



HAL
open science

Descripteurs locaux pour l'imagerie radar et applications

Flora Dellinger

► **To cite this version:**

Flora Dellinger. Descripteurs locaux pour l'imagerie radar et applications. Traitement des images [eess.IV]. Télécom ParisTech, 2014. Français. NNT : 2014ENST0037 . tel-01354286

HAL Id: tel-01354286

<https://pastel.hal.science/tel-01354286>

Submitted on 18 Aug 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.



EDITE - ED 130

T H È S E

pour obtenir le grade de docteur délivré par

TELECOM ParisTech

Spécialité « SIGNAL ET IMAGES »

présentée et soutenue publiquement par

Flora DELLINGER

le 1er juillet 2014

Descripteurs locaux pour l'imagerie radar et applications

Directeurs de thèse : **Florence Tupin et Yann Gousseau**

Co-encadrement de la thèse : **Julie Delon**

Jury

Jordi INGLADA
Sylvie LE HÉGARAT
Grégoire MERCIER
Jean-Yves TOURNERET
Florence TUPIN
Yann GOUSSEAU
Julie DELON
Julien MICHEL

Rapporteurs

Examineurs

Directeurs de thèse

Encadrant

Invité

TELECOM ParisTech

école de l'Institut Mines-Télécom - membre de ParisTech

46 rue Barrault 75013 Paris - (+33) 1 45 81 77 77 - www.telecom-paristech.fr

A mes grands-parents

Remerciements

Je souhaiterais remercier tout d'abord mes directeurs de thèse et encadrants, Florence Tupin, Yann Gousseau et Julie Delon, de m'avoir permis de faire cette thèse. Je vous remercie pour votre soutien, vos conseils et votre investissement. Ces années de thèse ont été très enrichissantes intellectuellement et je suis reconnaissante d'avoir pu bénéficier de vos connaissances et de votre rigueur scientifique.

Je remercie aussi tout particulièrement Julien Michel, mon responsable CNES, qui a toujours été très présent, même de Toulouse. Merci pour ton soutien, ton entrain et ta gentillesse.

Je remercie Jordi Inglada et Sylvie Le Hégarat d'avoir accepté d'être rapporteurs pour mon manuscrit, ainsi que Jean-Yves Tournet et Grégoire Mercier pour avoir accepté de faire parti de mon jury.

Je tiens à remercier aussi Jean-Marie Nicolas pour sa disponibilité et son enthousiasme sans faille à propos de l'imagerie radar.

Je souhaite remercier le CNES pour m'avoir financé pendant trois ans et l'équipe SI/AR du Centre Spatial de Toulouse pour m'avoir accueillie pendant neuf mois. La découverte de la ville rose n'aurait pas été si plaisante sans la sympathie et la gentillesse de toute l'équipe. Je tiens à remercier en particulier Myriam, Céline, Manuel et Julien, avec qui j'ai partagé de bons moments de rigolades et des discussions toujours intéressantes. Merci enfin à l'équipe des stagiaires avec qui j'ai pu découvrir les joies de la vie toulousaine.

Je remercie chaleureusement tous mes camarades doctorants de Télécom ParisTech pour leur bonne humeur et leur sympathie. J'ai apprécié particulièrement nos discussions passionnées sur toutes sortes de sujets, mais aussi nos bons moments de rigolades et nos friday beers sur la butte. Merci en particulier Sonia pour ton soutien sans faille, ton amitié et ta gentillesse. Merci aussi Hélène pour toutes ces joyeuses discussions et ta bienveillance. Et merci à tous mes camarades du bureau C32 et à la fine équipe du bureau C07.

Je tiens à remercier aussi mes amies qui ont été d'un grand soutien toutes ces années : Juliette, Anissa, Caroline, Diane, Louise et Dianna.

Enfin je remercie avec beaucoup de gratitude ma famille et surtout mes parents d'avoir été très présents et de m'avoir toujours soutenue durant toutes ces années d'étude.

Table des matières

Introduction	5
I Introduction	9
1 Contexte : imagerie satellitaire et nouvelles problématiques des images	
Haute Résolution en milieu urbain	11
1.1 Imagerie optique	11
1.1.1 Généralités	11
1.1.2 Principes d'acquisition	11
1.1.3 Résolution spatiale	13
1.1.4 Résolution spectrale	13
1.1.5 Radiométrie	14
1.1.6 Aspects géométriques	14
1.1.7 Statistiques des images	14
1.2 Imagerie radar	15
1.2.1 Généralités	15
1.2.2 Système radar	15
1.2.3 Formation des images	18
1.2.4 Résolution spatiale	19
1.2.5 Aspects géométriques	20
1.2.6 Statistiques des images radar	20
1.3 Images de télédétection Haute Résolution et nouvelles problématiques en milieu urbain	22
1.3.1 Images à Haute Résolution	22
1.3.2 Comparaisons optique/radar	23
1.3.3 Séries d'images multi-capteurs, multi-résolutions en milieu urbain	24
1.3.4 Conclusions	26
2 Présentation de l'algorithme SIFT et variantes	29
2.1 Détection de points-clés	30
2.1.1 Opérateurs différentiels appliqués à un espace-échelle	30
2.1.2 Détecteurs de points d'intérêt	41
2.2 Sélection des orientations	44
2.2.1 Maxima d'histogramme d'orientation	44
2.2.2 Sélection <i>a contrario</i> sur histogramme d'orientation	44
2.2.3 Sélection à l'aide de réponses d'ondelettes de Haar 2D	45
2.3 Extraction des descripteurs	45
2.3.1 SIFT	46
2.3.2 HOG	47
2.3.3 PCA-SIFT	48

2.3.4	GLOH	48
2.3.5	SIFT-circulaire	48
2.3.6	SURF	48
2.4	Mise en correspondance	49
2.4.1	Méthode NNDR	50
2.4.2	Méthode <i>a contrario</i>	50
2.5	Applications de l'algorithme SIFT en télédétection	52
2.6	Conclusions	53
3	Limites de l'algorithme SIFT sur les images bruitées	55
3.1	Détection de points-clés	55
3.1.1	Observations	55
3.1.2	Remarques théoriques	56
3.2	Calcul du gradient	60
3.2.1	Bruit gaussien	60
3.2.2	Bruit de speckle	61
3.3	Etat de l'art des adaptations de l'algorithme SIFT aux images radar	63
3.4	Conclusions	64
II	Un nouveau descripteur pour les données radar : SAR-SIFT	67
4	Calcul du gradient sur une image radar	69
4.1	Propositions dans la littérature	69
4.1.1	Différence des moyennes des logarithmes	70
4.1.2	Rapport des moyennes des intensités	70
4.1.3	Rapport de moyennes pondérées exponentiellement	73
4.1.4	Gradient radar basé sur le ROEWA	74
4.2	Proposition de nouveau gradient adapté aux images radar : Gradient par Ratio	75
4.2.1	Logarithme du ROA	75
4.2.2	Logarithme du ROEWA	76
4.2.3	Proposition du Gradient par Ratio	78
4.3	Conclusions	80
5	SAR-SIFT : un descripteur local pour les images radar	83
5.1	Présentation	83
5.1.1	Détection de points-clés	83
5.1.2	Sélection des orientations et extraction des descripteurs	91
5.2	Evaluations	91
5.2.1	Répétabilité des détecteurs de points-clés	93
5.2.2	Performances des mises en correspondance	94
5.2.3	Invariance à la rotation	96
5.3	Précision de l'appariement des points-clés	97
5.4	Comportement de SAR-SIFT avec des conditions d'acquisitions différentes .	100
5.5	Conclusions	104

III Applications des algorithmes SIFT et SAR-SIFT pour les images de télédétection	105
6 Estimation d'une déformation globale entre deux images radar	107
6.1 Présentation des algorithmes RANSAC et AC-RANSAC	107
6.2 Estimation de la déformation globale entre deux images	109
6.2.1 Etat de l'art sur le recalage d'images radar	109
6.2.2 Proposition d'une nouvelle méthode de recalage	111
6.3 Filtrage des faux appariements	114
6.4 Conclusions	118
7 Détection de changements par descripteurs locaux	119
7.1 Etat de l'art sur la détection de changements en télédétection	120
7.2 Détection de changements à l'aide d'algorithmes de type SIFT	121
7.2.1 Observations préliminaires	122
7.2.2 Détection de changements par densité de points-clés et approche a contrario	123
Conclusions et perspectives	133
IV Annexes	135
A Bases de données à disposition	137
A.1 Données radar	137
A.2 Données optiques	143
B Algorithme de type SIFT utilisé en imagerie optique	149
B.1 Détection de points-clés	149
B.2 Sélection des orientations	149
B.3 Extraction des descripteurs	150
C Calculs et démonstrations	151
C.1 Calcul du bruit résiduel d'une image bruitée après ajout de flou gaussien . .	151
C.2 Influence du bruit gaussien sur l'orientation du gradient d'un bord vertical ou horizontal	152
C.3 Estimateur MV du logarithme du contraste d'une rupture de réflectivité . .	153
C.4 Log-cumulants de la loi de Fisher	154
C.5 Invariance aux changements d'échelle de la matrice SAR-Harris	155
Publications	157
Bibliographie	159

Resumé

Dans cette thèse, nous étudions l'intérêt des descripteurs locaux pour les images satellites optiques et radar. Ces dernières années, la multiplication des satellites de télédétection haute résolution conduit à une forte augmentation des situations multi-capteurs, multi-résolutions et multi-angles d'incidence. Les descripteurs locaux, du fait de leurs invariances et de leur représentation compacte, présentent un intérêt certain pour la comparaison d'images satellitaires acquises dans des conditions différentes. Bien que facilement applicables aux images optiques, ils offrent des performances très limitées sur les images radar. Effectivement ces images sont très fortement bruitées par un bruit multiplicatif qui perturbe le traitement et l'interprétation des données.

Nous proposons ainsi un descripteur local original pour la comparaison d'images radar. Cet algorithme, appelé SAR-SIFT, repose sur la même structure que l'algorithme SIFT, à savoir une détection de points d'intérêts associée à l'extraction de descripteurs locaux. Afin d'adapter ces étapes aux statistiques des images radar, nous avons au préalable développé un nouvel opérateur différentiel, appelé Gradient par Ratio. Cet opérateur permet de calculer, de façon multi-échelles, une norme et une orientation du gradient robustes au bruit multiplicatif des images radar. Nous nous sommes ensuite appuyés sur ce gradient pour modifier les étapes de détection de points-clés et d'extraction de descripteurs. Les performances de l'algorithme SAR-SIFT proposé sont supérieures à celle de l'algorithme SIFT original pour des données radar. Nous avons ensuite testé le comportement de l'algorithme SAR-SIFT dans différentes situations d'acquisition.

Nous présentons par la suite deux applications pour la télédétection, basées sur les descripteurs locaux. En premier lieu, nous estimons une transformation globale entre deux images radar à l'aide de l'algorithme SAR-SIFT. L'estimation est réalisée à l'aide d'un algorithme de type RANSAC et en utilisant comme points homologues les points-clés mis en correspondance. Nous proposons d'utiliser cette méthode pour estimer deux types de transformation : une transformation affine entre deux images pour les zones plates, ou une translation résiduelle entre deux images géoréférencées. Enfin nous avons mené une étude prospective sur l'utilisation des descripteurs locaux (SIFT en imagerie optique et SAR-SIFT en imagerie radar) pour la détection de changements en télédétection. La méthode proposée consiste à comparer à l'aide d'une approche a contrario les densités de points-clés mis en correspondance aux densités de points-clés détectés pour mettre en évidence les zones de changement. L'intérêt de cette méthode est illustré sur des cas réels (paires d'images optiques et radar).

Abstract

In this thesis, we study the interest of local features for optical and SAR images. These past years, the multiplication of the number of high resolution remote sensing satellites has led to an increase of multi-sensors, multi-resolutions and multi-incidence angles situations. Local features, because of their invariances and their dense representation, offer a real interest for the comparison of satellite images acquired under different conditions. While it is easy to apply them to optical images, they offer limited performances on radar images. Indeed these images present a strong multiplicative noise that disturb the process and the reading of the datas.

We offer here an original local feature for the comparison of SAR images. This algorithm, called SAR-SIFT, relies on the same structure as the SIFT algorithm : the detection of local interest points associated with the extraction of local features. In order to adapt these steps to the statistics of SAR images, we have first developed a new differential operator, called Gradient by Ratio. This operator allows to compute a multi-scale magnitude and orientation of the gradient, which are robust to the multiplicative noise of the SAR images. Then we have used this gradient to modify the steps of the keypoints detection and the feature extraction. The performances of the algorithm SAR-SIFT are better than the ones of the original SIFT algorithm for the SAR images. We have then tested the behaviour of this new algorithm on situations with different acquisitions.

Consequently, we present two applications of for remote sensing, based on local features. First, we estimate a global transformation between two SAR images with the help of the SAR-SIFT algorithm. The estimation is realized thanks to a RANSAC-like algorithm and by using the matched keypoints as tie points. We propose to use this method to estimate two types of transformations : an affine transformation between two images on flat areas, or a residual translation between two georeferenced images. Finally, we have led a prospective study on the use of local features (SIFT for optical images and SAR-SIFT for SAR images) for change detection in remote sensing. The proposed method consists in comparing, with help of an a contrario approach, the densities of matched keypoints to the densities of detected keypoints, in order to point out changed areas. The interest of this method is illustrated on concrete cases (pairs of optical and radar images).

Introduction

Contexte

La dernière génération de satellites de télédétection offre une amélioration de la résolution spatiale des images, qui est fréquemment métrique ou submétrique. Les images obtenues permettent d'envisager de nouvelles applications de télédétection en Très Haute Résolution (THR). Citons par exemple la détection de changements après catastrophe, le suivi de la déforestation ou la reconnaissance de navires en mer. Toutefois, le niveau de détail entraîne une plus grande complexité des images, qui deviennent ainsi plus difficiles à traiter. Cette amélioration de la résolution s'est également accompagnée d'une augmentation du nombre de satellites de télédétection, conduisant à une hausse importante du nombre d'images Haute Résolution disponibles. La multiplication des situations multi-capteurs soulève de surcroît de nouvelles problématiques, comme la gestion des effets de parallaxes, des changements de résolution ou d'éclairement. Il devient ainsi essentiel de pouvoir comparer efficacement un grand nombre d'images, en gérant différentes modalités d'acquisitions d'images. Dans un cas général, la comparaison d'images pixel à pixel est limitée et repose sur des corrections géométriques et radiométriques parfois inaccessibles sans une connaissance précise de la scène. L'utilisation de descripteurs locaux pouvant intégrer des invariances est alors à privilégier.

Les descripteurs locaux représentent de façon locale les caractéristiques d'une image (géométrie, couleur, texture...). Leur extraction est souvent réalisée en des points intéressants de l'image (coin, centre d'un objet). Leur utilisation permet de s'affranchir des contraintes de recalage et de calibration. En outre, ils offrent une représentation compacte des données, utile pour la comparaison d'un grand nombre d'images. De nombreux descripteurs ont été développés pour les images naturelles (ou photographies) [Low04, KS04, BTG06] et sont donc relativement aisément applicables aux images optiques. Ils trouvent leur utilisation dans différentes applications comme la classification [YN13], la détection d'objets [SU09] ou le recalage d'images [JGKM14].

La situation est plus compliquée pour les images radar. Celles-ci sont fortement bruitées par un bruit multiplicatif, dit de speckle, qui perturbe le traitement et l'interprétation des données. Les performances des descripteurs locaux traditionnels sont très limitées sur ce type d'images. A l'instar de leur utilisation en imagerie optique, ils présentent pourtant un potentiel important pour de nombreuses applications. Les capteurs radar offrent effectivement un intérêt notable : leur acquisition est indépendante des conditions d'éclairage et de la couverture nuageuse. Ils sont par exemple d'une grande utilité dans l'étude des détections de changements après catastrophe. Dans un contexte où il est nécessaire d'obtenir des données le plus rapidement possible, les capacités d'acquisition des capteurs radar sont alors particulièrement profitables. Le choix des images avant et après catastrophe étant généralement restreint, on est souvent amené à effectuer des comparaisons d'images multi-capteurs, multi-résolutions et/ou multi-incidence. L'utilisation de descripteurs locaux est alors à privilégier. Plusieurs satellites radar fournissent d'ailleurs depuis quelques années des images à résolution métrique, autorisant une étude plus fine de ces change-

ments. Mentionnons les satellites jumeaux TerraSAR-X et TanDEM-X, ou la constellation Cosmo-SkyMed. Les descripteurs locaux peuvent être aussi d'un grand intérêt dans l'étude des très grandes séries multi-temporelles. Les deux satellites Sentinelle 1 ont pour objectif de fournir en accès libre des images radar à 5m de résolution sur l'ensemble de la Terre, pour des applications en environnement ou en sécurité (en particulier la gestion des crises). La quantité de données disponible sera particulièrement importante. Les descripteurs locaux peuvent trouver une utilité dans le stockage d'information et le traitement de données massives.

Au travers des exemples présentés et des applications mentionnées en télédétection optique, les descripteurs locaux sont d'un intérêt indéniable pour des applications en imagerie radar. Ils offrent un outil générique pour la comparaison d'images Haute Résolution. Toutefois, les adaptations de descripteurs locaux proposées dans la littérature pour l'imagerie radar ne produisent que de faibles améliorations [SSUR10, WYF12]. Il semble ainsi pertinent de développer une approche plus performante dédiée à ce type de données.

Contributions

Mes travaux de thèse se sont majoritairement concentrés sur le développement d'un descripteur local original pour les images radar. Ce descripteur est inspiré du descripteur SIFT [Low04], très populaire et efficace pour les images naturelles [MS05]. Comme SIFT, l'algorithme proposé, appelé SAR-SIFT, effectue une détection de points-clés, suivie d'une extraction de descripteurs locaux, codant la géométrie locale de l'image. Ces étapes ont été adaptées aux spécificités des images radar. Pour cela, une nouvelle notion de gradient a été développée au préalable. Le gradient par différence présente en effet des limites quand il est appliqué aux images radar, en raison du caractère multiplicatif du bruit. L'algorithme SAR-SIFT, basé sur ce nouveau gradient, est ainsi robuste au bruit de speckle. De la même façon que SIFT, il est invariant à la rotation et au changement d'échelle, et relativement robuste aux petites transformations affines. Il est utile pour comparer des images acquises avec des capteurs différents, des résolutions différentes et/ou des angles d'incidence différents. La contribution principale de ce travail de doctorat est le développement et la validation de ce descripteur local SAR-SIFT, dédié aux images radar et reposant sur l'adaptation de différentes étapes.

Ce descripteur a ensuite été exploité pour le développement d'applications en télédétection. Une application de recalage pour les images radar acquises dans des conditions différentes est présentée. Les changements d'angle d'incidence, entraînant des variations sur la géométrie des objets, compliquent particulièrement la comparaison des images. L'utilisation de descripteurs locaux permet de gérer ces effets et montre son efficacité dans l'extraction de points homologues pour l'estimation d'une transformation globale.

Enfin une méthode de détection de changements, basée sur les algorithmes SIFT et SAR-SIFT, a été développée pour comparer deux images optiques ou deux images radar, acquises dans différentes conditions. A l'instar de nombreuses autres méthodes de détection de changements, cette approche ne nécessite pas de pré-traitement ou de recalage. Ces deux méthodes représentant deux contributions applicatives de cette thèse.

Organisation du manuscrit

Le manuscrit est organisé en trois parties, suivies d'annexes.

La première partie décrit de façon précise le contexte d'étude et cerne les problématiques.

-
- Le Chapitre 1 décrit les caractéristiques des images de télédétection (optique et radar). Une revue des différentes problématiques posées par les images Haute Résolution en milieu urbain est proposée.
 - Le Chapitre 2 présente un état de l’art commenté sur les descripteurs locaux de type SIFT. Les différentes étapes de l’algorithme SIFT et ses variantes sont présentées en détail.
 - Le Chapitre 3 décrit les limites de l’algorithme SIFT sur les images bruitées, et plus particulièrement sur les images radar. Des études expérimentales et théoriques sont réalisées sur le comportement de l’étape de détection de points d’intérêt, ainsi que sur le calcul du gradient par différence sur les images bruitées. Ce chapitre permet d’identifier les étapes de l’algorithme SIFT qui doivent être adaptées à l’imagerie radar.

La deuxième partie présente les contributions principales de la thèse :

- Dans le Chapitre 4 nous proposons un nouveau calcul de gradient adapté aux images radar. Ce calcul se base sur des détecteurs de contour, tels que le ROA [Bov88, TLB88], développés pour les données radar.
- Nous présentons dans le Chapitre 5 un algorithme de type SIFT pour les images radar, appelé SAR-SIFT. Nous détaillons les étapes de l’algorithme puis nous étudions leurs performances respectives. Nous analysons enfin le comportement de l’algorithme pour des paires d’images présentant des conditions d’acquisition différentes.

La troisième partie propose deux modalités d’application des descripteurs locaux en télédétection :

- Dans le Chapitre 6, nous proposons d’utiliser l’algorithme SAR-SIFT pour estimer la transformation globale entre deux images radar. Nous utilisons les points-clés appariés comme points homologues afin d’estimer le modèle de déformation. Cette méthode trouve son utilité pour le recalage d’images acquises avec des incidences différentes.
- Dans le Chapitre 7 est présentée une étude prospective sur la détection de changements entre deux paires d’images optiques ou radar. La densité des points-clés détectés est comparée à la densité des points appariés à l’aide d’une approche a contrario. Nous proposons une première étude pour le choix des paramètres ainsi que plusieurs exemples d’applications.

Les annexes apportent des informations supplémentaires sur les travaux présentés :

- L’Annexe A présente les images satellites mises à disposition pendant cette thèse.
 - L’Annexe B décrit la variante de l’algorithme SIFT utilisée en imagerie optique et pour les comparaisons avec SAR-SIFT. Des détails sont données pour l’implémentation et les valeurs de paramètres choisis.
 - Les détails de certains calculs et démonstrations se trouvent dans l’Annexe C.
-

Première partie

Introduction

Chapitre 1

Contexte : imagerie satellitaire et nouvelles problématiques des images Haute Résolution en milieu urbain

Nous présentons dans ce chapitre les fondamentaux des images de télédétection optiques et radar. Nous détaillons ensuite les nouvelles problématiques posées par les images de nouvelle génération à haute résolution, et plus particulièrement en milieu urbain.

1.1 Imagerie optique

Cette section s'intéresse aux images satellites optiques. Elle s'appuie sur les références suivantes : [Gao08, Spo10, TNI14].

1.1.1 Généralités

Un capteur optique est un capteur dit passif : il mesure la lumière provenant du soleil qui est réfléchi sur la surface de la Terre. Ce signal lumineux est ensuite converti en signal électrique, puis numérisé. Sur un satellite de télédétection optique, plusieurs capteurs élémentaires sont assemblés sous différentes configurations afin d'obtenir une image.

1.1.2 Principes d'acquisition

1.1.2.1 Capteur élémentaire

Les capteurs optiques captent les photons provenant d'une direction, la ligne de visée, pendant un temps d'intégration T_{int} . À cause des lois de la diffraction, le système optique (lentilles et miroir) définit un cône centré autour d'un rayon perspectif (la ligne de visée) avec une ouverture angulaire $\delta\omega$. Cette ouverture est déterminée par l'ouverture du système optique (dimension du miroir ou de la lentille) et la longueur d'onde λ étudiée. Dans le cas de systèmes optiques circulaires, l'ouverture angulaire correspond au lobe principal d'une ouverture circulaire dont la réponse impulsionnelle, ou PSF (*Point Spread Function*), est définie par une fonction de Bessel. Pour un rayonnement monochromatique, on peut définir l'ouverture angulaire avec l'expression $\delta\omega = 1,22 \frac{\lambda}{L}$.

L'intersection de ce cône avec la surface de la Terre définit l'empreinte sol (ES) (Figure 1.1). Les ondes lumineuses réfléchies sur tous les objets contenus dans cette empreinte sol sont alors mesurées par le capteur élémentaire pendant le temps T_{int} . On obtient ainsi une valeur pour un pixel.

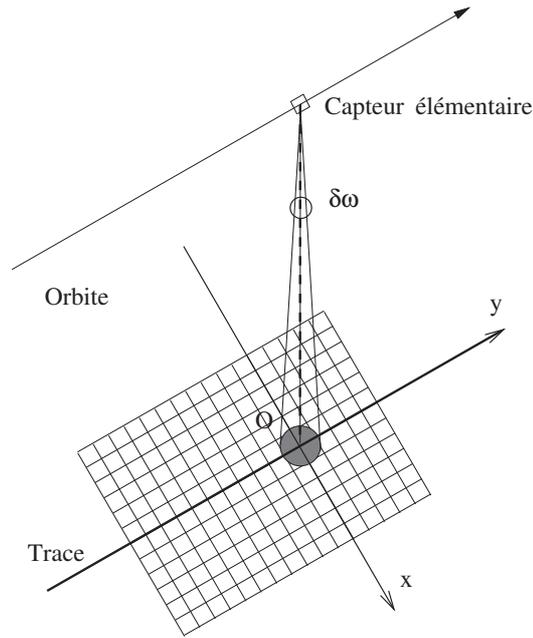


FIGURE 1.1 – Capteur élémentaire optique avec une ligne de visée verticale et une ouverture angulaire $\delta\omega$. La surface de la Terre est considérée comme plane. L’empreinte sol est la zone grise. Source : [TNI14].

On définit la résolution du capteur élémentaire, soit la distance minimale permettant de distinguer deux objets adjacents, par l’empreinte sol. Cette empreinte est déterminée par l’ouverture angulaire $\delta\omega$, mais aussi par la position du satellite (altitude et ligne de visée) et la surface de la Terre.

- Si l’on considère que la Terre présente une surface plane et que la ligne de visée est verticale, l’empreinte au sol est égale à $R\delta\omega$, avec R la distance entre le satellite et l’objet observé.
- Si la surface de la Terre présente une pente locale α (Figure 1.2), l’empreinte au sol est $\frac{R\delta\omega}{\cos \alpha}$.
- Dans le cas où la surface de la Terre est considérée comme plane mais que le capteur présente une visée oblique d’angle θ (Figure 1.3), l’empreinte au sol est $\frac{R\delta\omega}{\cos^2 \theta}$.

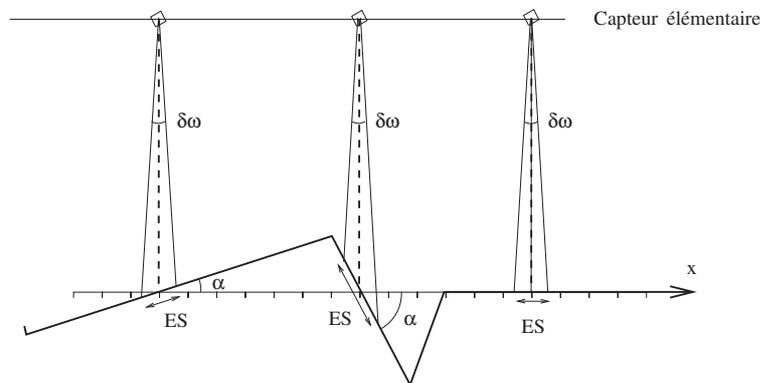


FIGURE 1.2 – Capteurs élémentaires optiques avec une ligne de visée verticale et différentes pentes locales. Source : [TNI14].

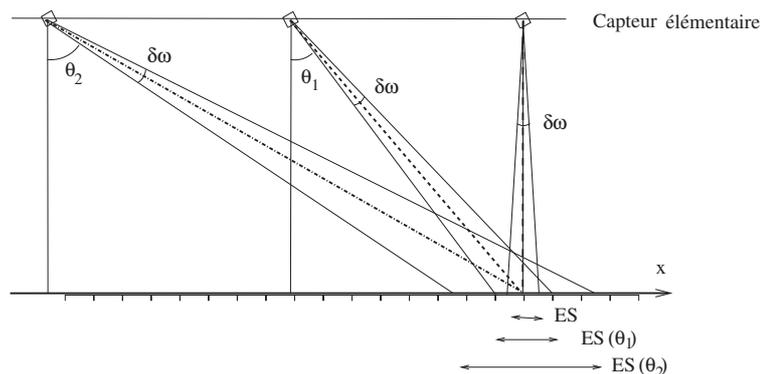


FIGURE 1.3 – Capteurs élémentaires optiques avec différentes lignes de visée obliques d’angle θ . La surface de la Terre est considérée comme plane. Source : [TNI14].

1.1.2.2 Système d’acquisition des images

Les images sont acquises selon deux directions perpendiculaires : la fauchée et la trace, cette dernière étant parallèle à la trajectoire du satellite. L’acquisition le long de la fauchée se fait à partir de plusieurs capteurs positionnés sur le satellite. Les capteurs peuvent être assemblés en ligne (*pushbroom*) ou en une matrice rectangulaire (*whiskbroom*). Dans le dernier cas, la fauchée est obtenue grâce à la rotation d’un miroir et plusieurs lignes sont acquises en même temps. L’acquisition selon la trace se fait grâce au mouvement du satellite. Pour le système *pushbroom*, elle se fait ainsi ligne par ligne. Cependant l’intervalle de temps Δt entre deux acquisitions doit être supérieur au temps d’intégration T_{int} des capteurs élémentaires.

L’acquisition par le mouvement permet aussi d’améliorer la sensibilité du système. La technique TDI (*time delay and integration*) consiste à assembler plusieurs lignes de capteurs de type *pushbroom* et à accumuler ainsi plusieurs expositions d’un même objet grâce au mouvement du satellite. Cela permet d’augmenter le temps d’intégration de chaque objet et pareillement le rapport signal sur bruit. Cette technologie est aujourd’hui la plus couramment utilisée.

1.1.3 Résolution spatiale

La résolution spatiale est déterminée par la construction du système.

- L’échantillonnage de la fauchée $\Delta x = \frac{fR}{d}$ dépend de la focale f du capteur, la distance R entre le sol et le satellite et la distance d entre deux capteurs.
- L’échantillonnage selon la trace $\Delta y = V_{sol}\Delta t$ dépend de la vitesse V_{sol} du satellite projetée au sol et du temps Δt entre deux acquisitions.

Les deux échantillonnages doivent cependant vérifier le critère de Shannon qui impose :

$$\Delta x \leq R\delta\omega \text{ et } \Delta y \leq R\delta\omega \quad (1.1)$$

En effet, $R\delta\omega$ correspond à l’empreinte sol du capteur sur une surface plane.

1.1.4 Résolution spectrale

Les résolutions sont différentes en fonction du type d’acquisition utilisé :

- En mode panchromatique, une large gamme de fréquences, correspondant souvent aux fréquences du visible, est observée. Les images obtenues sont en noir et blanc et sont très bien résolues spatialement.

- En mode multispectral, différentes bandes spectrales plus étroites sont observées indépendamment. Ces bandes correspondent souvent aux trois couleurs naturelles, mais il peut y avoir d'autres choix ainsi que l'association d'une bande dans le proche infrarouge, correspondant à la mesure de la végétation. En combinant ces bandes on peut ainsi obtenir des images en couleur, cependant moins bien résolues spatialement que les images panchromatiques. En effet, l'utilisation de filtres couleurs limite le nombre de photons reçu sur chaque capteur. Pour éviter un rapport signal sur bruit trop faible, il est nécessaire d'augmenter l'ouverture angulaire des capteurs et donc de diminuer la résolution spatiale.

1.1.5 Radiométrie

Le capteur élémentaire mesure l'énergie lumineuse reçue pendant le temps d'observation T_{int} , puis la quantifie en un signal électrique. Les niveaux de quantification possibles déterminent la dynamique de l'image et sont différents d'un capteur à l'autre. On définit la résolution radiométrique comme la possibilité de distinguer des petites variations d'intensité. Les capteurs actuels permettent d'obtenir des données codées entre 8 et 16 bits.

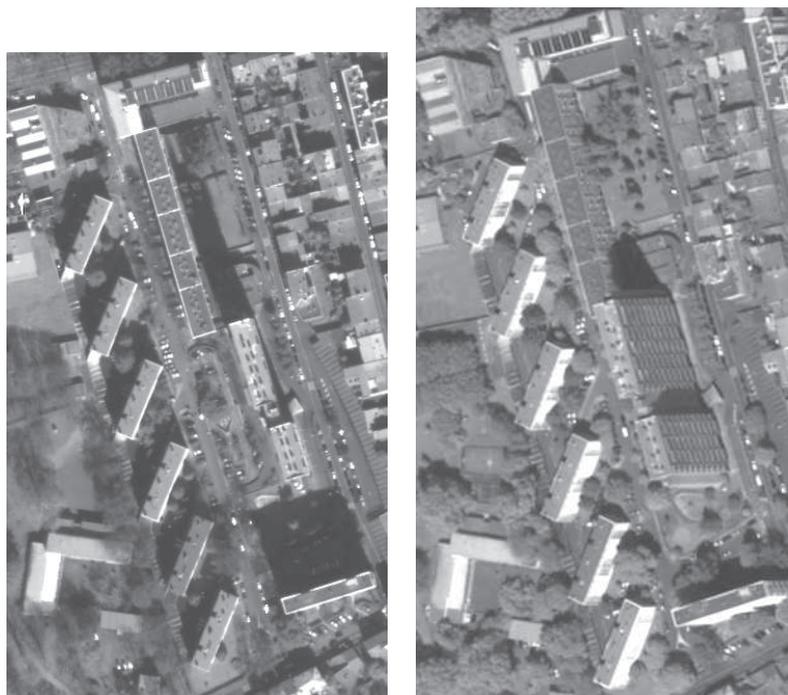
1.1.6 Aspects géométriques

L'empreinte sol des capteurs dépend de la pente locale et de l'angle de la ligne de visée (Section 1.1.2.1). Sur les zones à fort relief, la résolution n'est donc pas constante sur toute l'image. Afin de comparer différentes images, il est souvent nécessaire de les géoréférencer, c'est à dire de faire correspondre à chaque pixel de l'image une coordonnée géoréférencée (longitude, latitude altitude). Pour cela il est nécessaire de connaître l'élévation des terrains et d'avoir un modèle de transformation entre les positions des pixels de l'image (lignes et colonnes) et les coordonnées géographiques (longitude et latitude). Ces modèles sont souvent approchés par une transformation de type fonction rationnelle (RPM ou *Rational Polynomial Model*) décrite par ses coefficients RPC (*Rational Polynomial Coefficients*). Ces RPC peuvent être fournis directement avec les images optiques correspondantes ou déterminés à partir de points homologues extraits dans deux images (dont une orthorectifiée).

De la même façon, la représentation des bâtiments dépend de l'angle de la ligne de visée. En visée verticale (appelée prise de vue au nadir), les façades verticale des bâtiments ne sont pas présentes sur l'image et il n'y a pas de phénomène de parallaxe (Figure 1.4(a)). Il est alors difficile d'avoir des informations sur la hauteur des bâtiments. En visée oblique, les façades face à la trajectoire du satellite sont observables. Cependant certaines parties de l'image seront obstruées, et cela d'autant plus que l'angle de la ligne de visée par rapport au nadir et la hauteur des bâtiments sont importants (Figure 1.4(b)).

1.1.7 Statistiques des images

Les images obtenues par les satellites optiques présentent un faible bruit. C'est une combinaison d'un bruit électronique et d'un bruit de compression, que l'on peut modéliser par un bruit blanc gaussien additif. Des filtres standards, comme les filtres moyenneurs, médians ou gaussiens, peuvent être utilisés. Cependant le rapport signal à bruit est très bon et il est rarement nécessaire de filtrer ces images.



(a) Image Worldview, 50 cm de résolution et 2° par rapport au nadir. (b) Image Pleiades, 70 cm de résolution et 23° par rapport au nadir.

FIGURE 1.4 – Exemples d’images satellite optiques de la même zone avec des lignes de visée différentes.

1.2 Imagerie radar

Cette section présente les fondamentaux de l’imagerie radar à synthèse d’ouverture (RSO). Ces informations proviennent des références suivantes : [NTD10, Spo10, TNI14].

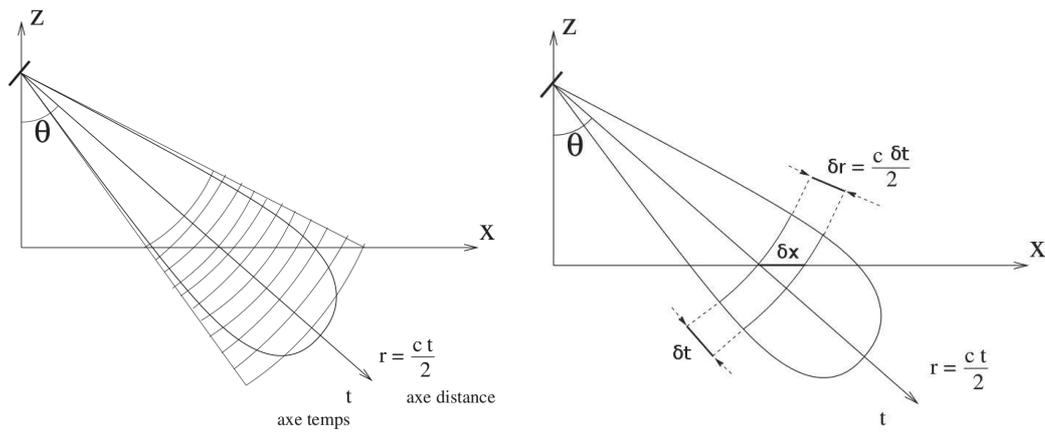
1.2.1 Généralités

Les systèmes radar sont basés sur le principe d’écholocation. Un capteur radar (RATio Detection And Ranging) émet une onde électromagnétique à l’aide d’une antenne puis mesure par la même antenne la rétrodiffusion de l’onde sur des objets à la surface de la Terre. Le capteur est dit actif parce qu’il mesure la rétrodiffusion d’un signal qu’il a lui-même émis. Le principal avantage de ce système est qu’il est indépendant des conditions temporelles et météorologiques. En effet un capteur radar peut obtenir des données de jour comme de nuit et à travers les nuages, à l’inverse des capteurs optiques. Cependant les ondes électromagnétiques utilisées sont cohérentes. Le signal mesuré est alors complexe, composé d’une phase et d’une amplitude, et ce signal d’amplitude est fortement bruité par un bruit que l’on peut modéliser comme multiplicatif.

1.2.2 Système radar

Principes du radar Comme la vitesse de propagation c des ondes électromagnétiques peut être supposée constante dans l’atmosphère, la mesure du temps de parcours t d’une onde permet d’obtenir la distance entre le satellite et l’objet rétrodiffusé :

$$r = \frac{ct}{2} \tag{1.2}$$



(a) Echantillonnage temporel du signal. (b) Correspondance entre échantillonnage temporel et spatial.

FIGURE 1.5 – Principe du Radar (observation dans un plan radial) : une antenne à une distance R émet un signal avec une ligne de visée oblique d’angle θ . Le lobe principal de l’antenne est présenté ici. Le signal présente un échantillonnage temporel δt correspondant à un échantillonnage spatial $\delta r = \frac{c \delta t}{2}$ et un pixel au sol $\delta x = \frac{\delta r}{\sin \theta}$. Source : [NTD10].

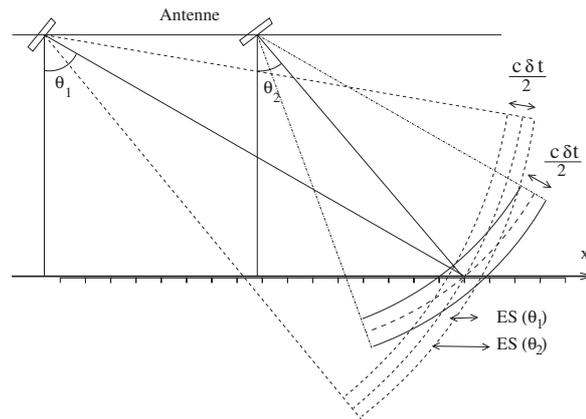


FIGURE 1.6 – Antennes radar avec différentes lignes de visée obliques d’angle θ_1 et θ_2 et visant le même point. La surface de la Terre est considérée comme plane. L’empreinte sol suit une loi en $\frac{1}{\sin \theta}$. Source : [TNI14].

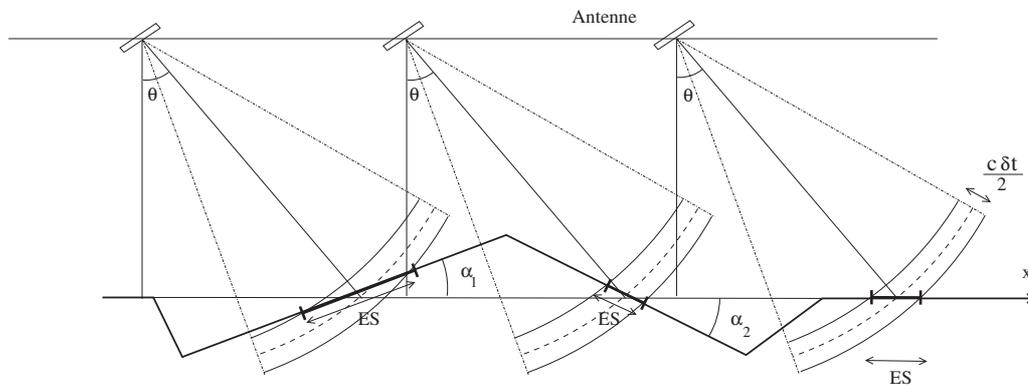


FIGURE 1.7 – Antennes radar avec une ligne de visée oblique de même angle θ et différentes pentes locales. Source : [TNI14].

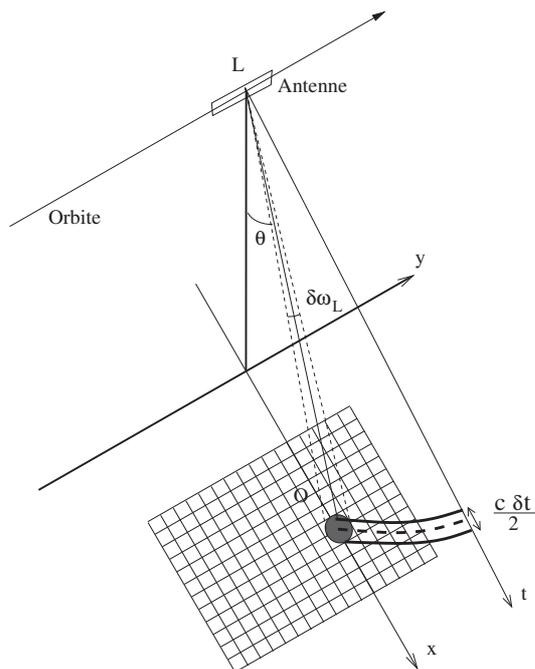


FIGURE 1.8 – Antenne radar avec une ligne de visée selon un angle θ et une extension azimutale $\delta\omega_L$. La surface de la Terre est considérée comme plane. La zone grise est l’empreinte sol. Source : [TNI14].

Le signal reçu par l’antenne est alors échantillonné temporellement avec une période δt , appelée case temps. Cet échantillonnage entraîne un échantillonnage spatial $\delta r = \frac{c\delta t}{2}$, appelé case distance (Figure 1.5). Toutes les contributions de la même case distance sont additionnées de façon cohérente (en complexe) dans un même échantillon temporel. Les échos de tous les objets à une même distance r du satellite seront confondus. L’intersection entre le lobe principal de l’antenne et la case distance correspond alors à l’empreinte sol δx .

La ligne de visée, appelée axe distance, est presque toujours oblique, avec un angle θ . Quand la visée est verticale (au nadir), les éléments de part et d’autre de la verticale sont confondus dans la même case distance, entraînant ainsi un repliement de l’image. On utilise ce type de visée avec un lobe très fin pour mesurer la distance entre le satellite et le sol. Cela correspond à un radar altimètre.

Si l’on considère la surface de la Terre plane, l’empreinte sol en distance δx est égale à $\frac{\delta r}{\sin\theta} = \frac{c\delta t}{2\sin\theta}$. Elle dépend de la résolution en distance δr et de l’angle de visée θ (Figure 1.6). En présence de pentes locales α (Figure 1.7), l’empreinte sol est égale à $\frac{c\delta t}{2\sin(\theta-\alpha)}$. On remarque que plus la pente locale s’oriente en direction du capteur, plus l’empreinte est grande.

