniciu and P. Meer. Mean shift: a robust approach toward feature space analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(5):603–619, 2002.
- [CM05] O. Chum and J. Matas. Matching with PROSAC - progressive sample consensus. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, volume 1, pages 220–226, San Diego (CA), USA, 2005.
- [CM10] B. Coulangue and L. Moisan. An aliasing detection algorithm based on suspicious colocalizations of Fourier coefficients. In *Proceedings of the IEEE International Conference on Image Processing (ICIP)*, pages 2013–2016, Hong Kong, 2010.
- [CMK03] O. Chum, J. Matas, and J.V. Kittler. Locally optimized RANSAC. In *Proceedings of the DAGM Symposium*, pages 236–243, Magdeburg, Germany, 2003.
- [CMP10] J. Cech, J. Matas, and M. Perdoch. Efficient sequential correspondence selection by cosegmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(9):1568–1581, 2010.

- [Coh95] L. Cohen. *Time-frequency analysis*. Prentice-Hall, 1995.
- [Col01] S. Coles. *An introduction to statistical modeling of extreme values*. Springer, 2001.
- [Cox55] D.R. Cox. Some statistical methods connected with series of events. *Journal of the Royal Statistical Society. Series B (Methodological)*, 17(2):129–164, 1955.
- [CRC16] B. Charrette, E. Royer, and F. Chausse. Vision-based robot localization based on the efficient matching of planar features. *Machine Vision and Applications*, 27(4):415–436, 2016.
- [CRS85] T. C. Chu, W. F. Ranson, and M. A. Sutton. Applications of Digital-Image-Correlation techniques to experimental mechanics. *Experimental Mechanics*, 25(3):232–244, 1985.
- [CS12] A. Chrysochoos and Y. Surrél. Basics of metrology and introduction to techniques. In M. Grédiac and F. Hild, editors, *Full-field measurements and identification in solid mechanics*, pages 1–29. Wiley, 2012.
- [CYG⁺16] B. Cornelis, H. Yang, A. Goodfriend, N. Ocon, J. Lu, and I. Daubechies. Removal of canvas patterns in digital acquisitions of paintings. *IEEE Transactions on Image Processing*, 2016. Early access article.
- [CZSP12] A. Cohen, C. Zach, S.N. Sinha, and M. Pollefeys. Discovering and exploiting 3D symmetries in structure from motion. In *Proceedings of the Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1514–1521, Providence (RI), USA, 2012.
- [DD95] D.F. DeMenthon and L.S. Davis. Model-based object pose in 25 lines of code. *International Journal of Computer Vision*, 15(1-2):123–141, 1995.
- [Del04] J. Delon. Midway image equalization. *Journal of Mathematical Imaging and Vision*, 21(2):119–134, 2004.
- [Des16] A. Desolneux. When the a contrario approach becomes generative. *International Journal of Computer Vision*, 116(1):46–65, 2016.
- [DFKE07] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian. Image denoising by sparse 3-D transform-domain collaborative filtering. *IEEE Transactions on Image Processing*, 16(8):2080–2095, 2007.
- [DGBB12] D. Delpueyo, M. Grédiac, X. Balandraud, and C. Badulescu. Investigation of martensitic microstructures in a monocrystalline Cu-Al-Be shape memory alloy with the grid method and infrared thermography. *Mechanics of Materials*, 45(1):34–51, 2012.
- [DKP09] F. Dibos, G. Koepfler, and S. Pelletier. Adapted windows detection of moving objects in video scenes. *SIAM Journal on Imaging Sciences*, 2(1):1–19, 2009.
- [DLR⁺08] S. Delpretti, F. Luisier, S. Ramani, T. Blu, and M. Unser. Multiframe sure-let denoising of timelapse fluorescence microscopy images. In *Proceedings of the IEEE International Symposium on Biomedical Imaging (ISBI)*, pages 149–152, Paris, France, 2008.
- [DMM00] A. Desolneux, L. Moisan, and J.-M. Morel. Meaningful alignments. *International Journal of Computer Vision*, 40(1):7–23, 2000.

-
- [DMM03] A. Desolneux, L. Moisan, and J.-M. Morel. Computational gestalts and perception thresholds. *Journal of Physiology-Paris*, 97(2–3):311–324, 2003.
- [DMM08] A. Desolneux, L. Moisan, and J.-M. Morel. *From Gestalt theory to image analysis: a probabilistic approach*. Interdisciplinary applied mathematics. Springer, 2008.
- [DMSD06] H. Deng, E. N. Mortensen, L. Shapiro, and T. G. Dietterich. Reinforcement matching using region context. In *Proceedings of the Beyond Patches workshop at the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, New York (NY), USA, 2006.
- [DS15] M. Delbracio and G. Sapiro. Removing camera shake via weighted Fourier burst accumulation. *IEEE Transactions on Image Processing*, 24(11):3293–3307, 2015.
- [DVJ03] D.J. Daley and D. Vere-Jones. *An introduction to the theory of point processes*, volume 1. Springer, 2nd edition, 2003.
- [DXW⁺14a] X. Dai, H. Xie, H. Wang, C. Li, Z. Liu, and L. Wu. The geometric phase analysis method based on the local high resolution discrete Fourier transform for deformation measurement. *Measurement Science and Technology*, 25(2):025402, 2014.
- [DXW14b] X. Dai, H. Xie, and Q. Wang. Geometric phase analysis based on the windowed Fourier transform for the deformation field measurement. *Optics and Laser Technology*, 58:119–127, 2014.
- [emv10] Standard 1288, standard for characterization of image sensors and cameras, release 3.0. Technical report, European Machine Vision Association (EMVA), 2010.
- [ET84] B. Ellingwood and A. Tallin. Structural serviceability: Floor vibrations. *Journal of Structural Engineering*, 110(2):401–418, 1984.
- [FACM03] P. Flandrin, F. Auger, and E. Chassande-Mottin. *Applications in time-frequency signal processing*, chapter Time-frequency reassignment: From principles to algorithms, pages 179–203. CRC Press, 2003.
- [FB81] M. Fischler and R. Bolles. Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, 24(6):381–395, 1981.
- [FCF⁺95] S.D. Forman, J.D. Cohen, M. Fitzgerald, W.F. Eddy, M.A. Mintun, and D.C. Noll. Improved assessment of significant activation in functional magnetic resonance imaging (fMRI): Use of a cluster-size threshold. *Magnetic Resonance in Medicine*, 33(5):636–647, 1995.
- [FH11] A. Flenner and G. Hewer. A Helmholtz principle approach to parameter free change detection and coherent motion using exchangeable random variables. *SIAM Journal on Imaging Sciences*, 4(1):243–276, 2011.
- [Fie87] D.J. Field. Relations between the statistics of natural images and the response properties of cortical cells. *Journal of the Optical Society of America*, A, 4:2379–2394, 1987.