Antennes radar Une antenne radar est de forme rectangulaire, de dimension $L \times l$. La ligne de visée est perpendiculaire à l’antenne et définit l’axe distance (slant range) spécifié précédemment (Figure 1.8). Le côté de dimension L est parallèle à la trajectoire du satellite et définit l’axe azimut. La troisième direction, perpendiculaire à l’axe azimut, est l’axe distance projeté au sol (ground range). L’ouverture angulaire de l’antenne est définie par le lobe principal. Les extensions angulaires selon respectivement les directions distance et azimutale sont :

$$\delta\omega_l = \frac{\lambda}{l} \text{ et } \delta\omega_L = \frac{\lambda}{L}, \quad (1.3)$$

avec λ la longueur d'onde de l'onde électromagnétique. Ces grandeurs définissent la taille du lobe d'antenne $R\delta\omega_l$ en distance et $R\delta\omega_L$ en azimut, avec R la distance au satellite. En distance, on effectue un échantillonnage temporel dans ce lobe. Afin d'améliorer la résolution, on procède à une modulation en fréquence du signal suivi d'un filtrage adapté. Les retours de plusieurs cibles proches peuvent alors être séparés. En azimut la résolution du lobe $\delta a = \frac{\lambda R}{L}$ est trop grossière et on effectue la synthèse d'ouverture. Puisqu'un objet est illuminé plusieurs fois par le satellite radar en mouvement, cette synthèse consiste à effectuer un traitement des données par transformée de Fourier et à simuler ainsi une antenne avec une très grande résolution.

Fréquences des ondes Différentes bandes fréquentielles peuvent être utilisées en radar, de 0.225 GHz à 94 GHz (soit des longueurs d'onde entre 133 et 0.3 cm). Les bandes intermédiaires entre 1 et 12 GHz (L, S, C et X) sont les plus utilisées. Le choix des bandes se fait en fonction de l'application du satellite. La bande X par exemple, autour de 8 GHz, permet d'obtenir des images à haute résolution et est très utilisée pour la détection d'objets et la cartographie (TerraSAR-X, CosmoSky-Med).

1.2.3 Formation des images

Acquisition des données A la différence des satellites optiques, les satellites radar ne présentent qu'un seul capteur. La fauchée, soit la largeur de l'image, est construite en analysant le signal sur un intervalle temporel T_F très grand devant δt . En effet, en échantillonnant ce signal temporellement en case temps, on effectue aussi un échantillonnage spatial en case distance. En supposant la surface de la Terre plane, la fauchée est égale à $L_F = \frac{cT_F}{2 \sin \theta}$. On peut ainsi acquérir simultanément une ligne de l'image. Comme pour les satellites optiques, les images sont ensuite construites lignes par lignes, en utilisant le mouvement du satellite et en effectuant des tirs radar à la fréquence FRI (Fréquence de Répétition des Impulsions) ou PRF (Pulse Repetition Frequency). Afin d'éviter de mélanger les signaux provenant de deux tirs différents, il est nécessaire que la période T_{FRI} , l'inverse de la FRI, soit supérieure à T_F . Ce mode d'acquisition est appelé *Stripmap*.

Il existe cependant d'autres modes d'acquisition sur les satellites récents. Le mode *SpotLight* par exemple maximise le temps d'illumination des objets observés en faisant varier en azimut la direction de visée de l'antenne. La zone couverte est réduite mais la résolution spatiale est améliorée.

L'acquisition peut se faire avec une visée à gauche ou à droite (selon le satellite), et en orbite descendante ou ascendante selon la trajectoire du satellite. Un bâtiment peut être observé de deux côtés différents. Dans ces situations, les deux acquisitions sont très différentes (Figure 1.9) : la géométrie et la texture des bâtiments peuvent fortement changer. Toutes les images présentées dans la suite du manuscrit sont acquises avec une visée à droite.

Typologie des données Les mesures obtenues par le capteur sont appelées données *Raw*. Différents traitements leur sont appliqués pour les représenter sous forme d'images. Les données sont alors représentées en *géométrie radar* (slant geometry), dans un repère composé des axes azimut et distance (slant range). Ce type de données est appelé *Single Look Complex* (SLC). Enfin des corrections géométriques et radiométriques peuvent être appliquées pour projeter les données en *géométrie terrain* (ground geometry), dans un repère composé des axes azimut et distance au sol (ground range).

Radiométrie Les données obtenues sont des données complexes (réponse du champ électromagnétique) composées de la partie réelle et imaginaire du signal. Elles sont co-



(a) Image Cosmo-SkyMed acquise en phase descendante. Le satellite est à droite de l'image. (b) Image TerraSAR-X acquise en phase ascendante. Le satellite est à gauche de l'image.

FIGURE 1.9 – Exemple de deux images radar acquises sous des orbites différentes. Les deux images ont la même résolution (1 m) et sont acquises avec une même visée à droite sous un même angle d'incidence (34°). Elles ne proviennent cependant pas du même capteur.

dées sur 32 bits, 16 bits pour chaque partie. Ces images présentent donc une dynamique importante caractéristique des images radar. A partir des parties réelle et imaginaire, on peut extraire l'amplitude et la phase du signal. L'amplitude donne des informations sur la rétrodiffusion des objets de la scène imagée. La valeur des pixels dépend de la réflectivité des objets rétrodiffusés : la réflectivité est la capacité d'un objet à réfléchir l'onde électromagnétique. Elle dépend de différents paramètres comme le type de matériau, la texture, la forme de la cible ou l'humidité.

1.2.4 Résolution spatiale

Comme présenté précédemment, la résolution en distance est $\delta r = \frac{c\delta t}{2}$. Cela correspond à la plus petite distance entre deux cibles selon l'axe des distances pouvant être séparées par le capteur radar. La résolution azimutale dépend de l'extension de l'antenne et de la distance au satellite : $\delta a = \frac{\lambda R}{L}$.

Cependant, des améliorations ont été apportées aux derniers satellites afin d'augmenter la résolution. Pour la résolution en distance, l'onde électromagnétique est compressée par une modulation linéaire en fréquence. La modulation est réalisée par un chirp défini par sa largeur de bande B_{utile} . Le signal reçu est ensuite démodulé. Après compression, la résolution obtenue en distance est $\delta r^* = \frac{c}{2B_{utile}}$. Pour la résolution en azimut, il faudrait de très grandes antennes pour obtenir une résolution fine. Pour pallier cela, on peut synthétiser de façon artificielle une antenne plus grande que celle utilisée avec une synthèse d'ouverture. En effet le lobe d'antenne étant large, une même cible est éclairée plusieurs fois par le satellite le long de sa trajectoire. Le signal est alors modulé en fréquence grâce au mouvement du capteur (effet Doppler). On obtient alors comme résolution en azimut $\delta a^* = \frac{L}{2}$. Ce type d'image est appelé RSO (Radar à Synthèse d'Ouverture) ou SAR (Synthetic Aperture Radar).

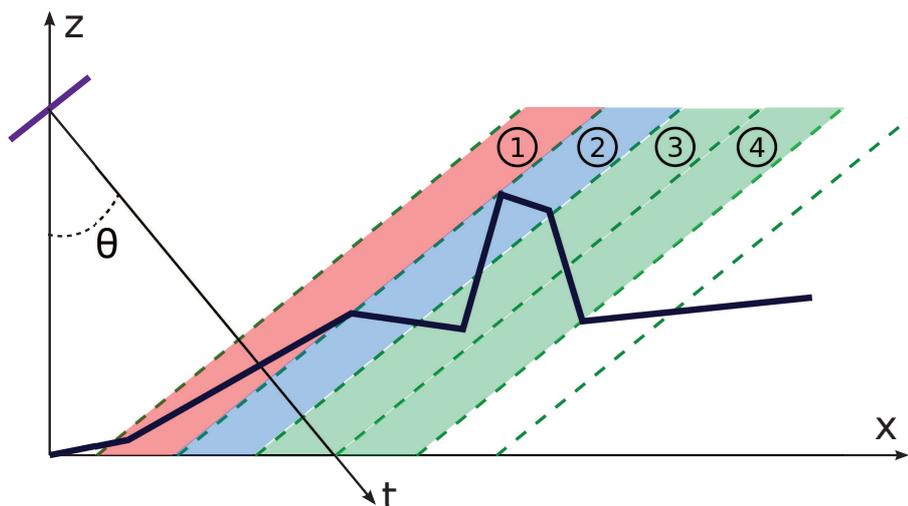


FIGURE 1.10 – Exemple schématique de distorsions géométriques observées sur les images Radar. La courbe noire présente un exemple de relief. La ligne de visée du satellite est oblique d'angle θ . Les lignes en pointillés verts définissent différentes cases temps ou cases distance. Trois distorsions sont observées : contraction (1, en rouge), recouvrement (2, en bleu) et ombrage (3 et 4, en vert). Notons que dans la case 3 il y a aussi recouvrement.

1.2.5 Aspects géométriques

Le système d'écholocation et la visée latérale entraîne des distorsions géométriques caractéristiques des images radar le long de l'axe des distances. Ces déformations compliquent l'interprétation et l'exploitation des images. On dénote trois types de distorsions (Figure 1.10) :

- la *contraction* (forshortening) : les pentes orientées face au capteur sont réduites et celles dans l'autre sens sont étirées.
- le *recouvrement* (layover) : les signaux provenant de deux objets différents mais à la même distance du satellite (donc dans la même case temps) sont mélangés.
- l'*ombrage* (shadow) : un objet caché par un autre objet peut ne pas être éclairé par le satellite, et donc ne pas apparaître sur l'image.

Ces distorsions sont influencées par la hauteur des objets, l'angle de la ligne de visée, les pentes locales et la position des objets.

La Figure 1.11 présente un extrait d'images radar avec des exemples de recouvrement et d'ombrage. On remarque que la tour à gauche recouvre même une route et que la tour en haut à droite recouvre une partie de l'ombre de celle à gauche. Notons que ces distorsions et occlusions sont aussi importantes sur une image optique avec un grand d'angle d'incidence. Les empreintes sol sont comparables pour une incidence θ pour le radar et $\pi/2 - \theta$ pour l'optique. Cela est illustré avec la Figure 1.16 présentant deux autres exemples d'images optiques de la même zone : une image avec de fortes ombres et une image avec un angle d'incidence de 23° .

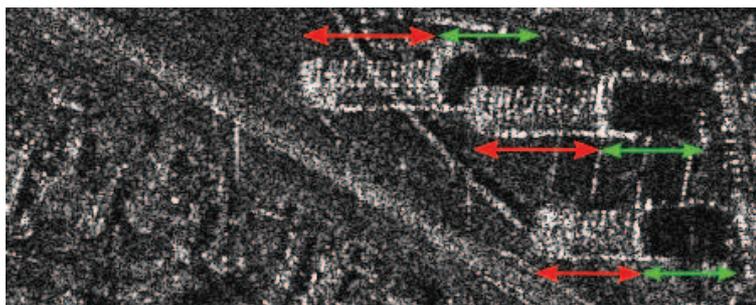
De la même façon que les images optiques (Section 1.1.6), il est possible de géoréférencer des images radar à l'aide d'un modèle numérique de terrain (MNT) et d'un RPM, qui peut être fourni avec les images.

1.2.6 Statistiques des images radar

Les images radar présentent la caractéristique d'être fortement bruitées par un bruit multiplicatif appelé *speckle* ou *chatoiement*. Cela est dû à la nature cohérente des ondes



(a) Image optique Quickbird à 60 cm de résolution



(b) Image radar TerraSAR-X de résolution 1 m et d'angle d'incidence de 34°.

FIGURE 1.11 – Exemple réel de distorsions géométriques sur une image radar. Les deux images représentent la même zone. Sur l'image Radar, le satellite est à gauche et les trois tours présentent un recouvrement (en rouge) suivi d'un ombrage (en vert).

électromagnétiques rétrodiffusées. Les images peuvent être observées en amplitude ou en intensité. Les valeurs des pixels en intensité représentent la réflectivité de la zone observée.

Modèles statistiques Dans le cas d'un speckle pleinement développé et sur des zones homogènes, la distribution de probabilité en intensité d'une image radar peut être décrite par une loi exponentielle :

$$p(I|R) = \frac{1}{R} \exp\left(-\frac{I}{R}\right) \quad I \geq 0 \quad (1.4)$$

avec R la réflectivité moyenne de la zone. L'espérance et l'écart-type de I sont égaux à R . La variabilité des valeurs dépend donc de la réflectivité de la zone. La distribution en amplitude $A = \sqrt{I}$ suit une loi de Rayleigh :

$$p(A|R) = \frac{2A}{R} \exp\left(-\frac{A^2}{R}\right) \quad A \geq 0. \quad (1.5)$$

Afin de réduire le speckle, on peut moyennner L images en intensité décorréelées pour obtenir une image en intensité L -vues. Ces images sont obtenues en découpant l'ouverture synthétique en L sous-ouvertures. L'image multi-vue obtenue aura donc une plus faible résolution. Pour une image L -vues, l'intensité suit une loi Gamma généralisée à deux paramètres :

$$p(I|R, L) = \frac{L^L}{R^L \Gamma(L)} I^{L-1} \exp\left(-\frac{LI}{R}\right) \quad (1.6)$$

et l'amplitude une loi de Raleigh-Nakagami :

$$p(A|R, L) = \frac{2L^L}{R^L \Gamma(L)} A^{2L-1} \exp\left(-\frac{LA^2}{R}\right), \quad (1.7)$$

avec Γ la fonction Gamma. Ces lois sont bien adaptées aux images basse résolution, mais elles sont insuffisantes pour les images à haute résolution et particulièrement les milieux urbains.

La loi de Fisher est cependant bien adaptée à ce type d'images. C'est une loi à trois paramètres μ , L et M :

$$F_{\mu,L,M}(I) = \frac{L}{M\mu} \frac{\Gamma(L+M)}{\Gamma(L)\Gamma(M)} \frac{\left(\frac{LI}{M\mu}\right)^{L-1}}{\left(1 + \frac{LI}{M\mu}\right)^{L+M}} \quad (1.8)$$

avec ici $\mu = R$. L'espérance et la variance de I sont égales à :

$$\begin{aligned} E[I] &= \frac{M}{M-1}\mu \\ Var[I] &= \frac{M^2(L+M-1)}{L(M-1)^2(M-2)}\mu^2. \end{aligned} \quad (1.9)$$

Estimation des paramètres des lois Plusieurs méthodes existent afin d'estimer les paramètres, comme la méthode des moments ou du maximum de vraisemblance. On peut aussi citer la méthode des log-moments [Nic02], valable pour des variables x positives. Pour une loi de probabilité $p(x)$ on définit le log moment d'ordre i :

$$\tilde{m}_{p,i} = \int_0^{+\infty} (\log x)^i p(x) dx. \quad (1.10)$$

Ces log-moments correspondent aux moments de la variable $\log x$. Connaissant la loi $p(x)$, on peut calculer $\tilde{m}_{p,i}$ et obtenir une formule dépendant des paramètres de $p(x)$. En estimant les log-moments à partir de données expérimentales, on peut en déduire les valeurs des paramètres. Avec les mêmes formules reliant les moments aux cumulants, on définit les log-cumulants $\tilde{\kappa}_{p,i}$ à partir des log-moments $\tilde{m}_{p,i}$. Cela permet d'obtenir par exemple l'espérance et la variance de $\log x$.

Filtrage des images radar De nombreux filtres ont été proposés pour le débruitage des images radar. Des filtres classiques peuvent être utilisés comme le filtre multi-vues, moyenneur ou médian. D'autres filtres prennent en compte les modèles statistiques des images radar. On peut citer par exemple les filtres de type MAP tel que le Gamma-MAP ou Fisher-MAP, ou les filtres de Lee et Kuan. Des filtres récents s'appuient sur des approches non-locales comme NL-SAR [DDT⁺13] ou SAR-BM3D [MFFL10, PPAV12]. On trouvera une revue récente de ces filtres dans [ALAB13].

1.3 Images de télédétection Haute Résolution et nouvelles problématiques en milieu urbain

1.3.1 Images à Haute Résolution

La dernière génération de satellites a permis une amélioration de la résolution spatiale des images. Les capteurs radar obtiennent une résolution de l'ordre du mètre (TerraSAR-X et Cosmo-SkyMed à 1 m en mode Spotlight) et les capteurs optiques de l'ordre de plusieurs dizaines de centimètres (Geoeye à 50 cm, Pleiades à 70 cm). Cette amélioration du niveau de détail entraîne cependant une augmentation de la complexité des images.

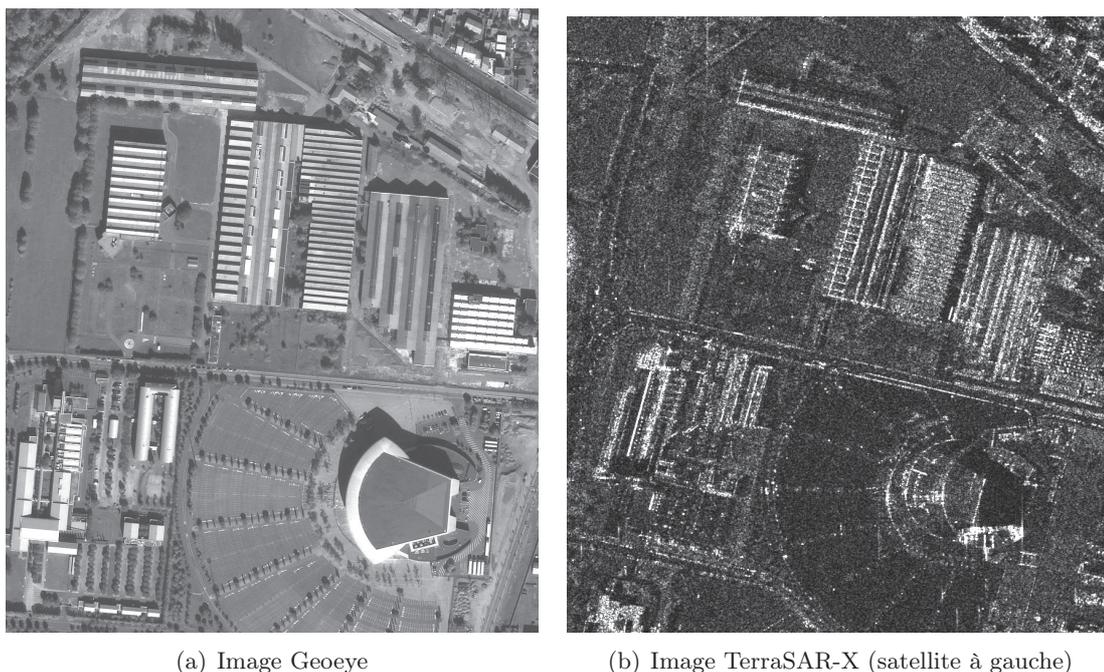


FIGURE 1.12 – Exemple d’une image optique et radar présentant la même zone.

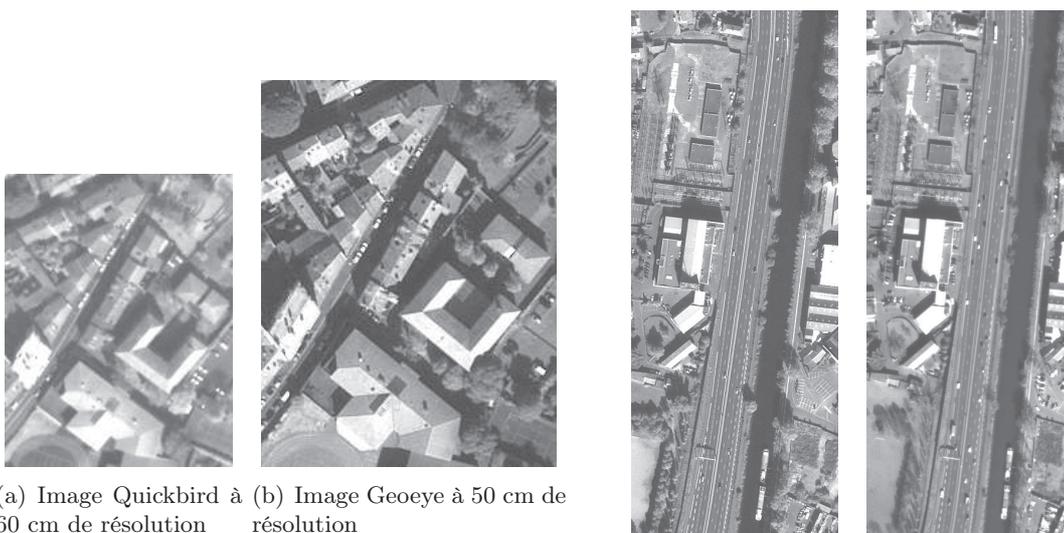


FIGURE 1.13 – Changement de résolution entre deux images optiques provenant de capteurs différents.

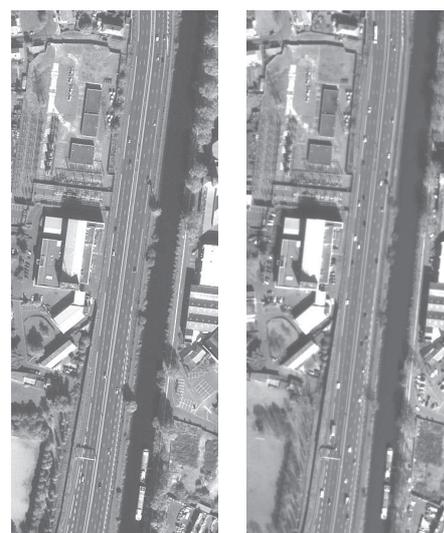
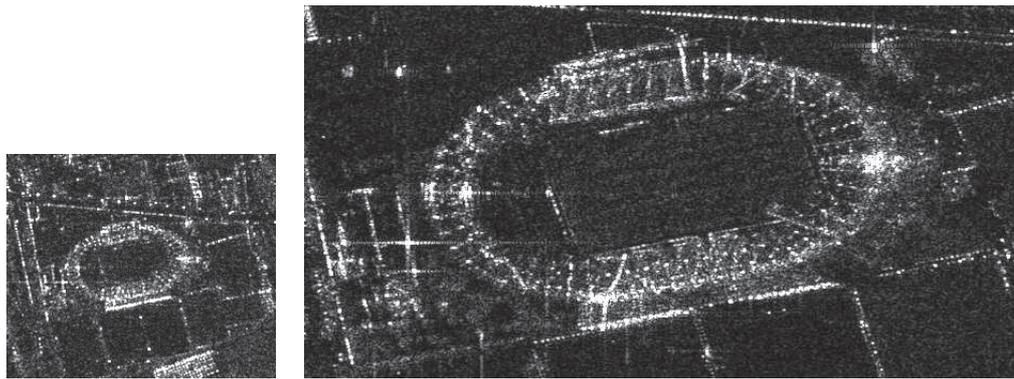


FIGURE 1.14 – Rotation entre deux images optiques provenant de capteurs différents.

1.3.2 Comparaisons optique/radar

Les images optiques ont des résolutions proches des images radar, mais elles offrent une très bonne précision des détails et un faible bruit. Leur mode d’acquisition étant proche des appareils photographiques, elles sont facilement interprétables. Cependant l’acquisition de ces images est fortement dépendante des conditions temporelles et météorologiques (nuit, nuages...). A l’inverse il est possible d’acquérir des images radar la nuit et à travers les nuages. Néanmoins ces images sont très bruitées, présentent des distorsions géométriques



(a) Mode Stripmap 3 m de résolution

(b) Mode Spotlight 1 m de résolution

FIGURE 1.15 – Changement de résolution entre deux images radar Cosmo-SkyMed.

liées à la visée latérale nécessaire et sont donc plus difficilement interprétables (Figure 1.12). Lorsqu'elles sont disponibles, les images optiques sont généralement préférées. Mais dans de nombreuses situations, et tout particulièrement la détection de changements après catastrophe, les images radar sont souvent les premières disponibles. Par ailleurs, les images radar sont utiles pour des applications spécifiques. Par exemple, elles sont intensivement utilisées pour les modèles numériques d'élévation et le suivi des mouvements de terrain grâce à l'interférométrie. Elles sont aussi très utiles pour la détection des surfaces inondées ou des pollutions marines difficilement visibles en optique. Par ailleurs la polarimétrie apporte des informations très fines sur la végétation.

1.3.3 Séries d'images multi-capteurs, multi-résolutions en milieu urbain

Dans des situations comme la détection de changements après catastrophe, le choix des images disponibles est souvent limité. On peut être amené à comparer des images provenant de différents capteurs, avec des résolutions et angles de prise de vue différents, ou même des modalités différentes. La comparaison entre une image optique et une radar (Figure 1.12) est un problème à part entière à cause de la différence de système d'acquisition : les objets sont en effet présentés avec une géométrie différente sur les images radar en raison des angles d'incidence élevés (par rapport à l'imagerie optique) qui sont utilisés. Cependant comparer des images haute résolution d'une même modalité n'est pas exempt de difficultés. La différence de capteurs et de conditions d'acquisition entraîne des problèmes récurrents :

Changement d'échelle Les images peuvent présenter une différence de résolution liée à des capteurs différents (Figure 1.13) ou des modes d'acquisitions différents, comme les modes Stripmap et Spotlight en radar (Figure 1.15). Ce changement d'échelle entraîne des différences pour la taille des objets ou le niveau de détails.

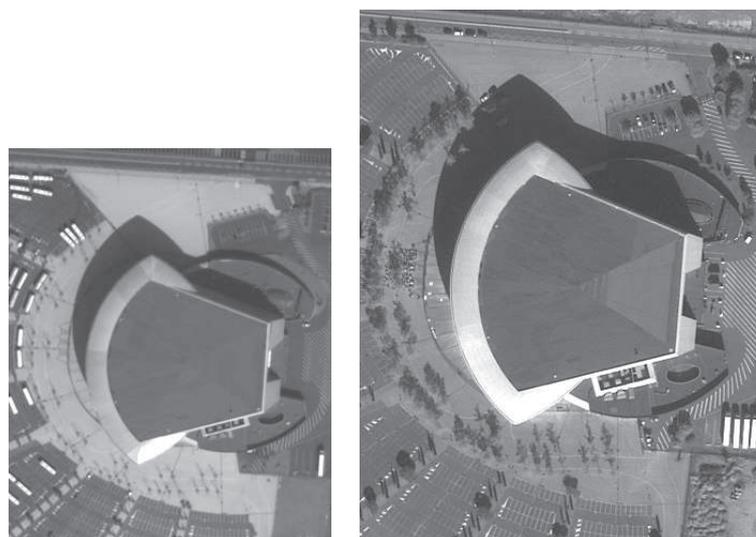
Rotation A cause de la trace du capteur et des paramètres d'acquisitions, les images peuvent présenter une rotation, souvent faible (Figure 1.14).

Parallaxes et distorsions géométriques Comme présenté précédemment, les changements d'angle de ligne de visée ont une forte influence sur la géométrie apparente des objets. Pour les images optiques, il se produit des phénomènes de parallaxe et d'occlusions (Figure 1.4 et 1.16). Les images radar présentent des distorsions géométriques (Section 1.2.5), plus ou moins fortes en fonction de l'angle d'incidence. Des exemples sont



(a) Image Worldview, 2° par rapport au nadir. (b) Image Pleïades, 23° par rapport au nadir.

FIGURE 1.16 – Différences d’ombres et de parallaxes entre deux images optiques. Pour l’image (b), le satellite est à droite de l’image. On observe bien les façades des tours à droite des toits.



(a) Image Quickbird (b) Image Geoeye

FIGURE 1.17 – Différence d’illumination entre deux images optiques.

présentés Figure 1.18. On peut notamment remarquer que les objets sont plus «étalés» quand l’angle d’incidence est grand (bâtiments 2 et 3, et alignement d’arbres 6) et que certaines lignes brillantes sont déplacées (bâtiments 1, 2, 4 et 5).

Des distorsions interviennent aussi si l’acquisition des images ne se fait pas du même côté : par exemple entre une orbite descendante et ascendante pour une visée à droite (Figure 1.9). Dans ces situations, les différences sont très fortes et la comparaison des images est particulièrement difficile.

Luminosité et dynamique La radiométrie des images dépendant des capteurs, des différences peuvent apparaître pour la luminosité de certains objets ou la dynamique des images. On peut observer un exemple de changement d’illumination sur la Figure 1.17 en optique et sur la Figure 1.18 (bâtiment 3) en radar.

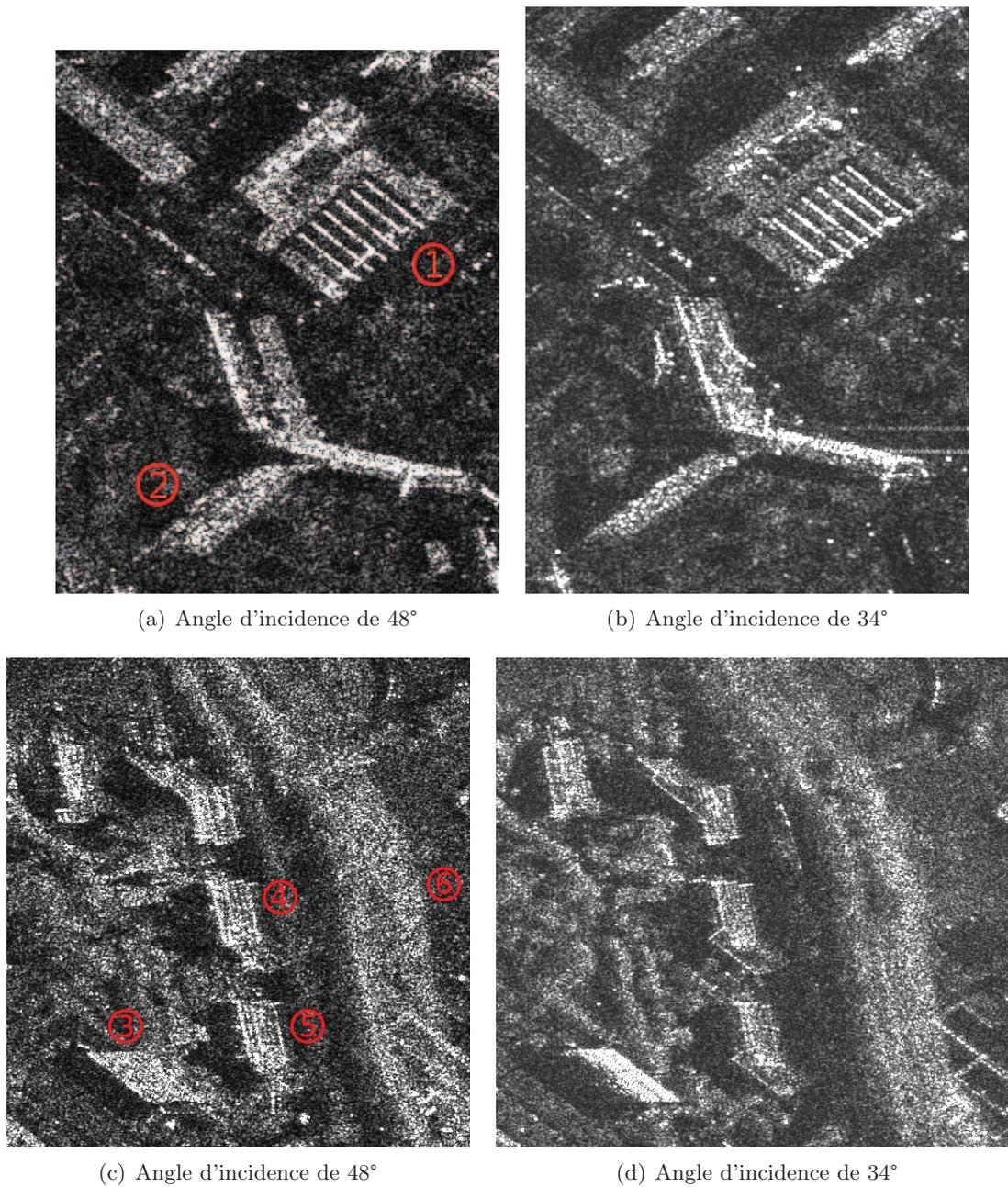


FIGURE 1.18 – Images radar CosmoSky-Med prises avec des angles d’incidence différents. Les numéros font référence à des objets présentant des distortions quand l’angle d’incidence varie.

Ombres Les ombres des bâtiments peuvent fortement varier en fonction de la position du soleil (déterminée par l’heure et la saison). Ce phénomène est d’autant plus important si le bâtiment est haut (Figure 1.16).

1.3.4 Conclusions

Du fait des problèmes présentés précédemment, la comparaison d’images haute résolution prises dans des conditions non identiques ne peut se faire par des approches pixelliques. L’utilisation de descripteurs locaux est alors à privilégier. Un descripteur local, ou caractéristique, décrit une zone de l’image de façon compacte et discriminante. L’information

codée peut être la couleur, la géométrie, la texture, la forme... Ces descripteurs sont souvent extraits à des positions particulières de l'image, après une étape de détection de points intéressants. Grâce à certaines normalisations, ils permettent souvent de gérer les problèmes de translation, rotation, changement d'échelle et d'illumination. En associant ces descripteurs à l'aide de distances, on peut ainsi mettre en correspondance différents objets d'une image. L'utilisation de descripteurs permet donc de gérer les problèmes liés à notre contexte.

De nombreux descripteurs ont été développés pour les images numériques et tout particulièrement l'algorithme SIFT [Low04]. Certains descripteurs ont déjà été appliqués aux images optiques, cependant leurs performances restent très limitées pour les images radar.

Nous présenterons dans la suite de cette partie un état de l'art commenté sur l'algorithme SIFT et ses variantes, suivi d'une étude préliminaire de la robustesse de cet algorithme sur les images bruitées.

Chapitre 2

Présentation de l'algorithme SIFT et variantes

L'algorithme SIFT (Scale-invariant feature transform), développé par David Lowe [Low99, Low04], est un algorithme d'analyse d'images très utilisé dans le domaine de la vision par ordinateur mais aussi en télédétection. Il est particulièrement utile pour la comparaison d'images, mais également pour la détection ou la reconnaissance d'objets. Il associe une détection de points d'intérêts et une extraction de descripteurs locaux. Il présente l'avantage d'être invariant par translation, rotation et changement d'échelle, ainsi que d'être robuste aux changements d'illumination et aux transformations affines. Il est ainsi utilisé pour de nombreuses applications comme la mise en correspondance entre images, la détection et la reconnaissance d'objets, l'indexation ou le recalage. Dans ce qui suit, nous nous restreignons aux applications des SIFTs à la mise en correspondance.

Différentes variantes de l'algorithme ont été développées, mais elles reposent toutes sur le même principe :

1. Dans un premier temps des points-clés sont extraits à partir d'un espace-échelle. La détection est invariante à différentes transformations (translation, changement d'échelle et d'illumination, rotation) et robuste au bruit (supposé gaussien la plupart du temps).
2. Une ou plusieurs orientations principales sont associées à chaque point. On obtient des points-clés caractérisés par quatre paramètres : une position (x, y) , une échelle σ et une orientation θ , assurant ainsi l'invariance à la translation, au changement d'échelle et à l'orientation.
3. A chaque point-clé est ensuite associé un descripteur qui caractérise la géométrie locale de l'image autour du point. En utilisant les quatre paramètres du point, ce descripteur est invariant à différentes transformations.
4. Enfin, dans le cas où l'on cherche à mettre en correspondance deux images, les descripteurs de deux images sont comparés (en calculant des distances par exemple) afin d'apparier leurs points-clés respectifs.

Chacune de ces quatre étapes sera présentée pour l'algorithme SIFT original ainsi que pour les différentes variantes proposées dans la littérature. S'en suivra un rapide état de l'art sur l'application de l'algorithme SIFT aux images de télédétection.

2.1 Détection de points-clés

La première étape consiste à extraire des points d'intérêts, dits points-clés, en appliquant un opérateur différentiel à un espace-échelle. Ces points sont caractérisés par une position (x, y) et un facteur d'échelle caractéristique σ qui donne une information sur la taille de la structure détectée. Une étape est souvent rajoutée afin de filtrer certains points non intéressants et de préciser la position et l'échelle des points.

Cette section présente au préalable l'espace-échelle gaussien et trois opérateurs différentiels. Plusieurs détecteurs proposés dans la littérature et basés sur ces opérateurs seront ensuite décrits.

2.1.1 Opérateurs différentiels appliqués à un espace-échelle

Soit une image en niveaux de gris :

$$I : (x, y) \in \Omega \subset \mathbb{R}^2 \mapsto I(x, y) \in \mathbb{R}$$

et un noyau gaussien de paramètre σ , avec $\sigma > 0$:

$$\mathcal{G}_\sigma : (u, v) \in \mathbb{R}^2 \mapsto \frac{1}{2\pi\sigma^2} \exp\left(-\frac{u^2 + v^2}{2\sigma^2}\right).$$

La famille de convolutions de cette image I avec le noyau \mathcal{G}_σ :

$$I_\sigma(x, y) := \mathcal{G}_\sigma * I(x, y) = \iint_{(u,v) \in \Omega} I(u, v) \mathcal{G}_\sigma(x - u, y - v) dudv \quad (2.1)$$

est une représentation en espace-échelle linéaire couramment utilisée. L'opération de convolution permet de lisser l'image à différentes échelles, et donc d'estomper des détails plus ou moins grands. On obtient ainsi une représentation de l'image en trois dimensions, avec la position (x, y) et l'échelle σ .

Afin d'extraire de cette représentation des points de façon robuste et répétable, différents opérateurs différentiels sont proposés.

2.1.1.1 Laplacien normalisé

Lindeberg [Lin98] propose un détecteur de *blob*¹ avec le laplacien normalisé :

$$\Delta_\sigma I_\sigma(x, y) := \sigma^2 \Delta I_\sigma(x, y) = \left\{ \left(\frac{x^2 + y^2}{\sigma^2} - 2 \right) \mathcal{G}_\sigma(x, y) \right\} * I(x, y). \quad (2.2)$$

Ce détecteur répond fortement aux structures de type blob, avec une relation entre l'échelle σ de détection et la taille des «blobs». Une normalisation du laplacien selon l'échelle est cependant nécessaire pour avoir une invariance au changement d'échelle. On appelle Laplacien de Gaussienne (Laplacian of Gaussian ou LoG) le filtre :

$$\Delta_\sigma \mathcal{G}(x, y) := \sigma^2 \Delta \mathcal{G}_\sigma(x, y) = \left(\frac{x^2 + y^2}{\sigma^2} - 2 \right) \mathcal{G}_\sigma(x, y).$$

Pour obtenir les points d'intérêt, on sélectionne alors les extrema locaux de $\Delta_\sigma I_\sigma(x, y)$ en espace (x, y) et en échelle σ .

1. Un *blob* est une «tache», une zone homogène pouvant avoir différentes tailles.

(a) Exemples de détection avec le Laplacien dans différents cas simples :

Un segment de largeur $2N$ et de contraste M (en 1D) :

$$\text{Soit } I(x) = \begin{cases} M & \text{si } -N \leq x \leq N \\ 0 & \text{sinon,} \end{cases}$$

la convolution gaussienne de I et son laplacien normalisé sont donnés par :

$$I_\sigma(x) = \frac{M}{2} \left(\operatorname{erf} \left(\frac{N-x}{\sigma\sqrt{2}} \right) + \operatorname{erf} \left(\frac{N+x}{\sigma\sqrt{2}} \right) \right)$$

$$\Delta_\sigma I_\sigma(x) = \frac{2MN}{\sqrt{2\pi\sigma^2}} \exp \left(-\frac{x^2 + N^2}{2\sigma^2} \right) \cosh \left(\frac{xN}{\sigma^2} \right) \left(\frac{x}{N} \tanh \left(\frac{xN}{\sigma^2} \right) - 1 \right).$$

La fonction erf est la fonction d'erreur de Gauss :

$$\operatorname{erf}(x) = \frac{2}{\pi} \int_0^x e^{-t^2} dt.$$

Les extrema locaux du laplacien normalisé selon x et σ satisfont :

$$\begin{cases} x = 0 \\ \sigma = N. \end{cases}$$

Un extremum est trouvé au centre du segment pour $\sigma = N$, ce qui donne une information sur la taille du segment.

Un coin : Considérons un coin infini de contraste M :

$$I(x, y) = \begin{cases} M & \text{si } x, y \geq 0 \\ 0 & \text{sinon.} \end{cases}$$

La convolution gaussienne de I et son laplacien normalisé sont représentés par :

$$I_\sigma(x, y) = \frac{M}{4} \left(1 + \operatorname{erf} \left(\frac{x}{\sigma\sqrt{2}} \right) \right) \left(1 + \operatorname{erf} \left(\frac{y}{\sigma\sqrt{2}} \right) \right)$$

$$\Delta_\sigma I_\sigma(x, y) = -\frac{M}{\sqrt{2\pi}} \left(\frac{x e^{-\frac{x^2}{2\sigma^2}}}{\sigma} \left(1 + \operatorname{erf} \left(\frac{y}{\sigma\sqrt{2}} \right) \right) + \frac{y e^{-\frac{y^2}{2\sigma^2}}}{\sigma} \left(1 + \operatorname{erf} \left(\frac{x}{\sigma\sqrt{2}} \right) \right) \right).$$

Les extrema locaux selon x , y et σ satisfont $x = y = a\sigma$, avec $a \simeq 1.13$. Pour chaque échelle on obtient un extremum sur la bissectrice du coin. En théorie, il n'y a pas d'extrema au sens strict en échelle, mais en raison de la discrétisation des images et des échelles on en obtient de nombreux à différentes échelles.

Une gaussienne de moyenne μ et d'écart-type ς : Dans le cas mono-dimensionnel et avec un contraste M , cette gaussienne s'écrit :

$$I(x) = \frac{M}{\sqrt{2\pi\varsigma^2}} \exp \left(-\frac{(x-\mu)^2}{2\varsigma^2} \right),$$

la convolution gaussienne de I et son laplacien normalisé sont représentés par :

$$I_\sigma(x) = \frac{M}{\sqrt{2\pi(\varsigma^2 + \sigma^2)}} \exp \left(-\frac{(x-\mu)^2}{2(\varsigma^2 + \sigma^2)} \right)$$

$$\Delta_\sigma I_\sigma(x) = \frac{\sigma^2 M}{\sqrt{2\pi(\varsigma^2 + \sigma^2)}^{\frac{3}{2}}} \left(\frac{(x-\mu)^2}{\varsigma^2 + \sigma^2} - 1 \right) \exp \left(-\frac{(x-\mu)^2}{2(\varsigma^2 + \sigma^2)} \right).$$

Les extrema locaux selon x et σ satisfont :

$$\begin{cases} x = \mu \\ \sigma = \varsigma\sqrt{2} \end{cases}$$

On trouve donc un extremum en μ avec comme échelle $\varsigma\sqrt{2}$.

On remarque que les extrema du LoG sont invariants aux changements de contraste affines de l'image. Ce détecteur présente donc l'avantage d'être invariant à l'illumination, mais aussi à la rotation et au changement d'échelle. En pratique on ne conserve cependant que les extrema locaux pour lesquels $\Delta_\sigma I_\sigma$ est supérieur à un seuil, il y a donc une dépendance au contraste.

(b) Calculs pratiques : Pour calculer le laplacien normalisé sur des images discrètes, il est nécessaire d'échantillonner le domaine des échelles avec une progression géométrique :

$$\sigma_k = \sigma_0 \cdot r^k \text{ avec } k \in \{0, \dots, k_{max}\}$$

avec k l'indice de l'échelle, σ_0 la plus petite échelle étudiée et r le rapport entre deux échelles successives. Les extrema du laplacien sont sélectionnés en observant les voisins dans les trois dimensions de chaque point (x, y, σ) . A cause de la discrétisation des images, on observe de nombreux extrema détectés autour des structures pour de faibles valeurs du laplacien. Pour pallier ce problème, on calcule les extrema de $|\Delta_\sigma I_\sigma(x, y)|$ et on seuille les faibles valeurs avec :

$$|\Delta_\sigma I_\sigma(x, y)| < t_l. \quad (2.3)$$

Lowe [Low04] propose comme valeur $t_l = 0.03$, considérant que l'image I a été préalablement normalisée dans l'intervalle $[0, 1]$. Cette opération limite cependant l'invariance au contraste.

(c) Deux approximations sont proposées dans [Low04] afin d'optimiser la convolution gaussienne et le calcul du laplacien. Pour la convolution gaussienne, l'espace-échelle est représenté par une pyramide : l'image I est sous-échantillonnée successivement par un facteur 2 à chaque fois que l'échelle σ_k atteint une puissance de 2. Les séries d'images successives ayant la même taille sont appelées octave. En réduisant la taille de l'image, on peut ainsi diminuer la complexité de la convolution gaussienne pour les grandes échelles. Une interpolation est ensuite réalisée afin d'améliorer la précision de la position des points détectés.

La deuxième approximation concerne le calcul du laplacien normalisé. Sachant que dans le cas discret on a comme approximation :

$$\mathcal{G}_{\sigma_{k+1}} \simeq \mathcal{G}_{\sigma_k} + (\sigma_{k+1} - \sigma_k) \frac{\partial \mathcal{G}_\sigma}{\partial \sigma} = \mathcal{G}_{\sigma_k} + \sigma_k(r - 1) \frac{\partial \mathcal{G}_\sigma}{\partial \sigma} \quad (2.4)$$

et que \mathcal{G}_σ vérifie l'égalité :

$$\frac{\partial \mathcal{G}_\sigma}{\partial \sigma} = \sigma \Delta_\sigma \mathcal{G}_\sigma, \quad (2.5)$$

on peut approximer le LoG par une différence de gaussiennes (Difference of Gaussians ou DoG) :

$$\mathcal{G}_{\sigma_{k+1}} - \mathcal{G}_{\sigma_k} \simeq (r - 1) \sigma_k^2 \Delta \mathcal{G}_{\sigma_k} = (r - 1) \Delta_{\sigma_k} \mathcal{G}_{\sigma_k}. \quad (2.6)$$

Le laplacien normalisé est alors estimé par :

$$\Delta_{\sigma_k} I_{\sigma_k} \simeq \frac{I_{\sigma_{k+1}} - I_{\sigma_k}}{r - 1}. \quad (2.7)$$

La pyramide laplacienne est ainsi obtenue par soustraction de deux images successives de la pyramide gaussienne.

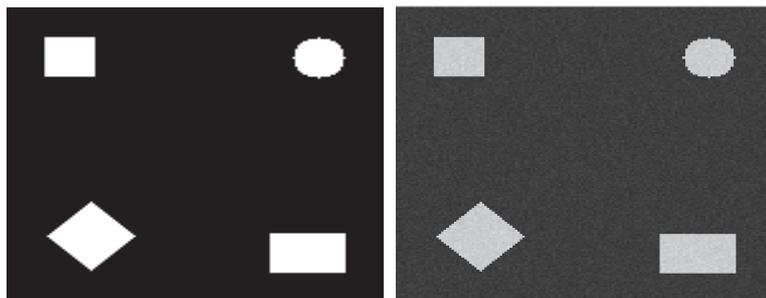
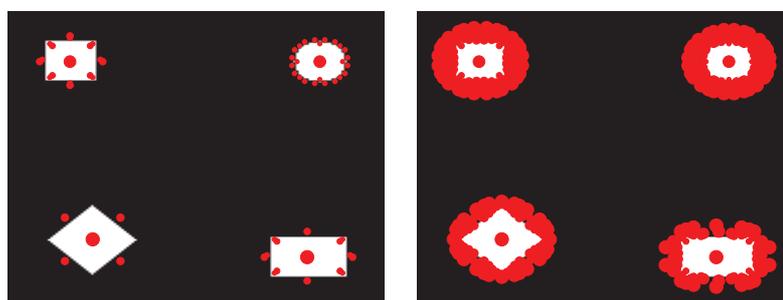
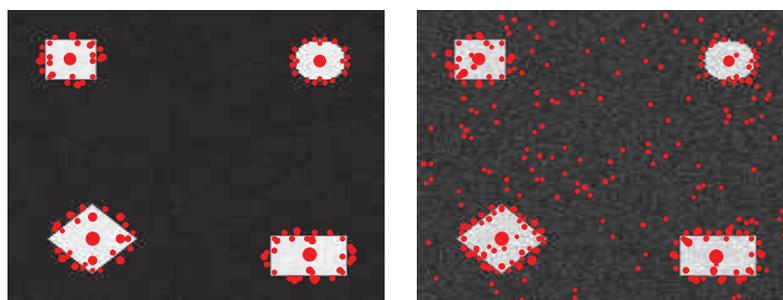


FIGURE 2.1 – Exemple jouet idéal, à gauche sans bruit et à droite avec ajout de bruit gaussien centré d'écart-type $\nu = 2$. Les objets ont pour niveau de gris $M_1 = 170$, et le fond $M_2 = 100$.



(a) Calcul sur l'image non bruitée, avec seuillage $t_l = 0.03$. (b) Calcul sur l'image non bruitée, sans seuillage t_l .



(c) Calcul sur l'image bruitée avec un écart-type $\nu = 2$, avec seuillage $t_l = 0.03$. (d) Calcul sur l'image bruitée avec un écart-type $\nu = 5$, avec seuillage $t_l = 0.03$.

FIGURE 2.2 – Extrema locaux en 3D du laplacien normalisé calculés sur l'image jouet bruitée et non bruitée de la Figure 2.1. Les échelles utilisées sont calculées avec $\sigma_0 = 0.63$, $r = 2^{1/3}$ et $k_{max} = 13$. Les extrema sont représentés à l'aide de points rouges, de taille proportionnelle à leur échelle σ .

(d) **Exemples :** Afin d'étudier le comportement du laplacien normalisé, on observe le calcul des extrema sur un exemple jouet avec et sans bruit (Figure 2.1). Les points détectés sont présentés Figure 2.2. Comme attendu, un extremum est détecté au centre de chacun des objets, à une échelle caractéristique de la taille de l'objet. On remarque que le seuillage sur les faibles valeurs du laplacien est nécessaire pour supprimer les nombreux extrema sur les contours des objets. Concernant les rectangles, des points sont bien détectés au niveau des coins mais ils sont décalés par rapport au coin. Cependant pour le losange, aucun point n'est détecté dans les coins. De plus, même avec l'utilisation du seuil, des points sont détectés à côté des bords de chacun des objets. Ceci est lié aux dérivées partielles secondes

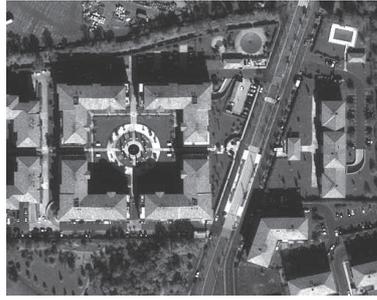


FIGURE 2.3 – Exemple d'une image satellite optique

qui entraînent des extrema autour des bords stricts. Ces points ne sont pas très informatifs et risquent de causer des erreurs lors de la mise en correspondance. En effet deux points d'un même bord peuvent représenter la même information.

La détection sur une image bruitée présente de nombreuses détections de points-clés long des bords. Comme expliqué précédemment, ces points ne sont pas intéressants. Leur localisation est d'ailleurs peu précise à cause de la présence de bruit. On remarque aussi que pour un bruit plus fort ($\nu = 5$), il y a de nombreuses fausses détections dans les zones homogènes. Ces erreurs ne sont pas filtrées par le seuil sur le laplacien t_l , mais un seuil plus strict risquerait de supprimer des points dans des zones caractéristiques.

La Figure 2.4 présente un exemple de calcul du laplacien normalisé et de ses extrema sur une image satellite (Figure 2.3). De nombreux points sont détectés sur des structures intéressantes. On observe cependant beaucoup de détections peu informatives (sur les zones homogènes, de végétation, de bord...), notamment pour les faibles échelles. Comme pour l'exemple jouet, ces détections ne sont pas pertinentes. Il est donc nécessaire d'utiliser un critère pour les filtrer.

(e) **Conclusions :** Le LoG est un détecteur de blob invariant aux changements de contraste affines, à la rotation et au changement d'échelle. Il nécessite cependant le recours à un seuillage pour limiter les détections qui lui font perdre son invariance au contraste. Il a la particularité d'être un très bon estimateur de l'échelle caractéristique de structures de type blobs [MS01], mais il présente une mauvaise précision concernant la position en espace des points d'intérêts et est peu robuste face à au bruit. Il est nécessaire d'associer un critère de filtrage à ce détecteur.

2.1.1.2 Critère de Harris multi-échelle

(a) **Critère de Harris :** Une autre sélection de points d'intérêt en espace-échelle [MS01] se base sur le détecteur de coin de Harris [HS88].

Ce détecteur repose sur le calcul de la matrice de Harris en un point (x, y) :

$$M_H(x, y) = w * \begin{bmatrix} \left(\frac{\partial I}{\partial x}(x, y)\right)^2 & \frac{\partial I}{\partial x} \cdot \frac{\partial I}{\partial y}(x, y) \\ \frac{\partial I}{\partial x} \cdot \frac{\partial I}{\partial y}(x, y) & \left(\frac{\partial I}{\partial y}(x, y)\right)^2 \end{bmatrix} \quad (2.8)$$

Avec w une pondération spatiale. Pour une pondération avec une fenêtre gaussienne \mathcal{G} de paramètre s fixé, on obtient :

$$M_H(x, y) = \mathcal{G}_s * \begin{bmatrix} \left(\frac{\partial I}{\partial x}(x, y)\right)^2 & \frac{\partial I}{\partial x} \cdot \frac{\partial I}{\partial y}(x, y) \\ \frac{\partial I}{\partial x} \cdot \frac{\partial I}{\partial y}(x, y) & \left(\frac{\partial I}{\partial y}(x, y)\right)^2 \end{bmatrix} \quad (2.9)$$

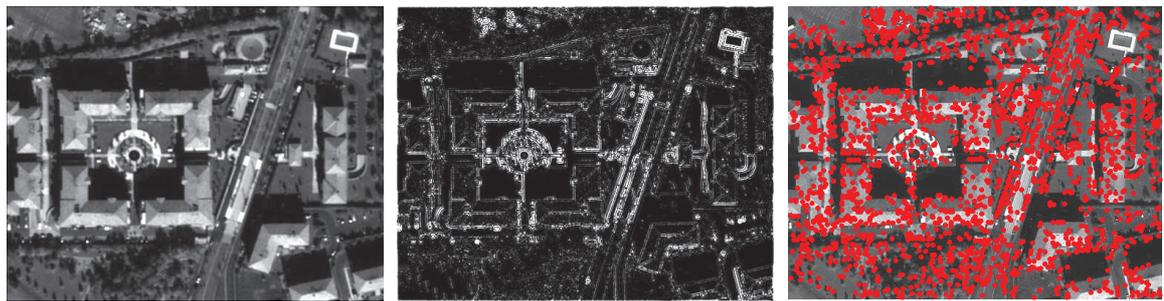
(a) Echelle $\sigma = 1$ (b) Echelle $\sigma = 2$ (c) Echelle $\sigma = 4$ (d) Echelle $\sigma = 8$

FIGURE 2.4 – Exemple de calcul du laplacien normalisé et de ses extrema sur l'image 2.3. Sur chaque ligne sont représentés de gauche à droite la convolution gaussienne, le laplacien normalisé en valeur absolue et la position spatiale des extrema pour une échelle σ . Les positions des extrema locaux (en espace et en échelle, calculés en 26-connexité) sont représentées à l'aide de points rouges. Les point candidats sont filtrés avec le test de l'Equation (2.3) pour $t_l = 0.03$.

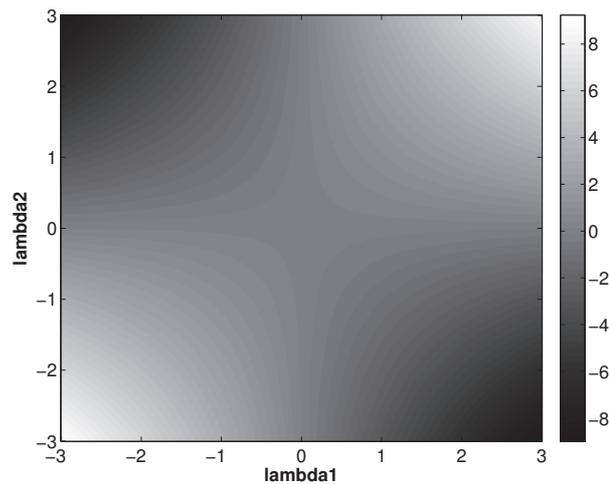


FIGURE 2.5 – Représentation de $\mathcal{C}_H(x, y)$ en fonction des valeurs propres λ_1 et λ_2 avec $k_1 = 0.04$.

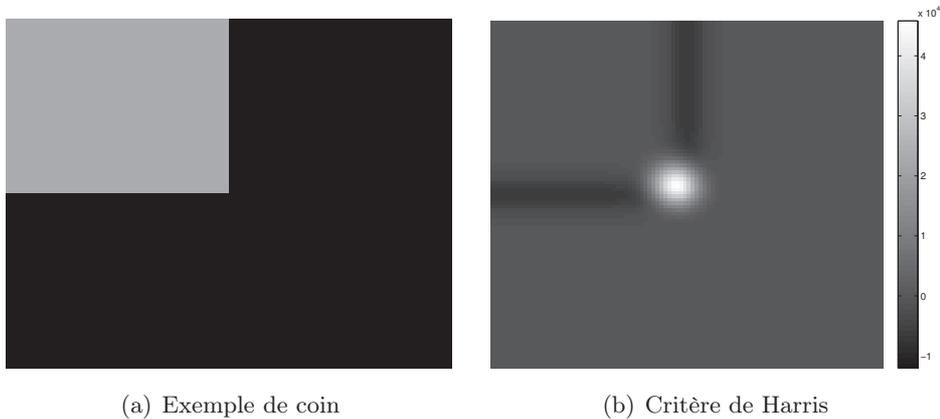


FIGURE 2.6 – Calcul du critère de Harris $\mathcal{C}_H(x, y)$ sur un coin pour $s = 5$ et $k_1 = 0.04$.

Cette matrice décrit la courbure de l'image localement :

- Si ses deux valeurs propres λ_1 et λ_2 ont conjointement des fortes valeurs et sont de même signe, il y a deux orientations principales autour du point : on suppose qu'on se trouve sur un coin.
- Si λ_1 est faible et λ_2 fort (ou inversement), il y a une orientation principale autour du point : on suppose qu'on se trouve sur un bord.
- Si λ_1 et λ_2 sont toutes les deux faibles, il n'y a pas d'orientation principale : on suppose qu'on se trouve sur une zone homogène.

Pour éviter de calculer les valeurs propres, Harris et Stephens [HS88] proposent d'utiliser le critère :

$$\mathcal{C}_H(x, y) = \det(M_H(x, y)) - k_1 \cdot \text{Tr}(M_H(x, y))^2 = \lambda_1 \lambda_2 - k_1 (\lambda_1 + \lambda_2)^2 \quad (2.10)$$

avec k_1 un paramètre à définir empiriquement.

Comme présenté Figure 2.5, ce critère prend de fortes valeurs quand les valeurs propres λ_1 et λ_2 sont conjointement importantes. En effectuant le test :

$$\mathcal{C}_H(x, y) > d_h \quad (2.11)$$

On peut ainsi détecter les coins d'une image. Le paramètre positif d_h est à définir empiriquement. Les valeurs 1500 [MS01] ou 2000 [Rab09] sont régulièrement utilisées pour les images numériques codées sur 8 bits. La Figure 2.6 présente un exemple de calcul du critère de Harris sur un coin. On observe bien que $\mathcal{C}_H(x, y)$ présente de fortes valeurs sur les zones que l'on souhaite identifier comme des coins, des valeurs négatives pour les bords et des valeurs nulles pour les zones homogènes. Un seuil d_h important permet donc de sélectionner les points sur le coin. Ce détecteur a l'avantage principal d'être invariant à la rotation. Il est cependant dépendant du contraste de l'image et du paramètre d'échelle s choisi.

(b) Invariance au changement d'échelle : Une adaptation multi-échelle du critère est proposée par [DSH00] pour obtenir une invariance au changement d'échelle. La matrice et le critère de Harris multi-échelle sont calculés sur I_σ et définis par :

$$M_H(x, y, \sigma) = \sigma^2 \mathcal{G}_{a_1 \sigma} * \begin{bmatrix} (\frac{\partial I_\sigma}{\partial x}(x, y))^2 & \frac{\partial I_\sigma}{\partial x} \cdot \frac{\partial I_\sigma}{\partial y}(x, y) \\ \frac{\partial I_\sigma}{\partial x} \cdot \frac{\partial I_\sigma}{\partial y}(x, y) & (\frac{\partial I_\sigma}{\partial y}(x, y))^2 \end{bmatrix} \quad (2.12)$$

$$\mathcal{C}_H(x, y, \sigma) = \det(M_H(x, y, \sigma)) - k_1 \cdot \text{Tr}(M_H(x, y, \sigma))^2.$$

Afin d'assurer l'invariance au changement d'échelle, la matrice de Harris est normalisée par σ^2 et l'écart-type de la pondération gaussienne \mathcal{G} est proportionnel à σ (a_1 étant un paramètre supérieur à 1). De la même façon que le critère initial, les coins sont détectés en effectuant le test :

$$\mathcal{C}_H(x, y, \sigma) > d_h. \quad (2.13)$$

Afin de sélectionner des points dans un contexte multi-échelle, Mikola, Jczyk et Schmid [MS01] proposent d'utiliser ce critère pour une analyse en espace-échelle. Comme pour le laplacien normalisé, une représentation en espace-échelle d'une image I est construite avec des échelles $\sigma_k = \sigma_0 r^k$ suivant une progression géométrique. Les extrema en espace satisfaisant le test de l'Equation (2.13) sont sélectionnés comme points d'intérêt. La sélection ne se fait pas en trois dimensions (espace et échelle) puisque en pratique $\mathcal{C}_H(x, y, \sigma)$ atteint rarement des extrema en échelle. Dans [MS01], la sélection se fait donc en deux dimensions et le LoG est ensuite utilisé pour déterminer l'échelle la plus pertinente de chaque point.

La sélection des extrema est invariante aux changements de contraste, mais pas l'étape de seuil d_h . En effet, $\mathcal{C}_H(x, y, \sigma)$ est dépendant du contraste de l'image. Le critère de Harris multi-échelle d'une image $I_2 = m \cdot I_1$ est $\mathcal{C}_{H,2}(x, y, \sigma) = m^4 \cdot \mathcal{C}_{H,1}(x, y, \sigma)$. Sur les zones homogènes bruitées, des erreurs liées au bruit pourront ainsi être filtrées. Cependant cette dépendance au contraste risque de supprimer des points sur des structures de faible contraste, le critère n'est ainsi pas invariant aux changements d'illumination. Dans [MS01] la valeur de d_h est fixée à 1500 pour une image sur 8 bits.

(c) Exemples : La Figure 2.7 présente le calcul des extrema du critère de Harris sur l'image jouet de la Figure 2.1. Comme pour le laplacien normalisé, un seuillage d_h sur les faibles valeurs du critère est nécessaire pour supprimer des erreurs liées à la discrétisation des images. On observe ici que pour chacun des objets, des points sont détectés au centre, mais aussi sur les coins. La détection sur des images bruitées avec un bruit gaussien plus ou moins important n'entraîne cependant pas de fausses détections sur les bords ou les zones homogènes. La position des points d'intérêt n'est que faiblement perturbée.

La Figure 2.8 présente un exemple de calcul du critère de Harris multi-échelle ainsi qu'une sélection des extrema locaux en 3D (en espace et en échelle) et en 2D (en espace). On remarque que dans les deux cas les extrema sont situés en grande partie sur des structures intéressantes et qu'il y a peu de détections parasites. On remarque également qu'il y a peu d'extrema locaux en 3D.

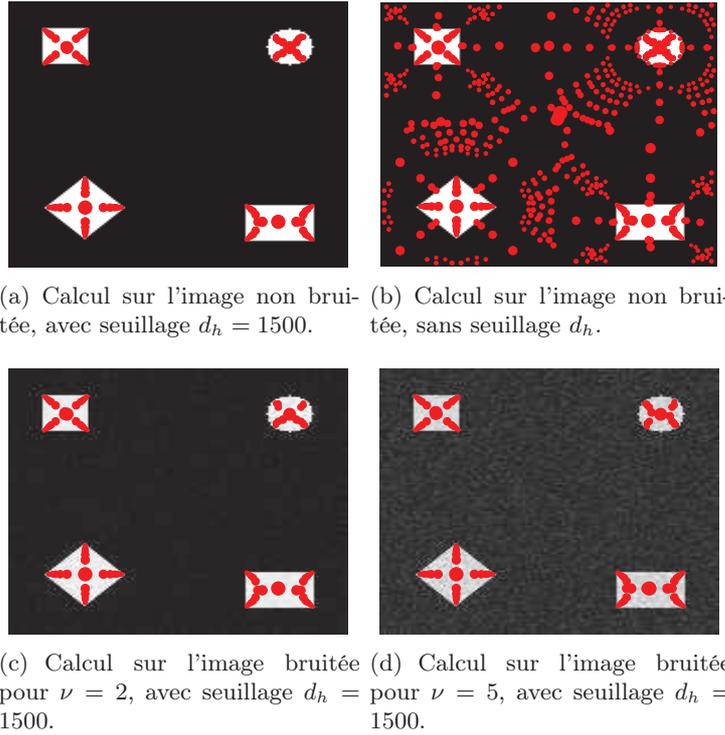


FIGURE 2.7 – Extrema locaux en 2D du critère de Harris multi-échelle calculé sur l'image jouet bruitée et non bruitée de la Figure 2.1. Les échelles utilisées sont calculées avec $\sigma_0 = 0.63$, $r = 2^{1/3}$ et $k_{max} = 13$. Les extrema sont représentés à l'aide de points rouges, de taille proportionnelle à leur échelle σ .

(d) Conclusions : Le critère de Harris multi-échelle est un détecteur de coin invariant à la rotation et au changement d'échelle. Il est plus précis que le laplacien normalisé pour la position des coins [MS02], mais est plus limité concernant l'estimation en échelle. Il est assez robuste face à du bruit, notamment grâce au filtrage sur les faibles valeurs. Sa dépendance au contraste risque cependant de supprimer des points sur des structures de faible contraste.

2.1.1.3 Critère basé sur la matrice hessienne

(a) Matrice hessienne : La détection de points d'intérêt peut aussi se faire à partir de la matrice hessienne :

$$H(x, y) = \begin{bmatrix} \frac{\partial^2 I}{\partial x^2}(x, y) & \frac{\partial^2 I}{\partial x \partial y}(x, y) \\ \frac{\partial^2 I}{\partial y \partial x}(x, y) & \frac{\partial^2 I}{\partial y^2}(x, y) \end{bmatrix}, \quad (2.14)$$

ou sa version multi-échelle :

$$H_\sigma(x, y) = \sigma^2 \begin{bmatrix} \frac{\partial^2 I_\sigma}{\partial x^2}(x, y) & \frac{\partial^2 I_\sigma}{\partial x \partial y}(x, y) \\ \frac{\partial^2 I_\sigma}{\partial y \partial x}(x, y) & \frac{\partial^2 I_\sigma}{\partial y^2}(x, y) \end{bmatrix}. \quad (2.15)$$

Cette matrice donne une information sur la courbure de l'image I_σ au point (x, y) : si les deux valeurs propres de H_σ sont proches, il y a des variations de gradient dans les deux directions. On peut ainsi détecter les structures de type blob et les coins. Comme pour la

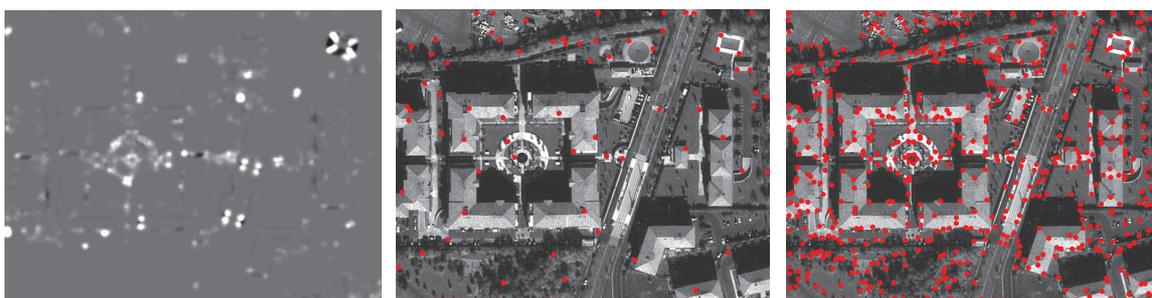
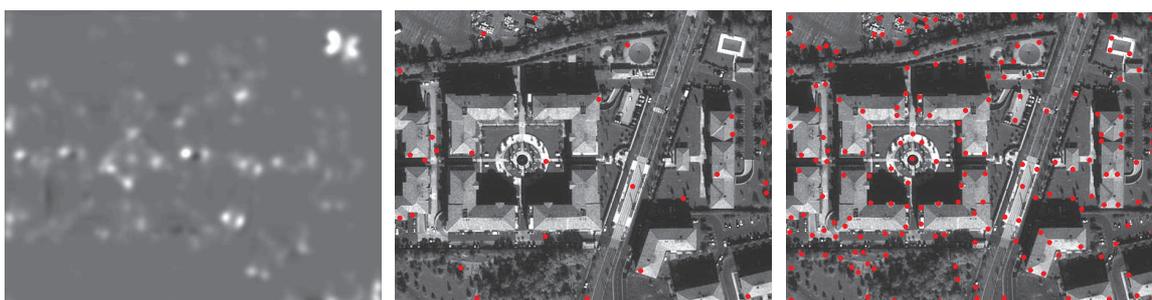
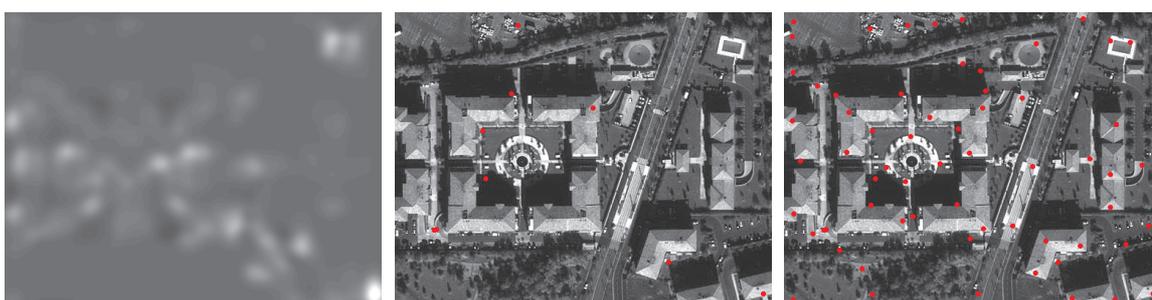
(a) Echelle $\sigma = 1$ (b) Echelle $\sigma = 2$ (c) Echelle $\sigma = 4$ (d) Echelle $\sigma = 8$

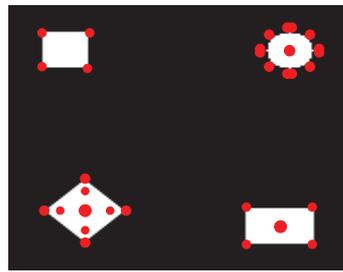
FIGURE 2.8 – Exemple de calcul du critère de Harris multi-échelle et de sélection des extrema sur l'image 2.3. Sur chaque ligne est représenté de gauche à droite le critère de Harris multi-échelle et la sélection des extrema locaux en 3D et en 2D pour une échelle σ . Les extrema locaux en 3D sont sélectionnés en espace et en échelle en 26-connexité et ceux en 2D seulement en espace en 8-connexité. Pour les deux situations, seuls les points vérifiant le test de l'Equation (2.13) avec $d_h = 2000$ sont conservés afin de ne garder que les points intéressants. La position de ces points est représentée en rouge.

matrice de Harris, Lindeberg [Lin98] propose d'observer la trace et le déterminant :

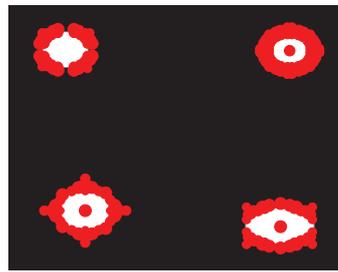
$$\begin{aligned} \text{Tr}(H_\sigma(x, y)) &= \sigma^2 \left(\frac{\partial^2 I_\sigma}{\partial x^2}(x, y) + \frac{\partial^2 I_\sigma}{\partial y^2}(x, y) \right) \\ \text{Det}(H_\sigma(x, y)) &= \sigma^4 \left(\frac{\partial^2 I_\sigma}{\partial x^2}(x, y) \cdot \frac{\partial^2 I_\sigma}{\partial y^2}(x, y) - \frac{\partial^2 I_\sigma}{\partial y \partial x}(x, y)^2 \right) \end{aligned} \quad (2.16)$$

Les points correspondant aux extrema simultanés de la trace et du déterminant sont sélectionnés. On remarque que la trace correspond exactement au LoG. La combinaison des extrema du DoH (Determinant of Hessian) avec ceux du LoG permet de supprimer les détections de points le long des bords, souvent obtenus avec le LoG. Mikolajczyk et Schmid [MS04a] proposent de sélectionner les extrema en 3D du DoH, puis d'utiliser le LoG (ou la trace de la matrice Hessienne) pour préciser l'échelle des points. De la même façon que le LoG, la sélection des extrema du DoH est indépendante du contraste. Cependant il est souvent nécessaire de seuiller les faibles valeurs du DoH pour supprimer les erreurs de détection liées au bruit. Cette opération limite l'indépendance au contraste.

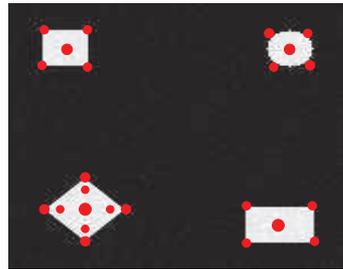
(b) Exemples : la Figure 2.9 présente un exemple de calcul du déterminant de la matrice hessienne pour différentes échelles σ sur un exemple jouet. On remarque que pour presque tous les objets, un point est détecté au centre pour une certaine échelle caractéristique, à la fois sur les images bruitées et non bruitées. Pour les objets de type rectangle, des points sont trouvés avec une bonne précision sur les coins mais aucun sur les bords. La



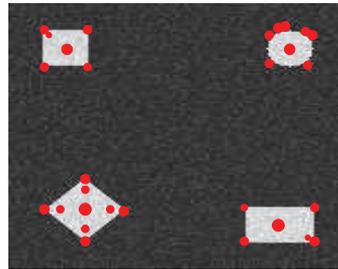
(a) Calcul sur l'image non bruitée, avec seuillage t_{he} .



(b) Calcul sur l'image non bruitée, sans seuillage t_{he} .



(c) Calcul sur l'image bruitée pour $\nu = 2$, avec seuillage t_{he} .



(d) Calcul sur l'image bruitée pour $\nu = 5$, avec seuillage t_{he} .

FIGURE 2.9 – Extrema locaux en 3D du déterminant de la matrice hessienne calculés sur l'image jouet bruitée et non bruitée de la Figure 2.1. Les échelles utilisées sont calculées avec $\sigma_0 = 0.63$, $r = 2^{1/3}$ et $k_{max} = 13$. Les extrema sont représentés à l'aide de points rouges, de taille proportionnelle à leur échelle σ . Pour les images 2.9(c) et 2.9(d), les valeurs de $|\text{Det}(H_\sigma(x, y))|$ ont été seuillées par $t_{he} = 0.03$ pour supprimer les fausses détections liées au bruit ou à la discrétisation, en considérant que l'image I est normalisée dans l'intervalle $[0, 1]$.

présence de bruit gaussien n'entraîne pas de fausses détections. La localisation des points d'intérêt est faiblement perturbée en présence de fort bruit ($\nu = 5$).

La Figure 2.10 présente un exemple de calcul du DoH ainsi qu'une sélection des extrema locaux en 3D. On trouve de nombreux extrema sur des structures intéressantes avec peu de fausses détections. Cependant le nombre de points détectés est assez faible.

(c) Conclusions : Le hessien est un détecteur de «blob» invariant à l'illumination, à la rotation et au changement d'échelle. Il permet une meilleure localisation des points d'intérêt que le critère de Harris [MS02], mais détecte moins de points. Il est assez robuste face à du bruit. Son association avec le LoG (ou la trace de la matrice hessienne) permet ainsi de limiter les fausses détections, notamment le long des bords.

2.1.2 Détecteurs de points d'intérêt

Différents détecteurs, basés sur les opérateurs différentiels précédemment présentés, sont définis dans la littérature. Ces détecteurs sont tous invariants par rotation et changement d'échelle. Ils sont composés d'une sélection en espace-échelle, suivie d'un ou plusieurs filtrages des points candidats. Cette deuxième étape est nécessaire pour supprimer les points de bords et les fausses détections liées au bruit. En effet, l'étape de filtrage est souvent dépendante du contraste de l'image, contrairement aux sélections en espace-échelle. On obtient ainsi des points caractérisés par trois paramètres (x, y, σ) , leur position et leur échelle.

2.1.2.1 Points-clés de l'algorithme SIFT

Le détecteur de point-clé de l'algorithme SIFT [Low99, Low04] est un détecteur de «blob» basé sur une pyramide DoG. La pyramide est construite avec les deux approximations présentées précédemment (différence de gaussiennes et sous-échantillonnage). Les extrema en 3D sont sélectionnés. Leur position sous-pixellique est précisée à l'aide d'une interpolation de la pyramide. Les points de faible contraste sont supprimés à l'aide d'un seuil sur la différence de gaussienne. Enfin un critère basé sur la matrice hessienne est utilisé pour filtrer les points de bord :

$$\frac{\text{Tr}(H_\sigma(x, y))^2}{\text{Det}(H_\sigma(x, y))} < t_r \quad (2.17)$$

En définissant λ_1 et λ_2 comme les valeurs propres de $H_\sigma(x, y)$ et r le rapport $\frac{\lambda_1}{\lambda_2}$, on peut écrire :

$$\frac{\text{Tr}(H_\sigma(x, y))^2}{\text{Det}(H_\sigma(x, y))} = \frac{(\lambda_1 + \lambda_2)^2}{\lambda_1 \lambda_2} = \frac{(1 + r)^2}{r} \quad (2.18)$$

Ce rapport atteint sa valeur minimale pour $r = 1$, donc quand les deux valeurs propres sont égales. Comme expliqué précédemment, si λ_1 et λ_2 ont des valeurs proches, la courbure de l'image est importante en ce point. Le critère de l'Equation (2.17), avec $t_r = \frac{(1+r_0)^2}{r_0}$ pour une valeur de r_0 légèrement supérieure à 1, permet donc de supprimer les points de bord.

2.1.2.2 Harris-Laplace

Le détecteur appelé Harris-Laplace [MS01, MS04a] se base sur le critère de Harris multi-échelle. Ce critère est appliqué à l'espace-échelle gaussien et les extrema en espace sont choisis. Les points de faible contraste et de bord sont filtrés avec un seuil sur les valeurs du critère. Un critère de sélection est ensuite appliqué afin de vérifier qu'un point \mathbf{x} a bien été détecté à sa bonne échelle caractéristique : les points pour lesquels le LoG n'atteint

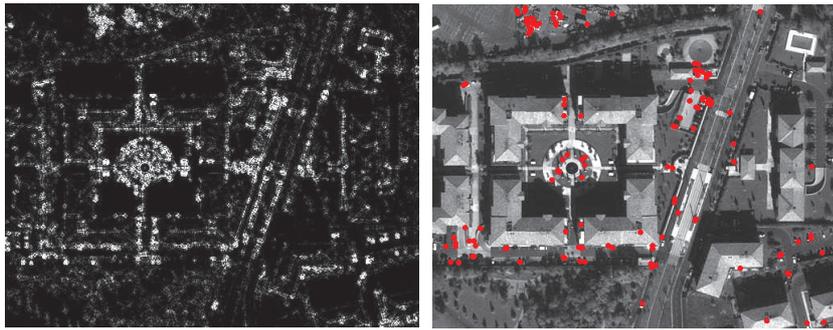
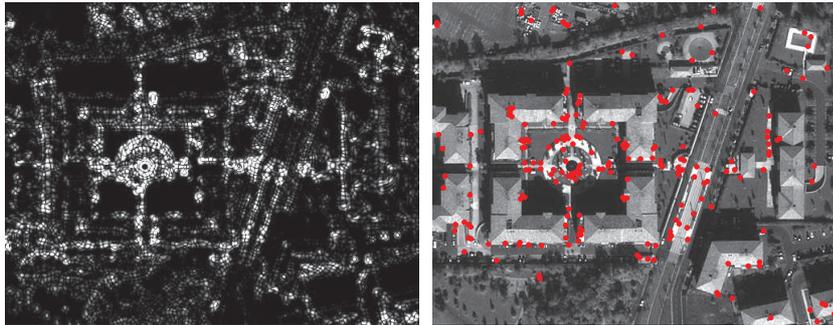
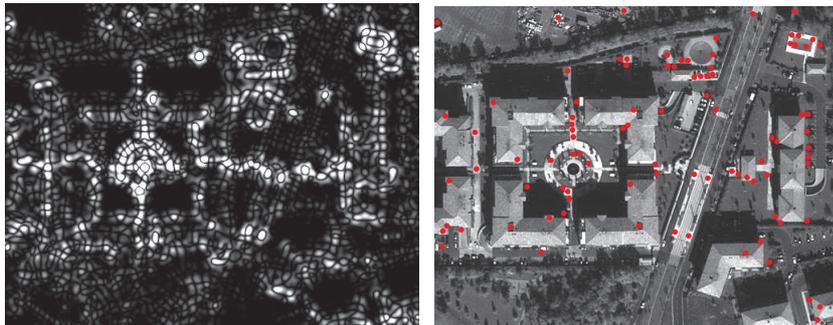
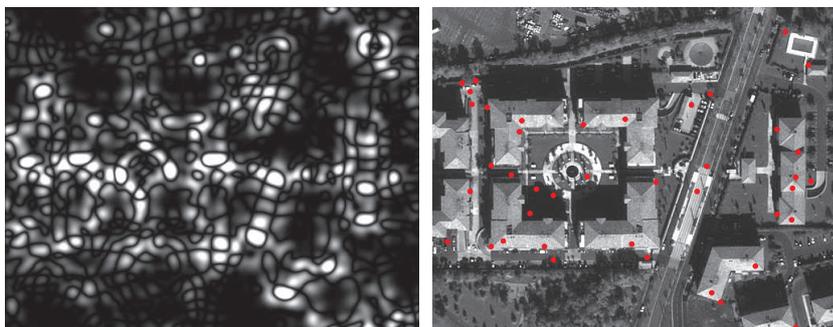
(a) Echelle $\sigma = 1$ (b) Echelle $\sigma = 2$ (c) Echelle $\sigma = 4$ (d) Echelle $\sigma = 8$

FIGURE 2.10 – Exemple de calcul du DoH et sélection des extrema en 3D sur l'image 2.3. Sur chaque ligne est représentée à gauche le hessien et à droite la sélection des extrema locaux en 3D pour une échelle σ . Les extrema locaux en 3D sont sélectionnés en espace et en échelle en 26-connexité. Seuls les points vérifiant $|\text{Det}(H_\sigma(x, y))| < t_{he}$ avec $t_{he} = 0.03$ sont conservés, en considérant que l'image I est normalisée dans l'intervalle $[0, 1]$.

pas un extremum en échelle en \mathbf{x} et est inférieur à un certain seuil sont supprimés. Le Harris-Laplace est un détecteur de coin.

2.1.2.3 Hessian-Laplace

Le détecteur appelé Hessian-Laplace [MS04a] reprend la même idée que le détecteur de Harris-Laplace mais en utilisant l'opérateur différentiel DoH. Les points sont sélectionnés comme extrema en espace du déterminant de la matrice hessienne et extrema en échelle de la trace de la matrice hessienne, qui est égale au LoG. Des seuils sur le déterminant et la trace de la matrice hessienne permettent de supprimer les points de faible contraste. Ce détecteur est un détecteur de «blob».

2.1.2.4 LoG-Harris

Un autre détecteur proposé par [Rab09] associe une sélection des extrema en 3D du LoG (avec seuillage des faibles valeurs) et un filtrage des points de bord à l'aide d'un seuil sur le critère de Harris multi-échelle. Dans la suite, on appellera ce détecteur LoG-Harris.

2.1.2.5 Détecteur Fast-Hessian

Le détecteur Fast-Hessian [BTG06] sélectionne aussi les maxima du hessien. Cependant le calcul de ce déterminant est optimisé : les dérivées secondes sont approximées en utilisant des filtres moyenneurs dont la taille du masque augmente avec l'échelle σ .

2.1.2.6 Détecteur de régions affines

Certains détecteurs inspirés des SIFT ont rajouté une invariance aux changements affines. Le détecteur A-SIFT [YM09,MY09] extrait 6 paramètres pour chaque point-clé au lieu de 4, les deux supplémentaires définissant les angles d'orientation de l'appareil photo. Différentes prises de vue de l'image sont simulées en faisant varier ces paramètres. Les points sont ensuite sélectionnés en appliquant l'algorithme SIFT à ces images. Le Harris-Affine [MS04a] et le Hessien-Affine [MTS⁺05] sont des améliorations des détecteurs Harris-Laplace et Hessian-Laplace. Dans les deux cas, une région elliptique autour du point-clé est estimée à l'aide d'une méthode itérative basée sur la matrice d'auto-corrélation.

2.1.2.7 Exemples de détecteurs non basés sur des opérateurs différentiels

D'autres détecteurs proposés dans la littérature ne reposent pas sur des opérateurs différentiels. Kadir et Brady [KB01] proposent par exemple un détecteur multi-échelle basé sur un calcul de l'entropie sur des voisinages de différentes tailles. Le détecteur SUSAN [SB97] repère les coins ou les bords en observant, sur un masque circulaire, le nombre de pixels avec des valeurs proches du point central. Le détecteur FAST [RD05] reprend le même concept en observant 16 pixels sur un cercle autour du point étudié et en cherchant si au moins 12 pixels contigus ont tous des valeurs supérieures ou inférieures au point central. Enfin certains algorithmes de mise en correspondance de descripteurs sélectionnent les points d'intérêt avec des méthodes plus simples. Dalal et Triggs [DT05] utilisent une grille dense avec un espacement uniforme des points. Belongie et al. [BMP02] choisissent d'utiliser les positions de points de contour obtenus par un détecteur de contour.

2.2 Sélection des orientations

Après avoir sélectionné des points dans un espace-échelle, une ou plusieurs orientations leur sont assignées en observant la géométrie de leur voisinage. Ceci permet ensuite d'assurer l'invariance à la rotation. Plusieurs méthodes sont proposées dans la littérature.

2.2.1 Maxima d'histogramme d'orientation

Lowe [Low04] propose d'observer les orientations du gradient autour du point d'intérêt. Pour chaque échelle σ , l'amplitude et l'orientation du gradient de l'image I_σ sont calculées avec les formules suivantes :

$$g_n(x, y) = \sqrt{(I_\sigma(x+1, y) - I_\sigma(x-1, y))^2 + (I_\sigma(x, y+1) - I_\sigma(x, y-1))^2}$$

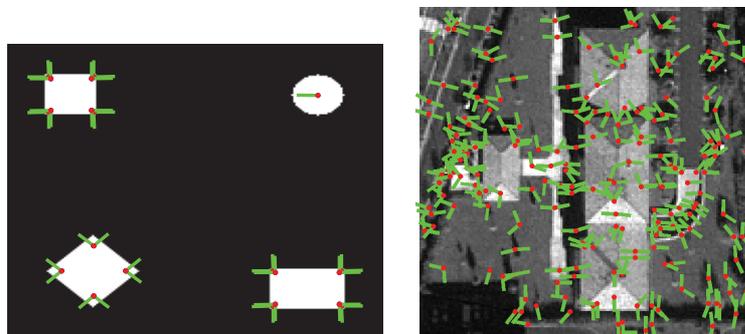
$$g_t(x, y) = \begin{cases} \arctan\left(\frac{I_\sigma(x, y+1) - I_\sigma(x, y-1)}{I_\sigma(x+1, y) - I_\sigma(x-1, y)}\right) \\ \frac{\pi}{2}, \text{ si } I_\sigma(x+1, y) - I_\sigma(x-1, y) = 0. \end{cases} \quad (2.19)$$

Pour chaque point-clé (x, y, σ) , un histogramme d'orientation du gradient g_t est construit sur un voisinage centré en (x, y) et de taille proportionnelle à σ . L'histogramme est quantifié sur 36 classes. Les contributions en chaque pixel sont pondérées par la norme du gradient g_n ainsi que par une fenêtre Gaussienne circulaire de paramètre $\rho = 1.5\sigma$ centrée sur (x, y) . Les orientations choisies correspondent aux maxima locaux de l'histogramme, pour des valeurs supérieures à 80% de la valeur maximale. On obtient ainsi une ou plusieurs orientations θ . Plusieurs points-clés détectés peuvent avoir la même position et la même échelle, mais une orientation différente. Cela permet de considérer les différentes orientations d'une structure, notamment les coins.