- [FKE07] A. Foi, V. Katkovnik, and K. Egiazarian. Pointwise shape-adaptive DCT for high-quality denoising and deblocking of grayscale and color images. *IEEE Transactions on Image Processing*, 16(5):1395–1411, 2007.
- [Fla98] P. Flandrin. Separability, positivity, and minimum uncertainty in time–frequency energy distributions. *Journal of Mathematical Physics*, 39(8):4016–4040, 1998.
- [FM06] H. Faraji and W.J. MacLean. CCD noise removal in digital images. *IEEE Transactions on Image Processing*, 15(9):2676–2685, 2006.
- [FREM04] S. Farsiu, M.D. Robinson, M. Elad, and P. Milanfar. Fast and robust multiframe super resolution. *IEEE Transactions on Image Processing*, 13(10):1327–1344, 2004.
- [FSBM10] M.J. Fadili, J.-L. Starck, J. Bobin, and Y. Moudden. Image decomposition and separation using sparse representations: An overview. *Proceedings of the IEEE*, 98(6):983–994, 2010.
- [FTKE08] A. Foi, M. Trimeche, V. Katkovnik, and K. Egiazarian. Practical Poissonian-Gaussian noise modeling and fitting for single-image raw-data. *IEEE Transactions on Image Processing*, 17(10):1737–1754, 2008.
- [FWL12] J. Fehrenbach, P. Weiss, and C. Lorenzo. Variational algorithms to remove stationary noise: Applications to microscopy imaging. *IEEE Transactions on Image Processing*, 21(10):4420–4430, 2012.
- [FZB02] H. Foroosh, J.B. Zerubia, and M. Berthod. Extension of phase correlation to subpixel registration. *IEEE Transactions on Image Processing*, 11(3):188–200, 2002.
- [GL06] I. Gordon and D.G. Lowe. What and where: 3D object recognition with accurate pose. In J. Ponce, M. Hebert, C. Schmid, and A. Zisserman, editors, *Toward Category-Level Object Recognition*, volume 4170 of *Lecture Notes in Computer Science*, pages 67–82. Springer, 2006.
- [GLH⁺15] R. Gras, H. Leclerc, F. Hild, S. Roux, and J. Schneider. Identification of a set of macroscopic elastic parameters in a 3D woven composite: Uncertainty analysis and regularization. *International Journal of Solids and Structures*, 55:2–16, 2015.
- [GM98] F. Guichard and F. Malgouyres. Total variation based interpolation. In *Proceedings of the European Signal Processing Conference (EUSIPCO)*, volume 3, pages 1741–1744, Island of Rhodes, Greece, 1998.
- [GM09] B. Grosjean and L. Moisan. A-contrario detectability of spots in textured backgrounds. *Journal of Mathematical Imaging and Vision*, 33(3):313–337, 2009.
- [GP98] D.C. Ghiglia and M.D. Pritt. *Two dimensional phase unwrapping: theory, algorithms and software*. Wiley and Sons, 1998.
- [GT13] M. Grédiac and E. Toussaint. Studying the mechanical behaviour of asphalt mixtures with the grid method. *Strain*, 49(1):1–15, 2013.
- [GTO14] J. Gilles, G. Tran, and S. Osher. 2D empirical transforms. Wavelets, ridgelets, and curvelets revisited. *SIAM Journal on Imaging Sciences*, 7(1):157–186, 2014.
- [GW06] R.C. Gonzalez and R.E. Woods. *Digital Image Processing (3rd edition)*. Prentice-Hall, 2006.