2.2.2 Sélection *a contrario* sur histogramme d'orientation

Une autre méthode [Rab09] extrait les modes principaux de l'histogramme d'orientation, en utilisant un algorithme *a contrario* de segmentation d'histogramme [DMM03]. Jusqu'à deux orientations peuvent être sélectionnées par point-clé.

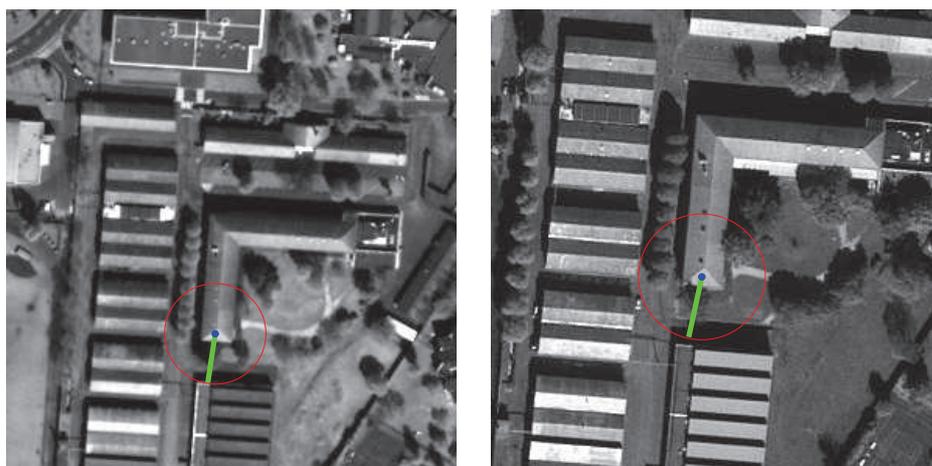
Deux exemples de calcul d'orientations de point-clé sont représentés Figure 2.11.



(a) Image jouet non bruitée de la Figure 2.1

(b) Image satellite optique

FIGURE 2.11 – Exemples de calcul d'orientations. La position des points-clés est représentée par des points rouges, les orientations par des traits verts.



(a) Extrait d'une image du satellite Quick-bird (60 cm de résolution). Ici $\sigma = 4$ et $\theta = 1.7$.
 (b) Extrait d'une image du satellite Geoeye (50 cm de résolution). Ici $\sigma = 5$ et $\theta = 1.8$.

FIGURE 2.12 – Exemple de deux points-clés représentant un même objet sur deux images satellites optiques différentes. La position des points-clés est repérée par un point-bleu, leur orientation θ par un trait vert et leur échelle σ par un cercle rouge (diamètre proportionnel à σ). Les deux images ont ici la même taille, mais pas la même résolution. Les images ne présentent qu'une très faible rotation.

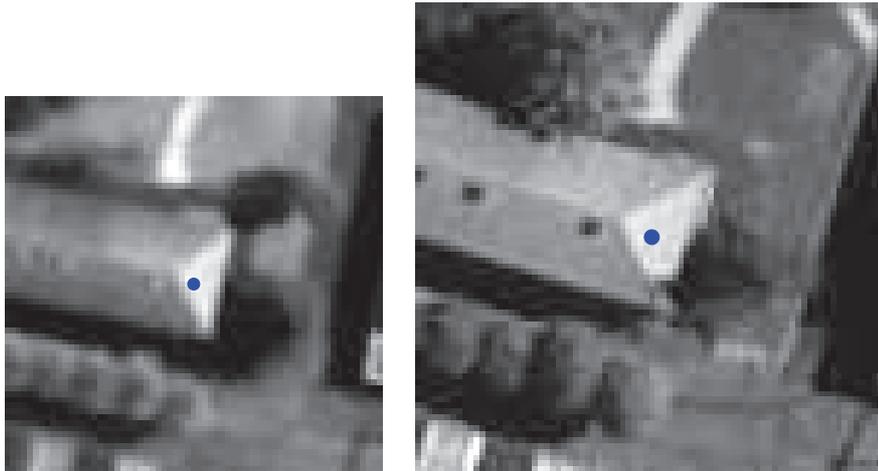
2.2.3 Sélection à l'aide de réponses d'ondelettes de Haar 2D

Bay et al [BTG06] proposent une autre méthode basée sur les ondelettes de Haar. Pour un point-clé (x, y, σ) , les réponses d'ondelettes de Haar 2D pour les directions horizontales et verticales sont calculées sur l'image I_σ sur un voisinage circulaire de rayon 6σ centré autour de (x, y) . La taille de ces ondelettes est σ . Après pondération des réponses par une Gaussienne de paramètre 2.5σ centrée autour du point-clé, les réponses verticales sont représentées en fonction des réponses horizontales. Les sommes des réponses horizontales et verticales sont ensuite calculées à l'intérieur d'une fenêtre glissante couvrant un angle de $\frac{\pi}{3}$. Pour chaque orientation de la fenêtre, un vecteur de deux dimensions est obtenu. L'orientation correspondant au vecteur le plus long est choisie comme orientation principale du point (x, y, σ) .

2.3 Extraction des descripteurs

On obtient ainsi des points-clés caractérisés par quatre paramètres : leur position, leur échelle et leur orientation principale. Chaque point-clé (x, y, σ, θ) définit une région d'intérêt invariante à la translation, rotation et au changement d'échelle. Un exemple d'extraction de point-clé est présenté Figure 2.12 : les deux points sont extraits sur deux images différentes mais représentent une même structure. En effet leur position et leur orientation sont très proches. L'échelle des points, représentée par les cercles rouges, représente la même région. Le rapport entre les échelles $\frac{4}{5} = 0.8$ est en effet à peu près égal à l'inverse du rapport des résolutions $\frac{50}{60} = 0.83$.

La Figure 2.13 présente les régions d'intérêt définies par les deux points-clés précédents. Ce sont des patches, extraits des images de Figure 2.12, centrés autour des positions (x, y) , de tailles proportionnelles à l'échelle σ et ayant subi une rotation de $-\theta$. Ces deux extraits, bien que provenant d'images de résolution et de dynamique différentes, montrent la même



(a) Point-clé de l'image du satellite Quickbird. (b) Point-clé de l'image du satellite Geoeye.

FIGURE 2.13 – Extraction des régions d'intérêt des points-clés de Figure 2.12. La position du point-clé est représentée par un point bleu.

zone d'intérêt. Définir la taille comme proportionnelle à σ permet d'assurer l'invariance au changement d'échelle, et pivoter l'image autour de θ permet d'obtenir une invariance à la rotation. Par la suite nous appellerons ce type de patch «région d'intérêt normalisée».

En comparant ces zones d'intérêt, on peut associer deux points-clés provenant d'images différentes. Différentes méthodes sont proposées dans la littérature pour coder l'information géométrique contenue dans ces zones. On peut utiliser directement les patches extraits autour des points-clés et les comparer à l'aide d'une distance euclidienne, d'une mesure de corrélation ou d'un maximum de vraisemblance. Cependant ces patches peuvent être de grande taille et le temps de calcul est alors important. Ils sont également peu robustes aux petites distorsions locales et trop discriminants : certains détails peuvent être présents seulement sur l'image mieux résolue, les ombres et illuminations peuvent être différentes et un décalage de quelques pixels ne peut être toléré. Ces techniques sont adaptées aux situations où les images sont prises dans les mêmes conditions (même capteur, même résolution). Mais pour des situations plus complexes, elles ne sont pas assez robustes et entraînent de nombreuses erreurs.

De nombreux articles proposent de décrire ces zones d'intérêt de façon compacte et robuste à différentes transformations, tout en étant la plus discriminante possible. Ces régions d'intérêt normalisées sont alors représentées par un vecteur, appelé descripteur, de dimension réduite par rapport aux patches. Le descripteur SIFT [Low99, Low04] est un descripteur très discriminant et présentant de très bonnes performances [MS05]. Il est invariant à la translation, au changement d'échelle et d'illumination et à la rotation. Ce descripteur et quelques variantes sont présentés dans ce qui suit :

2.3.1 SIFT

Le descripteur SIFT [Low99, Low04] propose de représenter la géométrie locale autour des points-clés par des histogrammes d'orientation du gradient. La Figure 2.14 illustre le calcul du descripteur. Pour un point-clé (x, y, σ, θ) , on sélectionne l'image de la pyramide DoG (Section (c)) correspondant à l'échelle σ et on calcule la norme et l'orientation du gradient sur un voisinage autour de (x, y) . Afin d'assurer l'invariance à la rotation, les coordonnées des pixels et les valeurs d'orientations du gradient subissent une rotation de $-\theta$

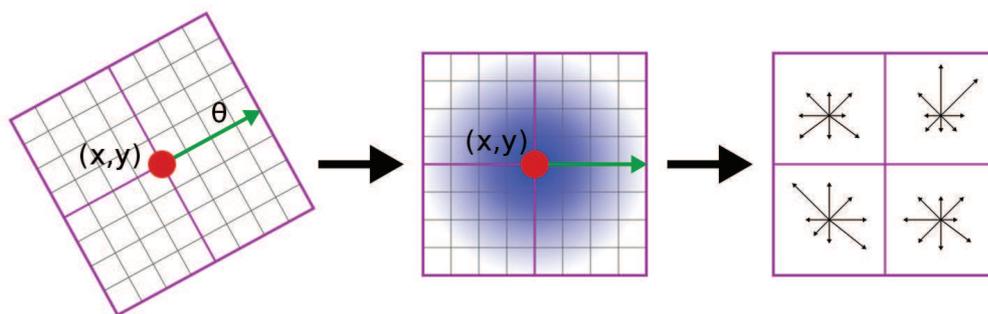


FIGURE 2.14 – Schéma de calcul du descripteur SIFT d'un point-clé (x, y, σ, θ) . L'image I_σ est pivotée pour s'aligner sur θ et un patch centré autour de (x, y) est défini. Ici pour des raisons pratiques on représente (à gauche) un patch de taille 8x8 pixels, divisé en quatre secteurs de 4x4 pixels. Les normes et orientations du gradient sont calculées en chaque pixel, et les valeurs de la norme sont pondérées par une fenêtre Gaussienne (représentée en bleu sur la figure du centre). Des histogrammes d'orientation du gradient sont ensuite construits sur chaque secteur avec 8 directions (à droite).

autour du point-clé. Un patch centré autour de (x, y) est ensuite extrait, de la même façon que les régions d'intérêt normalisées présentées précédemment. La taille des patches est fixe ici (16x16 pixels) mais comme les images de la pyramide DoG sont sous-échantillonnées en fonction de σ , la taille de la zone d'intérêt dépend effectivement de l'échelle caractéristique. Les valeurs de la norme sont pondérées à l'aide d'une fenêtre Gaussienne centrée en (x, y) et de paramètre égal à la moitié de la largeur du patch. Ceci permet de donner plus de poids aux pixels proches du point-clé. Le patch est ensuite divisé en 16 secteurs de taille 4x4 pixels. Sur chacun de ces secteurs, un histogramme d'orientation du gradient pondéré par la norme est construit avec 8 directions. Un schéma du calcul du descripteur SIFT est représenté Figure 2.14. Ces 16 histogrammes sont concaténés en un vecteur, par la suite normalisé à l'unité, que l'on nomme descripteur SIFT.

Cette configuration permet de gérer les effets de bords liés à de faibles erreurs pour la détection de point-clé. En effet si deux points-clés présentent un faible décalage de quelques pixels et/ou une faible erreur d'orientation, l'utilisation d'histogrammes calculés sur un bloc de pixels permet d'atténuer ces imprécisions, contrairement à l'utilisation directe des patches. La normalisation du vecteur est nécessaire pour obtenir une invariance aux changements affines de contraste sur la région d'intérêt. Ce descripteur est donc bien invariant par translation, changement d'échelle, rotation et changement d'illumination localement affine.

2.3.2 HOG

Le descripteur HOG (Histograms of Oriented Gradients) [DT05] est une version simplifiée du descripteur SIFT. Ces descripteurs sont ici calculés sur une grille dense et non sur une sélection de points d'intérêt. Les HOG sont extraits de la même façon que les SIFT, mais ils ne présentent pas d'orientation et sont tous calculés pour une même échelle. La grille de détection est construite de façon à ce que plusieurs descripteurs se superposent. Les régions d'intérêt peuvent avoir une forme carrée comme les SIFT, mais des formes circulaires ou rectangulaires sont aussi proposées. Ce détecteur est très efficace pour la détection de piétons [DT05].

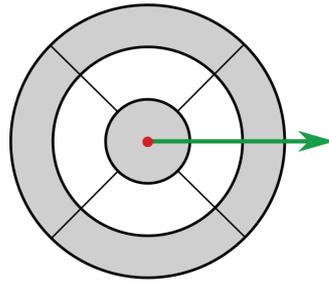


FIGURE 2.15 – Masque du descripteur SIFT-circulaire, découpé en 9 régions.

2.3.3 PCA-SIFT

Le descripteur PCA-SIFT [KS04] est très proche de SIFT mais présente une plus petite dimensionalité. La sélection des patchs est la même que pour SIFT, mais avec une taille de 41×41 pixels. Ensuite pour chaque patch est calculé le gradient horizontal et vertical. On obtient ainsi un vecteur de dimension $2 \times 39 \times 39 = 3042$, normalisé pour une invariance à l'illumination. La méthode d'analyse en composante principale permet ensuite de réduire la dimension du vecteur à 20. Le vecteur obtenu est appelé PCA-SIFT. C'est un descripteur très discriminant [MS05] malgré sa faible dimension.

2.3.4 GLOH

Le descripteur GLOH (Gradient location-orientation histogram) [MS05] est une amélioration de SIFT plus robuste et discriminante. La région d'intérêt normalisée est ici circulaire et divisée en 17 secteurs avec une grille de positions log-polaires, plutôt qu'en sous-régions carrées. Les histogrammes d'orientation sont calculés sur chaque secteur comme pour SIFT, mais avec 16 directions. La dimension du vecteur obtenu est réduite de 272 à 128 ou 64 avec une PCA.

2.3.5 SIFT-circulaire

Pour ce descripteur [Rab09], la région d'intérêt normalisée est définie comme un masque circulaire (rayon de 12σ) avec une division log-polaire (Figure 2.15), comme le GLOH. Les images de gradient sont calculées sur l'image $I_{0.4\sigma}$. En effet, il a été observé que sur l'image I_σ la plupart des détails ont été effacés par le flou Gaussien et les histogrammes sont plats. Par contre il n'y a pas ici de pondération Gaussienne pour la construction des histogrammes. Du fait de la géométrie de ce descripteur, l'intérêt de cette pondération est très limité. Comme pour le descripteur SIFT, les histogrammes d'orientation du gradient sont ensuite calculés sur chaque secteur, puis concaténés en un vecteur normalisé. On appellera dans la suite ce descripteur SIFT-circulaire.

Un exemple d'extraction du descripteur SIFT-circulaire est présenté sur la Figure 2.16 pour trois points-clés différents. On remarque que les descripteurs des deux premiers points-clés, représentant la même structure, sont assez proches, alors qu'ils sont assez différents du descripteur du dernier point-clé, présentant un autre objet. La distance euclidienne entre les deux premiers descripteurs est égale à 0.095, alors que celle entre le premier et dernier descripteur est égale à 0.21.

2.3.6 SURF

Le descripteur SURF (Speeded Up Robust Features) [BTG06] est inspiré du descripteur SIFT, mais présente un temps de calcul optimisé. Il repose sur un calcul d'ondelettes de

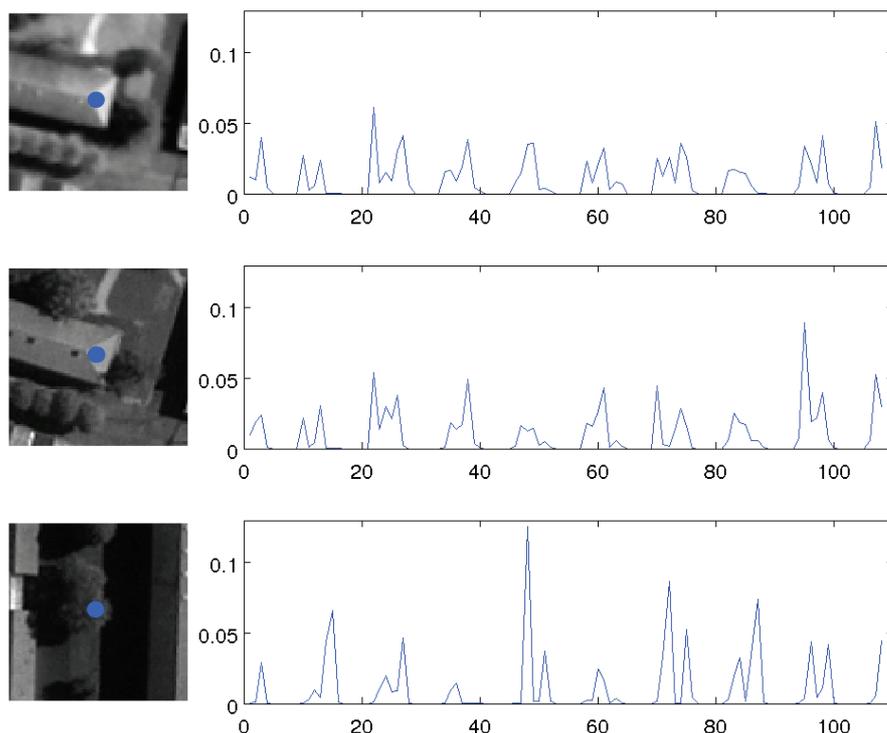


FIGURE 2.16 – Exemple d'extraction de descripteurs pour 3 points-clés différents. À gauche sont présentées les régions d'intérêt normalisées de ces points-clés (les points bleu représentent leur position), à droite les descripteurs SIFT-circulaires correspondant. Ces descripteurs sont calculés sur 9 secteurs et avec des histogrammes de 12 directions. Les deux premiers points-clés sont extraits de deux images satellites différentes mais présentent une même structure, les deux derniers proviennent de la même image mais représentent des structures différentes.

Haar plutôt que sur un calcul du gradient.

Pour chaque point-clé, une région d'intérêt normalisée, du même type que celles de la Figure 2.13 et de taille 10σ , est extraite sur I . Les réponses en ondelettes de Haar 2D pour la direction horizontale, D_x , et la direction verticale, D_y , sont calculées avec une taille d'ondelette de 2σ . Les réponses D_x et D_y sont pondérées par une Gaussienne centrée en (x, y) et de paramètre 3.3σ . La région d'intérêt est ensuite divisée en 4×4 sous-régions. Sur chacune d'elles est extrait un vecteur de 4 dimensions : $\mathbf{v} = (\sum D_x, \sum D_y, \sum |D_x|, \sum |D_y|)$. Tous ces vecteurs sont rassemblés dans un même vecteur de dimensionnalité 64. Ce vecteur est ensuite normalisé à l'unité pour obtenir l'invariance au contraste. Le vecteur obtenu est appelé descripteur SURF.

2.4 Mise en correspondance

On s'intéresse maintenant à la mise en correspondance de deux images I_1 et I_2 afin de déterminer si un ou plusieurs objets sont communs aux deux images. Pour cela, des points-clés $\alpha^{i=1..N_1}$ et $\beta^{j=1..N_2}$ sont extraits respectivement sur les images I_1 et I_2 . A chaque point-clé α^i (ou β^j) est ensuite associé un descripteur a^i (ou b^j) décrivant la géométrie locale. En

associant les descripteurs, on peut appairer des points-clés provenant des deux images. Du fait de leur invariance à différentes transformations, des descripteurs représentant un même objet mais avec des tailles, orientation et/ou positions différentes devraient être semblables. Une mesure de dissimilarité entre descripteurs associée à un critère d'appariement permet donc d'appairer les points-clés de deux images.

Différentes mesures de dissimilarité peuvent être utilisées, les plus courantes étant la distance euclidienne, la distance L1 ou la distance L1 cumulée. Concernant les critères d'appariement, deux méthodes sont présentées dans ce qui suit.

2.4.1 Méthode NNDR

La méthode Nearest Neighbord Direct Ratio (NNDR) [Low04] utilise la restriction au plus proche voisin et un critère de suppression des fausses mises en correspondance.

Cette méthode suit deux étapes. Premièrement, pour chaque descripteur a^i , on calcule la distance avec tous les descripteurs b^j . Le descripteur a^i est ensuite apparié à b^{j_1} , le descripteur présentant la plus faible distance $\|b^{j_1} - a^i\|$. Cette étape est la restriction au plus proche voisin (NN : Nearest Neighbor). Dans un second temps, la correspondance $a^i \longleftrightarrow b^{j_1}$ est conservée si le rapport $\frac{\|b^{j_1} - a^i\|}{\|b^{j_2} - a^i\|}$ est inférieur à un seuil th , b^{j_2} étant le deuxième plus proche voisin de a^i . L'idée est que, s'il n'y a pas de descripteur correspondant à a^i dans I_2 , ses deux plus proches voisins sont dus au hasard et les distances $\|b^{j_1} - a^i\|$ et $\|b^{j_2} - a^i\|$ devraient être proches. Cette étape est le filtrage du rapport des distances (DR : Distance Ratio).

Ce filtrage repose sur l'hypothèse qu'un descripteur a^i ne peut avoir qu'un seul descripteur correspondant dans l'image I_2 . En effet, dans le cas où l'on cherche à reconnaître deux instances d'un même objet, le descripteur a^i peut avoir deux descripteurs correspondants dans I_2 , soit idéalement ses deux plus proches voisins b^{j_1} et b^{j_2} . Les distances $\|b^{j_1} - a^i\|$ et $\|b^{j_2} - a^i\|$ étant logiquement proches, leur ratio sera proche de 1. La mise en correspondance $a^i \longleftrightarrow b^{j_1}$ ne sera alors pas conservée, de même que a^i ne pourra être associé à b^{j_2} . Cette méthode, bien qu'efficace, n'autorise donc pas l'association d'un même descripteur a^i à plusieurs descripteurs b^j . En particulier, elle ne permet pas de reconnaître plusieurs instances d'une même structure.

La Figure 2.17 présente un exemple de mise en correspondance de points-clés utilisant ce critère. Le patch extrait de la Figure 2.12(a) est correctement reconnu dans la Figure 2.12(b). On remarque que les points-clés présentés Figure 2.12 sont correctement mis en correspondance.

2.4.2 Méthode *a contrario*

Cette méthode [RDG09], basée sur une approche *a contrario* [DMM07], permet de calculer automatiquement des seuils adaptatifs de mise en correspondance. Elle autorise également la mise en correspondance de plusieurs occurrences d'un même objet.

A chaque descripteur a^i de I_1 , on souhaite associer un, plusieurs ou aucun descripteurs b^j de I_2 . Pour comparer deux descripteurs, on suppose que la mesure de dissimilarité $D(a^i, b^j)$ s'écrit :

$$D(a^i, b^j) = \sum_{m=1}^M d_m(a_m^i, b_m^j) \quad (2.20)$$

avec a_m^i et b_m^j les histogrammes du secteur m des descripteurs a^i et b^j , et d_m une distance. M est le nombre total de secteurs. L'idée de l'algorithme *a contrario* est de limiter le nombre d'erreurs en calculant la probabilité de fausses alarmes de chaque mise en correspondance.

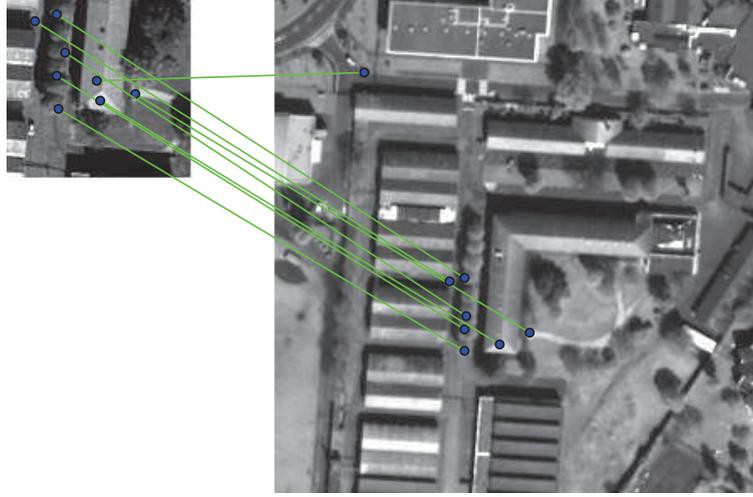


FIGURE 2.17 – Exemple de mise en correspondance entre deux images (extraits de Figure 2.12). Le seuil th est fixé à 0.8 pour le filtrage des mises en correspondance. 8 mises en correspondance sont obtenues, avec seulement une erreur.

On définit l'hypothèse \mathcal{H}_0 suivante : pour chaque descripteur a^i fixé et quelque soit le descripteur \mathbf{b} de I_2 , les distances $d_m(a_m^i, \mathbf{b}_m)$ sont des variables indépendantes.

Soit la probabilité sous \mathcal{H}_0 que $D(a^i, \mathbf{b}) \leq \alpha$:

$$P_{\mathcal{H}_0}(D(a^i, \mathbf{b}) \leq \alpha) = P_{\mathcal{H}_0} \left(\sum_{m=1}^9 d_m(a_m^i, \mathbf{b}_m) \leq \alpha \right) \quad (2.21)$$

Soit p_m la densité de la distribution de $d_m(a_m^i, \mathbf{b}_m)$. Alors sous \mathcal{H}_0 , la densité de $D(a^i, \mathbf{b})$ est $p_1 * \dots * p_M$. On peut donc calculer :

$$F_i(\alpha) := P_{\mathcal{H}_0}(D(a^i, \mathbf{b}) \leq \alpha) = \int_0^\alpha p_1 * \dots * p_M(x) dx \quad (2.22)$$

On apprend p_m empiriquement en faisant varier \mathbf{b} dans l'ensemble $\{b^1, b^2, \dots, b^j, \dots, b^{N_2}\}$. Les nombres N_1 et N_2 correspondent aux nombre de points-clés des images I_1 et I_2 . En construisant l'histogramme des valeurs obtenues, on a donc la distribution empirique des distances $d_m^i(a_m^i, b_m^j)$ quand j parcourt $\{1, \dots, N_2\}$.

On obtient ainsi une fonction $F_i(\alpha)$ croissante de α , et on définit le «nombre de fausses alarmes» pour deux descripteurs a^i et b^j comme :

$$NFA(a^i, b^j) := N_1 N_2 F_i(D(a^i, b^j)) \quad (2.23)$$

Deux descripteurs a^i et b^j sont alors mis en correspondance si $NFA(a^i, b^j)$ est inférieur à un seuil ε . Ce seuil permet d'assurer que, sous l'hypothèse \mathcal{H}_0 , le nombre moyen de fausses mises en correspondance, lorsque l'on compare tous les descripteurs de I_1 à ceux de I_2 , est inférieur à ε [RDG09]. Une distance maximale entre deux descripteurs est cependant prise en compte pour supprimer rapidement des mises en correspondances aberrantes afin de limiter le temps de calcul.

Un exemple de mise en correspondance est présenté Figure 2.18. On remarque que certains points-clés de l'image de gauche sont mis en correspondance avec plusieurs points de l'image de droite. Ainsi, les six occurrences d'un même modèle de bâtiment sont reconnues.

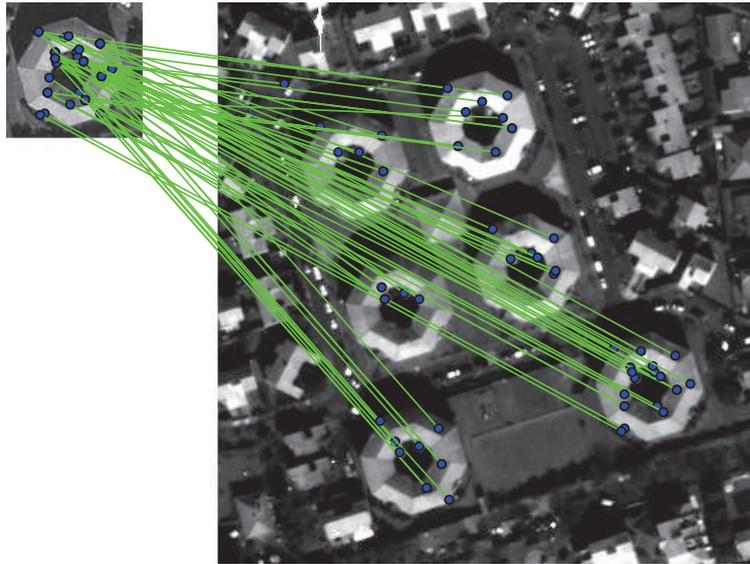


FIGURE 2.18 – Exemple de mise en correspondance *a contrario* entre deux images. Les deux extraits proviennent d'images satellites optiques de même résolution (50cm) mais de capteurs différents (Geoeye à gauche et Worldview à droite). Le seuil ε est ici fixé à 0.8. 28 points-clés de l'image de gauche sont mis en correspondance avec 69 de l'image de droite.

2.5 Applications de l'algorithme SIFT en télédétection

L'algorithme SIFT présenté est un algorithme efficace pour la détection et la reconnaissance de points d'intérêt [MS05]. Ses invariances à la translation, rotation, changement d'échelle, ainsi que sa robustesse aux changements d'illumination et aux transformations affines en font un choix intéressant pour les images de télédétection. Cet algorithme a déjà été appliqué à des images optiques, celles-ci ont en effet des caractéristiques similaires aux images naturelles. Des exemples d'applications sont présentés dans ce qui suit.

Recalage Plusieurs méthodes [LWLC09, HPHZ12, SME11] proposent d'utiliser les points-clés SIFT comme points de contrôle afin d'estimer une déformation. Li et al. [LWLC09] prennent en compte la spécificité des images de télédétection en considérant que la déformation entre deux images peut être limitée à une translation, une rotation et un changement d'échelle. Ils introduisent alors un nouveau critère de mise en correspondance avec des restrictions sur l'échelle et l'orientation des points. Huo et al. [HPHZ12] proposent une amélioration de cette méthode en ajoutant une approche multi-échelle et l'utilisation de l'algorithme RANdom SAMple Consensus (RANSAC). Sedaghat et al. [SME11] adaptent l'algorithme SIFT pour obtenir des points-clés avec une distribution spatiale uniforme et filtrent les mauvaises mises en correspondance en appliquant un modèle projectif. Joglekar et al. [JGKM14] utilisent les descripteurs SIFT pour estimer la matrice épipolaire entre deux images stéréo.

Classification Yang et Newsam [YN13] représentent les descripteurs SIFT sous forme de *sac de mots*, ou *bag of words* (BoW), afin de classifier des images entre occupation et utilisation du sol (land-use/land-cover). Risojevic and Babic [RB13] reprennent la même idée mais effectuent une fusion entre les BoW des descripteurs SIFT et un descripteur de texture basé sur les filtres de Gabor.

Détection d'objets Sirmacek et Unsalan [SU09] réalisent la détection de bâtiments à l'aide des points-clés SIFT et de la théorie des graphes. Les zones urbaines sont extraites à l'aide d'une méthode de mise en correspondance de graphes multiples. Les bâtiments sont ensuite sélectionnés séparément avec une méthode de graph cut. Tao et al. [TTCT11] effectuent la détection d'aéroports en combinant un clustering effectué sur les points-clés SIFT et une segmentation de l'image.

Détection de changements Li et al. [LXWH06] proposent de détecter des changements en zone urbaine en comparant les lignes extraites sur des images pré-recalées. Les erreurs liées à la parallaxe et aux ombres sont supprimées à l'aide de l'algorithme SIFT. L'algorithme de détection de changements GeoCDX [KCS⁺13] utilise les SIFT pour recalculer des images multi-capteurs et multi-résolutions tout en mesurant une erreur moyenne de recalage. Les changements sont détectés en comparant des caractéristiques (entropie, coefficient de dissymétrie et lignes), tout en considérant une tolérance spatiale.

2.6 Conclusions

L'algorithme SIFT est donc pertinent pour étudier la détection de changement entre images optiques. Il présente des invariances (rotation, translation et changement d'échelle) et robustesses (changement d'illumination et transformations affines) intéressantes pour la comparaison d'images optiques haute résolution. Ses performances sont cependant très faibles pour les images radar. La présence de speckle perturbe en effet les traitements. Des améliorations ont été proposées pour l'adapter, mais restent limitées (Section 3.3). Considérant le large champ d'application offert par l'algorithme SIFT en imagerie optique, il serait particulièrement intéressant d'obtenir un algorithme de type-SIFT performant et adapté aux images radar.

Le prochain chapitre présentera les limites de l'algorithme SIFT sur les images bruitées (bruit gaussien et speckle), puis détaillera les propositions d'adaptation des SIFT aux images radar présentes dans la littérature. Dans le reste du manuscrit, nous utiliserons comme algorithme de type SIFT l'implémentation décrite dans l'Annexe B.

Chapitre 3

Limites de l'algorithme SIFT sur les images bruitées

L'algorithme SIFT, présenté dans le chapitre précédent, voit ses performances diminuer sur les images bruitées. Afin d'adapter l'algorithme aux images radar, nous souhaitons mettre en évidence les étapes affectées par le bruit. Pour cela nous étudierons l'influence du bruit (gaussien et speckle) sur la détection de points-clés puis sur le calcul du gradient (nécessaire à l'assignation d'orientation et l'extraction de descripteurs). Enfin un état de l'art sur les adaptations de l'algorithme SIFT aux images radar sera présenté.

3.1 Détection de points-clés

3.1.1 Observations

Dans un premier temps, nous étudions l'influence du bruit gaussien sur le nombre de points-clés détectés. Pour cela, nous ajoutons du bruit gaussien sur quatre images numériques (Figure 3.1) et détectons les points-clés avec la méthode LoG-Harris. On suppose que le bruit originel de ces images est négligeable. L'implémentation décrite en Annexe B est utilisée, mais sans la pondération du contraste de l'image pour le critère de Harris. Nous observons, pour chaque écart-type ς de bruit gaussien ajouté, le nombre total de points-clés détectés sur les quatre images pour une échelle σ donnée. La Figure 3.3 présente l'évolution de ce nombre de points-clés en fonction de l'échelle σ des points-clés puis en fonction de l'écart-type ς du bruit.

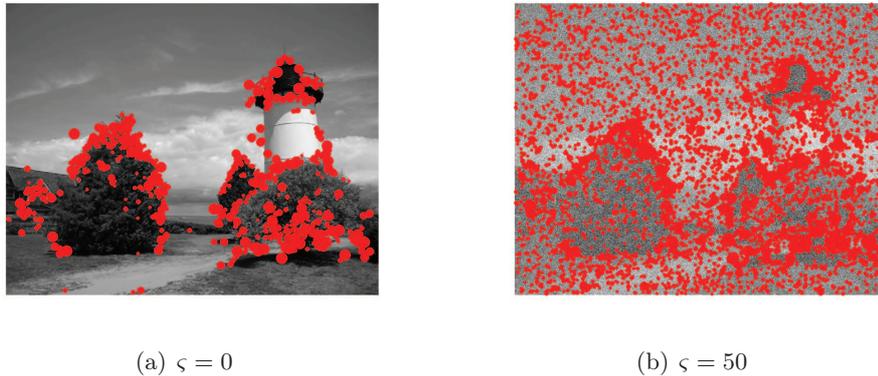
On remarque que les échelles ς supérieures à 2 ne sont pas affectées par la présence de bruit, contrairement aux petites échelles. En effet, les quatre premières échelles présentent une augmentation croissante du nombre de points-clés, allant de pair avec l'augmentation de ς . Ceci est encore plus marqué pour les échelles à 0.8 et 1, où ce nombre est multiplié par respectivement 7 et 6. On peut noter que les variances de bruit utilisées sont fortes, mais ceci est à mettre en perspective avec le fort bruit de speckle des images radar.

Les détections sur la version bruitée et non bruitée d'une image sont présentées Figure 3.2. On remarque que sur l'image non bruitée on ne trouve des points que sur les objets (phare, arbres). Cependant sur l'image fortement bruitée, il y a des détections sur l'ensemble de l'image : objets mais aussi zones homogènes (ciel, herbe). On trouve donc de nombreuses fausses détections. Celles-ci semblent être réparties spatialement de façon uniforme.

L'ajout de bruit gaussien entraîne donc une explosion du nombre de détections sur l'ensemble de l'image. Celles-ci se concentrent surtout dans les petites échelles.



FIGURE 3.1 – Exemples d'images numériques

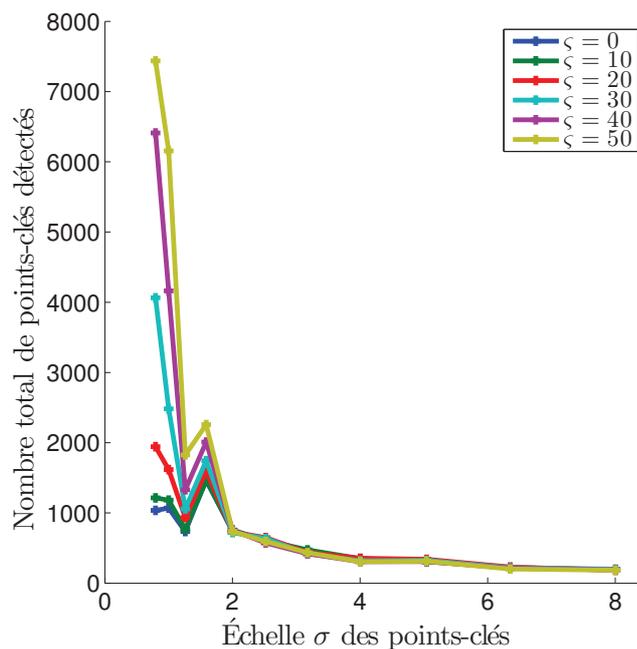
(a) $\varsigma = 0$ (b) $\varsigma = 50$ FIGURE 3.2 – Détection de points-clés sur une image non bruitée (à gauche) et bruitée (à droite) avec un bruit gaussien centré d'écart-type ς . Les points-clés sont représentés en rouge, leur taille est proportionnelle à leur échelle σ .

3.1.2 Remarques théoriques

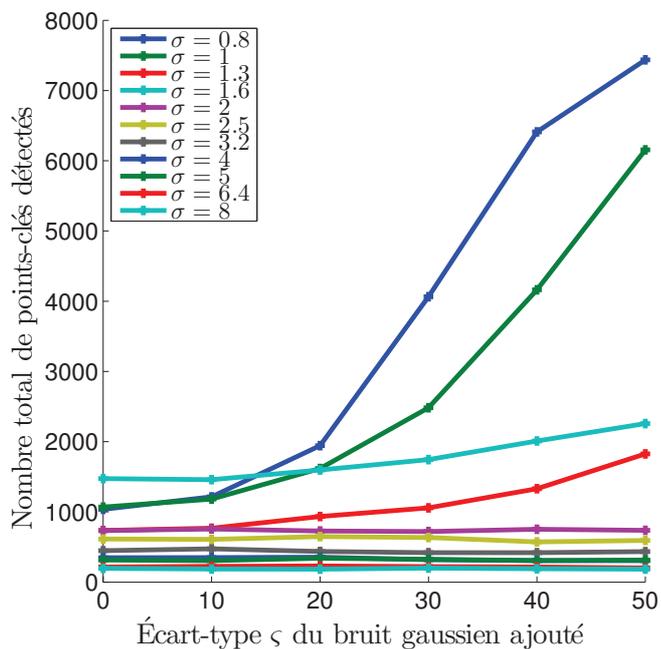
Les détecteurs de points-clés présentés en Section 2.1 présentent tous la même construction : une sélection d'extrema locaux à l'aide d'un opérateur différentiel appliqué à un espace-échelle, puis un ou plusieurs filtrages des points candidats. Nous étudions de façon théorique l'influence du bruit sur ces deux étapes.

Opérateurs appliqués à un espace-échelle Les points-clés sont extraits à partir des extrema en espace et/ou en échelle de ces opérateurs (LoG, trace et déterminant de Harris, critère de Harris). Comme présenté en Section 2.1, la sélection de ces extrema est indépendante du contraste. Sur une zone homogène bruitée, le pourcentage de fausses détections sera donc le même quelque soit l'intensité du bruit et le contraste de l'image. Par exemple pour la méthode LoG-Harris, on mesure expérimentalement que ce pourcentage est autour de 3.5% pour du bruit gaussien et de 3.1% pour du speckle. La Figure 3.4 présente la répartition selon les échelles du nombre d'extrema du LoG en 3D pour une image ne présentant qu'un bruit gaussien. Les points extraits n'ont pas été filtrés. On peut noter que le nombre d'extrema détectés diminue de façon inversement proportionnelle à l'échelle.

Sur les images naturelles, les zones homogènes ne sont jamais idéales et il y a toujours un faible bruit (lié à la texture, au capteur photographique...). Quelque soit l'intensité du bruit, il y aura donc toujours des fausses détections sur les zones homogènes bruitées. Cependant comme on peut l'observer sur la Figure 3.2(a), il n'y a pas de détections sur les zones homogènes. Ces points ont donc été probablement supprimés par l'étape de filtrage.



(a) Nombre de points-clés en fonction de l'échelle σ . Les couleurs correspondent à différentes valeurs de l'écart-type ζ du bruit.



(b) Nombre de points-clés en fonction de l'écart-type ζ du bruit. Les couleurs correspondent aux différentes échelles σ des points-clés.

FIGURE 3.3 – Les deux courbes présentent l'évolution du nombre de points-clés détectés par la méthode LoG-Harris sur les images de la Figure 3.1 auxquelles a été ajouté du bruit gaussien. Cette évolution est présentée en fonction de l'échelle σ des points, puis en fonction de l'écart-type ζ du bruit. Ces nombres sont obtenus en additionnant le nombre de points-clés détectés sur les quatre images.

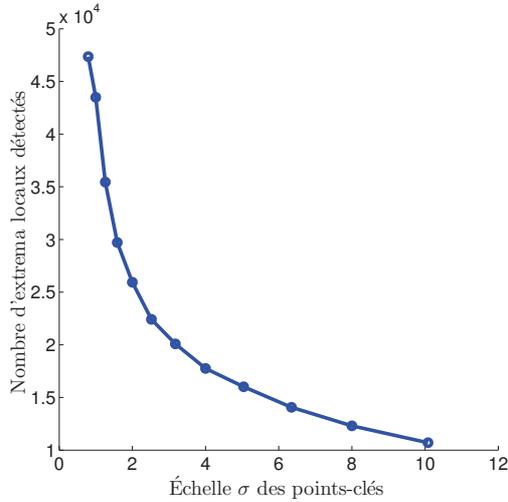


FIGURE 3.4 – Répartition en échelle des extrema en 3D du LoG d'une image de taille 1000×1000 ne présentant qu'un bruit gaussien de moyenne $m = 100$ et d'écart-type $\varsigma = 5$.