-
- [HBB92] F. Hlawatsch and G.F. Boudreaux-Bartels. Linear and quadratic time-frequency signal representations. *IEEE Signal Processing Magazine*, 9(2):21–67, 1992.
- [HBLG02] M.A. Herraiez, D.R. Burton, M.J. Lalor, and M.A. Gdeisat. Fast two-dimensional phase-unwrapping algorithm based on sorting by reliability following a noncontinuous path. *Applied Optics*, 41(35):7437–7444, 2002.
- [HCH10] E. Hsiao, A. Collet, and M. Hebert. Making specific features less discriminative to improve point-based 3D object recognition. In *Proceedings of the Conference on Computer Vision and Pattern Recognition (CVPR)*, San Francisco (CA), USA, 2010.
- [HDD⁺92] H. Hoppe, T. DeRose, T. Duchamp, J. McDonald, and W. Stuetzle. Surface reconstruction from unorganized points. *ACM SIGGRAPH Computer Graphics*, 26(2):71–78, 1992.
- [HHH09] A. Hyvärinen, J. Hurri, and P.O. Hoyer. *Natural Image Statistics: A Probabilistic Approach to Early Computational Vision*. Springer, 2009.
- [HK94] G.E. Healey and R. Kondepudy. Radiometric CCD camera calibration and noise estimation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 16(3):267–276, 1994.
- [HMdC87] F. Heitz, H. Maitre, and C. de Couessin. Application of autoregressive models to fine arts painting analysis. *Signal Processing*, 13(1):1–14, 1987.
- [HR11] J.A. Hesch and S.I. Roumeliotis. A direct least-squares (DLS) method for PnP. In *Proceedings of the International Conference on Computer Vision (ICCV)*, pages 383–390, Barcelona, Spain, 2011.
- [HR12] F. Hild and S. Roux. Comparison of local and global approaches to digital image correlation. *Experimental Mechanics*, 52(9):1503–1519, 2012.
- [HSK98] M.J. Hÿtch, E. Snoeck, and R. Kilaas. Quantitative measurement of displacement and strain fields from HREM micrographs. *Ultramicroscopy*, 74(3):131–146, 1998.
- [HSL⁺98] N.E. Huang, Z. Shen, S.R. Long, M.C. Wu, H.H. Shih, Q. Zheng, N.-C. Yen, C.C. Tung, and H.H. Liu. The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 454(1971):903–995, 1998.
- [HT05] G.A.A. Hudhud and M.J. Turner. Digital removal of power frequency artifacts using a Fourier space median filter. *IEEE Signal Processing Letters*, 12(8):573–576, 2005.
- [Hyt06] H.T. Hytti. Characterization of digital image noise properties based on raw data. In *Proceedings of SPIE-IS&T Electronic Imaging, Image Quality and System Performance III*, volume 6059, San Jose (CA), USA, 2006.
- [HZ00] R. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, 2000.
- [ICST15] G.-M. Ionita, D. Coltuc, S. G. Stanciu, and D. E. Tranca. Automatic moiré pattern removal in microscopic images. In *Proceedings of the 19th International Conference on System Theory, Control and Computing (ICSTCC)*, pages 776–779, Cheile Gradistei, Romania, 2015.

- [iee10] A review of the literature on light flicker: Ergonomics, biological attributes, potential health effects, and methods in which some LED lighting may introduce flicker. Technical report, IEEE Standards P1789, 2010.
- [IZFB09] A. Irschara, C. Zach, J.-M. Frahm, and H. Bischof. From structure-from-motion point clouds to fast location recognition. In *Proceedings of the International Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2599–2606, Miami (FL), USA, 2009.
- [Jal12] K. Jalalzai. *Regularization of inverse problems in image processing*. PhD thesis, École Polytechnique, Palaiseau (France), 2012.
- [JCP⁺14] A. Jezierska, C. Chaux, J.-C. Pesquet, H. Talbot, and G. Engler. An EM approach for time-variant Poisson-Gaussian model parameter estimation. *IEEE Transactions on Signal Processing*, 62(1):17–30, 2014.
- [JPTC14] A. Jezierska, J.-C. Pesquet, H. Talbot, and C. Chaux. Iterative Poisson-Gaussian noise parametric estimation for blind image denoising. In *Proceedings of the IEEE International Conference on Image Processing (ICIP)*, Paris, France, 2014.
- [JSHV10] H. Jégou, C. Schmid, H. Harzallah, and J. Verbeek. Accurate image search using the contextual dissimilarity measure. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(1):2–11, 2010.
- [JXH14] X. Jin, Z. Xu, and K. Hirakawa. Noise parameter estimation for Poisson corrupted images using variance stabilization transforms. *IEEE Transactions on Image Processing*, 23(3):1329–1339, 2014.
- [KEA06] V. Katkovnik, K. Egiazarian, and J. Astola. *Local approximation techniques in signal and image processing*. SPIE publications, 2006.
- [Kem07] Q. Kemaο. Two-dimensional windowed Fourier transform for fringe pattern analysis: Principles, applications and implementations. *Optics and Lasers in Engineering*, 45(2):304–317, 2007.
- [Kem15] Q. Kemaο. Applications of windowed Fourier fringe analysis in optical measurement: A review. *Optics and Lasers in Engineering*, 66:67–73, 2015.
- [KK01] Y. Kanazawa and K. Kanatani. Do we really have to consider covariance matrices for image features? In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, volume 2, pages 301–306, Los Alamitos (CA), USA, 2001.
- [KS14] M. Kushnir and I. Shimshoni. Epipolar geometry estimation for urban scenes with repetitive structures. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 36(12):2381–2395, 2014.
- [KvdG80] J.D. Krol and W.A. van de Grind. The double-nail illusion: experiments on binocular vision with nails, needles, and pins. *Perception*, 9(6):651–669, 1980.
- [KZ13] C. Knaus and M. Zwicker. Dual-domain image denoising. In *Proceedings of the IEEE International Conference on Image Processing (ICIP)*, pages 440–444, Melbourne, Australia, 2013.
- [KZ15] C. Knaus and M. Zwicker. Dual-domain filtering. *SIAM Journal on Imaging Sciences*, 8(3):1396–1420, 2015.