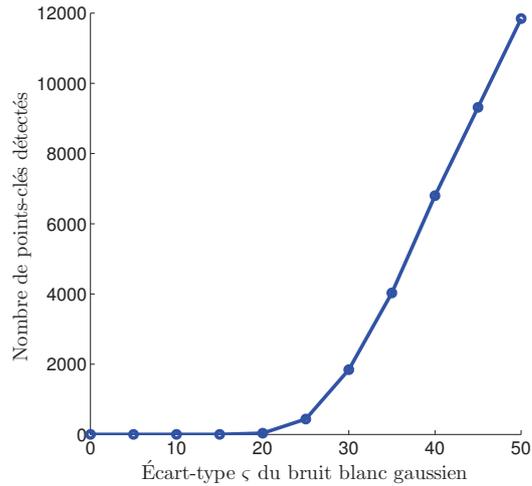
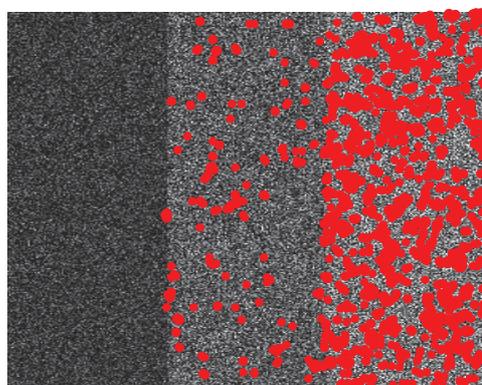
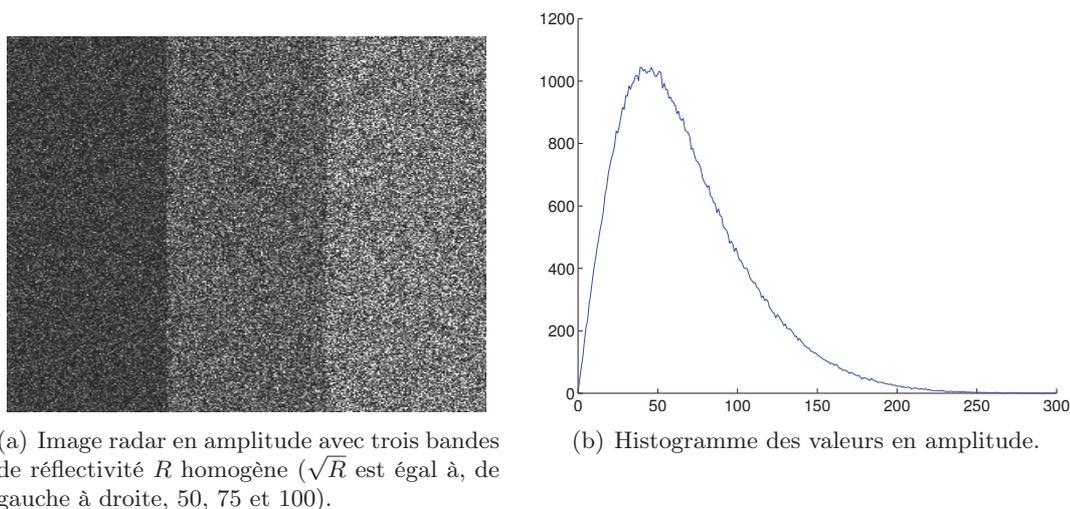


FIGURE 3.5 – Nombre de points-clés détectés sur une image homogène de moyenne $m = 100$ et de taille 1000×1000 en fonction de l'écart-type ς du bruit gaussien ajouté. Les points-clés détectés sont des extrema en 3D du LoG, suivi d'un filtrage à l'aide du critère de Harris (seuil $t_h = 2000$).

Filtrage des points-clés Après la sélection de points candidats, un ou plusieurs filtrages sont nécessaires afin de supprimer les points de bords, mais aussi les fausses détections liées au bruit. Ces filtrages impliquent toujours un seuil sur un critère dépendant du contraste local de l'image (critère de Harris, seuil sur le LoG ou DoH). Pour un faible niveau de bruit (cas de la plupart des images numériques), la majorité des erreurs sont filtrées avec un seuil raisonnable. Ceci explique l'absence de fausses détections sur les zones homogènes pour la plupart des images numériques (comme sur la Figure 3.2(a)). En cas de fort bruit, un seuil important sera cependant nécessaire. Il faut donc trouver un compromis entre supprimer les erreurs liées au bruit et garder des points intéressants de faible contraste. On peut remarquer sur la Figure 3.3 que le seuil utilisé est robuste jusqu'à un bruit gaussien d'écart-type autour de 10-20. Au delà, le seuil n'est pas suffisant. L'augmentation du nombre de points-clés se produit en premier pour les petites échelles, puisque c'est à ce niveau que se produisent le plus de fausses détections (Figure 3.4).

La Figure 3.5 illustre la robustesse du critère de Harris face au bruit. Des points-clés sont extraits sur une image présentant un bruit gaussien de moyenne $m = 100$ et différentes valeurs d'écart-type ς . La sélection se fait avec les extrema en 3D du LoG, suivi d'un filtrage sur le critère de Harris ($t_h = 2000$). On observe l'évolution du nombre de points-clés détectés en fonction de ς . Jusqu'à $\varsigma = 20$, aucun point n'est détecté. Puis le nombre de points-clés augmente rapidement. Le critère de Harris dépend donc bien du contraste local de l'image. Le filtrage avec un seuil $t_h = 2000$ est robuste face à un bruit blanc gaussien, mais tant que l'écart-type du bruit est inférieur à 20.

Deux raisons expliquent l'augmentation des détections pour les petites échelles observée en Section 3.4. Premièrement, comme présenté sur la Figure 3.4, les fausses détections parmi les extrema 3D du LoG ont lieu principalement aux petites échelles. Or, le critère de Harris est plus efficace pour les grandes échelles que pour les petites. En effet, l'ajout de flou gaussien d'écart-type σ à une image présentant un bruit gaussien d'écart-type ς permet de réduire l'écart-type du bruit résiduel à $\nu = \frac{\varsigma^2}{4\pi\sigma^2}$ (Annexe C.1). Plus le flou gaussien



(c) Détection de points-clés avec la méthode LoG-Harris (seuil $t_h = 2000$).

FIGURE 3.6 – Exemple de détection de points-clés et d’histogramme des valeurs en amplitude calculées sur une image radar.

est important, plus l’écart-type ν diminue. Pour les grandes échelles, le bruit résiduel est faible et donc les fausses détections sont plus facilement filtrées par le critère de Harris.

Cas des images radar Les images radar sont fortement perturbées par du bruit de type multiplicatif. Si on considère que les valeurs d’une image suivent une loi Gamma en intensité, l’écart-type d’une zone homogène est égal à la réflectivité R de la zone étudiée. En amplitude, les valeurs suivent alors une loi de Rayleigh avec comme écart-type $\sqrt{\frac{R\pi}{4}}$. Plus une zone aura une réflectivité importante, plus le bruit sera important. Il n’est donc pas aisé de fixer un seuil pour le filtrage des erreurs.

La Figure 3.6 présente un histogramme des valeurs en amplitude ainsi que la détection de points-clés d’une image jouet. L’image étudiée présente trois zones homogènes mais avec des valeurs de réflectivité différentes. Sur l’histogramme, on remarque bien que les valeurs sont fortement étalées par rapport aux valeurs de réflectivité. On observe beaucoup de fausses détections de points-clés, mais avec une répartition non uniforme. En effet, il y a plus de fausses détections sur la bande de forte réflectivité que sur les deux autres bandes. Il n’y a d’ailleurs aucune fausse détection sur la bande de faible réflectivité. Le taux de fausses détections n’est donc pas constant.

A cause du caractère multiplicatif du speckle, les détecteurs de points-clés présentés en

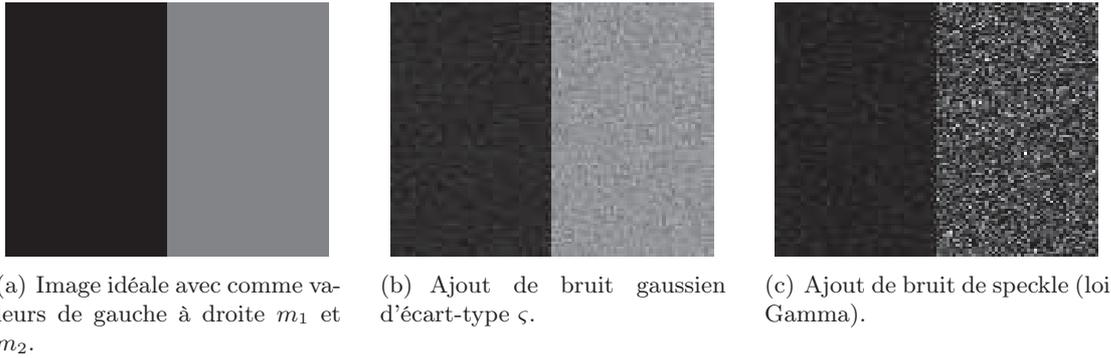


FIGURE 3.7 – Modèle de bord.

Section 2.1 ne sont pas adaptés aux images radar. Ils entraînent un taux de fausse détection non constant. Il est donc nécessaire d'étudier les caractéristiques du bruit de speckle pour définir un nouveau détecteur de points-clés adapté aux images radar.

3.2 Calcul du gradient

Les étapes d'assignation d'orientation et d'extraction de descripteurs se basent principalement sur le calcul du gradient. Hors en cas de fort bruit gaussien ou de speckle, ce calcul est perturbé. Prenons l'exemple du calcul du gradient sur un bord vertical (Figure 3.7).

3.2.1 Bruit gaussien

Soit une image I idéale d'un bord vertical séparant deux bandes de valeurs m_1 et m_2 (Figure 3.7(a)). On ajoute à I un bruit gaussien $B \sim \mathcal{N}(0, \varsigma^2)$ (Figure 3.7(b)). On choisit de calculer les valeurs de gradients horizontaux et verticaux d'une image I avec les formules suivantes :

$$g_x(x, y) = \frac{I(x+1, y) - I(x-1, y)}{2}$$

$$g_y(x, y) = \frac{I(x, y+1) - I(x, y-1)}{2}.$$

Ce calcul de gradient est appelé *gradient par différence*. On obtient les valeurs de normes et d'orientation du gradient avec :

$$g_m(x, y) = \sqrt{g_x(x, y)^2 + g_y(x, y)^2}$$

$$\theta(x, y) = \arctan\left(\frac{g_y(x, y)}{g_x(x, y)}\right).$$

Les composantes verticales et horizontales du gradient suivent respectivement les lois normales $\mathcal{N}(\frac{m_2-m_1}{2}, \frac{\varsigma^2}{2})$ et $\mathcal{N}(0, \frac{\varsigma^2}{2})$. On observe la distribution des valeurs de normes g_m et d'orientation g_t sur la ligne de bord pour différentes valeurs de m_1 , m_2 et ς (Figure 3.8). On remarque que les valeurs de normes et d'orientations semblent suivre une loi unimodale. De façon empirique, on calcule que les valeurs de normes ont pour moyenne $n_m = \frac{|m_1-m_2|}{2}$ et pour écart-type $\varsigma_m = \frac{\varsigma}{2}$. Pour les valeurs d'orientations, les lois sont centrées en zéro et l'écart-type est égal à $\sigma_t = \frac{\varsigma\sqrt{2}}{|m_2-m_1|}$. La distribution empirique des valeurs d'orientation est conforme à la densité de probabilité calculée théoriquement en Annexe C.2 :

$$f_\theta(t) = \frac{\exp(-\frac{m^2}{4\varsigma^2})}{2\pi} + \frac{m \cos t}{4\sqrt{\pi}\varsigma} \exp\left(\frac{-m^2 \sin^2 t}{4\varsigma^2}\right) \left(1 + \operatorname{erf}\left(\frac{m \cos t}{2\varsigma}\right)\right),$$

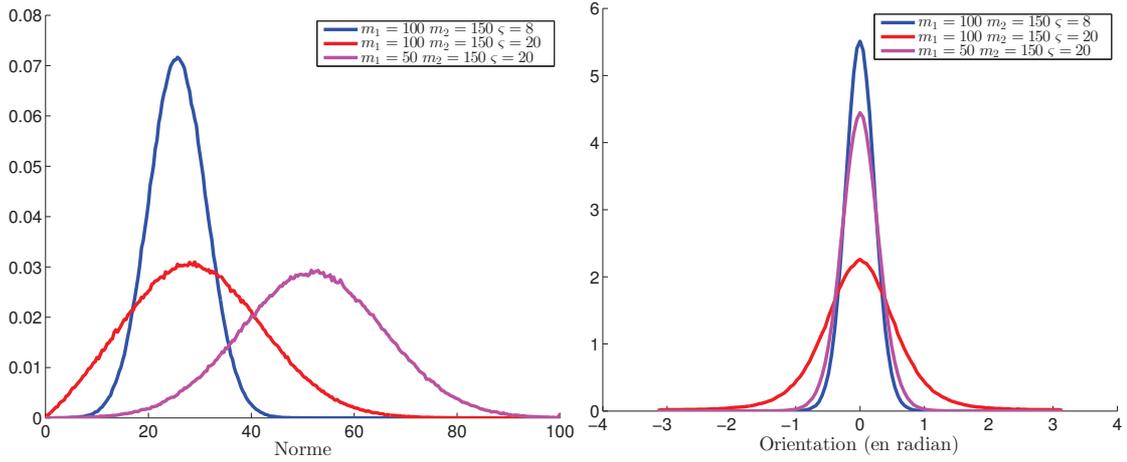


FIGURE 3.8 – Répartition des valeurs de normes et d'orientations du gradient sur un bord vertical en présence de bruit gaussien pour différentes valeurs de m_1 , m_2 et ζ .

avec $m = m_2 - m_1$ le contraste de l'image non bruitée. Comme on peut s'y attendre, les valeurs de normes et d'orientations sont donc d'autant plus étalées que le bruit de l'image est important. En cas de fort bruit gaussien, le calcul du gradient est donc très instable.

3.2.2 Bruit de speckle

On ajoute à une image de bord idéale de réflectivités R_1 et R_2 un bruit de speckle suivant une loi Gamma (Figure 3.7(c)). On souhaite étudier le comportement de la composante verticale du gradient par différence sur la ligne de bord.

On utilise comme estimateur du gradient la différence des intensités moyennées :

$$d = \bar{I}_1 - \bar{I}_2. \quad (3.1)$$

Les moyennes arithmétiques des intensités \bar{I}_1 et \bar{I}_2 sont calculées sur les zones de réflectivités respectives R_1 et R_2 , et avec respectivement N_1 et N_2 pixels (Fig. 3.9). L'estimateur d est l'estimateur au maximum de vraisemblance (MV) de la différence des réflectivités $R_1 - R_2$ [Fjo99]. En effet l'estimateur MV de la réflectivité R_i d'une zone homogène est la moyenne arithmétique des intensités [OBW96] :

$$\hat{R}_i = \bar{I}_i = \frac{1}{N_i} \sum_{k=1}^{N_i} I_i(k). \quad (3.2)$$

La quantité d peut être considérée comme un estimateur du gradient par différence selon la direction verticale. La distribution de d pour une image radar monovue est [TLB88, Fjo99] :

$$p_d(d) = \begin{cases} \frac{\alpha^{N_1+N_2-1}}{\Gamma(N_1)\Gamma(N_2)\beta_1^{N_1}\beta_2^{N_2}} e^{\frac{d}{\beta_2}} \sum_{k=0}^{N_2-1} \binom{N_2-1}{k} \Gamma(N_1+N_2-k-1) \left(-\frac{d}{\alpha}\right)^k & d < 0 \\ \frac{\alpha^{N_1+N_2-1}}{\Gamma(N_1)\Gamma(N_2)\beta_1^{N_1}\beta_2^{N_2}} e^{-\frac{d}{\beta_1}} \sum_{k=0}^{N_1-1} \binom{N_1-1}{k} \Gamma(N_1+N_2-k-1) \left(\frac{d}{\alpha}\right)^k & d \geq 0 \end{cases} \quad (3.3)$$

avec $\beta_1 = \frac{R_1}{N_1}$, $\beta_2 = \frac{R_2}{N_2}$ et $\frac{1}{\alpha} = \frac{1}{\beta_1} + \frac{1}{\beta_2}$.

L'espérance et la variance de la différence des intensités moyennées sont [Fjo99] :

$$\begin{aligned} E[d] &= R_1 - R_2 \\ Var[d] &= \frac{R_1^2}{N_1} + \frac{R_2^2}{N_2}. \end{aligned} \quad (3.4)$$

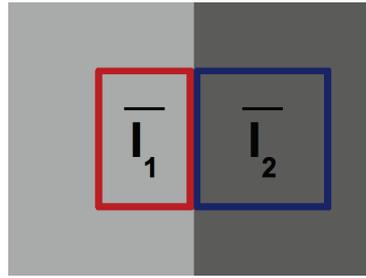


FIGURE 3.9 – Schéma de calcul de l'estimateur du gradient d . L'image présente un bord séparant deux zones de réflectivités R_1 et R_2 . Les moyennes \bar{I}_1 et \bar{I}_2 sont calculées respectivement sur les carrés de couleurs rouge et bleu, de tailles respectives N_1 et N_2 pixels.

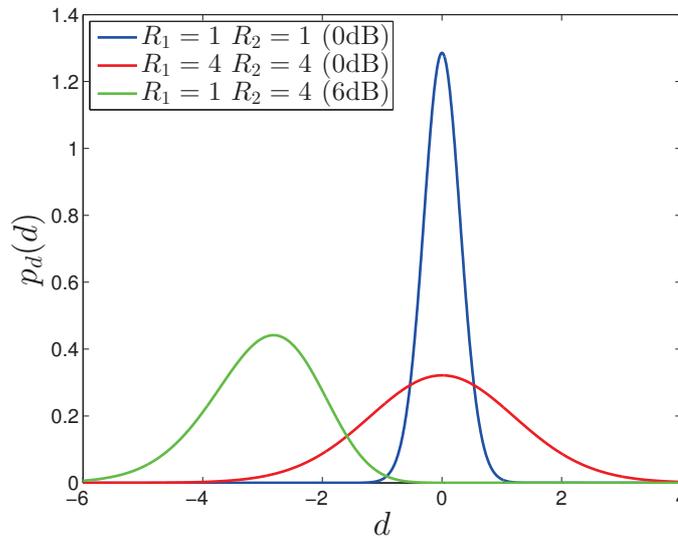


FIGURE 3.10 – Distribution de la différence des intensités moyennes à travers un bord et sur une zone homogène pour des fenêtres symétriques ($N_1 = N_2 = 20$) avec différents contrastes.

L'estimateur d est donc non biaisé, mais sa variance dépend des réflectivités R_1 et R_2 . Sur une zone homogène, la variance augmente alors avec la réflectivité. Ceci est confirmé par la représentation des distributions de d pour différentes valeurs de R_1 et R_2 (Figure 3.10).

Dans le cas d'une application de détection de contour, le but est de repérer les discontinuités de réflectivités. L'espérance de d sur une zone homogène étant nulle, les valeurs de d sont centrées autour de zéro. Deux seuils (un positif et un négatif) sont appliqués à d pour séparer les zones de bord des zones homogènes. Ces seuils doivent permettre de détecter correctement les contours tout en ayant le moins de fausses détections possibles sur les zones homogènes.

On souhaite détecter le contour de la Figure 3.7(b) avec comme valeurs de réflectivités $R_1 = 1$ et $R_2 = 4$. Pour cela on observe la distribution de d sur un bord avec $R_1 = 1$ et $R_2 = 4$, et sur deux zones homogènes de réflectivité $R_0 = 1$ et $R_0 = 4$ (Fig. 3.10). On remarque que le seuil $t_1 = -1$ sur d permet un taux de fausses alarmes quasi-nul sur la zone de réflectivité $R_0 = 1$ et un taux de détection du bord presque maximal. Cependant ce seuil entraîne un taux de fausses alarmes important sur la zone de réflectivité $R_0 = 4$. Cela peut être observé directement sur la Figure 3.11. Pour un seuil donné, le taux de fausse détection dépend de la réflectivité de la zone. La détection de contour n'est donc pas à taux de fausses alarmes constant (TFAC).

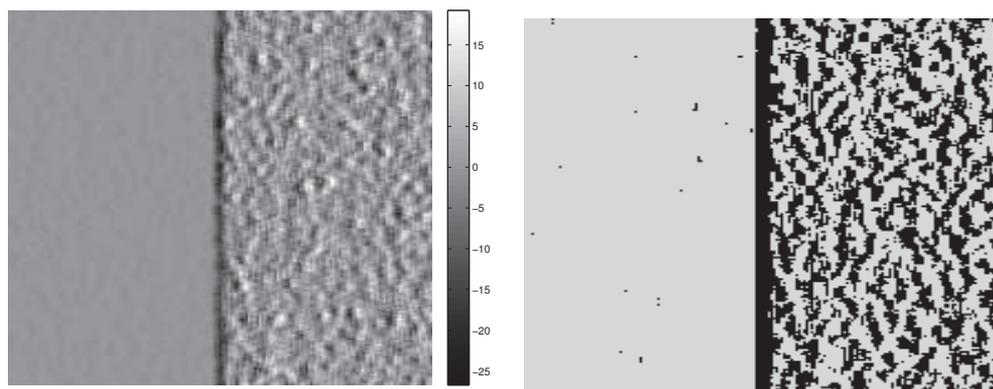
(a) Différence des intensités moyennées d (b) Seuillage de d avec $t_1 = -1$

FIGURE 3.11 – Calcul de la différence des intensités moyennées d de l'image 4.1(b) avec $R_1 = 1$ et $R_2 = 4$, et seuillage de d avec $t = -1$. Des fenêtres symétriques ont été utilisées pour calculer d avec $N_1 = N_2 = 20$ pixels.

On ne peut donc utiliser cet estimateur pour le calcul du gradient, il entraînerait plus de fortes valeurs sur les zones de fortes réflectivités que sur les faibles. L'ajustement des seuils en fonction de la radiométrie locale compliquerait son utilisation et nous préférons donc développer une autre solution.

3.3 Etat de l'art des adaptations de l'algorithme SIFT aux images radar

Comme expliqué précédemment, les algorithmes de détection de points-clés et de calcul du gradient utilisés pour les images numériques sont difficilement applicables aux images radar. Le caractère multiplicatif du speckle entraîne des problèmes récurrents sur les zones de fortes réflectivités. On trouvera plus de fausses détections pour les points-clés et plus de fortes valeurs de gradient sur ces zones que sur les zones de faibles réflectivités.

Des améliorations ont été proposées dans la littérature afin d'augmenter les performances de l'algorithme SIFT sur ce type d'images. On peut les regrouper en différentes catégories :

Simplifications de l'algorithme On peut contraindre l'algorithme en supprimant des étapes afin d'obtenir de meilleures performances. Plusieurs auteurs [SSRS10, SSRU09, WYF12, FHPK13] s'accordent à dire que les détections de points-clés aux petites échelles sont instables, comme présenté en Section 3.1. Ils proposent pour cela l'algorithme SIFT-OCT qui saute la première octave de la pyramide DoG (Section (c)) et ne sélectionne donc que des points avec une échelle σ supérieure à 2. Cela permet en effet de réduire le nombre de fausses détections, puisque comme présenté en Section 3.1.1 l'augmentation des détections avec le bruit se produit à ces échelles. Une autre simplification consiste à supprimer l'assignation d'orientation [SSUR10, FHPK13] dans les situations où les images ont été acquises dans les mêmes conditions (même angle de prise de vue et même orientation). Dans ces cas-là, les images ne présentent qu'une faible rotation et l'assignation d'orientation peut entraîner des erreurs, si ce calcul n'est pas robuste face au bruit. Cependant, cela limite le champ d'application de l'algorithme.

Améliorations de l'algorithme Les étapes de l'algorithme peuvent être aussi modifiées. Dans [LWW08], les valeurs d'intensité de l'image sont seuillées pour obtenir des points-clés distribués spatialement de façon uniforme et la taille de la région utilisée pour calculer le descripteur est étendue. Cela limite cependant le caractère distinctif du descripteur et empêche l'application de l'algorithme à des images présentant de grands changements. Dans [LWGZ09], une nouvelle pyramide avec un sous-échantillonnage progressif est utilisée pour la détection de points-clés et le descripteur SIFT est remplacé par une version améliorée du descripteur Shape Context [BMP02]. Ce nouvel algorithme est plus rapide que l'algorithme SIFT, mais il ne présente pas de meilleures performances. Dans l'algorithme BF-SIFT [WYF12], un espace-échelle anisotropique est utilisé à la place de l'espace-échelle gaussien. Un nouveau calcul du gradient est proposé par Suri et al. [SSUR10] pour construire les descripteurs mais il présente certaines erreurs. Ceci sera discuté plus précisément en Section 4.1.4.

Réduction de l'influence du speckle à l'aide d'un pré-filtrage Plusieurs auteurs proposent de pré-filtrer les images afin de réduire l'influence du bruit. Le débruitage peut s'effectuer avec un filtre multi-vues [XTPG11, WHR07], une transformée en ondelettes [LY08] ou un filtre ISEF (Infinite Symmetric Exponential Filter) [SSRU09]. Cette étape de débruitage est coûteuse en temps de calcul et peut causer une perte d'information ou la création d'artefacts.

Suppression des fausses mises en correspondance Certaines études suggèrent d'améliorer les performances en rejetant les mises en correspondance aberrantes. Lv et al. [LYWW11] divisent l'image maître en quatre sous-régions et ainsi contraignent les relations spatiales entre les points-clés mis en correspondance. Cet algorithme permet de gérer des situations de changement de résolution, mais se retrouve limité pour des cas de rotation. L'algorithme BF-SIFT ne conserve que les mises en correspondance réalisées dans les deux sens (image I_1 contre image I_2 puis inversement). Pour une application de recalage, Fan et al. [FHPK13] estiment la déformation entre les images à l'aide des meilleures mises en correspondance mais ils imposent une transformation restreinte. Xiaoping et al. [XTPG11] limitent la zone de recherche en effectuant un pré-recalage manuel. Plusieurs algorithmes utilisent un autre algorithme ou des informations capteur afin d'estimer une transformation affine entre les images et supprimer les points en dehors du modèle : Wu et Yang [WY11] utilisent la mesure de similarité CRA (cluster reward algorithm) ; Liu et Yu [LY08] les données capteur ; Wessel et al. [WHR07] les informations sur l'orbite et un modèle numérique de terrain ; Suri et al. [SSRU09] l'information mutuelle.

De nombreux algorithmes se basent donc sur des informations de prise de vue qui ne sont pas toujours connues (les données capteur) ou s'appuient sur un autre algorithme. Leurs performances sont meilleures que celles de SIFT, mais le nombre de mises en correspondances correctes est insuffisamment dense et ne permet pas d'envisager une autre application que le recalage. De plus, ces algorithmes ne prennent pas en compte les caractéristiques statistiques du bruit de speckle.

3.4 Conclusions

L'arrivée de nouveaux capteurs optiques et radar pour les satellites de télédétection a eu pour conséquence la multiplication des images disponibles ainsi que l'augmentation de leur résolution. De nouvelles problématiques sont apparues, tout particulièrement pour la comparaison d'images multi-capteurs, multi-résolutions (Section 1). L'algorithme SIFT

et ses variantes, présentés en Section 2, a prouvé son efficacité pour de nombreuses applications en télédétection optiques (Section 2.5). Ses invariances (rotation, changement d'échelle, translation) et robustesses (changements d'illumination, transformations affines) sont particulièrement adaptées à la comparaison d'images à haute résolution.

Cependant, cet algorithme présente des limitations pour les images radar. En effet, l'étape de détection de points-clés entraîne de nombreuses fausses détections, d'autant plus sur les zones de fortes réflectivités, et le calcul du gradient par différence n'est pas adapté aux images présentant un bruit multiplicatif. Les améliorations proposées dans la littérature ne permettent pas une amélioration notable des performances.

Nous proposons dans la partie suivante un nouvel algorithme de type SIFT, efficace pour les images radar. Pour développer un tel algorithme, il est nécessaire de prendre en compte les spécificités des images radar. Dans un premier temps nous présenterons un nouveau calcul du gradient, robuste face au bruit de speckle (Section 4). Ce gradient nous sera ensuite utile pour adapter différentes étapes de l'algorithme SIFT : un nouveau détecteur de point-clé, un nouveau calcul d'orientation et un descripteur local adapté aux images radar seront développés (Section 5).

Deuxième partie

Un nouveau descripteur pour les données radar : SAR-SIFT

Chapitre 4

Calcul du gradient sur une image radar

Les étapes de calcul de l'orientation et des descripteurs de l'algorithme SIFT se basent sur un histogramme de l'orientation du gradient pondérée par la norme. Le gradient par différence, habituellement utilisé pour des images naturelles, n'est cependant pas adapté aux images radar à cause de la présence de bruit multiplicatif [TLB88] (Section 3.2.2). En nous basant sur les travaux réalisés pour des applications de détection de contour [TLB88, FLMCC98], un autre estimateur est proposé pour évaluer le gradient d'une image radar.

Dans le reste du chapitre, le speckle sera toujours de type monovue et on suppose que les valeurs en intensité de l'image suivent une loi exponentielle (Section 1.2.6).

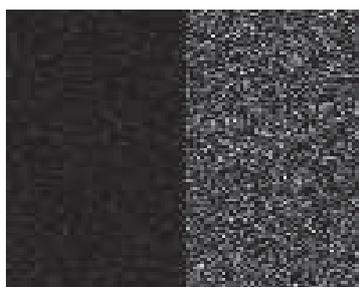
4.1 Propositions dans la littérature

De nombreux travaux ont été réalisés afin de définir une détection de contour à taux de fausses alarmes constant (TFAC) pour les images radar. Dans les cas étudiés dans cette section, la détection consiste à appliquer un ou deux seuils sur une quantité mesurée. Nous présentons ici les différentes quantités envisagées dans une optique de détection d'un bord vertical (Figure 4.1(b)).

Nous reprenons les mêmes notations que dans la Section 3.2.2 : le bord vertical sépare deux zones de réflectivités R_1 et R_2 , avec comme valeurs d'intensité des pixels I_1 et I_2 . Les quantités envisagées se basent sur la comparaison de N_1 pixels I_1 et N_2 pixels I_2 .



(a) Image de réflectivité idéale avec R_1 à gauche et R_2 à droite.



(b) Image d'amplitude avec du speckle monovue.

FIGURE 4.1 – Modèle de bord.

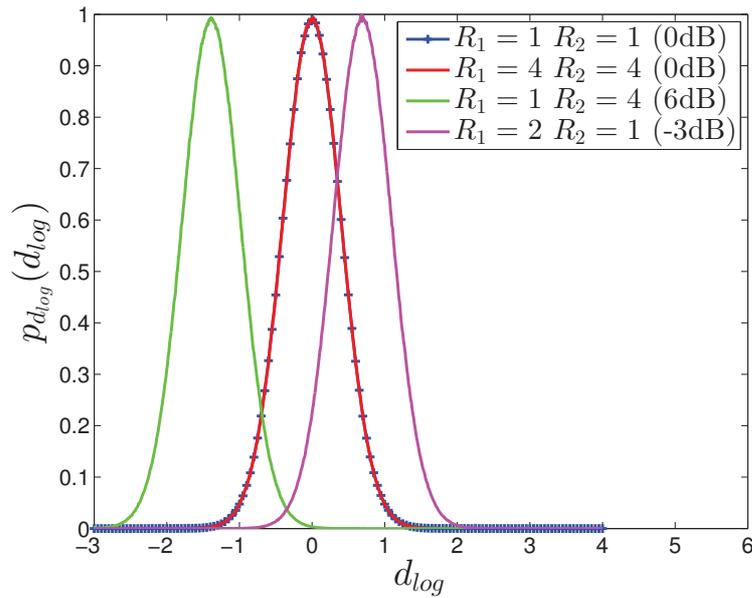


FIGURE 4.2 – Distribution de la différence des moyennes des logarithmes de l’intensité à travers un bord et sur une zone homogène pour des fenêtres symétriques ($N_1 = N_2 = 20$) et avec différents contrastes.

4.1.1 Différence des moyennes des logarithmes

Le speckle étant un bruit multiplicatif, le bruit d’une image radar devient additif quand on calcule le logarithme de l’image. Soit alors la différence des moyennes des logarithmes [Fjo99] :

$$d_{log} = \overline{\log I_1} - \overline{\log I_2}. \quad (4.1)$$

Les moyennes arithmétiques du logarithme des intensités $\overline{\log I_1}$ et $\overline{\log I_2}$ sont calculées sur les zones de réflectivités respectives R_1 et R_2 et avec respectivement N_1 et N_2 pixels, de la même façon que les moyennes $\overline{I_1}$ et $\overline{I_2}$ définies en Section 3.2.2 (voir Fig. 3.9).

Il est difficile de calculer la distribution de d_{log} , mais on peut la représenter à l’aide de simulations (Figure 4.2).

On observe que la variance de la distribution est constante, que l’on soit sur une zone de bord ou sur une zone homogène, et quelque soient les réflectivités. La présence d’un bord entraîne la translation de la distribution de $T = \log R_1 - \log R_2$. La détection de contour est donc TFAC, et la quantité d_{log} peut être utilisée comme estimateur du gradient.

Cette quantité est équivalente à un rapport des moyennes géométriques des intensités. Cependant l’estimateur optimal de la réflectivité est la moyenne arithmétique et non géométrique des intensités [OBW96]. Il est donc plus intéressant d’utiliser un estimateur considérant la moyenne arithmétique.

4.1.2 Rapport des moyennes des intensités

Afin de détecter le bord entre deux zones homogènes de réflectivité R_1 et R_2 , on peut définir le test d’hypothèse suivant :

$$\begin{cases} \mathcal{H}_0 : R_1 = R_2 = R_0 \\ \mathcal{H}_1 : R_1 \neq R_2. \end{cases} \quad (4.2)$$

Le cas \mathcal{H}_0 définit l’absence de bord. Dans notre contexte, le rapport de vraisemblance

s'exprime comme :

$$\lambda = \frac{P((I_1(i))_{1 \leq i \leq N_1}, (I_2(j))_{1 \leq j \leq N_2} | \mathcal{H}_0, R_0)}{P((I_1(i))_{1 \leq i \leq N_1} | \mathcal{H}_1, R_1) P((I_2(j))_{1 \leq j \leq N_2} | \mathcal{H}_1, R_2)} \quad (4.3)$$

Comme les valeurs des réflectivités sont inconnues, on calcule le rapport de vraisemblance généralisé en remplaçant R_0 , R_1 et R_2 par les estimateurs au sens du Maximum de Vraisemblance (MV). La moyenne arithmétique des intensités étant l'estimateur MV de la réflectivité, on obtient [OBW96] les log-vraisemblances généralisées des hypothèses \mathcal{H}_0 et \mathcal{H}_1 comme :

$$\lambda_1 = -N_1 \log \bar{I}_1 - N_2 \log \bar{I}_2 - (N_1 + N_2) \quad (4.4)$$

$$\lambda_0 = -(N_1 + N_2) \log \frac{N_1 \bar{I}_1 + N_2 \bar{I}_2}{N_1 + N_2} - (N_1 + N_2). \quad (4.5)$$

La différence des log-vraisemblances est donnée par :

$$\lambda_D \equiv \lambda_1 - \lambda_0 = -N_1 \log \bar{I}_1 - N_2 \log \bar{I}_2 + (N_1 + N_2) \log \frac{N_1 \bar{I}_1 + N_2 \bar{I}_2}{N_1 + N_2}. \quad (4.6)$$

Le test $\lambda_D > 0$ permet d'identifier les zones de bord. On peut montrer que λ_D ne dépend que du rapport des intensités moyennes $r = \frac{\bar{I}_1}{\bar{I}_2}$:

$$\lambda_D = N_1 \log r - (N_1 + N_2) \log \left(\frac{N_1}{N_1 + N_2} (r - 1) + 1 \right) \quad (4.7)$$

Selon le test d'hypothèse présenté, la mesure du rapport r permet alors d'évaluer la présence de bord. Par ailleurs, ce rapport r est l'estimateur MV du rapport des réflectivités $C = \frac{R_1}{R_2}$ [Fjo99].

La distribution de ce rapport pour une image radar monovue le long d'un bord suit une loi de Fisher, ou loi Beta de deuxième espèce, [Fjo99] :

$$p_r(r) = \frac{\Gamma(N_1 + N_2)}{\Gamma(N_1) \Gamma(N_2)} \frac{\frac{1}{r} \left(\frac{\beta_2 r}{\beta_1} \right)^{N_1}}{\left(1 + \frac{\beta_2 r}{\beta_1} \right)^{N_1 + N_2}}, \quad (4.8)$$

avec comme paramètres $L = N_1$, $M = N_2$ et $\mu = \frac{\beta_1 N_1}{\beta_2 N_2} = \frac{R_1}{R_2} = C$ (Eq. 1.8). L'espérance et la variance du rapport de deux intensités moyennées sont alors :

$$\begin{aligned} E[r] &= \frac{N_2}{N_2 - 1} C \\ Var[r] &= \frac{N_2^2 (N_1 + N_2 - 1)}{N_1 (N_2 - 1)^2 (N_2 - 2)} C^2. \end{aligned} \quad (4.9)$$

Le rapport r est biaisé mais le biais ne dépend que du nombre de pixels N_2 utilisé pour calculer la moyenne \bar{I}_2 . La variance de r ne dépend pas de la réflectivité pour une zone homogène ($C = 1$). L'utilisation de r permet donc une détection de contour à taux de fausses alarmes constant.

On observe Figure 4.3 la distribution de ce rapport le long d'un bord et sur une zone homogène pour différentes valeurs du contraste. La variance de la distribution augmente bien avec le rapport des intensités moyennes, mais elle est indépendante de la réflectivité pour une zone homogène. Considérant la détection d'un contour avec $R_1 = 1$ et $R_2 = 2$,

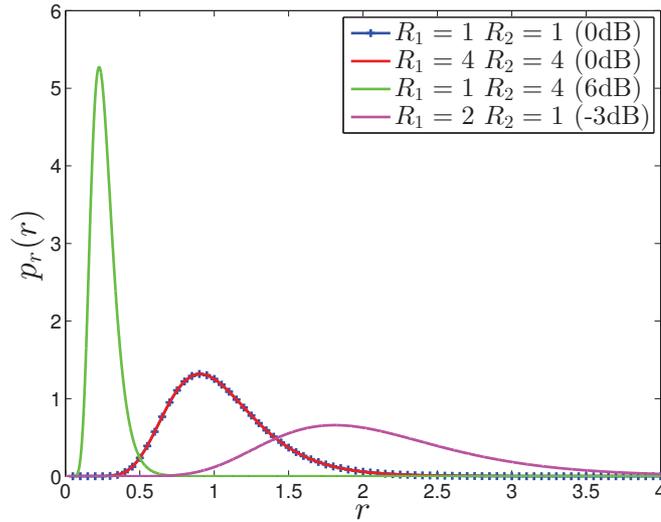


FIGURE 4.3 – Distribution du rapport des moyennes des intensités à travers un bord et sur une zone homogène pour des fenêtres symétriques ($N_1 = N_2 = 20$) et avec différents contrastes.

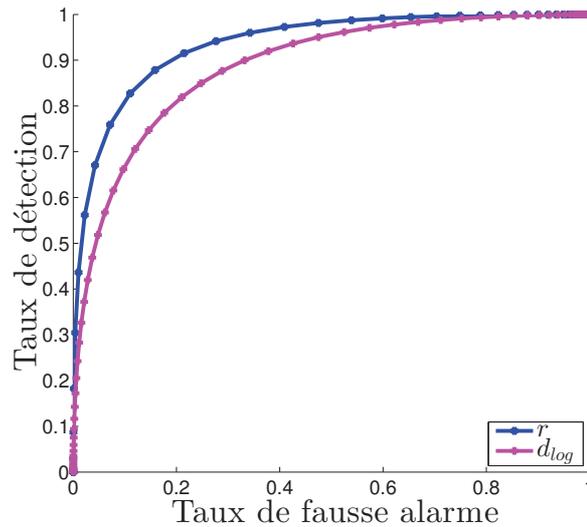


FIGURE 4.4 – Courbes ROC des opérateurs d_{log} et r pour la détection d'un contour séparant deux zones de réflectivités $R_1 = 1$ et $R_2 = 2$.

on présente Figure 4.4 les courbes ROC (Receiver Operating Characteristic) des opérateurs d_{log} et r . En faisant varier le seuil sur r ou d_{log} , on calcule le taux de détection en fonction du taux de fausses alarmes à l'aide des distributions présentées précédemment. On remarque bien que l'opérateur r présente de meilleures performances. Cet opérateur est donc plus adapté aux images radar que l'opérateur d_{log} .

Plusieurs détecteurs de contour ont été proposés à partir de ce rapport. Le Ratio of Averages (ROA) [TLB88, Bov88] par exemple calcule les moyennes d'intensités de part et d'autres d'une direction i (Figure 4.5(a)) puis sélectionne le minimum entre le rapport et son inverse afin d'être invariant au sens de variation de la réflectivité :

$$r_{min}\{i\} = \min \left\{ \frac{\overline{I_1}\{i\}}{\overline{I_2}\{i\}}, \frac{\overline{I_2}\{i\}}{\overline{I_1}\{i\}} \right\}. \quad (4.10)$$

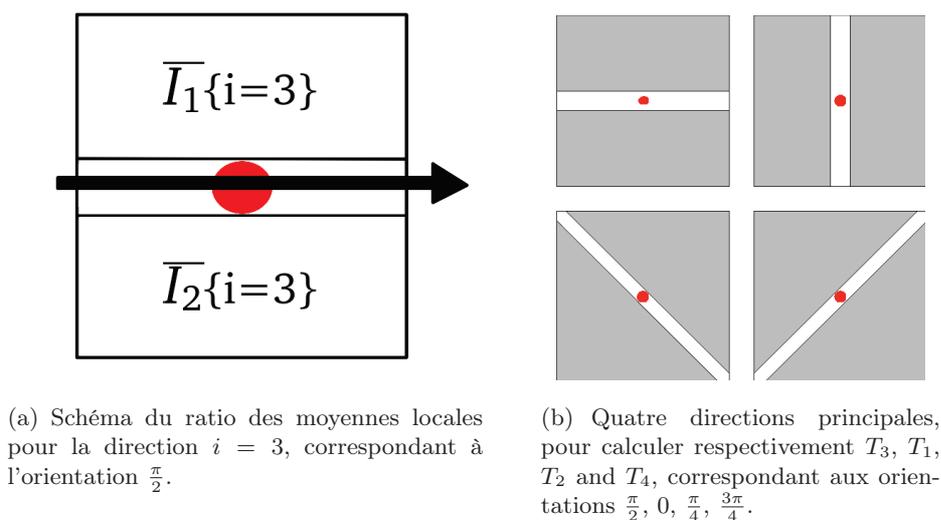


FIGURE 4.5 – Schéma du détecteur ROA [TLB88, Bov88].

Ce ratio normalisé $r_{min}\{i\}$ est calculé selon plusieurs directions i (Figure 4.5(b)). Et le minimum :

$$r_0 = \min_i(r_{min}\{i\}) \quad (4.11)$$

correspondant à la direction de bord la plus probable, est choisi pour construire l'image de contour de l'image originale. On obtient ainsi une estimation de la norme du gradient de l'image.

Une estimation de l'orientation du gradient peut être obtenue en choisissant pour chaque pixel :

$$t_0 = (\operatorname{argmin}_i(r_{min}\{i\}) - 1) \times \frac{\pi}{4}. \quad (4.12)$$

Néanmoins cette méthode limite le nombre de directions possibles. Plus de quatre directions peuvent être envisagées mais cela entraîne alors une augmentation du temps de calcul.

4.1.3 Rapport de moyennes pondérées exponentiellement

Pour diminuer le taux de fausses alarmes et augmenter le taux de détection, il faut diminuer la variance de la distribution et donc augmenter le nombre de pixels N_1 et N_2 utilisés pour calculer les moyennes. Néanmoins dans un modèle de contours multiples, on risque de ne pas détecter les structures fines. Dans des situations multi-contours, il faut donc trouver un compromis entre l'atténuation du speckle et la résolution spatiale de la détection de contour.