-
- [LCBM12] M. Lebrun, M. Colom, A. Buades, and J.-M. Morel. Secrets of image denoising cuisine. *Acta Numerica*, 21(1):475–576, 2012.
- [LF05] V. Lepetit and P. Fua. Monocular model-based 3D tracking of rigid objects: A survey. *Foundations and Trends in Computer Graphics and Vision*, 1(1):1–89, 2005.
- [LF06] V. Lepetit and P. Fua. Keypoint recognition using randomized trees. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(9):1465–1479, 2006.
- [LMNF09] V. Lepetit, F. Moreno-Noguer, and P. Fua. EPnP: An accurate $O(n)$ solution to the PnP problem. *International Journal of Computer Vision*, 81(2):155–166, 2009.
- [Loh33] R. W. Lohman. Neon tube characteristics. *Electrical Engineering*, 52(5):304–307, 1933.
- [Low04] D. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2):91–110, 2004.
- [LS04] Z. Lin and H.-Y. Shum. Fundamental limits of reconstruction-based super-resolution algorithms under local translation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(1):83–97, 2004.
- [LSB⁺15] S. Lynen, T. Sattler, M. Bosse, J. Hesch, M. Pollefeys, and R. Siegwart. Get out of my lab: Large-scale, real-time visual-inertial localization. In *Proceedings of the Conference Robotics Science and Systems (RSS)*, Roma, Italy, 2015.
- [LSC⁺15] H. Lim, S.N. Sinha, M.F. Cohen, M. Uyttendaele, and H.J. Kim. Real-time monocular image-based 6-dof localization. *International Journal of Robotics Research*, 34(4-5):476–492, 2015.
- [LSHF12] Y. Li, N. Snavely, D. Huttenlocher, and P. Fua. Worldwide pose estimation using 3D point clouds. In *Proceedings of the European Conference on Computer Vision (ECCV)*, volume 1, pages 15–29, Florence, Italy, 2012.
- [LSK⁺08] C. Liu, R. Szeliski, S.B. Kang, C.L. Zitnick, and W.T. Freeman. Automatic estimation and removal of noise from a single image. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30(2):299–314, 2008.
- [LTO14] X. Liu, M. Tanaka, and M. Okutomi. Practical signal-dependent noise parameter estimation from a single noisy image. *IEEE Transactions on Image Processing*, 23(10):4361–4371, 2014.
- [Luc74] L. B. Lucy. An iterative technique for the rectification of observed distributions. *Astronomical Journal*, 79(6):745–754, 1974.
- [LvGRM14] J. Lezama, R. Grompone von Gioi, G. Randall, and J.-M. Morel. A contrario detection of good continuation of points. In *Proceedings of the IEEE International Conference on Image Processing (ICIP)*, Paris, France, 2014.
- [LXZG15] H. Liu, R. Xiong, J. Zhang, and W. Gao. Image denoising via adaptive soft-thresholding based on non-local samples. In *Proceedings of the Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 484–492, Boston (MA), USA, 2015.
- [LYT⁺14] Z. Liu, L. Yuan, X. Tang, M. Uyttendaele, and J. Sun. Fast burst images denoising. *ACM Transactions on Graphics*, 33(6):232:1–232:9, 2014.

- [MA10] D.M. Mount and S. Arya. ANN: a library for approximate nearest neighbor searching, 2010. <https://www.cs.umd.edu/~mount/ANN/>.
- [Mal99] S. Mallat. *A wavelet tour of signal processing (2nd edition)*. Academic Press, 1999.
- [Mar82] D. Marr. *Vision*. The MIT Press, 1982.
- [MC04] J. Matas and O. Chum. Randomized RANSAC with $T_{d,d}$ test. *Image and Vision Computing*, 22(10):837–842, 2004.
- [MCUP04] J. Matas, O. Chum, M. Urban, and T. Pajdla. Robust wide-baseline stereo from maximally stable extremal regions. *Image and Vision Computing*, 22(10):761–767, 2004.
- [MF13] M. Makitalo and A. Foi. Optimal inversion of the generalized Anscombe transformation for Poisson-Gaussian noise. *IEEE Transactions on Image Processing*, 22(1):91–103, 2013.
- [MF14] M. Makitalo and A. Foi. Noise parameter mismatch in variance stabilization, with an application to Poisson-Gaussian noise estimation. *IEEE Transactions on Image Processing*, 23(12):5348–5359, 2014.
- [MGL13] A. Myaskovskey, Y. Gousseau, and M. Lindenbaum. Beyond independence: An extension of the a contrario decision procedure. *International Journal of Computer Vision*, 101(1):22–44, 2013.
- [Mil13] P. Milanfar. A tour of modern image filtering: New insights and methods, both practical and theoretical. *IEEE Signal Processing Magazine*, 30(1):106–128, 2013.
- [ML14] M. Muja and D.G. Lowe. Scalable nearest neighbor algorithms for high dimensional data. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 36(11):2227–2240, 2014.
- [MMM12] P. Moulon, P. Monasse, and R. Marlet. Adaptive structure from motion with a contrario model estimation. In *Proceedings of the Asian Conference on Computer Vision (ACCV)*, pages 257–270, Daejeon, Korea, 2012.
- [MMP15] D. Mishkin, J. Matas, and M. Perdoch. MODS: Fast and robust method for two-view matching. *Computer Vision and Image Understanding*, 141:81–93, 2015.
- [MN13] J. Molimard and L. Navarro. Uncertainty on fringe projection technique: A Monte-Carlo-based approach. *Optics and Lasers in Engineering*, 51(7):840–847, 2013.
- [Moi01] L. Moisan. Extrapolation de spectre et variation totale pondérée. In *Actes du colloque sur le traitement du signal et des images (GRETSI)*, Toulouse (France), 2001.
- [MP07] P. Moreels and P. Perona. Evaluation of features detectors and descriptors based on 3D objects. *International Journal of Computer Vision*, 73(3):263–284, 2007.
- [MS04a] K. Mikolajczyk and C. Schmid. Scale & affine invariant interest point detectors. *International Journal of Computer Vision*, 60(1):63–86, 2004.
- [MS04b] L. Moisan and B. Stival. A probabilistic criterion to detect rigid point matches between two images and estimate the fundamental matrix. *International Journal of Computer Vision*, 57(3):201–218, 2004.