Le Ratio of Exponentially Weighted Average (ROEWA) [FLMCC98] est une amélioration du ROA : les moyennes sont calculées avec une pondération exponentielle. Pour la direction présentée Figure 4.5(a), on obtient par exemple au point (a, b) :

$$r_{e,\alpha}\{i = 3\}(a, b) = \frac{\overline{I_{1,\alpha}\{i = 3\}}(a, b)}{\overline{I_{2,\alpha}\{i = 3\}}(a, b)}. \quad (4.13)$$

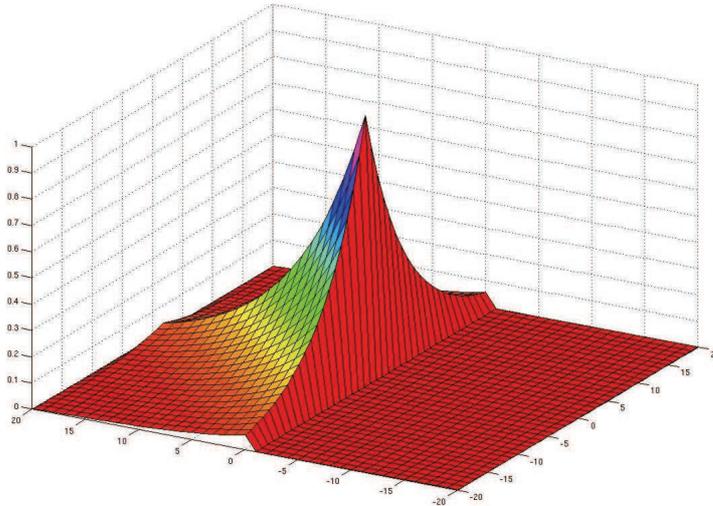


FIGURE 4.6 – Filtre exponentiel pour le calcul des moyennes pondérées

Avec :

$$\begin{aligned}\overline{I_{1,\alpha}\{i=3\}}(a,b) &= \int_{x=\mathbb{R}} \int_{y=\mathbb{R}^+} I(a+x,b+y) \times e^{-\frac{|x|+|y|}{\alpha}} dx dy \\ \overline{I_{2,\alpha}\{i=3\}}(a,b) &= \int_{x=\mathbb{R}} \int_{y=\mathbb{R}^-} I(a+x,b+y) \times e^{-\frac{|x|+|y|}{\alpha}} dx dy.\end{aligned}\quad (4.14)$$

L'utilisation d'un filtre exponentiel (Fig. 4.6) est adapté au modèle multi-contours des images. Le paramètre α (toujours supérieur à 1) permet un lissage plus ou moins important de l'image. La détection de contour à l'aide de la quantité $r_{e,\alpha}$ est toujours TFAC.

Le ratio est normalisé pour être invariant au sens de variation de la réflectivité :

$$r_{max,\alpha}\{i\} = \max \left\{ r_{e,\alpha}\{i\}, \frac{1}{r_{e,\alpha}\{i\}} \right\}. \quad (4.15)$$

Cette opération est réalisée pour la direction horizontale et verticale afin d'obtenir :

$$\begin{aligned}r_{x,\alpha} &= r_{max,\alpha}\{i=1\} \\ r_{y,\alpha} &= r_{max,\alpha}\{i=3\}.\end{aligned}\quad (4.16)$$

Par analogie avec les détecteurs de contours des images optiques basés sur le gradient, on obtient l'image de contour avec :

$$r_{0,\alpha} = \sqrt{r_{x,\alpha}^2 + r_{y,\alpha}^2}. \quad (4.17)$$

4.1.4 Gradient radar basé sur le ROEWA

En se basant sur le détecteur de contour ROEWA, Suri et al. [SSUR10] proposent d'estimer le gradient horizontal et vertical par $r_{x,\alpha}$ et $r_{y,\alpha}$. Et donc de définir la norme du gradient par $r_{0,\alpha}$ et l'orientation par :

$$t_{0,\alpha} = \arctan \left(\frac{r_{y,\alpha}}{r_{x,\alpha}} \right). \quad (4.18)$$

Les ratios normalisés $r_{x,\alpha}$ et $r_{y,\alpha}$ donnent bien une information sur la variation de la réflectivité selon la direction horizontale et verticale, mais ils entraînent des erreurs

importantes s'ils sont utilisés pour calculer l'orientation du gradient. En effet, $r_{x,\alpha}$ et $r_{y,\alpha}$ étant toujours positifs, $t_{0,\alpha}$ ne peut prendre que des valeurs entre 0 et $\frac{\pi}{2}$. Aussi en calculant le gradient le long du bord de la Figure 4.1(b) (avec $R_1 < R_2$), on obtient :

$$\begin{aligned} r_{x,\alpha} &= \frac{R_2}{R_1} \\ r_{y,\alpha} &= 1. \end{aligned} \quad (4.19)$$

L'orientation est alors :

$$t_{0,\alpha} = \arctan\left(\frac{R_1}{R_2}\right) \quad (4.20)$$

et dépend donc des valeurs des réflectivités, alors qu'en théorie elle est constante et égale à zéro.

Les ratios normalisés $r_{x,\alpha}$ et $r_{y,\alpha}$ ne devraient donc pas être utilisés comme estimateurs du gradient.

4.2 Proposition de nouveau gradient adapté aux images radar : Gradient par Ratio

Afin d'obtenir un bon estimateur de l'orientation du gradient sur les images radar, nous étudions le comportement du logarithme du ROA et du ROEWA. Cela nous permet ensuite de proposer un nouveau gradient adapté aux images radar : le Gradient par Ratio (GR).

4.2.1 Logarithme du ROA

Le logarithme du rapport r des moyennes des intensités :

$$r_l = \log(r) = \log\left(\frac{\bar{I}_1}{\bar{I}_2}\right) \quad (4.21)$$

a l'avantage d'être l'estimateur MV de $C_l = \log\left(\frac{R_1}{R_2}\right)$, contrairement à d_{log} (voir Annexe C.3). La distribution de cet estimateur se calcule facilement à partir de celle de r :

$$p_{r_l}(r_l) = \frac{\beta_1 \Gamma(N_1 + N_2)}{\beta_2 \Gamma(N_1)\Gamma(N_2)} \frac{\left(\frac{\beta_2}{\beta_1} e^{r_l}\right)^{N_1-1}}{\left(1 + \frac{\beta_2}{\beta_1} e^{r_l}\right)^{N_1+N_2}}. \quad (4.22)$$

En effet le rapport r suit une loi de Fisher (Eq. 1.8) avec comme paramètres $L = N_1$, $M = N_2$ et $\mu = \frac{\beta_1 N_1}{\beta_2 N_2} = \frac{R_1}{R_2}$ (Section 4.1.2). Comme présenté Section 1.2.6, les moments centrés de r_l peuvent être obtenus à partir des log-cumulants de r . En calculant les log-cumulants de la loi de Fisher (voir Annexe C.4), on obtient alors l'espérance et la variance de r_l :

$$E[r_l] = \log\left(\frac{R_1}{R_2}\right) + (\Psi(N_1) - \log(N_1)) - (\Psi(N_2) - \log(N_2)) \quad (4.23)$$

$$Var[r_l] = \Psi(1, N_1) + \Psi(1, N_2).$$

Dans le cas où $N = N_1 = N_2$, on obtient :

$$\begin{aligned} E[r_l] &= \log\left(\frac{R_1}{R_2}\right) \\ Var[r_l] &= 2\Psi(1, N). \end{aligned} \quad (4.24)$$

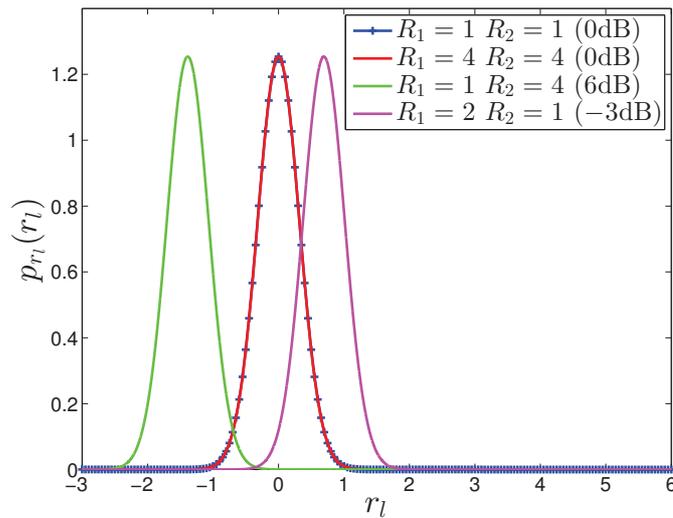


FIGURE 4.7 – Distribution du logarithme du rapport des moyennes de l'intensité à travers un bord et sur une zone homogène pour des fenêtres symétriques ($N_1 = N_2 = 20$) avec différents contrastes

La quantité r_l est alors l'estimateur MV mais aussi non biaisé de $C_l = \log\left(\frac{R_1}{R_2}\right)$. De plus, sa variance dépend seulement de N et non de C , contrairement à r . Elle peut alors être considérée comme un bon estimateur du gradient.

La Figure 4.7 présente la distribution de r_l pour différentes valeurs de réflectivité. On observe bien que r_l est centré autour de zéro pour une zone homogène et que quelque soient les valeurs des réflectivités, la variance de la distribution est constante.

4.2.2 Logarithme du ROEWA

Afin de détecter les contours multiples, il peut être plus intéressant d'utiliser le logarithme de $r_{e,\alpha}$ plutôt que celui de r :

$$r_{e_l,\alpha} = \log(r_{e,\alpha}) \quad (4.25)$$

On représente Figure 4.8 les distributions de $r_{e_l,\alpha}$ pour deux valeurs de α à l'aide de simulations. La quantité $r_{e_l,\alpha}$ comme r_l , suit une loi semblable à une gaussienne, dont la variance est aussi indépendante des réflectivités. Cette variance n'est cependant pas la même que pour r_l et dépend du paramètre α . L'utilisation d'une plus forte valeur de α entraîne une diminution de la variance. Ceci est cohérent puisque l'image est alors plus lissée. Néanmoins, on observe les mêmes espérances pour $r_{e_l,\alpha}$ et r_l avec les mêmes valeurs de réflectivités.

L'espérance de $r_{e_l,\alpha}$ semble être égale à celle de r_l :

$$E[r_{e_l}] = E[r_l] = C_l = \log\left(\frac{R_1}{R_2}\right) \quad (4.26)$$

On vérifie cette formule à l'aide de simulations en calculant la moyenne de $r_{e_l,\alpha}$ sur N échantillons d'images simulées. Le pourcentage d'erreur par rapport à la valeur attendue de l'espérance en fonction du nombre d'échantillons est représenté Figure 4.9. On observe qu'à partir de $N = 1000$ échantillons l'erreur reste inférieure à 1%. Expérimentalement, l'espérance de $r_{e_l,\alpha}$ correspond bien à C_l .

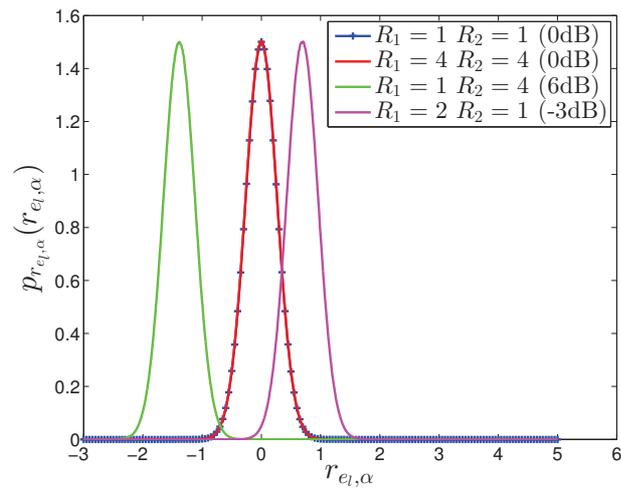
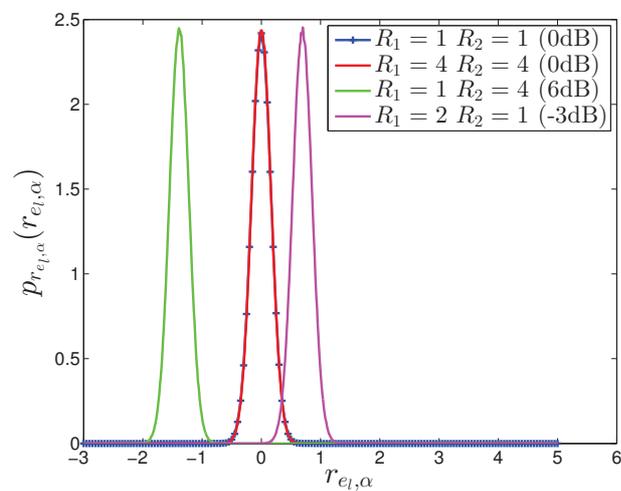
(a) $\alpha = 2$ (b) $\alpha = 6$

FIGURE 4.8 – Distribution du logarithme du rapport des moyennes de l'intensité pondérées exponentiellement ($\alpha = 2$ et $\alpha = 6$) à travers un bord et sur une zone homogène avec différents contrastes

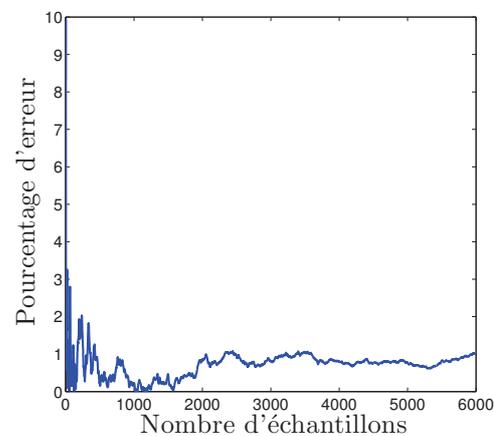
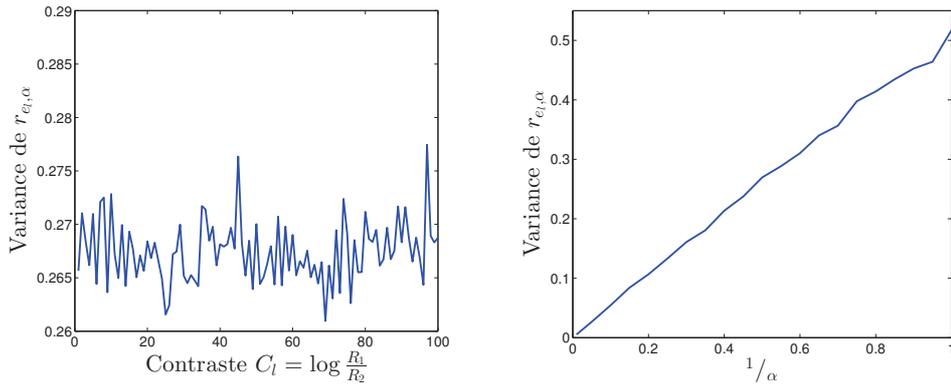


FIGURE 4.9 – Pourcentage d'erreur sur le calcul empirique de l'espérance de $r_{ei,\alpha}$ par rapport à la valeur attendue en fonction du nombre d'échantillons (avec $C_l = 2.3$ et $\alpha = 2$).



(a) Variance de $r_{e_l, \alpha}$ en fonction du contraste C_l , pour $\alpha = 2$.

(b) Variance de $r_{e_l, \alpha}$ en fonction de l'inverse du paramètre de pondération exponentielle α , pour $C_l = 2.3$.

FIGURE 4.10 – Evolution de la variance de $r_{e_l, \alpha}$ en fonction du contraste C_l ou du paramètre de pondération exponentielle α (calculée avec $N = 10000$ échantillons à chaque fois)

Concernant la variance de $r_{e_l, \alpha}$, on représente Figure 4.10 son évolution en fonction de différentes valeurs du contraste C_l ou du paramètre de pondération exponentielle α . On observe que la variance est indépendante de la réflectivité de la zone mais dépend linéairement de l'inverse du paramètre α . Comme on pouvait l'espérer, la variance ne dépend que du paramètre α . La quantité $r_{e_l, \alpha}$ semble donc être un bon estimateur du gradient.

4.2.3 Proposition du Gradient par Ratio

Définition On propose ici de définir comme composantes horizontales et verticales du gradient :

$$\begin{aligned} G_{x, \alpha} &= \log(r_{x, \alpha}) = r_{e_l, \alpha} \{i = 1\} \\ G_{y, \alpha} &= \log(r_{y, \alpha}) = r_{e_l, \alpha} \{i = 3\} \end{aligned} \quad (4.27)$$

et de calculer la norme et l'orientation par les formules habituelles :

$$\begin{aligned} G_{n, \alpha} &= \sqrt{G_{x, \alpha}^2 + G_{y, \alpha}^2} \\ G_{t, \alpha} &= \arctan\left(\frac{G_{y, \alpha}}{G_{x, \alpha}}\right). \end{aligned} \quad (4.28)$$

Avantages de cet estimateur du gradient La quantité r_l est l'estimateur MV et non biaisé de $C_l = \log\left(\frac{R_1}{R_2}\right)$. La pondération exponentielle permet de lisser l'image à différentes échelles, sans pour autant modifier l'espérance de l'estimateur. On peut alors comparer cet estimateur au gradient par différence appliqué sur une image avec un flou Gaussien.

L'utilisation du logarithme permet d'éviter les problèmes de calcul d'orientations présentés Section 4.1.4. $G_{x, \alpha}$ et $G_{y, \alpha}$ peuvent prendre ici des valeurs positives et négatives. Ainsi toutes les valeurs possibles d'orientations sont prises en compte. D'autre part, l'orientation du gradient est bien égale à zéro le long d'un bord vertical (Section 4.1.4). On obtient en effet comme valeurs de gradient :

$$\begin{aligned} G_{x, \alpha} &= \log(R_1) - \log(R_2) \\ G_{y, \alpha} &= 0. \end{aligned} \quad (4.29)$$

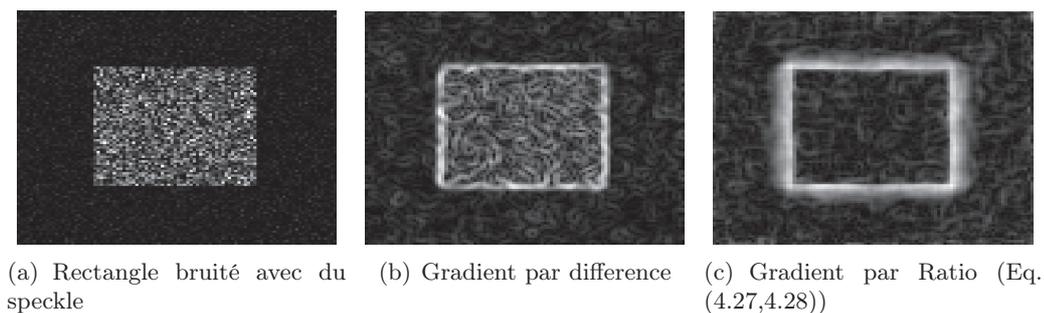


FIGURE 4.11 – Exemple d’un rectangle bruité avec du speckle et sa norme du gradient avec deux méthodes de calcul différents.

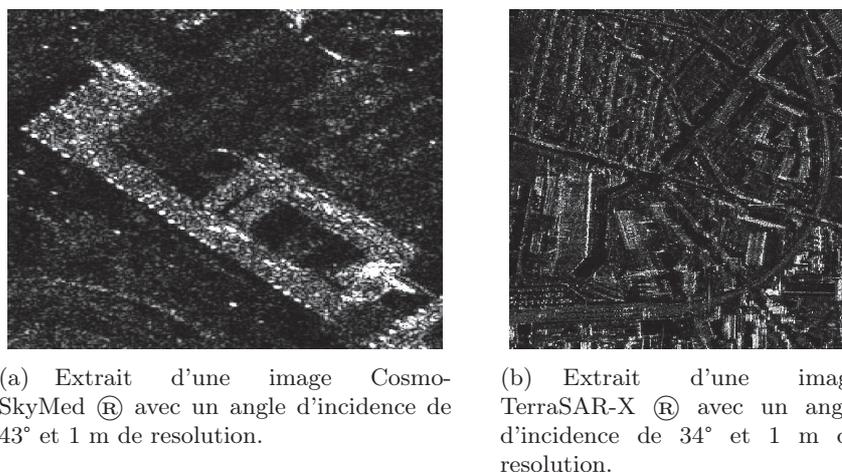


FIGURE 4.12 – Extraits d’image radar

Comme on calcule un ratio plutôt qu’une différence, on peut noter que ce calcul de gradient est invariant au changement de contraste multiplicatif.

La Figure 4.11 présente la norme du gradient d’un rectangle bruité avec du speckle, calculée avec le Gradient par différence et le Gradient par Ratio. On remarque que l’utilisation du GR, contrairement au Gradient par différence, n’entraîne pas plus de fortes valeurs de gradient sur les zones de fortes réflectivités. Cet estimateur du gradient est donc bien adapté aux images radar.

Un exemple de calcul du GR sur une image multi-contours bruitée est présenté Figure 4.13 pour deux valeurs de α différentes. Le speckle est mieux réduit pour une forte valeur de α , cependant les contours proches sont plus atténués. Faire varier le paramètre α permet donc d’étudier des structures de différentes tailles.

Exemples de calcul du Gradient La Figure 4.14 présente les Gradient par différence et Gradient par Ratio de l’image de la Figure 4.12(a). Le Gradient par différence semble présenter une meilleure réduction du bruit de speckle, mais les valeurs du gradient (norme, composantes horizontale et verticale) sont plus importantes sur les zones de fortes réflectivités que sur celles de faibles réflectivités. On remarque que le Gradient par Ratio ne présente pas ce problème.

L’influence de α est illustrée Figure 4.15 : le Gradient par Ratio de l’image de la Figure 4.12(b) est calculé pour deux valeurs de α . Les détails sont mieux conservés pour $\alpha = 2$. Pour $\alpha = 6$, l’image est plus lissée. Les contours sont donc plus nets et on distingue bien les grandes structures. Faire varier ce paramètre permet donc d’obtenir des images de

gradient à différentes échelles.

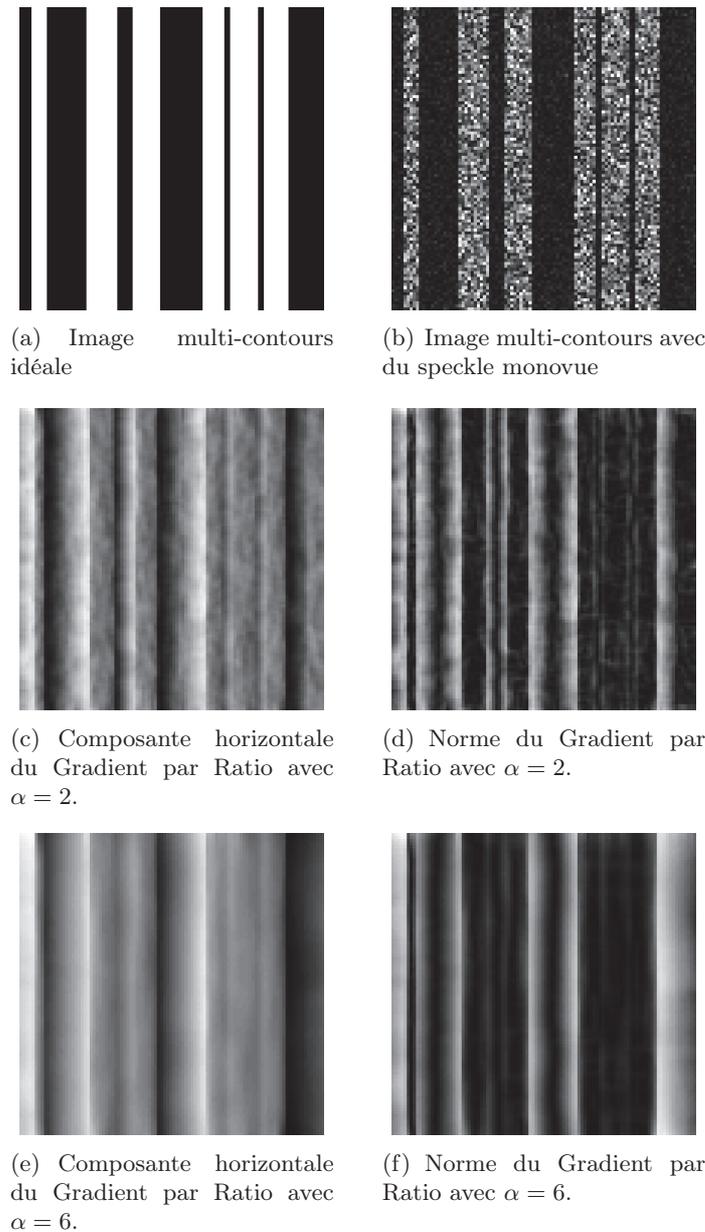
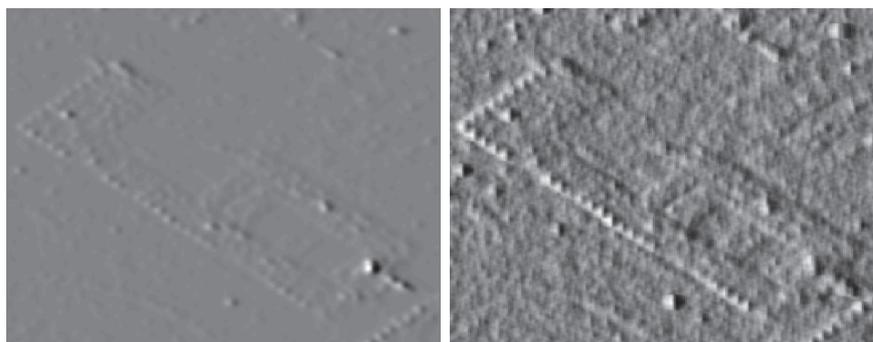


FIGURE 4.13 – Exemple de calcul du Gradient par Ratio sur une image multi-contours avec du speckle monovue.

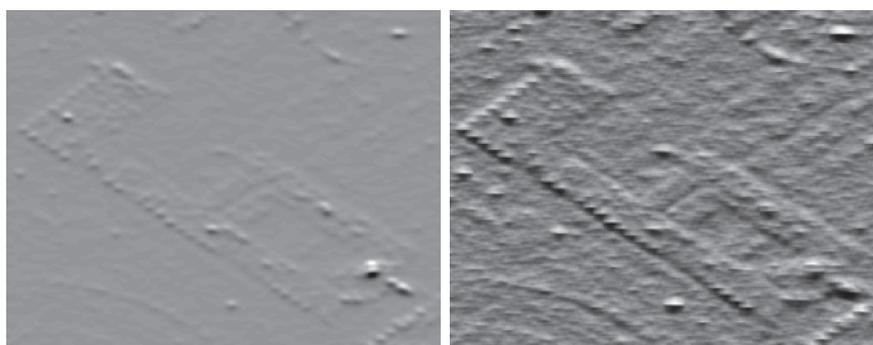
4.3 Conclusions

Nous avons présenté un nouveau calcul du gradient adapté aux images radar, permettant de calculer une norme et une orientation du gradient robuste face au bruit de speckle. Ce gradient permet aussi d'obtenir des images de gradient à différentes échelles.

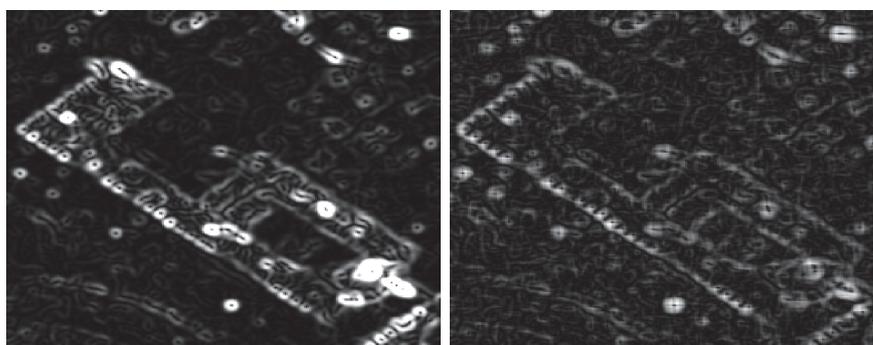
Ce Gradient par Ratio nous permet maintenant de développer un algorithme de type SIFT pour les images radar. Le chapitre suivant présente ce nouvel algorithme appelé SAR-SIFT. Les étapes de détection de points-clés, d'assignation d'orientations et d'extractions de descripteurs ont été adaptées aux caractéristiques des images radar. La suite du chapitre exposera une évaluation expérimentale de l'algorithme ainsi qu'une étude des performances.



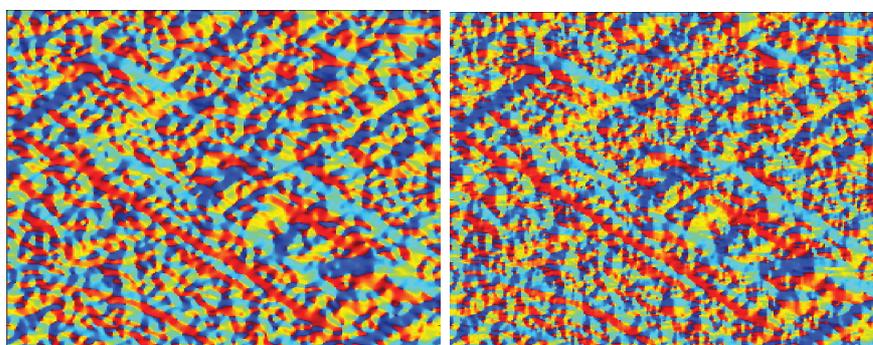
(a) Composante horizontale



(b) Composante verticale

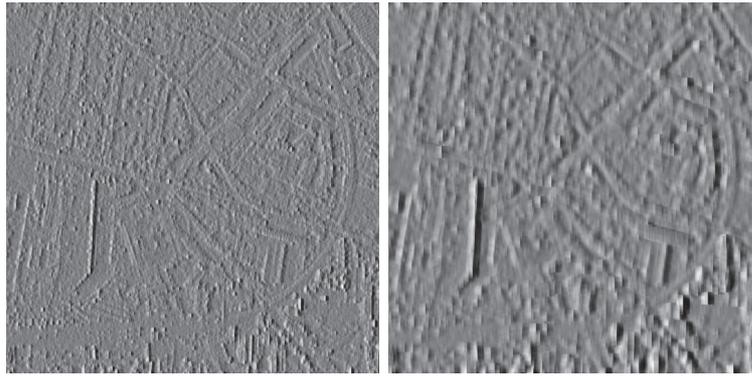


(c) Norme

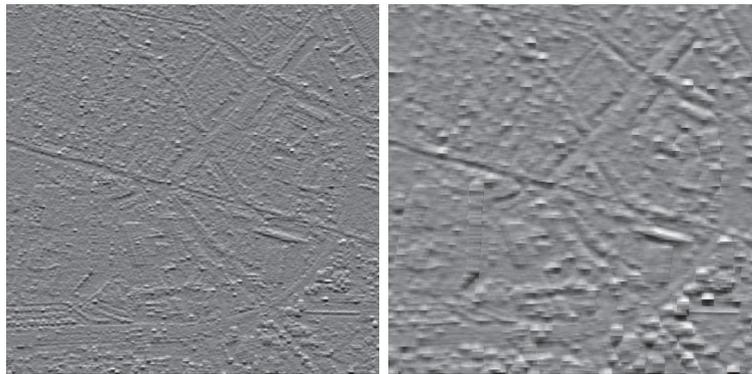


(d) Orientation

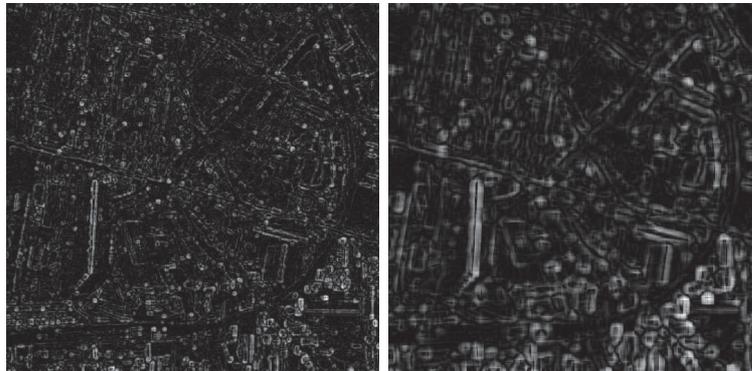
FIGURE 4.14 – Calcul du gradient appliqué sur l'image de Fig. 4.12(a) : à gauche Gradient par différence avec un flou Gaussien de $\sigma = 2$ et à droite Gradient par Ratio avec $\alpha = 2$.



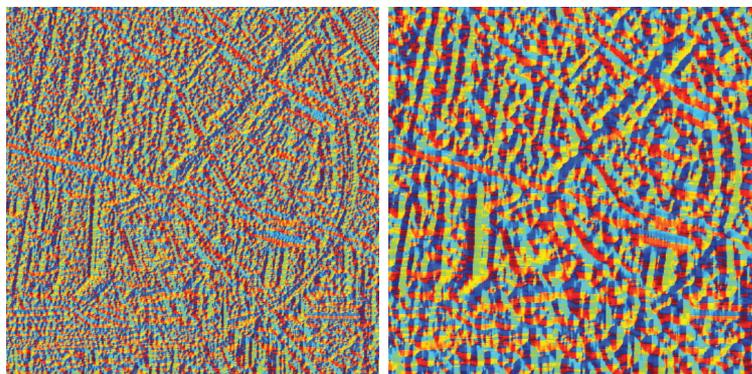
(a) Composante horizontale



(b) Composante vertical



(c) Norme



(d) Orientation

FIGURE 4.15 – Calcul du Gradient par Ratio de l'image de Fig. 4.12(b) avec $\alpha = 2$ (à gauche) et $\alpha = 6$ (à droite).

Chapitre 5

SAR-SIFT : un descripteur local pour les images radar

5.1 Présentation

Nous présentons ici un algorithme original pour coder la géométrie des images radar. Cet algorithme suit le même principe que l'algorithme SIFT, avec une détection de points-clés suivie d'une extraction de descripteurs locaux. Le schéma de calcul de cet algorithme, que nous avons appelé SAR-SIFT, est présenté en Figure 5.12.

5.1.1 Détection de points-clés

5.1.1.1 Introduction

Concernant la détection de points-clés sur les images radar, une première idée est de considérer l'algorithme SIFT-OCT [SSRU09] (Section 3.3) qui consiste à omettre les petites échelles pour la détection. Une autre approche peut être de considérer le détecteur LoG-Harris (Section 2.1.2.4) appliqué au logarithme de l'image. Cela permet de faire face à un bruit gaussien au lieu d'un bruit multiplicatif, et ainsi de ne plus avoir le problème du grand nombre de fausses détections sur les zones de fortes réflectivités. Bien que tentantes par leur simplicité, ces approches ne sont pas assez robustes face à un fort bruit typique des images radar et n'améliorent pas grandement les performances du détecteur LoG-Harris original.

La Figure 5.1 présente un exemple de calcul des points-clés sur une image synthétique d'un rectangle bruitée avec du speckle monovue. Trois méthodes sont testées : LoG-Harris sur l'image d'amplitude, LoG-Harris avec suppression des petites-échelles (SIFT-OCT) et LoG-Harris sur le logarithme de l'image. Le calcul des points-clés sur l'image non bruitée avec LoG-Harris est présenté à titre de comparaison. La détection avec LoG-Harris entraîne de nombreuses fausses détections sur la zone de forte réflectivité. La suppression des petites échelles permet d'en supprimer un grand nombre. Avec l'application de LoG-Harris sur le logarithme de l'image, on trouve toujours un nombre conséquent de fausses détections mais elles sont réparties sur l'ensemble de l'image. Dans les trois cas on trouve des points-clés situés sur les coins (ce qui est attendu), mais ils semblent mal localisés et peu stables. Les deux approches envisagées (SIFT-OCT et LoG-Harris sur le logarithme de l'image) ne présentent pas une amélioration significative des performances de LoG-Harris.

Parmi les opérateurs présentés en Section 2.1.1, le LoG et le DoH reposent sur un calcul des dérivées secondes. Comme le gradient présente une forte variance sur les images avec du bruit multiplicatif (Section 3.2.2), ces opérateurs ne semblent pas facilement adaptables aux images radar. Le critère de Harris multi-échelle [DSH00] (Section 2.1.1.2) repose lui

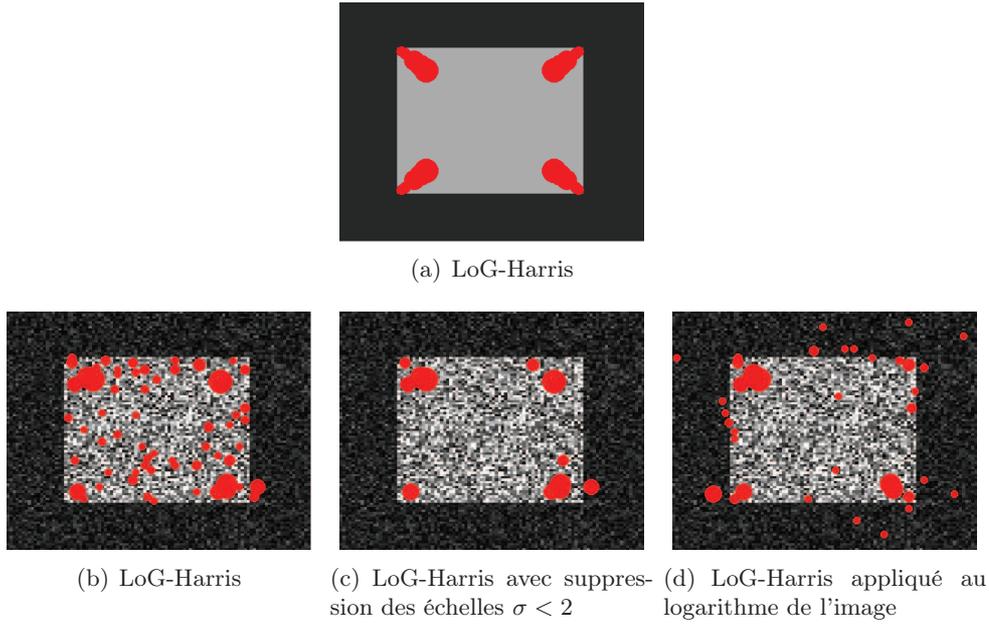


FIGURE 5.1 – Calcul des points-clés à l'aide de différentes méthodes sur une image de rectangle idéale (a) et bruitée (b), (c) et (d). Le nombre de points-clés détectés est de 21 (a), 62 (b), 21 (c) et 38 (d).

sur les dérivées premières. A partir du calcul du gradient adapté aux images radar présenté au Chapitre 4, nous proposons une nouvelle approche inspirée de ce critère Harris multi-échelle.

5.1.1.2 Détection de points-clés SAR-Harris

(a) **Critère SAR-Harris** Pour rappel (Section 2.1.1.2), la matrice et le critère de Harris multi-échelle d'une image I sont définis par :

$$M_H(x, y, \sigma) = \sigma^2 \mathcal{G}_{a_1 \sigma} \star \begin{bmatrix} \left(\frac{\partial I_\sigma}{\partial x}(x, y)\right)^2 & \frac{\partial I_\sigma}{\partial x} \cdot \frac{\partial I_\sigma}{\partial y}(x, y) \\ \frac{\partial I_\sigma}{\partial x} \cdot \frac{\partial I_\sigma}{\partial y}(x, y) & \left(\frac{\partial I_\sigma}{\partial y}(x, y)\right)^2 \end{bmatrix} \quad (5.1)$$

$$\mathcal{C}_H(x, y, \sigma) = \det(M_H(x, y, \sigma)) - k_1 \cdot \text{Tr}(M_H(x, y, \sigma))^2.$$

Ce critère permet de sélectionner des points situés sur des coins à l'aide du test :

$$\mathcal{C}_H(x, y, \sigma) > d_h. \quad (5.2)$$

En s'inspirant de cette définition et du Gradient par Ratio défini en Section 4, nous proposons un nouveau détecteur de coins adapté aux images radar. Nous obtenons ainsi la nouvelle matrice et le nouveau critère SAR-Harris multi-échelle, respectivement définis comme :

$$M_{SH}(x, y, \alpha) = \mathcal{G}_{a_2 \alpha} \star \begin{bmatrix} (G_{x,\alpha})^2 & G_{x,\alpha} \cdot G_{y,\alpha} \\ G_{x,\alpha} \cdot G_{y,\alpha} & (G_{y,\alpha})^2 \end{bmatrix} \quad (5.3)$$

$$\mathcal{C}_{SH}(x, y, \alpha) = \det(M_{SH}(x, y, \alpha)) - k_2 \cdot \text{Tr}(M_{SH}(x, y, \alpha))^2$$

avec $a_2 > 1$ et k_2 des paramètres à définir. Les composantes du GR $G_{x,\alpha}$, $G_{y,\alpha}$ sont calculées sur l'image I à l'aide de l'Equation (4.27), α étant le paramètre de pondération exponentielle. De la même façon que le critère de Harris multi-échelle, cet opérateur permet de sélectionner des points situés sur des coins en effectuant le test :

$$\mathcal{C}_{SH}(x, y, \alpha) > d_{SH} \quad (5.4)$$

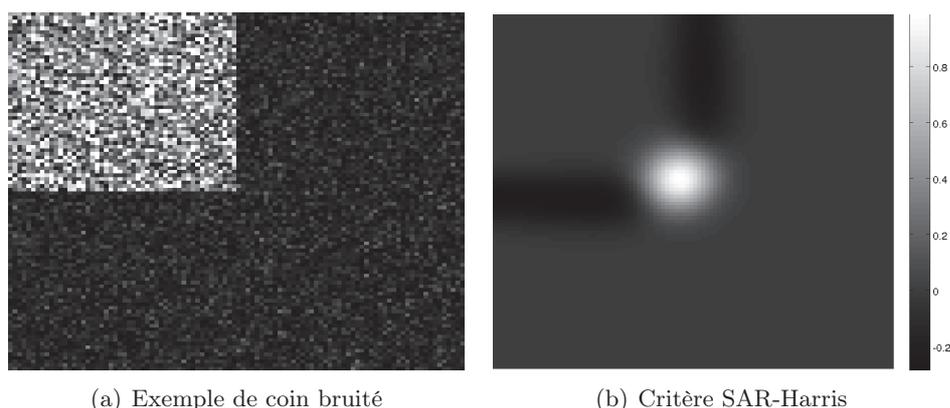


FIGURE 5.2 – Calcul du critère SAR-Harris $\mathcal{C}_{SH}(x, y, \alpha)$ sur un coin en présence de speckle pour $\alpha = 5$.

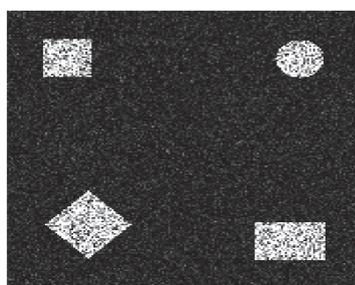


FIGURE 5.3 – Exemple d'image jouet avec du speckle

avec d_{SH} un paramètre dont le choix sera discuté au paragraphe (d). Comme la pondération exponentielle pour le calcul du GR permet d'obtenir des images de gradient à différentes échelles, on peut s'affranchir de l'étape de flou et calculer directement le gradient sur l'image d'origine. La normalisation σ^2 de la matrice de Harris n'est pas nécessaire ici pour assurer l'invariance au changement d'échelle (voir Annexe C.5). Nous utiliserons pour la suite les valeurs $a_2 = \sqrt{2}$ et $k_2 = 0.04$ comme pour le critère de Harris original.

La Figure 5.2 présente un exemple de calcul de ce critère sur une image synthétique d'un coin avec du bruit de speckle. On remarque que le critère répond de la même façon que le critère de Harris appliqué à une image non bruitée (Fig. 2.6) : les valeurs sont fortes au niveau du coin, négatives sur les bords et proche de zéro sur les zones homogènes. Ce détecteur est donc efficace pour détecter les coins. On peut observer Figure 5.4 la réponse du critère pour différentes échelles α sur l'image jouet de Figure 5.3. Ce critère présente des extrema sur les coins, mais aussi au centre d'objets à certaines échelles.

Nous proposons d'utiliser ce critère pour développer un détecteur de points-clés adapté aux images radar. L'idée est de s'inspirer du détecteur LoG-Harris en construisant un espace-échelle en calculant $\mathcal{C}_{SH}(x, y, \alpha)$ à différentes échelles et en sélectionnant les extrema. Pour cela, nous étudions dans un premier temps la réponse du critère sur un carré.