-
- [MSB95] F. Murthag, J.L. Starck, and A. Bijaoui. Image restoration with noise suppression using a multiresolution support. *Astronomy and astrophysics*, 112:179–189, 1995.
- [MTS⁺06] K. Mikolajczyk, T. Tuytelaars, C. Schmid, A. Zisserman, J. Matas, F. Schafalitzky, T. Kadir, and L. Van Gool. A comparison of affine region detectors. *International Journal of Computer Vision*, 65(1/2):43–72, 2006.
- [MY09] J.-M. Morel and G. Yu. ASIFT: A new framework for fully affine invariant image comparison. *SIAM Journal on Imaging Sciences*, 2(2):438–469, 2009.
- [MY11] J.-M. Morel and G. Yu. Is SIFT scale invariant? *AIMS Inverse Problems and Imaging*, 5(1):115–136, 2011.
- [Nis05] D. Nistér. Preemptive RANSAC for live structure and motion estimation. *Machine Vision and Applications*, 16(5):321–329, 2005.
- [Nou11] N. Noury. *Mise en correspondance a contrario de points d'intérêt sous contraintes géométrique et photométrique*. PhD thesis, Université Henri-Poincaré, Nancy (France), 2011.
- [NS06] D. Nister and H. Stewenius. Scalable recognition with a vocabulary tree. In *Proceedings of the International Conference on Computer Vision and Pattern Recognition (CVPR)*, volume 2, pages 2161–2168, New York (NY), USA, 2006.
- [OB06] B. Ochoa and S. Belongie. Covariance propagation for guided matching. In *Proceedings of the Workshop on Statistical Methods in Multi-Image and Video Processing (SMVP)*, Graz, Austria, 2006.
- [OCLF10] M. Ozuysal, M. Calonder, V. Lepetit, and P. Fua. Fast keypoint recognition using random ferns. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 32(3):448–461, 2010.
- [OT01] A. Oliva and A. Torralba. Modeling the shape of the scene: A holistic representation of the spatial envelope. *International Journal of Computer Vision*, 42(3):145–175, 2001.
- [PBM14] G. Palma, I. Bloch, and S. Muller. Detection of masses and architectural distortions in digital breast tomosynthesis images using fuzzy and a contrario approaches. *Pattern Recognition*, 47(7):2467–2480, 2014.
- [PGvG12] V. Patraucean, P. Gurdjos, and R. Grompone von Gioi. A parameterless line segment and elliptical arc detector with enhanced ellipse fitting. In *Proceedings of the European Conference on Computer Vision (ECCV)*, volume II, pages 572–585, Florence, Italy, 2012.
- [PH14] S. Pyatykh and J. Hesser. Image sensor noise parameter estimation by variance stabilization and normality assessment. *IEEE Transactions on Image Processing*, 23(9):3990–3998, 2014.
- [PL00] T. Papadopoulos and M.I.A. Lourakis. Estimating the Jacobian of the Singular Value Decomposition: theory and applications. In *Proceedings of the European Conference on Computer Vision (ECCV)*, volume 1, pages 554–570, London, UK, 2000.
- [PLZ⁺07] N.N. Ponomarenko, V.V. Lukin, M.S. Zriakhov, A. Kaarna, and J. Astola. An automatic approach to lossy compression of AVIRIS images. In *Proceedings of*