(b) Réponse du critère SAR-Harris sur un carré Nous souhaitons utiliser le critère SAR-Harris comme espace-échelle pour la détection de points-clés en sélectionnant les extrema locaux en échelle α et en espace (x, y) . Pour que l'information d'échelle soit pertinente, un objet et sa version sous-échantillonnée devraient donc être détectés à des échelles

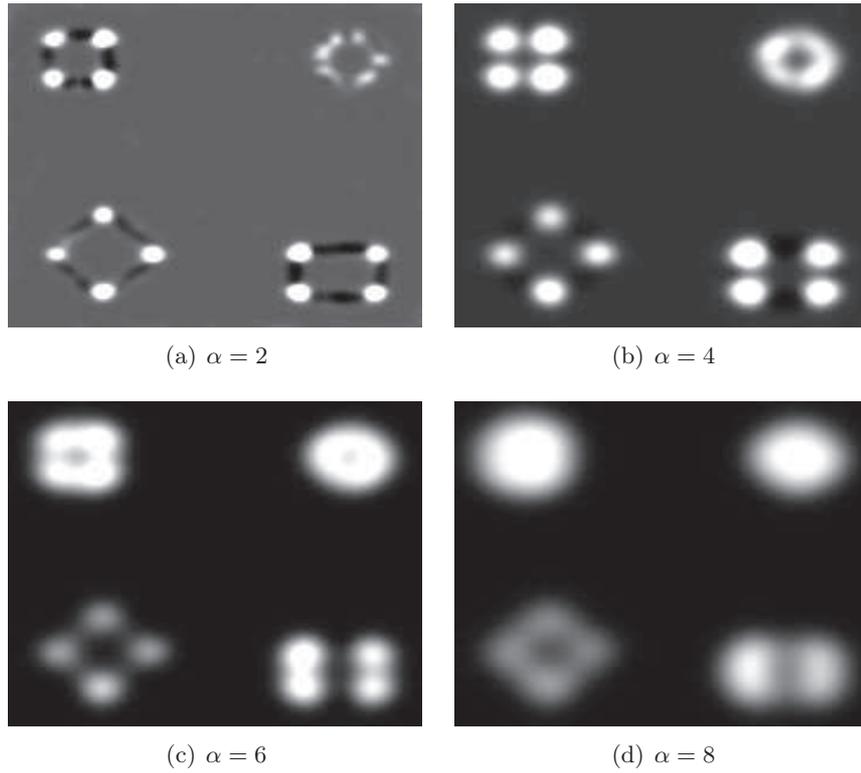


FIGURE 5.4 – Réponse du critère SAR-Harris $\mathcal{C}_{SH}(x, y, \alpha)$ sur une image jouet (Fig. 5.3) en présence de speckle pour différentes valeurs de α .

différentes. Il faut donc vérifier expérimentalement que, au centre de l'objet, l'échelle pour laquelle le critère de Harris est maximal dépend de sa taille de manière approximativement linéaire.

Pour cela, nous observons la réponse du critère au centre de carrés de différentes tailles, tout en faisant varier l'échelle α . On représente Figure 5.5(a) les courbes obtenues pour différentes tailles de carrés (t : demi-taille des carrés). On remarque que le maximum selon l'échelle n'est pas atteint pour les mêmes valeurs de α quand t varie. L'échelle correspondant à ce maximum dépend linéairement de la longueur du carré (Fig. 5.5(b)). La sélection des extrema en échelle permet donc de sélectionner l'échelle caractéristique de l'objet étudié.

Cependant nous avons remarqué expérimentalement que le critère SAR-Harris atteint rarement des extrema locaux en trois dimensions. Cette situation a aussi été observée pour le critère de Harris multi-échelle [MS01]. Il est ainsi préférable de se limiter à une sélection des extrema en 2D et à un seuillage des faibles valeurs. Plusieurs points-clés peuvent alors être détectés à la même position (x, y) mais avec des échelles α différentes. Le seuil d_{SH} permet de conserver les points (x, y, α) correspondant à une valeur maximale du critère pour les différentes valeurs de α détectées. L'échelle de ces points est alors caractéristique de la taille de l'objet détecté.

(c) Détection de points-clés SAR-Harris Nous proposons de remplacer l'espace-échelle LoG par une représentation multi-échelle de l'image, obtenue en calculant le critère SAR-Harris (Eq. 5.3) à différentes échelles $\alpha_n = \alpha_0 \cdot c^m$ (avec α_0, c des réels fixes et m prenant des valeurs entières dans $\{0, \dots, m_{max}\}$). Les extrema locaux en espace sont ensuite sélectionnés à chaque échelle en 8-connexité. Les positions sous-pixelles des points-clés sont obtenues à l'aide d'une interpolation bilinéaire de $\mathcal{C}_{SH}(x, y, \alpha)$ autour des extrema

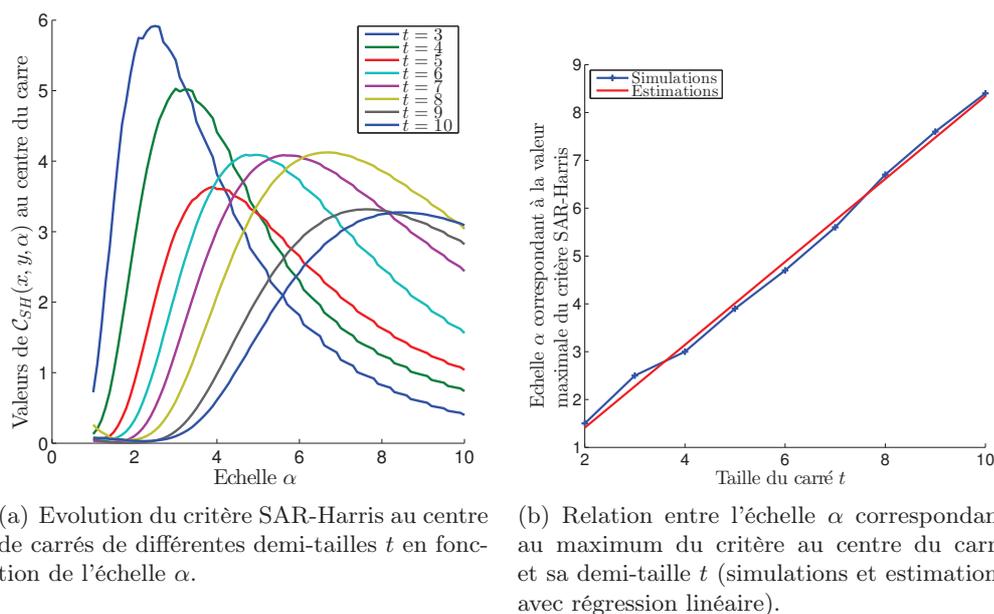


FIGURE 5.5 – Relation entre la taille d'un carré et l'échelle à laquelle il serait détecté.

locaux. Un seuil d_{SH} sur $C_{SH}(x, y, \alpha)$ permet de filtrer les points de bords ou de faible contraste (Eq. 5.4). Nous obtenons alors des points caractérisés par leur position (x, y) et leur échelle α .

Cette approche, que nous définissons comme le détecteur SAR-Harris, regroupe les deux étapes du détecteur LoG-Harris de façon à éviter l'utilisation de dérivées secondes. Elle a l'avantage d'être invariante à un changement multiplicatif du contraste global de l'image, de la même façon que le GR (Section 4.2.3).

(d) Choix des paramètres

Seuil sur le critère SAR-Harris Le seuil d_{SH} doit permettre de filtrer les détections liées au bruit et les détections sur les bords. Cependant il ne doit pas être très important pour ne pas supprimer des points intéressants mais faiblement contrastés. Pour cela nous étudions la réponse du critère SAR-Harris sur une zone homogène, un bord et un coin. La Figure 5.6 présente la répartition des valeurs du critère pour ces trois situations. Les données sont mesurées pour différentes valeurs de α et de contraste local $C = \frac{R_1}{R_2}$ en intensité. En effet, le critère SAR-Harris est invariant au contraste global de l'image, il ne dépend que du contraste local.

Les observations sont conformes à ce qu'on attend de ce critère. Les valeurs sont proches de zéro sur une zone homogène, alors qu'elles sont négatives sur un bord et largement positives sur un coin. Dans les trois cas l'augmentation de la valeur de α permet de resserrer les distributions. En effet, l'image est ainsi plus lissée et on trouve donc plus souvent les mêmes valeurs de gradient. On note que les valeurs sur les coins et les bords se rapprochent de zéro quand les contrastes locaux C diminuent. Dans ces situations, les courbes de répartition ont tendance à se chevaucher et il sera alors difficile de sélectionner les coins.

Afin de déterminer le seuil d_{SH} optimal, nous observons alors les probabilités de fausses alarmes et de détection du critère pour différentes valeurs de α et C . On cherche un compromis entre un fort taux de détection (TD) sur les coins et un faible taux de fausses alarmes (TFA) sur les zones homogènes et les bords. On observe les seuils d_{SH} permettant d'obtenir un TD de 99% et un TFA de 0.1%, en faisant varier α et C (Figure 5.7). Il y a

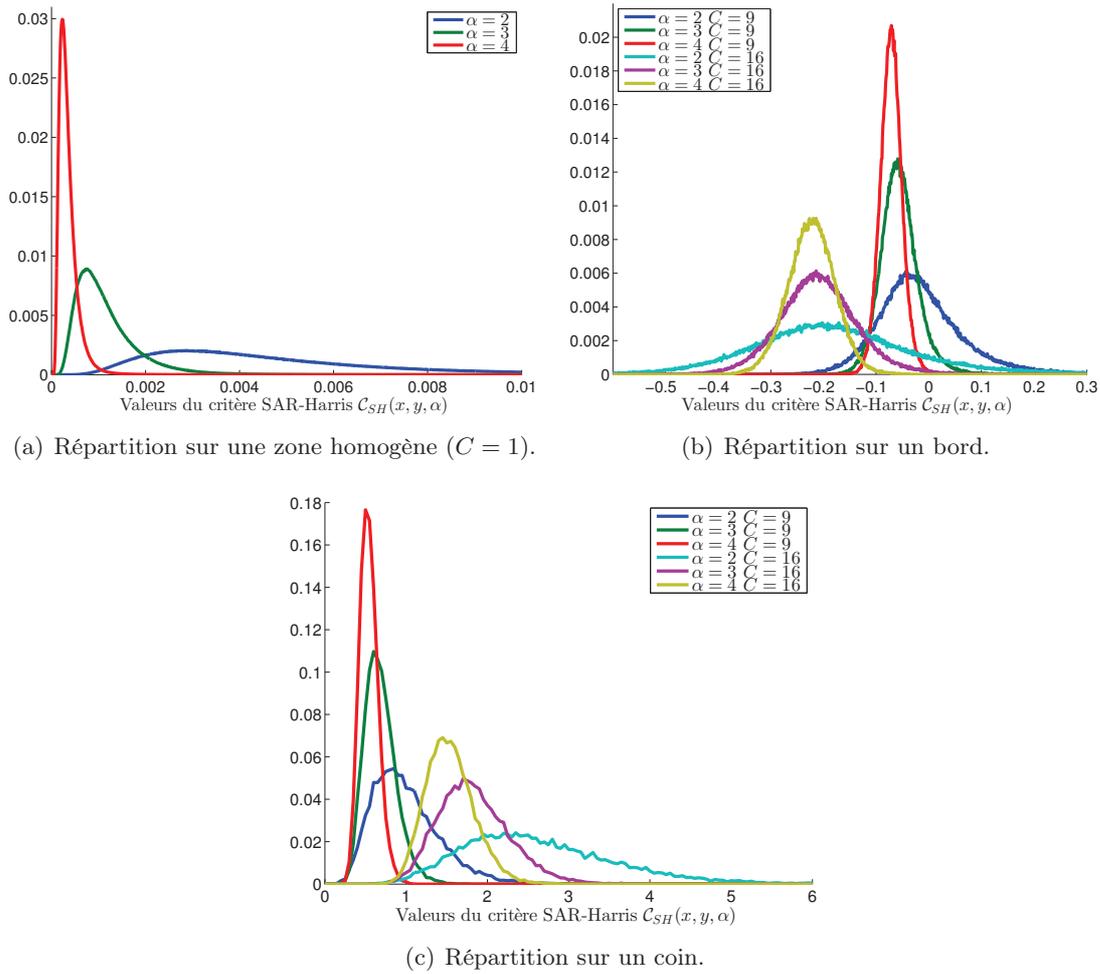


FIGURE 5.6 – Histogrammes normalisés des valeurs du critère SAR-Harris $\mathcal{C}_{SH}(x, y, \alpha)$ en fonction de α et du contraste de la zone $C = \frac{R_1}{R_2}$, pour différentes configurations géométriques. Pour les trois configurations, $R_1 = 100$.

chevauchement entre les seuils pour les coins et bords avec un contraste $C < 16$, et tout particulièrement pour les bords avec de petites échelles ($\alpha < 4$). On choisit comme seuil $d_{SH} = 0.8$, afin d'éviter en grande majorité les fausses détections, tout en maintenant un fort taux de détection sur les coins pour $C \geq 16$.

Échelles On choisit d'utiliser les mêmes échelles α que dans l'algorithme SIFT, à l'exception des plus faibles. En effet comme présenté en Section 3.1, la première octave ($\sigma < 2$) est à l'origine de nombreuses fausses détections. Nous venons également de voir que plus l'échelle α est proche de 1, plus les valeurs du critère SAR-Harris sont étalées. Cela entraîne donc une diminution du TD et une augmentation du TFA. Nous choisissons finalement les valeurs $\alpha_n = \alpha_0 \cdot c^m$ pour $m \in \{0, \dots, m_{max}\}$, avec $\alpha_0 = 2$, $c = 2^{1/3}$ et $m_{max} = 7$.

(e) **Exemples** La Figure 5.8 présente le calcul des points-clés avec la méthode SAR-Harris sur un carré avec du speckle. On trouve bien des points-clés sur les coins, comme attendu, et aucune fausse détection. Sur la Figure 5.9, on observe la détection de points-clés sur l'image jouet de la Figure 5.3 avec et sans le seuil d_{SH} . Le seuil est efficace pour

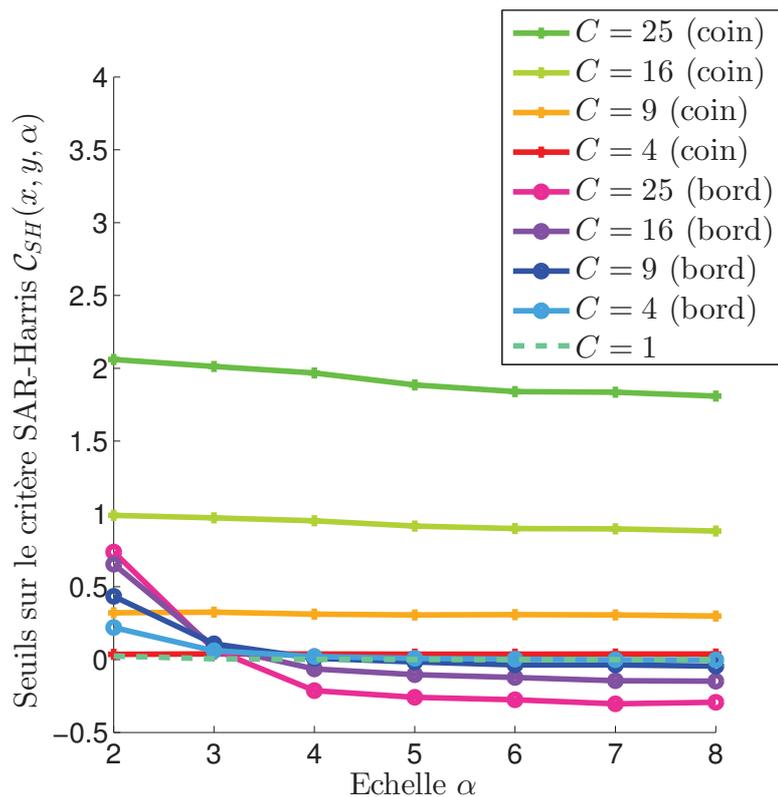


FIGURE 5.7 – Seuils permettant d’avoir un taux de détection de 99% et un taux de fausses alarmes de 0,1% en fonction de l’échelle α pour différentes valeurs de contraste C .

supprimer les fausses détections. Des points-clés sont détectés au niveau des coins des objets, mais finalement assez peu en leurs centres.

La Figure 5.10 présente un exemple de détection de points-clés pour plusieurs détecteurs. Comme attendu, on observe que les points-clés détectés avec la méthode LoG-Harris appliquée sur l’image d’amplitude se trouvent principalement sur les zones de fortes réflectivités. De nombreuses fausses détections se produisent sur les zones homogènes de forte radiométrie. Concernant les points-clés obtenus avec la méthode LoG-Harris sur le logarithme, on trouve aussi beaucoup de fausses détections sur les zones homogènes mais elles apparaissent autant sur les zones de fortes que faibles réflectivités. En revanche, les points-clés détectés avec la méthode SAR-Harris sont en grande majorité positionnés sur les coins et points brillants. Le nombre de fausses détections est aussi très faible.

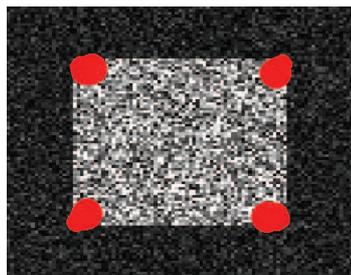


FIGURE 5.8 – Calcul des points-clés sur le carré de Fig. 5.1(a) à l’aide de la méthode SAR-Harris. Plusieurs points-clés sont superposés au niveau des coins.

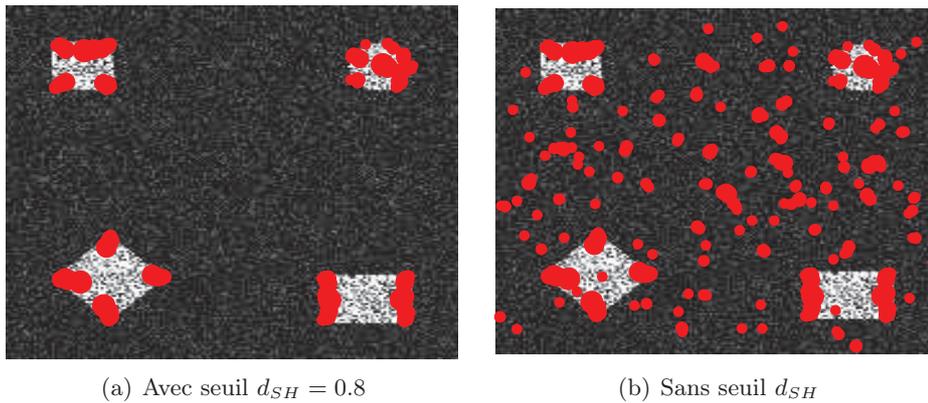


FIGURE 5.9 – Calcul des points-clés de l'image jouet de Fig. 5.3 avec la méthode SAR-Harris, avec et sans le seuil d_{SH} .

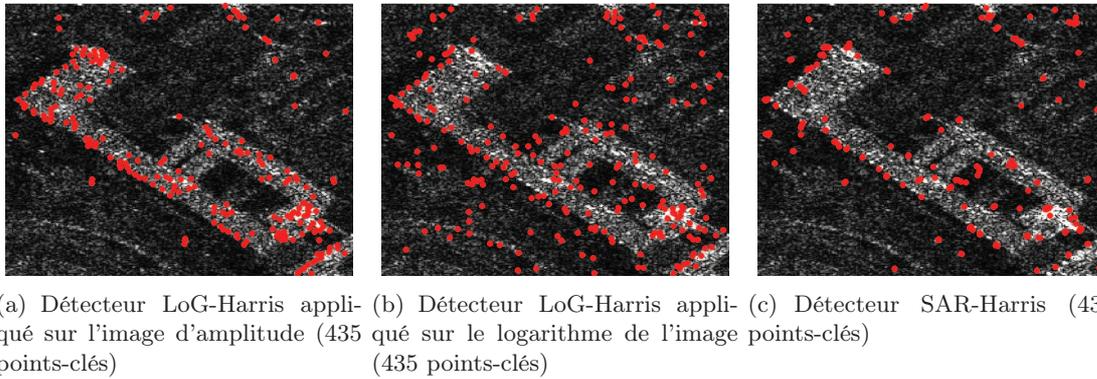


FIGURE 5.10 – Détection de points-clés sur l'image de Fig. 4.12(a) avec le détecteur LoG-Harris, appliqué sur l'image d'amplitude (Fig. 5.10(a)) et le logarithme (Fig. 5.10(b)), et le détecteur SAR-Harris (Fig. 5.10(c)). Le seuil d_h a été ajusté pour obtenir le même nombre de points-clés à chaque détection.

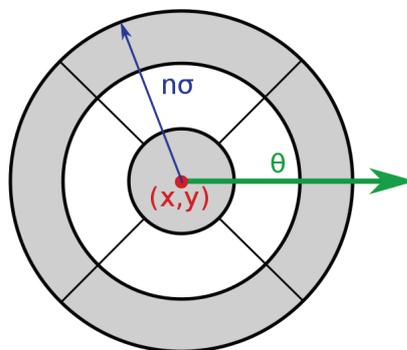


FIGURE 5.11 – Masque du descripteur circulaire, découpé en 9 secteurs. Le voisinage est recentré sur l'orientation θ du point-clé. Le rayon du descripteur $n\sigma$ dépend de l'échelle du point-clé ($n = 12$ ici). Les coefficients appliqués à $n\sigma$ pour obtenir les rayons des cercles intérieurs et extérieurs sont respectivement 0.25 et 0.73.

5.1.2 Sélection des orientations et extraction des descripteurs

Dans l'algorithme SIFT original, ces deux étapes de calcul d'orientations et d'extractions de descripteurs reposent sur des histogrammes d'orientations du gradient. Ces histogrammes sont calculés sur un voisinage carré de chaque point-clé.

Ici nous proposons d'utiliser le Gradient par Ratio (GR), présenté en Section 4, à la place du gradient par différence pour construire ces histogrammes. Dans la suite nous appellerons les échelles des points-clés σ comme dans l'algorithme SIFT, plutôt que α .

Assignment des orientations En un point-clé (x, y, σ) , un histogramme d'orientation du gradient est construit sur un voisinage circulaire de rayon 6σ autour du point-clé étudié. L'histogramme est construit sur 36 classes. Les modes principaux de l'histogramme sont sélectionnés avec la méthode *a contrario* [Rab09] présentée en Section 2.2.2. Jusqu'à deux orientations peuvent être choisies pour chaque point : une si un seul mode de l'histogramme est significatif, deux si deux modes ou plus sont significatifs.

Extraction des descripteurs Au lieu d'un voisinage carré divisé en secteurs de 4×4 pixels, nous utilisons un voisinage circulaire et des secteurs log-polaires, comme les descripteurs GLOH [MS05] et SIFT-circulaire [Rab09] (Section 2.3.5). La Figure 5.11 présente le schéma de calcul (le même que pour SIFT-circulaire). Autour du point (x, y) , nous considérons un voisinage circulaire de rayon 12σ , divisé en 9 secteurs log-polaires. Sur chacun des secteurs, on construit sur 12 classes un histogramme d'orientation du gradient pondéré par la norme du gradient, en utilisant le GR. Les orientations sont calculées relativement à θ , l'orientation du point-clé. Tous ces secteurs sont ensuite regroupés dans un vecteur normalisé. Le descripteur obtenu est appelé *descripteur Ratio*. Dans le cas où deux orientations θ ont été sélectionnées pour le même point-clé, on calcule deux descripteurs.

Ce descripteur est construit de la même façon que le descripteur SIFT-circulaire. La différence se trouve dans le calcul du gradient : gradient par différence pour SIFT-circulaire et GR pour Ratio.

5.2 Evaluations

Dans cette section, nous évaluons les performances du nouvel algorithme SAR-SIFT en comparaison avec l'algorithme SIFT original. Nous utilisons ici l'implémentation décrite en Annexe B. Le détecteur SAR-Harris est donc comparé au détecteur LoG-Harris (Section 2.1.2.4), et le descripteur Ratio au descripteur SIFT-circulaire (Section 2.3.5). Nous étudions tout d'abord la stabilité et la robustesse des détecteurs de points-clés en mesurant leur taux de répétabilité. Ensuite des courbes ROC (Receiver Operating Characteristic) permettent d'évaluer les performances des détecteurs et des descripteurs, puis de mesurer la pertinence de l'utilisation du GR plutôt que du gradient par différence pour le calcul des orientations.

Ici, nous souhaitons apprécier le comportement de SAR-SIFT face à du bruit de speckle. L'étude est réalisée sur des paires d'images acquises dans les mêmes conditions. Nous utilisons 18 paires d'images de taille 512×512 . Ces extraits proviennent de deux images TerraSAR-X acquises dans les mêmes conditions (Données 2 présentées en Annexe A.1). Pour chaque paire, les images présentent un recalage sous-pixelique et aucun changement temporel apparent. Seule la réalisation de bruit diffère. Deux paires de ce type sont présentées en Figure 5.13.

Les paramètres utilisés pour l'algorithme SAR-SIFT sont ceux présentés en Section 5.1. Pour l'algorithme SIFT, ils sont détaillés en Annexe B. Cependant comme les images radar ont des dynamiques différentes, le seuil d_H sera adapté à chaque image afin d'obtenir

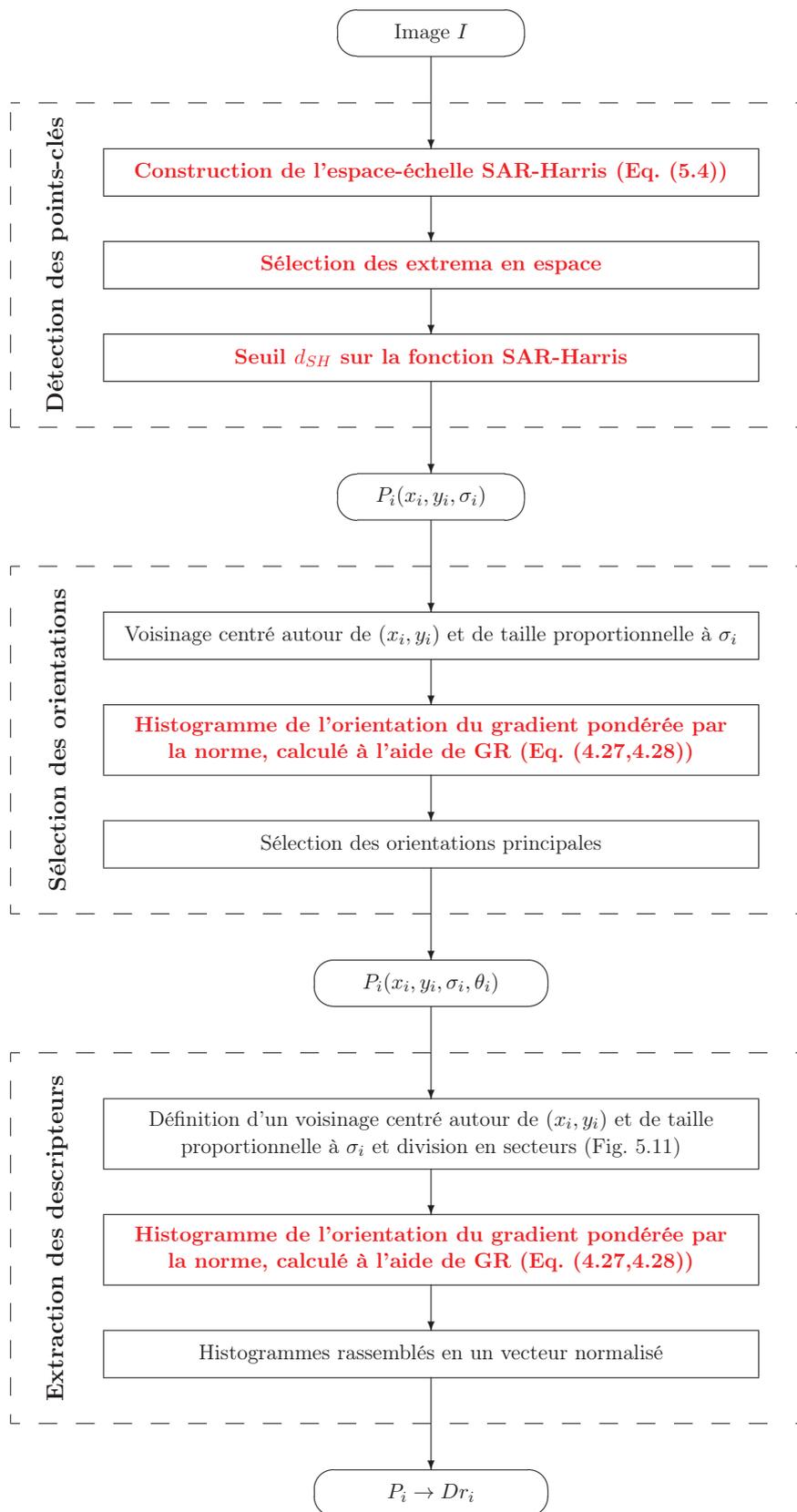
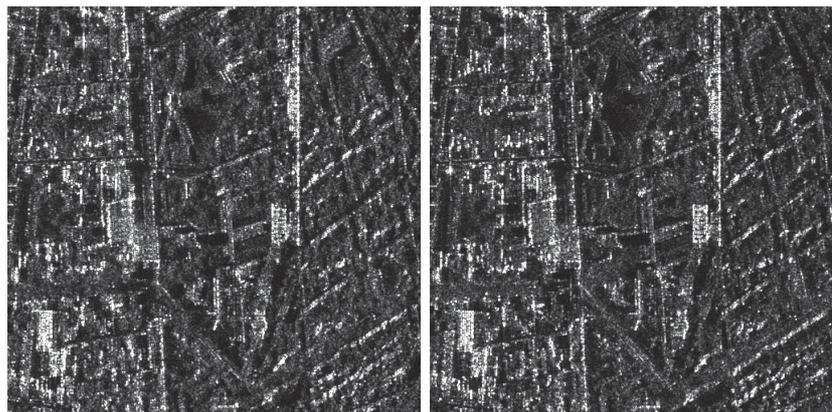
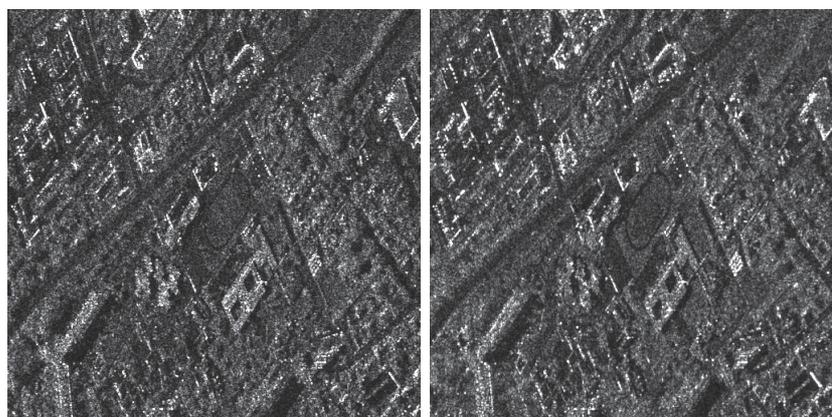


FIGURE 5.12 – Schéma de l’algorithme SAR-SIFT : détection de points-clés SAR-Harris, sélection des orientations et extraction des descripteurs à l’aide de GR. Les contributions apportées apparaissent en rouge.



(a) Paire 1



(b) Paire 2

FIGURE 5.13 – Exemple de deux paires d’images recalées utilisées pour les expériences de la Section 5.2.

environ le même nombre de points-clés détectés quelque soit la méthode de détection (à 5% près).

5.2.1 Répétabilité des détecteurs de points-clés

Le critère de répétabilité [SMB00] donne une mesure de la stabilité des détecteurs de points-clés, en présence de perturbations sur les images. Ici la perturbation étudiée est le bruit de speckle. On souhaite que la détection SAR-Harris soit indépendante de la réalisation du bruit. Le nouveau détecteur SAR-Harris est comparé au détecteur LoG-Harris, appliqué sur l’image d’amplitude et le logarithme.

La méthode de calcul du critère est la suivante. Soit une paire d’images recalées. Pour chaque point-clé de la première image, nous cherchons le plus proche détecté sur l’autre image avec le même détecteur. Ensuite nous mesurons, pour différents seuils u , le nombre de points-clés répétés sur l’autre image à une distance inférieure à u . La Figure 5.14 présente, pour chaque détecteur, le pourcentage de points-clés répétés sur l’ensemble des paires en fonction de l’erreur de localisation u . Les seuils d_H ont été adaptés afin d’obtenir en moyenne le même nombre de points-clés que le détecteur SAR-Harris. Nous obtenons pour tout le jeu de données un total de 25032 points-clés extraits avec le détecteur LoG-Harris, 24729 points-clés quand le logarithme de l’image est utilisé et 21253 avec le détecteur SAR-Harris. La densité de points-clés est donc presque la même pour chaque détecteur.

On remarque que le détecteur SAR-Harris est plus performant que les deux détecteurs

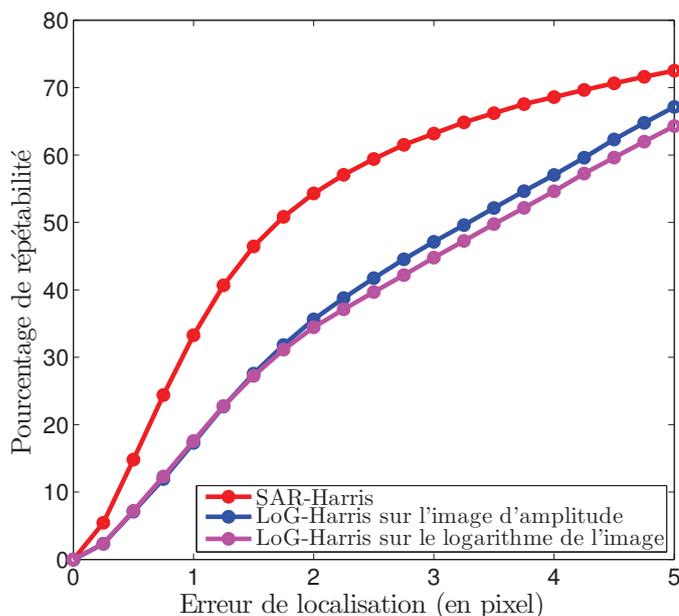


FIGURE 5.14 – Taux de répétabilité des points-clés, construits sur 18 paires d’images, en fonction de l’erreur de localisation. Les points-clés sont extraits avec trois détecteurs différents : le détecteur LoG-Harris appliqué sur l’image d’amplitude et le logarithme, et le détecteur SAR Harris.

LoG-Harris. Par exemple pour une erreur de localisation $u = 1.5$ pixels, plus de 50% des points-clés extraits par le détecteur SAR-Harris sont répétés. Ce taux est seulement de 30% pour les points extraits avec les deux autres méthodes. Nous n’observons pas d’améliorations des performances du détecteur LoG-Harris quand il est appliqué au logarithme de l’image.

Le détecteur SAR-Harris est donc robuste face à du bruit de speckle.

5.2.2 Performances des mises en correspondance

Des courbes ROC sont construites pour différentes combinaisons de détecteurs de points-clés et de types de descripteurs. Les points-clés de deux images différentes sont associés à l’aide de leurs descripteurs respectifs. Nous utilisons ici la méthode d’association NNDR, décrite en Section 2.4.1. Cette méthode est constituée de deux étapes : l’association du plus proche voisin (étape NN) et le filtrage du rapport des distances (étape DR).

Soit $M_k(P_{1,i}, P_{2,j})$ une mise en correspondance entre un point $P_{1,i}(x_{1,i}, y_{1,i}, \sigma_{1,i}, \theta_{1,i})$ d’une image I_1 et un point $P_{2,j}(x_{2,j}, y_{2,j}, \sigma_{2,j}, \theta_{2,j})$ d’une image I_2 . Considérant la déformation T de l’image I_1 par rapport à l’image I_2 , la mise en correspondance M_k est définie comme correcte si :

$$\|T(x_{1,i}, y_{1,i}) - (x_{2,j}, y_{2,j})\|_2 < t_1 \cdot \min(\sigma_1, \sigma_2), \quad (5.5)$$

Le paramètre t_1 a été fixé empiriquement à 5. Ici les paires d’images sont recalées avec une précision sous-pixelle, donc il n’y a pas de déformation entre les images et T est la fonction identité.

Dans un premier temps, tous les points $P_{1,i}$ sont associés à leur plus proche voisin $P_{2,j}$ (étape NN). On définit alors les quantités $\#CM_{all}$ et $\#FM_{all}$ comme le nombre total de mises en correspondance correctes et fausses sur toutes les paires. Les mises en correspondances M_k sont ensuite filtrées avec un seuil th sur le rapport des distances au premier et

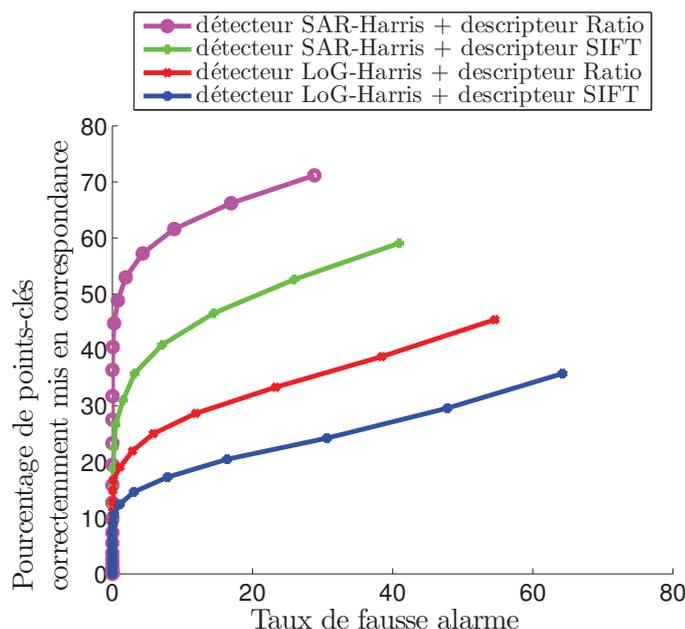


FIGURE 5.15 – Courbes ROC, construites sur 18 paires d’images, afin d’évaluer les performances des détecteurs LoG-Harris et SAR-Harris, ainsi que des descripteurs Ratio et SIFT-circulaire.

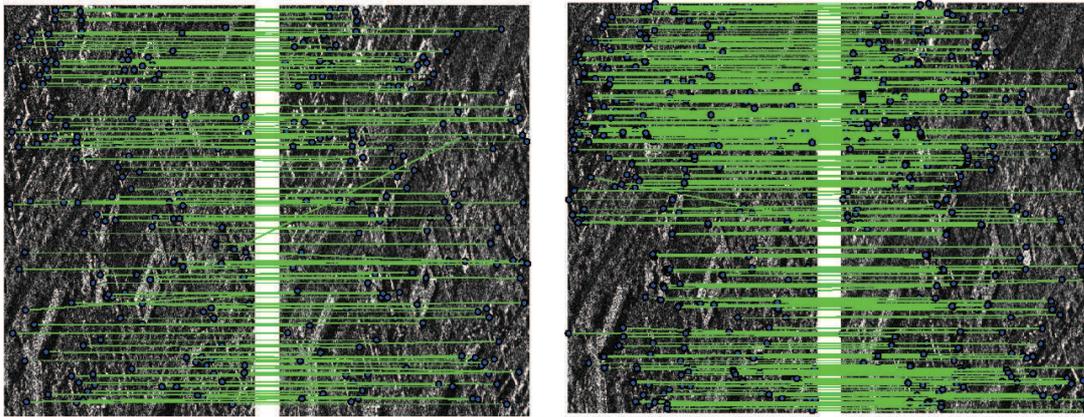
deuxième plus proche voisin (étape DR). On nomme $\#CM$ et $\#FM$ le nombre de mises en correspondances correctes et fausses obtenus pour ce seuil. Pour obtenir des courbes ROC, on représente le pourcentage de points-clés correctement associés $\frac{\#CM}{\#CM_{all} + \#FM_{all}}$ en fonction du taux de fausses alarmes $\frac{\#FM}{\#CM + \#FM}$ en faisant varier th .

Nous comparons ici deux détecteurs : LoG-Harris sur l’image d’amplitude (Section 2.1.2.4) et SAR-Harris (Section 5.1.1.2), ainsi que deux descripteurs, SIFT-circulaire (Section 2.3.5) et Ratio (Section 5.1.2). Les résultats des quatre situations ainsi considérées sont représentés en Figure 5.15. Comme les images ne présentent pas de rotation, aucune orientation n’est assignée aux points-clés pour cette expérience. Les distances L^1 et L^2 ont été testées pour le calcul des similarités entre descripteurs. Nous avons observé que les résultats avec la distance L^1 sont toujours meilleurs que ceux avec la distance euclidienne. Pour simplifier la lecture, seuls les résultats avec la distance L^1 sont affichés.

Nous observons que les meilleures performances sont obtenues avec une combinaison du détecteur SAR-Harris et du descripteur Ratio. En effet, pour un taux de fausses alarmes de 1%, presque 50% des associations sont correctes. Pour les autres configurations, ce taux est inférieur à 30%. Ce taux de fausses alarmes est le pourcentage de fausses associations parmi toutes les mises en correspondance obtenues avec un seuil th . Nous remarquons que l’utilisation du détecteur SAR-Harris avec le descripteur SIFT-circulaire améliore déjà grandement les performances de l’algorithme. L’utilisation du détecteur LoG-Harris avec le descripteur Ratio apporte une amélioration notable mais moins importante.

Pour cette expérience, les taux de fausses alarmes 1% et 0.1% sont obtenus pour des seuils respectifs $th = 0.88$ et $th = 0.83$.

On montre 5.16 un exemple de mises en correspondance entre les deux images de Fig. 5.13(b), effectuées avec l’algorithme SIFT original (détecteur LoG-Harris + descripteur SIFT-circulaire) et l’algorithme SAR-Harris (détecteur SAR-Harris + descripteur Ratio). Ce deuxième algorithme permet d’obtenir plus de trois fois plus de bons appariements



(a) Détecteur LoG-Harris + descripteur SIFT-circulaire : 526 mises en correspondances dont 1 mises en correspondance vraie et 2 fausses.
 (b) Détecteur SAR-Harris + descripteur Ratio : 526 mises en correspondances dont 1 mises en correspondance vraie et 2 fausses.

FIGURE 5.16 – Exemple de mises en correspondance entre les deux images de Fig. 5.13(b). Les appariements avec $th < 0.85$ sont affichés.

pour un même seuil $th = 0.85$.

Pour résumer, le détecteur SAR-Harris est plus stable et robuste face au bruit que le détecteur LoG-Harris et le descripteur Ratio surpasse le descripteur SIFT-circulaire. La combinaison de SAR-Harris et du descripteur Ratio est utilisée dans l'algorithme SAR-SIFT (Fig. 5.12). Cet algorithme atteint ainsi de meilleurs résultats que l'algorithme SIFT original sur les images radar.

5.2.3 Invariance à la rotation

L'invariance à la rotation est obtenue grâce aux orientations assignées à chaque point-clé. Pour l'algorithme SAR-SIFT, nous avons choisi d'utiliser une méthode *a contrario* avec le calcul du Gradient par Ratio. Nous souhaitons évaluer l'intérêt de cette méthode par rapport à la méthode des maxima de l'histogramme (Section 2.2.1), mais aussi apprécier l'opportunité d'utiliser le GR plutôt que le gradient par différence pour le calcul des histogrammes.

Pour cela, nous construisons des courbes ROC à l'aide de paires d'images qui diffèrent d'une rotation. Quatre méthodes d'assignation d'orientation sont testées, soit les différentes combinaisons entre les deux méthodes d'assignation et les deux calculs de gradient possibles. Nous utilisons la méthode SAR-Harris pour le calcul des points-clés et le descripteur Ratio pour l'extraction des descripteurs. Les paires d'images utilisées sont les mêmes que précédemment, à la différence que pour chaque paire, une image a subi une rotation de 90° . Cela permet d'effectuer une rotation sans pour autant perturber la géométrie de l'image ou effectuer une interpolation qui pourrait modifier les propriétés statistiques des pixels. Les points-clés sont ensuite reprojétés sur l'image d'origine et la classification des mises en correspondance se fait de la même façon que dans la section précédente (Eq. 5.5).