- the International Geoscience and Remote Sensing Symposium (IGRSS)*, pages 472–475, Barcelona, Spain, 2007.
- [PT97] X. Pennec and J.-Ph. Thirion. A framework for uncertainty and validation of 3D registration methods based on points and frames. *International Journal of Computer Vision*, 25(3):203–229, 1997.
- [PVG^V+04] M. Pollefeys, L. Van Gool, M. Vergauwen, F. Verbiest, K. Cornelis, J. Tops, and R. Koch. Visual modeling with a hand-held camera. *International Journal of Computer Vision*, 59(3):207–232, 2004.
- [RDF⁺71] T. C. Rindfleisch, J. A. Dunne, H. J. Frieden, W. D. Stromberg, and R. M. Ruiz. Digital processing of the Mariner 6 and 7 pictures. *Journal of Geophysical Research*, 76(2):394–417, 1971.
- [RDG08a] J. Rabin, J. Delon, and Y. Gousseau. Circular Earth Mover’s Distance for the comparison of local features. In *Proceedings of the International Conference on Pattern Recognition (ICPR)*, Tampa (FL), USA, 2008.
- [RDG08b] J. Rabin, J. Delon, and Y. Gousseau. A contrario matching of SIFT-like descriptors. In *Proceedings of the International Conference on Pattern Recognition (ICPR)*, Tampa (FL), USA, 2008.
- [RDG09] J. Rabin, J. Delon, and Y. Gousseau. A statistical approach to the matching of local features. *SIAM Journal on Imaging Sciences*, 2(3):931–958, 2009.
- [RDGM10] J. Rabin, J. Delon, Y. Gousseau, and L. Moisan. MAC-RANSAC: a robust algorithm for the recognition of multiple objects. In *Proceedings of the Symposium on 3D Data Processing, Visualization and Transmission (3DPVT)*, Paris, France, 2010.
- [Rét10] J. Réthoré. A fully integrated noise robust strategy for the identification of constitutive laws from digital images. *International Journal for Numerical Methods in Engineering*, 84(6):631–660, 2010.
- [RFH⁺07] F. Rousseau, S. Faisan, F. Heitz, J. P. Armspach, Y. Chevalier, F. Blanc, J. de Seze, and L. Rumbach. An a contrario approach for change detection in 3D multimodal images: Application to multiple sclerosis in MRI. In *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 2069–2072, 2007.
- [RFP09] R. Raguram, J.-M. Frahm, and M. Pollefeys. Exploiting uncertainty in random sample consensus. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, pages 2074–2081, Kyoto, Japan, 2009.
- [Ric72] W. H. Richardson. Bayesian-based iterative method of image restoration. *Journal of the Optical Society of America*, 62(1):55–59, 1972.
- [RKBW13] P. Roudot, C. Kervrann, J. Boulanger, and F. Waharte. Noise modeling for intensified camera in fluorescence imaging: Application to image denoising. In *Proceedings of the IEEE International Symposium on Biomedical Imaging (ISBI)*, pages 600–603, San Francisco (CA), USA, 2013.
- [RKWB12] P. Roudot, C. Kervrann, F. Waharte, and J. Boulanger. Lifetime map reconstruction in frequency-domain fluorescence lifetime imaging microscopy. In *Proceedings of the IEEE International Conference on Image Processing (ICIP)*, pages 2537–2540, Orlando (FL), USA, 2012.

-
- [RLSP06] F. Rothganger, S. Lazebnik, C. Schmid, and J. Ponce. 3D object modeling and recognition using local affine-invariant image descriptors and multi-view spatial constraints. *International Journal of Computer Vision*, 66(3):231–259, 2006.
- [RMF15] M. Rais, J.-M. Morel, and G. Facciolo. Iterative gradient-based shift estimation: To multiscale or not to multiscale? In *Proceedings of the Iberoamerican Congress in Pattern Recognition, Image Analysis, Computer Vision, and Applications (CIARP)*, pages 416–423, Montevideo, Uruguay, 2015.
- [RMLHM10] A. Robin, L. Moisan, and S. Le Hegarat-Masclé. An a-contrario approach for subpixel change detection in satellite imagery. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(11):1977–1993, 2010.
- [RO15] I. Rey-Otero. *Anatomy of the SIFT method*. PhD thesis, École Normale Supérieure de Cachan, Cachan (France), 2015.
- [ROF92] L.I. Rudin, S. Osher, and E. Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D: Nonlinear Phenomena*, 60(1–4):259–268, 1992.
- [RSSS11] R. Roberts, S.N. Sinha, R. Szeliski, and D. Steedly. Structure from motion for scenes with large duplicate structures. In *Proceeding of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Colorado Springs (CO), USA, 2011.
- [RVU08] S. Ramani, C. Vonesch, and M. Unser. Deconvolution of 3D fluorescence micrographs with automatic risk minimization. In *Proceedings of the IEEE International Symposium on Biomedical Imaging (ISBI)*, pages 732–735, Paris, France, 2008.
- [SBS00] H.W. Schreier, J.R. Braasch, and M.A. Sutton. Systematic errors in digital image correlation caused by intensity interpolation. *Optical Engineering*, 39(11):2915–2921, 2000.
- [SBS07] G. Schindler, M. Brown, and R. Szeliski. City-scale location recognition. In *Proceedings of the International Conference on Computer Vision and Pattern Recognition (CVPR)*, Minneapolis (MN), USA, 2007.
- [Sch06] R.A. Schowengerdt. *Remote sensing: models and methods for image processing*. Academic Press, 2006.
- [SCW88] R. Srinivasan, M. Cannon, and J. White. Landsat data destriping using power spectral filtering. *Optical Engineering*, 27(11):939–943, 1988.
- [SED05] J.-L. Starck, M. Elad, and D.L. Donoho. Image decomposition via the combination of sparse representations and a variational approach. *IEEE Transactions on Image Processing*, 14(10):1570–1582, 2005.
- [sen05] Sensicam QE - 1288 datasheet. Technical report, PCO Imaging, 2005.
- [SH15] M. A. Sutton and F. Hild. Recent advances and perspectives in digital image correlation. *Experimental Mechanics*, 55:1–8, 2015.
- [Sha48] C.E. Shannon. A mathematical theory of communication. *Bell System Technical Journal*, 27:379–423, 623–656, 1948.
- [Sik95] T. Sikora. Low complexity shape-adaptive DCT for coding of arbitrarily shaped image segments. *Signal Processing: Image Communication*, 7(4–6):381–395, 1995.

- [Sna10] N. Snavely. Bundler: Structure from Motion (SfM) for unordered image collections. <http://www.cs.cornell.edu/~snavely/bundler/>, 2010.
- [SOS09] M. Sutton, J.-J. Orteu, and H. Schreier. *Image Correlation for Shape, Motion and Deformation Measurements*. Springer, 2009.
- [SSG03] H. Shao, T. Svoboda, and L. Van Gool. ZuBuD – Zurich buildings database for image based recognition. Technical Report 260, Computer Vision Lab, ETHZ, 2003.
- [SSS06] N. Snavely, S.M. Seitz, and R. Szeliski. Photo tourism: Exploring photo collections in 3D. *ACM Transactions on Graphics*, 25(3):835–846, 2006.
- [SSS08] N. Snavely, S.M. Seitz, and R. Szeliski. Modeling the world from internet photo collections. *International Journal of Computer Vision*, 80(2):189–210, 2008.
- [Sur00] Y. Surrel. *Photomechanics*, volume 77 of *Topics in Applied Physics*, chapter Fringe analysis, pages 55–102. Springer, 2000.
- [SWC⁺14] Q. Shan, C. Wu, B. Curless, Y. Furukawa, C. Hernandez, and S.M. Seitz. Accurate geo-registration by ground-to-aerial image matching. In *Proceedings of the International Conference on 3D Vision (3DV)*, volume 1, pages 525–532, 2014.
- [SZ03] J. Sivic and A. Zisserman. Video Google: A text retrieval approach to object matching in videos. In *Proceedings of the International Conference on Computer Vision (ICCV)*, volume 2, pages 1470–1477, Nice, France, 2003.
- [TA77] A.N. Tikhonov and V.A. Arsenin. *Solutions of ill-posed problems*. Winston and Sons, 1977.
- [Tak90] M. Takeda. Spatial-carrier fringe-pattern analysis and its applications to precision interferometry and profilometry: An overview. *Industrial Metrology*, 1(2):79–99, 1990.
- [TB10] B. Triggs and P. Bendale. Epipolar constraints for multiscale matching. In *Proceedings of the British Machine Vision Conference (BMVC)*, Aberystwyth, UK, 2010.
- [TBH⁺15] S. Tourbier, X. Bresson, P. Hagmann, J.-Ph. Thiran, R. Meuli, and M. Bach Cuadra. An efficient total variation algorithm for super-resolution in fetal brain MRI with adaptive regularization. *NeuroImage*, 118:584–597, 2015.
- [TC05] B.J. Tordoff and R. Cipolla. Uncertain RanSaC. In *Proceedings of the IAPR Workshop on Machine Vision Applications (MVA)*, Tsukuba, Japan, 2005.
- [TCR14] T.H. Thai, R. Cogramne, and F. Reiraint. Camera model identification based on the heteroscedastic noise model. *IEEE Transactions on Image Processing*, 23(1):250–263, 2014.
- [TEC09] D. Teyssieux, S. Euphrasie, and B. Cretin. Thermal detectivity enhancement of visible and near infrared thermography by using super-resolution algorithm: Possibility to generalize the method to other domains. *AIP Journal of Applied Physics*, 105(6):064911–064914, 2009.
- [TH86] Q. Tian and M.N. Huhns. Algorithms for subpixel registration. *Journal of Computer Vision, Graphics, and Image Processing*, 35(2):220–233, 1986.
- [TM83] M. Takeda and K. Mutoh. Fourier transform profilometry for the automatic measurement of 3-d object shapes. *Applied Optics*, 22(24):3977–3982, 1983.

-
- [TM98] C. Tomasi and R. Manduchi. Bilateral filtering for gray and color images. In *Proceedings of the International Conference on Computer Vision (ICCV)*, pages 839–846, Bombay, India, 1998.
- [TM05] B.J. Tordoff and D.W. Murray. Guided-MLESAC: faster image transform estimation by using matching priors. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(10):1523–1535, 2005.
- [TMAM11] M. Tepper, P. Musé, A. Almansa, and M. Mejail. Automatically finding clusters in normalized cuts. *Pattern Recognition*, 44(7):1372–1386, 2011.
- [TO03] A. Torralba and A. Oliva. Statistics of natural image categories. *Network*, 14(3):391–412, 2003.
- [TZ00] P. Torr and A. Zisserman. MLESAC: A new robust estimator with application to estimating image geometry. *Computer Vision and Image Understanding*, 78:138–156, 2000.
- [UVLC13] M.L. Uss, B. Vozel, V.V. Lukin, and K. Chehdi. Image informative maps for component-wise estimating parameters of signal-dependent noise. *IS&T / SPIE Journal of Electronic Imaging*, 22(1):013019, 2013.
- [Vap00] V. Vapnik. *The nature of statistical learning theory*. Springer, 2000.
- [vdSvH96] A. van der Schaaf and J.H. van Hateren. Modelling the power spectra of natural images: Statistics and information. *Vision Research*, 36(17):2759–2770, 1996.
- [vGJMR10] R. Grompone von Gioi, J. Jakubowicz, J.-M. Morel, and G. Randall. LSD: a fast line segment detector with a false detection control. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 32(4):722–732, 2010.
- [VS06] A. Vedaldi and S. Soatto. Local features, all grown up. In *Proceedings of the Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1753–1760, New York (NY), USA, 2006.
- [WACS11] C. Wu, S. Agarwal, B. Curless, and S.M. Seitz. Multicore bundle adjustment. In *Proceedings of the International Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3057–3064, Providence (RI), USA, 2011.
- [WB91] S.D. Whitehead and D.H. Ballard. Learning to perceive and act by trial and error. *Machine Learning*, 7(1):45–83, 1991.
- [WCL⁺08] C. Wu, B. Clipp, X. Li, J.-M. Frahm, and M. Pollefeys. 3D model matching with viewpoint-invariant patches (VIP). In *Proceedings of the International Conference on Computer Vision and Pattern Recognition (CVPR)*, Anchorage (AK), USA, 2008.
- [Wei98] J. Weickert. *Anisotropic diffusion in image processing*. Teuber, 1998.
- [Wer23] M. Wertheimer. Untersuchungen zur Lehre von der Gestalt, II. *Psychologische Forschung*, 4:301–350, 1923.
- [Wie64] N. Wiener. *Extrapolation, Interpolation, and Smoothing of Stationary Time Series*. The MIT Press, 1964.
- [WKR07] B. Williams, G. Klein, and I. Reid. Real-time SLAM relocalisation. In *Proceedings of the International Conference on Computer Vision (ICCV)*, Rio de Janeiro, Brazil, 2007.

- [WL10] J. Wang and D.C. Liu. 2D FFT for periodic noise removal on strain images. In *Proceedings of the 4th International Conference on Bioinformatics and Biomedical Engineering (ICBBE)*, Chengdu, China, 2010.
- [WLLD15] L. Wittevrongel, P. Lava, S.V. Lomov, and D. Debruyne. A self adaptive global digital image correlation algorithm. *Experimental Mechanics*, 55(2):361–378, 2015.
- [Wor03] K.J. Worsley. Developments in random field theory. In J. Ashburner, K. Friston, and W. Penny, editors, *Human brain function, Part II*. Elsevier, 2nd edition, 2003.
- [Wu11] C. Wu. VisualSFM: A visual structure from motion system. <http://homes.cs.washington.edu/~ccwu/vsfm/>, 2011.
- [WWN⁺12] Z. Wei, J. Wang, H. Nichol, S. Wiebe, and D. Chapman. A median-Gaussian filtering framework for moiré pattern noise removal from X-ray microscopy image. *Micron: the international research and review journal for microscopy*, 43(2-3):170–176, 2012.
- [XCYQ08] J. Xiao, J. Chen, D.-Y. Yeung, and L. Quan. Structuring visual words in 3D for arbitrary-view object localization. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 725–737, Marseille, France, 2008.
- [XLF07] F. Xue, Q. Liu, and J. Froment. An a contrario approach for parameters estimation of a motion-blurred image. In *Proceedings of the International Conference on Energy Minimization Methods in Computer Vision and Pattern Recognition (EMMCVPR)*, pages 2332–2336, Ezhou, China, 2007.
- [YM11] G. Yu and J.-M. Morel. ASIFT: An algorithm for fully affine invariant comparison. *Image Processing On Line (IPOL)*, 1, 2011.
- [YS11] G. Yu and G. Sapiro. DCT image denoising: a simple and effective image denoising algorithm. *Image Processing On Line (IPOL)*, 1, 2011.
- [ZF05] X. Zhang and J. Froment. Total variation based Fourier reconstruction and regularization for computer tomography. In *Nuclear Science Symposium Conference Record*, volume 4, pages 2332–2336, 2005.
- [ZGS⁺09] B. Zeisl, P. Georgel, F. Schweiger, E. Steinbach, and N. Navab. Estimation of location uncertainty for scale invariant feature points. In *Proceedings of the British Machine Vision Conference (BMVC)*, London, UK, 2009.
- [ZK06] W. Zhang and J. Kosecka. Generalized RANSAC framework for relaxed correspondence problems. In *Proceedings of the International Symposium on 3D Data Processing, Visualization and Transmission (3DPVT)*, Chapel Hill (NC), USA, 2006.
- [ZP01] M. Zibulevsky and B.A. Pearlmutter. Blind source separation by sparse decomposition in a signal dictionary. *Neural Computation*, 13(4):863–882, 2001.

Résumé. Ce mémoire présente de manière synthétique une sélection de travaux relevant de l'analyse et du traitement des images numériques. Il est constitué de trois parties. La première partie traite d'un problème à la base de nombreuses applications de la vision par ordinateur : l'appariement de points d'intérêt entre deux images d'une part, et entre une image et un modèle non-structuré de scène d'autre part. La deuxième partie porte sur la notion de bruit. Deux problèmes sont envisagés : la caractérisation et l'élimination d'un bruit périodique se manifestant par une texture régulière couvrant l'image, et l'estimation des paramètres du bruit blanc affectant toute image numérique. Enfin, la troisième partie présente des contributions à l'analyse d'images en mécanique des solides expérimentale. Le but est d'améliorer l'estimation de champs de déplacement et de déformation à la surface d'un matériau soumis à un effort de traction ou de compression, tout en quantifiant les performances métrologiques.

Abstract. This document presents a synthetic overview of works in the fields of image analysis and processing. It is made of three parts. The first part addresses a key problem of computer vision, namely interest point matching, whether between two images or between an image and an unstructured scene model. The second part deals with image noise. Two problems are addressed: characterization and removal of periodic noise giving a repetitive texture covering the whole image, and estimation of white noise affecting any digital image. The third part presents contributions to image analysis in experimental solid mechanics. The aim is to enhance the estimation of displacement and strain fields on the surface of a specimen under a tensile or compressive test, while quantifying the metrological performances.