Les résultats sont présentés sur la Figure 5.17. On remarque que les quatre calculs d'orientations offrent des résultats assez proches pour de faibles taux de fausses alarmes. Le calcul des orientations avec la méthode *a contrario* et le GR offre cependant les meilleurs performances. Pour un taux de fausses alarmes de 1%, cette méthode offre 13,3% de bons appariements, contre 9,94% quand on choisit d'utiliser le gradient par différence. L'utilisation du GR ne fournit ainsi qu'une modeste amélioration. Ce résultat s'explique

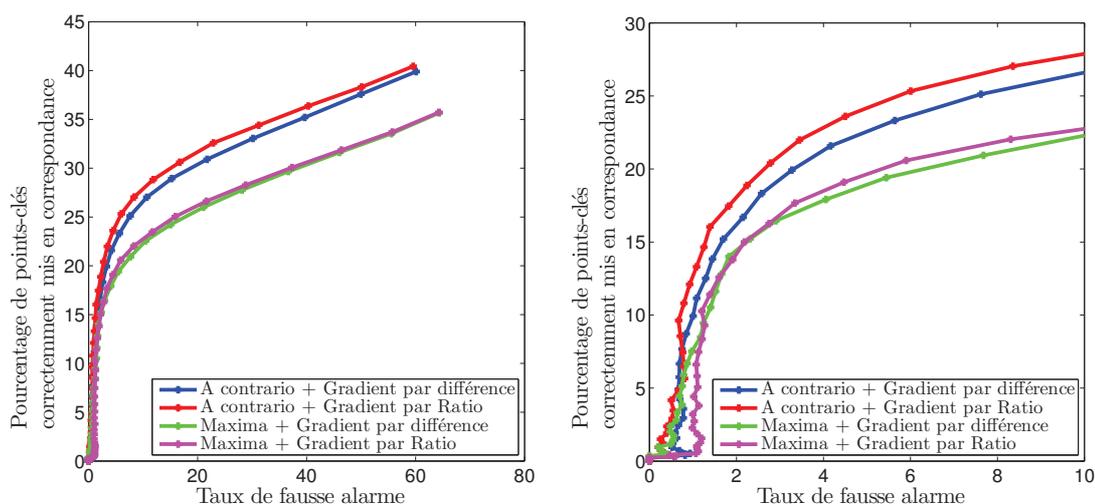


FIGURE 5.17 – Courbes ROC, construites sur 18 paires d’images présentant une rotation, afin d’évaluer les performances du calcul des orientations des points-clés. La figure de droite présente un recentrage des résultats pour de faibles taux de fausses alarmes.

par le fait que, dans le cas du gradient par différence, les histogrammes d’orientation du gradient sont calculés sur des images avec un flou gaussien et sur un voisinage important (de rayon 6σ). Par conséquent, l’influence du speckle est fortement atténuée et le calcul d’orientation est assez fiable. L’avantage apporté par le GR est assez faible pour cette étape.

Par contre, pour le calcul du descripteur Ratio, les histogrammes sont calculés sur de plus petites zones (12 secteurs sur un voisinage de rayon 12σ). L’intérêt du GR est donc plus évident pour l’étape d’extraction des descripteurs.

En conclusion, les trois premières étapes de l’algorithme SAR-SIFT (détecteur SAR-Harris, calcul des orientations avec GR et descripteur Ratio) offrent de meilleurs résultats que celles de l’algorithme SIFT original.

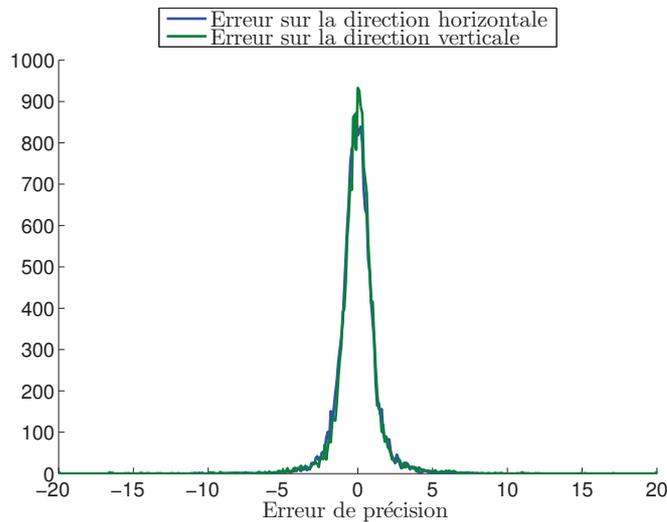
5.3 Précision de l’appariement des points-clés

Cette section étudie la précision de la mise en correspondance des points-clés de l’algorithme SAR-SIFT, à l’aide d’une expérience synthétique avec transformations simulées. On considère une mise en correspondance, supposée correcte, entre un point-clé $kp_{1,i}$ d’une image I_1 et un point-clé $kp_{2,j}$ d’une image I_2 . La transformation entre les images I_1 et I_2 est une similitude. D’autre part, les images ne présentent pas la même réalisation de bruit. On étudie la distance entre le point-clé $kp_{1,i}$ et le projeté $kp_{2,j}$ sur l’image I_1 . En observant les distances pour un ensemble d’appariement, on obtient une idée de leur précision.

Présentation de l’expérience L’étude est réalisée à partir d’une image radar débruitée I_0 (100 vues) de taille 409x409 (Figure 5.18). Des transformations T_i sont appliquées à I_0 pour obtenir des images I_{T_i} . Nous avons choisi de nous limiter à des similitudes, soit une rotation d’angle ϑ et un changement d’échelle de facteur ρ . Les points-clés, orientations et descripteurs sont extraits sur une version bruitée de I_0 et I_{T_i} en utilisant l’algorithme SAR-SIFT présenté en Section 5.1. La méthode d’appariement utilisée est le calcul de la distance L^1 entre descripteurs et la méthode NNDR. Les points-clés $kp_{2,j}$ extraits sur l’image I_{T_i} sont reprojétés sur l’image I_0 . On calcule ensuite le vecteur de déplacement



FIGURE 5.18 – Image 100 vues du CNES à Toulouse.

FIGURE 5.19 – Distribution de l’erreur de précision des points-clés appariés pour $\rho = 2$ et $\vartheta = 0$.

entre les points-clés $kp_{2,j}^r$ reprojétés et leurs correspondants $kp_{1,i}$. Les points-clés sont tous reprojétés sur l’image d’origine I_0 , les distances sont donc calculées avec la même échelle.

Deux expériences sont réalisées. Dans la première on étudie le changement d’échelle. Le facteur d’échelle ρ varie de 0,5 à 2 et il n’y a pas de rotation. Dans la deuxième, on s’intéresse à la rotation. L’angle de rotation ϑ (en radian) varie de 0 à π et il n’y a pas de changement d’échelle. Pour chaque valeur de ρ ou ϑ , on réalise 20 fois le même test avec des réalisations de bruit différentes pour I_{Ti} . On regroupe les mises en correspondances pour former un ensemble $M_1(\rho, \vartheta)$ de mises en correspondances.

Observations Pour chaque valeur de ρ et ϑ , on observe la distribution des composantes horizontales et verticales du vecteur de déplacement des points-clés de tous les appariements de $M_1(\rho, \vartheta)$. On choisit comme seuil $th = 0.8$ pour limiter le nombre de fausses alarmes.

La Figure 5.19 représente un exemple de ces distributions pour $\rho = 2$ et $\vartheta = 0$. On

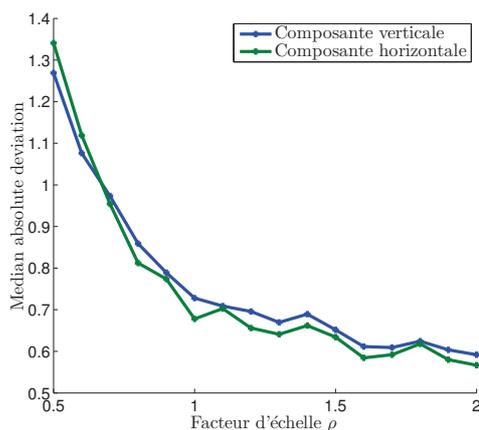


FIGURE 5.20 – Influence du facteur d'échelle ρ sur le MAD des composantes horizontale et verticale de l'erreur de précision entre points-clés appariés.

remarque que la distribution des deux composantes est centrée en zéro. Le seuil th a permis de supprimer la quasi totalité des outliers. On est passé de 25% de faux appariements (norme de l'erreur de précision supérieure à 10 pixels) à 0.15%. L'écart-type de la distribution donne une idée de la précision des points-clés mis en correspondance. Pour le déterminer, nous avons choisi d'utiliser l'estimateur Median Absolute Deviation (MAD), plus robuste à un fort nombre d'outliers que le calcul traditionnel de l'écart-type. Pour une distribution $\mathbf{X} = X_i, i \in \{1, \dots, N\}$, le MAD se calcule de la façon suivante

$$\text{MAD}(\mathbf{X}) = \text{median}(|X_i - \text{median}(X_j)|). \quad (5.6)$$

Pour cet exemple, le MAD donne 0.59 et 0.57 pour les composantes horizontale et verticale, quand on obtient 2.62 et 2.66 avec le calcul traditionnel de l'écart-type. Les valeurs du MAD sont plus conformes à l'allure de la distribution. On mesure aussi ici que 84% des points-clés sont appariés avec une erreur inférieure à 2 pixels. La précision des points-clés mis en correspondance est donc très élevée ici.

Les études suivantes présentent l'évolution du MAD en fonction de ρ et ϑ .

Influence du changement d'échelle La Figure 5.20 présente l'évolution du MAD en fonction du facteur d'échelle. Il n'y a pas eu de rotation des images ici. Les valeurs du MAD sont sensiblement les mêmes pour les composantes verticale et horizontale du vecteur de déplacement entre points-clés appariées. Elles sont assez faibles et varient de 1.35 à 0.57. Quand la taille de l'image diminue (faible valeur de ρ), le MAD augmente fortement. Inversement, il diminue faiblement quand la taille de l'image augmente (forte valeur de ρ).

Le changement de taille d'une image influe donc peu sur la précision des mises en correspondance, sauf quand l'image devient trop petite. En effet, dans ce contexte, moins de points-clés sont détectés. Cela pourrait être amélioré en utilisant des échelles plus faibles dans la pyramide de détection des points-clés mais on risque alors de se retrouver avec beaucoup de fausses détections instables. Aussi, même avec une détection sous pixelique des points-clés, la précision de la détection est moindre sur une image de taille deux fois plus petite en nombre de pixels. A l'inverse, plus la taille de l'image augmente, plus on a de points-clés. Cela peut expliquer la légère amélioration de la précision.

La précision des points-clés mis en correspondance est donc très importante, tant que la taille d'une image et donc son contenu sont significatifs.

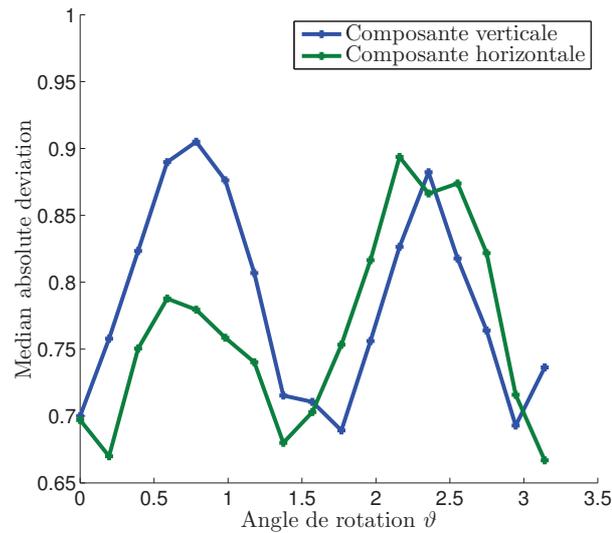


FIGURE 5.21 – Influence de l’angle de rotation ϑ sur le MAD des composantes horizontale et verticale de l’erreur de précision entre points-clés appariés.

Influence de la rotation La Figure 5.21 présente l’influence de la rotation sur le Median Absolute Difference des composantes du vecteur de déplacement entre points-clés appariés. Les valeurs du MAD restent faibles, entre 0.67 et 0.9. Elles présentent cependant une augmentation en $\pi/4$ et en $3\pi/4$, à la fois pour la composante verticale et pour celle horizontale. Ces rotations correspondent à une plus forte déformation de l’image, ce qui peut donc diminuer la précision des points-clés mis en correspondance. Les performances dans ces cas ($\pi/4$ et $3\pi/4$) restent cependant acceptables (valeurs du MAD inférieures à 1).

Conclusions Les transformations étudiées (rotation et changement d’échelle) dégradent peu les performances de la précision de la mise en correspondance des points-clés. L’écart-type des composantes vectorielles des distances entre points-clés appariés (après reprojection) est assez faible (inférieur à 1). Cependant quand la taille de l’image devient trop petite, la précision est fortement dégradée. Dans une moindre mesure, la rotation dégrade la précision pour des angles en $\pi/4 + k\pi$.

5.4 Comportement de SAR-SIFT avec des conditions d’acquisitions différentes

Les sections précédentes ont présenté une validation expérimentale de l’efficacité de l’algorithme SAR-SIFT face à du bruit de speckle, ainsi qu’une étude sur la précision des mises en correspondance en cas de transformation de type similitude. Nous analysons ici le comportement de l’algorithme dans des situations plus complexes et en particulier avec des modes d’acquisition différents.

Nous concentrons notre étude sur des paires d’images présentant des différences de résolution et/ou d’angles d’incidence (Section 1.2). Afin d’avoir une situation de référence, nous considérons aussi des paires d’images acquises dans les mêmes conditions. L’algorithme devrait présenter de bons résultats pour les situations avec une différence de résolution. La situation est plus compliquée quand il y a changement d’angle d’incidence. La géométrie de l’image change et le signal SAR reçu peut être différent (Section 1.2.5). On s’attend alors à une diminution du nombre de bons appariements de points-clés. Cette section étudie la

Jeu	Capteur	Angle d'incidence	Résolution	Orbite	Nombre	Taille
<i>a</i>	Cosmo-SkyMed®	48°	1 m	Descendante	4	2048 × 2048
<i>b</i>	Cosmo-SkyMed®	34°	1 m	Descendante	2	2097 × 1914
<i>c</i>	Cosmo-SkyMed®	43°	3 m	Descendante	6	646 × 550
<i>d</i>	TerraSAR-X	34°	2 m	Ascendante	2	1500 × 3000
<i>e</i>	TerraSAR-X	34°	1 × 2 m	Ascendante	1	750 × 1500

TABLE 5.1 – Images utilisées : caractéristiques, nombre pour chaque jeu et taille en pixels. Toutes les images ont été acquises en mode SpotLight ou Stripmap et avec une visée à droite.

variabilité de ce nombre dans différentes conditions.

Les jeux d'images disponibles sont présentés dans le Tableau 5.1, avec leurs caractéristiques (capteur, angle d'incidence, résolution, orbite), nombre d'images et taille en pixels. Pour les images acquises dans les mêmes conditions, les mises en correspondance sont calculées entre toutes les images de chaque jeu de données respectifs *a*, *b* et *c*. Pour obtenir une situation avec un changement de résolution, un moyennage multi-vue (Section 1.2.6) est calculé dans la direction azimuth des images du jeu *d* afin d'obtenir un facteur d'échelle de deux avec les images du jeu *e*. La Figure 5.22 montre une image de chaque jeu. Les zones imagées offrent un très faible relief. Les paires d'images affichent peu de changements.

Afin d'obtenir une vérité terrain, des grilles de déformations ont été estimées pour chaque paire d'image étudiée. Elles ont été obtenues à partir d'un recalage utilisant les paramètres capteur fournis par les agences spatiales [NTF⁺12], associé à une correction manuelle. Les mises en correspondance sont évaluées en utilisant l'Equation 5.5. Le paramètre t_1 est fixé à 5, à l'exception des situations avec des changements d'angle d'incidence où il est fixé à 7 à cause des effets de recouvrement (Section 1.2.5). En effet, à cause des distorsions le long de l'axe des distances, les tailles des bâtiments peuvent varier et un seuil plus lâche devrait être considéré.

Nous observons en premier la capacité de l'algorithme à associer les points-clés correspondant en utilisant seulement l'étape NN (Section 2.4.1). Nous souhaitons vérifier que l'algorithme peut obtenir un nombre suffisant de bons appariements, sans considérer l'étape DR. Cette deuxième étape n'est qu'une étape de filtrage et le nombre de bons appariements dépend fortement de la première étape. Le Tableau 5.2 présente le pourcentage moyen de points-clés correctement appariés pour chaque configuration.

Ensuite, on examine, à l'aide de courbes ROC, l'aptitude de l'étape DR à filtrer un maximum de faux appariements tout en conservant un maximum de bons. En utilisant les mêmes notations que dans la Section 5.2.2, on trace le pourcentage de bons appariements obtenus $\frac{\#CM}{\#CM_{all}}$ en fonction du pourcentage de faux appariements $\frac{\#FM}{\#CM+\#FM}$ pour différentes valeurs de th (Fig. 5.23). Le pourcentage $\frac{\#CM}{\#CM_{all}}$ donne le nombre de bons appariements obtenus avec le seuil th parmi tous les bons appariements sélectionnés lors de l'étape NN.

Comme on peut s'y attendre, l'algorithme présente de très bonnes performances dans les situations présentant les mêmes conditions d'acquisition ou un changement de résolution. L'étape NN permet d'obtenir un fort pourcentage de points-clés correctement appariés, respectivement 61% et 27%. Parmi ces associations correctes, respectivement environ 80% et 45% peuvent être sélectionnées avec l'étape DR pour un taux de fausses alarmes de 1%.

Quand on considère une différence d'angle d'incidence, les résultats sont bien plus

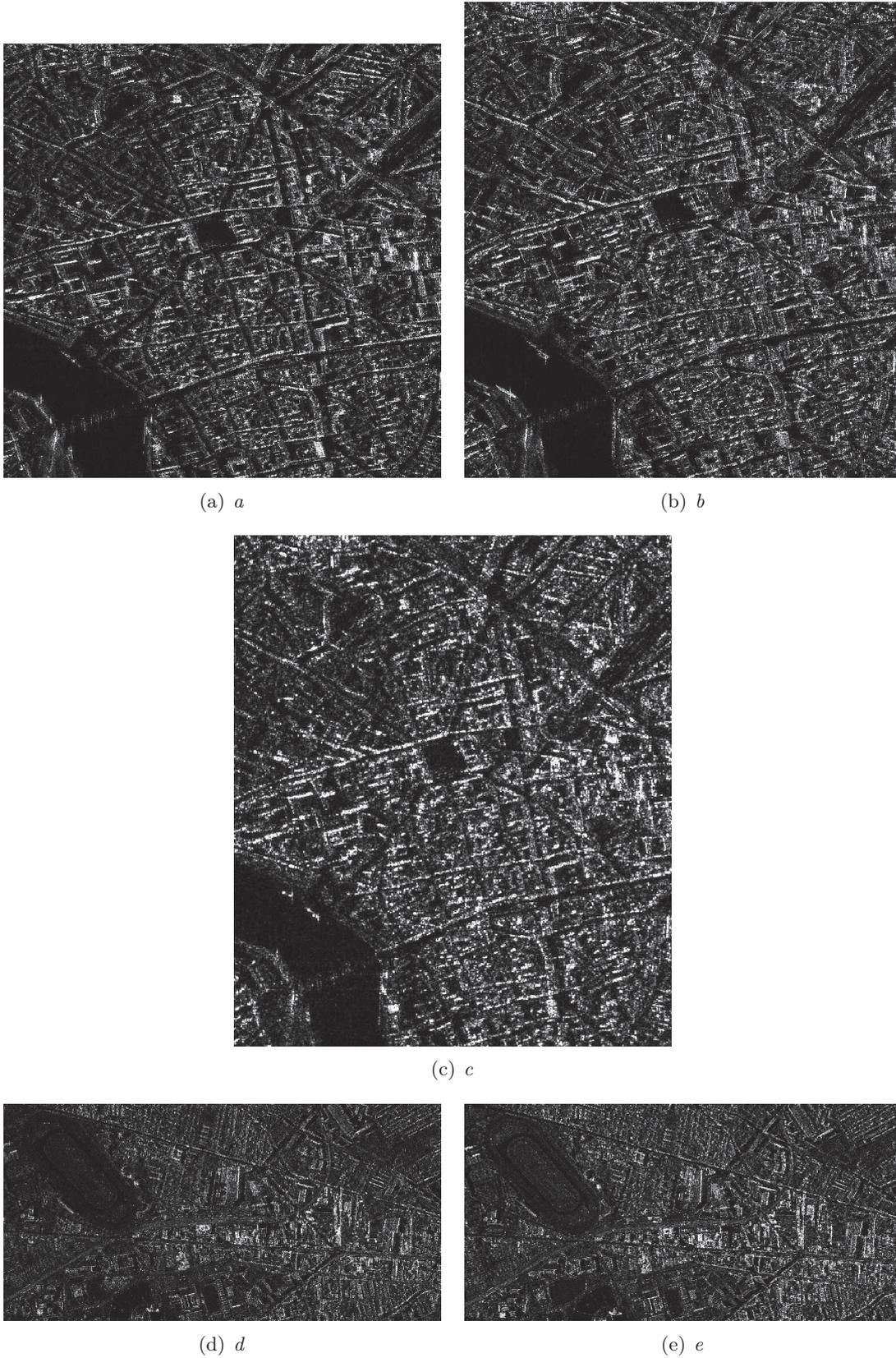


FIGURE 5.22 – Images du Tableau 5.1. Pour une meilleure lisibilité, les images ne sont pas à la même échelle.

Facteur d'échelle	Différence entre les angles d'incidence	Pourcentage de points-clés correctement appariés
-	-	61%
2	-	27%
3	5°	9.4%
3	9°	6.7%
-	14°	4.5%

TABLE 5.2 – Pourcentage moyen de points-clés correctement appariés avec l'étape NN pour différentes conditions (changement de résolution et/ou différence d'angles d'incidence).

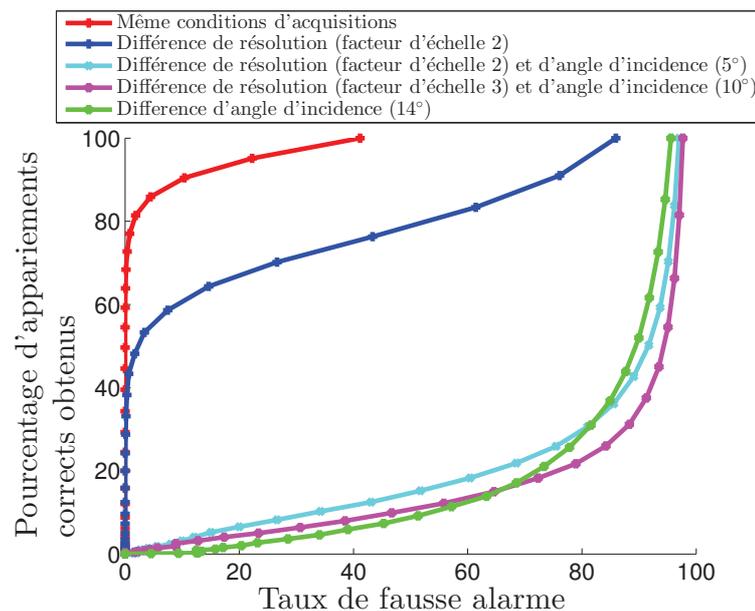


FIGURE 5.23 – Courbes ROC construites avec les images du Tableau 5.1 afin d'évaluer le comportement de l'algorithme SAR-SIFT dans des conditions d'acquisitions différentes.

faibles. Seulement 5% à 10% des points-clés sont correctement appariés avec l'étape NN. Cela correspond à environ 500 à 2500 appariements, en fonction de l'image (les différences sont liées à la taille des images). Ces nombres seraient tout à fait utilisables, seulement l'étape DR échoue à sélectionner les bons appariements. Les taux de fausses alarmes sont alors très importants. On peut toutefois faire remarquer que l'algorithme SIFT présente aussi des performances limitées quand il est appliqué à des paires d'images optiques avec des différences d'angle d'incidence non négligeables. Il est intéressant de constater que l'influence du changement de résolution est limité, contrairement à la différence d'angle d'incidence. En effet, la mise en correspondance entre les jeux a et b , avec une différence d'angle d'incidence de 14° et la même résolution, présente de plus faibles résultats que la mise en correspondance entre les jeux b et c , ou entre les jeux a et c , les deux situations présentant un changement d'échelle de 3 et une différence d'angle de respectivement 9° et 5°.

Sur les images acquises avec les mêmes conditions ou avec un changement de résolution, la géométrie des images change peu et SAR-SIFT réussit à appairer un grand nombre de points-clés. Cependant, avec une différence d'angle d'incidence, le signal radar varie et cela entraîne des distorsions qui ne sont pas prises en compte dans les invariances de SAR-SIFT.

Plus cette différence d'angle est importante, plus les variations des images seront grandes et plus grande sera la difficulté pour SAR-SIFT de trouver des points-clés correspondants.

5.5 Conclusions

En s'inspirant du Gradient par Ratio présenté en Section 4, nous avons présenté un nouvel algorithme de type SIFT, adapté aux images radar. Les trois premières étapes de l'algorithme ont été modifiées afin de s'adapter aux spécificités des images radar. Un nouveau détecteur de points-clés, SAR-Harris, et un nouveau descripteur, Ratio, ont notamment été proposés. Les performances de ce nouvel algorithme surpassent en tout point l'algorithme SIFT original. Cependant, il présente des limitations quand il est appliqué à des images acquises sous des angles d'incidence différents.

Dans la partie suivante, nous présentons des applications en télédétection pour les algorithmes de type SIFT (utilisation de SIFT en optique et SAR-SIFT en radar). Le chapitre 6 s'intéresse à l'estimation de transformation globale et au filtrage des appariements dans les situations de changement d'angles d'incidence pour les images radar. Enfin le chapitre 7 propose une application de détection de changements pour les images radar et optiques.

Troisième partie

Applications des algorithmes SIFT et
SAR-SIFT pour les images de
télédétection

Chapitre 6

Estimation d'une déformation globale entre deux images radar

En première approximation, on considère que deux images satellites représentant la même scène peuvent être reliées par une déformation globale rigide. Sur des scènes approximativement planes, cette déformation est souvent modélisée par une transformation affine [ZF03]. Les images radar peuvent présenter des distorsions locales (Section 1.2.5) mais la relation globale entre deux images radar peut être décrite par ce modèle, sous réserve qu'elles soient acquises avec la même orbite (ascendante ou descendante). Pour les zones avec du relief, il est nécessaire d'utiliser un modèle numérique de terrain (MNT) et d'intégrer l'information de hauteur dans le modèle.

Les appariements obtenus avec l'algorithme SAR-SIFT peuvent être utiles pour estimer la transformation affine entre deux images. Cependant, lorsque le nombre de faux appariements est trop important (Section 5.2.2), nous ne pouvons estimer directement les transformations. L'algorithme RANSAC (RANdom Sample Consensus) [FB81] permet d'estimer de façon itérative un modèle donné à partir d'un ensemble de données contenant des valeurs aberrantes. D'autre part, il fournit en sortie un sous-ensemble de données satisfaisant le modèle estimé. RANSAC peut alors être utilisé pour estimer une transformation globale à partir d'un ensemble d'appariements, tout en sélectionnant un sous-ensemble d'appariements cohérents avec cette transformation. Nous utiliserons cet algorithme pour deux applications : l'estimation de la déformation globale entre deux images et le filtrage des faux appariements.

Dans un premier temps, nous détaillerons l'algorithme RANSAC ainsi que AC-RANSAC, une version *a contrario* de RANSAC plus performante dans des situations affichant un grand pourcentage de valeurs aberrantes. Puis nous montrerons l'intérêt de ce type d'algorithme pour les deux applications proposées.

6.1 Présentation des algorithmes RANSAC et AC-RANSAC

Nous présentons ici les algorithmes RANSAC et AC-RANSAC, puis nous indiquons leur utilisation dans notre contexte.

Algorithme RANSAC Considérons un ensemble de données $C = \{c_i, i = 1, \dots, N\}$. Cet ensemble est composé de données suivant un même modèle, appelées *inliers*, (ici les appariements corrects) mais aussi de données aberrantes, appelées *outliers* (ici les faux appariements). Ces outliers perturbent l'estimation du modèle. Par exemple, une estimation par moindres carrés ne résistera pas à un pourcentage d'outliers trop conséquent. L'algorithme RANSAC [FB81] consiste à tester itérativement différentes estimations du modèle

en sélectionnant aléatoirement un faible nombre de données c_i . Le modèle satisfaisant le plus grand nombre de données est choisi. On suppose ensuite que les données satisfaisant ce modèle sont correctes. Cet algorithme nécessite cependant de fixer plusieurs paramètres, comme le nombre d'itérations, le nombre minimal d'inliers obtenus pour l'estimation du modèle ou le seuil de tolérance pour la distance des données au modèle. En outre, il n'est pas toujours efficace quand le pourcentage d'outliers est important.

Algorithme AC-RANSAC L'algorithme AC-RANSAC [MS04b, RDGM10] est une version a contrario de RANSAC qui s'intéresse plus particulièrement aux appariements de points entre deux images suivant un même modèle. Il offre l'avantage de ne nécessiter la sélection que d'un seul paramètre, le nombre d'itérations i_{max} . Par ailleurs, cette méthode produit de bons résultats même en présence d'un fort pourcentage d'outliers (jusqu'à 90%).

Soit un ensemble d'appariements de points $C = \{(P_{1,k}, P_{2,k}) | k = 1, \dots, N\}$ entre deux images I_1 et I_2 . On se limite ici à la recherche d'une transformation affine. Cette méthode repose sur l'hypothèse nulle \mathcal{H}_0 selon laquelle la distribution des appariements est uniforme. On suppose donc que les points $P_{1,k}$ et $P_{2,k}$ sont des variables aléatoires mutuellement indépendantes mais aussi que tous les points $P_{1,i}$ et $P_{2,j}$ sont répartis uniformément sur respectivement les images I_1 et I_2 . Un ensemble d'appariements sera considéré comme significatif s'il est peu probable qu'il apparaisse sous cette hypothèse \mathcal{H}_0 .

Soit un sous-ensemble S' de C constitué de n appariements sélectionnés aléatoirement. On estime la matrice $T_{S'}$ de la transformation affine entre I_1 et I_2 à partir des appariements de S' . Le nombre n est choisi comme le nombre minimal de données nécessaires à l'estimation de la transformation. Pour une transformation affine, n est égal à 3. Considérant un sous-ensemble S de C , avec $S' \cap S = \emptyset$, on définit la rigidité de S sous la transformation $T_{S'}$ par :

$$\alpha(S, T_{S'}) := \max_{(P_{1,k}, P_{2,k}) \in S} \max \left(\frac{\pi}{A_2} d(T_{S'} P_{1,k}, P_{2,k})^2, \frac{\pi}{A_1} d(P_{1,k}, T_{S'}^{-1} P_{2,k})^2 \right). \quad (6.1)$$

Les quantités A_1 et A_2 sont les aires respectives des images I_1 et I_2 . L'opérateur d est la distance euclidienne. L'algorithme AC-RANSAC consiste à chercher un ensemble S pour lequel le nombre de fausse-alarme $NFA(S, S')$ soit inférieur à ε . Un tel ensemble est dit ε -significatif. Le nombre de fausses alarmes est défini par :

$$NFA(S, T_{S'}) := (N - n) \binom{N}{K} \binom{N - K}{n} (\alpha(S, T_{S'}))^K, \quad (6.2)$$

avec K le nombre d'appariements de S . On fixe le seuil ε à 1 et le nombre n à 3 (transformation affine) [RDGM10]. Le seul paramètre à choisir est donc le nombre d'itérations i_{max} .

A chaque itération, on extrait aléatoirement n appariements de C pour former l'ensemble S' et on estime la transformation $T_{S'}$. On cherche ensuite l'ensemble S dans $C \setminus S'$ permettant de minimiser $NFA(S, S')$. On considère que cet ensemble est significatif si $NFA(S, S') < 1$. Cette opération est réalisée pour i_{max} itérations et on sélectionne l'ensemble S présentant le plus petit nombre de fausses alarmes $NFA(S, S')$. On obtient ainsi un ensemble d'appariements S que l'on suppose comme corrects et la transformation associée $T_{S'}$ entre les deux images. Si aucun ensemble ne satisfait $NFA(S, S') < 1$, on suppose que les points ne suivent aucun modèle. Plus de détails sur l'algorithme et l'implémentation se trouvent dans [MS04b, RDGM10].

Approche proposée Nous proposons d'appliquer l'algorithme AC-RANSAC à l'ensemble des appariements obtenus avec l'étape NN de l'algorithme SAR-SIFT. L'étape de filtrage DR n'étant pas efficace quand les images présentent des différences d'angles d'incidence (Section 5.2.2), cela permet d'obtenir le plus grand nombre possible de bons appariements. Cependant, pour accélérer le temps de calcul, les appariements utilisés pour estimer la déformation sont sélectionnés aléatoirement parmi les appariements présentant un rapport des distances inférieure au seuil $th = 0.9$. Cette valeur est plus élevée que les choix classiques, plus proches de $th = 0.8$ [Low99]. Le but ici est seulement de réduire le nombre d'appariements obtenus avec l'étape NN afin que l'algorithme AC-RANSAC n'ait pas à traiter un trop grand nombre de cas aberrants, tout en conservant un nombre d'appariements suffisant pour la mesure précise de la transformation. La valeur 0.9 est un bon compromis. Nous avons choisi $i_{max} = 10000$ comme nombre d'itération. La transformation recherchée étant une transformation affine, nous devons estimer les coefficients de la transformation polynomiale suivante :

$$\begin{aligned}x_2 &= a_1 + a_2 \cdot x_1 + a_3 \cdot y_1 \\y_2 &= b_1 + b_2 \cdot x_1 + b_3 \cdot y_1.\end{aligned}\tag{6.3}$$

L'algorithme AC-RANSAC nous donne en sortie une estimation des coefficients a_i et b_i , ainsi qu'un ensemble de mises en correspondance que l'on suppose comme correctes. Pour plus de lisibilité, nous appellerons par la suite cette méthode SAR-SIFT + AC-RANSAC.

Nous présentons maintenant son intérêt pour l'estimation de la transformation globale puis pour le filtrage des faux appariements dans des situations de changement d'angle incidence.

6.2 Estimation de la déformation globale entre deux images

Nous exposons ici l'intérêt d'utiliser SAR-SIFT + AC-RANSAC pour rechercher la déformation globale entre deux images radar acquises dans des conditions différentes. Nous produisons dans un premier temps un état de l'art sur le recalage d'images radar. Cet état de l'art ne se veut pas exhaustif mais donne une vision générale des techniques actuelles.

6.2.1 Etat de l'art sur le recalage d'images radar

En traitement d'images, le recalage consiste à rechercher la transformation géométrique optimale reliant deux images d'une même zone, afin de pouvoir les comparer ou les fusionner. Le recalage peut s'effectuer entre deux images acquises selon la même modalité (deux images radar) ou selon des modalités différentes (une image optique et une image radar). En télédétection, ce problème peut être envisagé sous deux angles. On peut chercher tout simplement à recalculer une image sur une autre, comme pour le recalage d'images naturelles. Une autre approche consiste à utiliser les données capteur fournies par les agences spatiales afin de géoreferencer les images. Ces informations nous permettent de relier les positions des pixels (ligne et colonne) aux coordonnées des objets (latitude, longitude et hauteur). On peut ainsi projeter indépendamment deux images dans la même géométrie. Cette méthode n'est pas à proprement parler du recalage puisque nous ne recherchons pas de transformation. Les données capteur comportent toutefois des imprécisions et il est souvent nécessaire d'effectuer un ajustement avec des méthodes de recalage traditionnelles. Nous aborderons ici les deux types d'approche. On distingue toutefois en premier lieu le cas particulier du recalage d'images en configuration interférométrique.

Recalage d'images en configuration interférométrique La comparaison d'une paire d'images dans cette configuration est aisée. En effet elles présentent très peu de changements puisqu'elles ont été acquises par un même capteur (ou un capteur similaire) et dans des conditions presque équivalentes. Les déformations sont ainsi souvent réduites à une translation. En outre, les techniques d'interférométrie nécessitent un recalage inférieur au dixième de pixel. Ce problème a donc été très étudié et les techniques de recalage sont aujourd'hui abouties. La méthode la plus couramment utilisée est la maximisation de la cohérence complexe. Cela permet d'estimer des transformations rigides et donne de très bons résultats sur ce type d'images. Cette technique est utilisée par exemple dans la toolbox Nest de l'ESA.

La situation est plus compliquée pour les images acquises dans des conditions différentes, par exemple avec un angle d'incidence ou un capteur différent. Un changement d'angle entraîne des déformations - d'autant plus notables que le relief est important - ainsi qu'une modification de la géométrie des objets (Section 1.3). Le modèle de déformation entre les images n'est plus une simple translation. Par ailleurs, les différences de géométrie compliquent la comparaison des images. L'utilisation de la corrélation croisée n'est pas efficace et de nouvelles méthodes ont ainsi dû être développées.

Utilisation des données capteur Les méthodes utilisant les données capteur sont dites «en aveugle», c'est à dire qu'elles ne se basent pas sur les données propres de l'image. Elles consistent à reprojeter indépendamment des images dans la même géométrie. Elles s'appuient fortement sur des informations annexes, comme les paramètres capteurs fournis par les agences spatiales ou les modèles numériques de terrain (MNT). Les paramètres capteurs [SBM⁺06, NTF⁺12] nous informent sur la trajectoire du satellite et permettent ainsi d'obtenir la localisation précise en longitude et latitude de chaque pixel. L'utilisation d'un MNT est cependant nécessaire pour prendre en compte le relief à la surface de la Terre. Ces données n'étant pas toujours disponibles, on peut se satisfaire de l'altitude moyenne de la zone étudiée si elle est globalement plate. Ce n'est par contre pas suffisant sur une zone avec un fort relief. La localisation des pixels permet ensuite de reprojeter l'image étudiée dans la géométrie terrestre. Ainsi deux images géoréférencées devraient être recalées de façon précise. Néanmoins, le recalage n'est pas toujours parfait. Il peut en particulier exister des distorsions si le MNT n'est pas parfait ou indisponible. D'autre part, il existe souvent un décalage pixelique dû à une imprécision sur le temps de départ des acquisitions. Il est alors nécessaire d'utiliser des méthodes de recalage traditionnelles pour estimer la déformation résiduelle.

Les agences spatiales peuvent aussi fournir directement des modèles de déformation de type RPM (Ratio Polynomial Model) [ZFL⁺10] pour le géoréférencement (Section 1.1.6). Ces modèles nécessitent aussi de connaître l'élévation du terrain. Néanmoins, ces RPM comme les informations capteurs ne sont pas toujours fournis avec une grande précision, ni même mis à disposition. Les agences spatiales utilisent ces méthodes pour fournir des images géoréférencées avec une très grande précision. Cependant, en l'absence de ces informations annexes, d'autres méthodes, basées sur la comparaison des images, doivent être envisagées.

Recalage par comparaisons d'images L'utilisation de méthodes de recalage traditionnelles est utile quand on souhaite affiner le recalage obtenu par géoréférencement ou que l'on a pas accès à des données annexes, de type données capteur ou MNT. On cherche alors à estimer la transformation géométrique entre deux images. Selon le type d'images comparées, cette transformation peut être définie par une transformation affine, une projection, un quotient de polynômes ou d'autres transformations.

Afin d'estimer les coefficients de la transformation, une méthode courante consiste à sélectionner des points homologues extraits sur une paire d'images. Ces points peuvent être sélectionnés de façon manuelle, en repérant des points stables comme des carrefours de route ou même en plaçant des coins réflecteurs sur la zone étudiée avant l'acquisition [LZbPQ10]. Cette extraction prend beaucoup de temps et présente une part d'erreur. De nombreuses méthodes proposent de sélectionner automatiquement des points et de les apparier. On peut citer l'extraction de jonctions de routes et de carrefours [DGL04] à l'aide d'une méthode automatique de détection de lignes. Les points sont ensuite mis en correspondance en considérant des informations comme le nombre de routes adjacentes, leurs orientations ou le type de jonction. D'autres méthodes se basent sur l'appariement de points de bords par l'optimisation d'une mesure de similarité basée sur la somme de gradients [KA06, CCS14]. L'algorithme ARRSI [WC07] s'appuie sur la congruence de phase. Les valeurs maximales de la phase de la transformée de Fourier sont détectées comme points significatifs. Leur comparaison se fait à l'aide de moments calculés sur des patches extraits de l'image de phase. Enfin l'algorithme SIFT ou des variantes [WYF12, FHPK13] peuvent être utilisés pour trouver des points homologues.

Afin d'évaluer l'alignement entre deux images, d'autres méthodes se basent sur le calcul de similarité entre les intensités de pixels. On peut citer par exemple l'utilisation du maximum de vraisemblance [LL04] ou de l'information mutuelle [SR10].

6.2.2 Proposition d'une nouvelle méthode de recalage

Nous proposons d'utiliser la méthode SAR-SIFT + AC-RANSAC pour recalibrer des images dans deux situations particulières :

- L'estimation de la transformation affine entre deux images radar acquises avec des angles d'incidence différents. Rappelons que dans ce cas, la déformation entre deux images peut être modélisée en première approximation par une transformation affine [ZF03]. Cette approximation n'est cependant valable que pour les points au sol, ainsi que dans les zones de faible relief.
- L'estimation de la translation résiduelle entre deux images géoréférencées au moyen des paramètres capteurs.

6.2.2.1 Estimation de la transformation affine entre deux images

Le Tableau 6.1 décrit la paire d'images utilisée pour évaluer la précision du recalage. Les images ont la même résolution mais une différence d'angle d'incidence de 14° . Comme il n'y avait pas de vérité-terrain à disposition, nous avons extrait manuellement 30 points de contrôle sur la paire d'images. Ils sont ensuite utilisés pour obtenir un recalage manuel ainsi qu'une mesure de la précision du recalage. La Figure 6.1 affiche la paire utilisée ainsi que la position des points de contrôle.

Dix des points de contrôle sont sélectionnés de façon aléatoire afin d'évaluer les paramètres de l'Eq. (6.3) et obtenir un recalage manuel. Les vingt autres points sont utilisés pour mesurer la précision des deux méthodes de recalage, SAR-SIFT + AC-RANSAC et

Capteur	Angle d'incidence	Résolution	Orbite	Taille
Cosmo-SkyMed®	48°	1 m	Descendante	2048×2048
Cosmo-SkyMed®	34°	1 m	Descendante	2048×2048

TABLE 6.1 – Paire d'images utilisées pour évaluer la précision de la méthode de recalage SAR-SIFT + AC-RANSAC. Les images sont acquises sur la zone de Toulouse, France, et ne présentent qu'un faible relief.

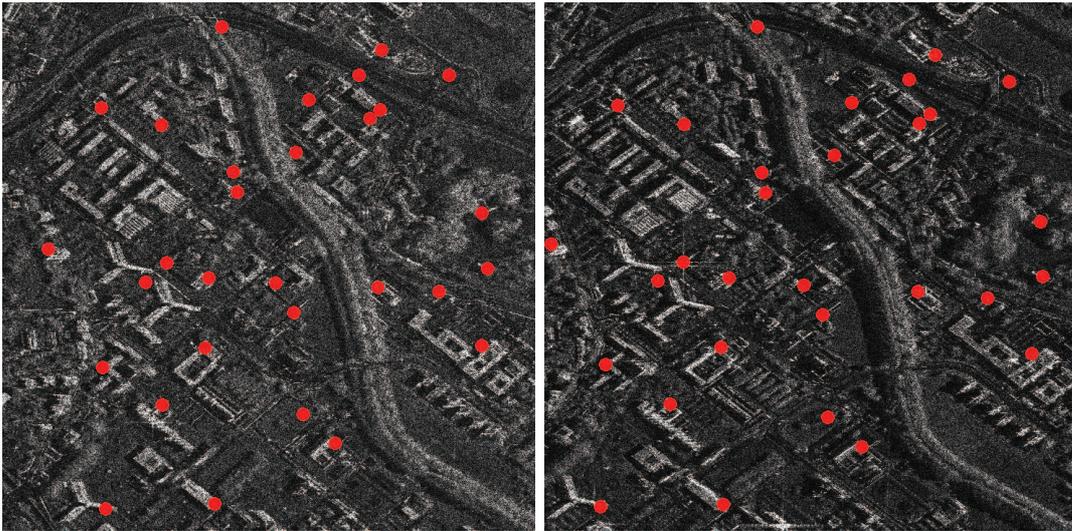


FIGURE 6.1 – Images du Tableau 6.1 (à gauche 48° et à droite 34°) avec les points de contrôle utilisés pour évaluer la précision du recalage.

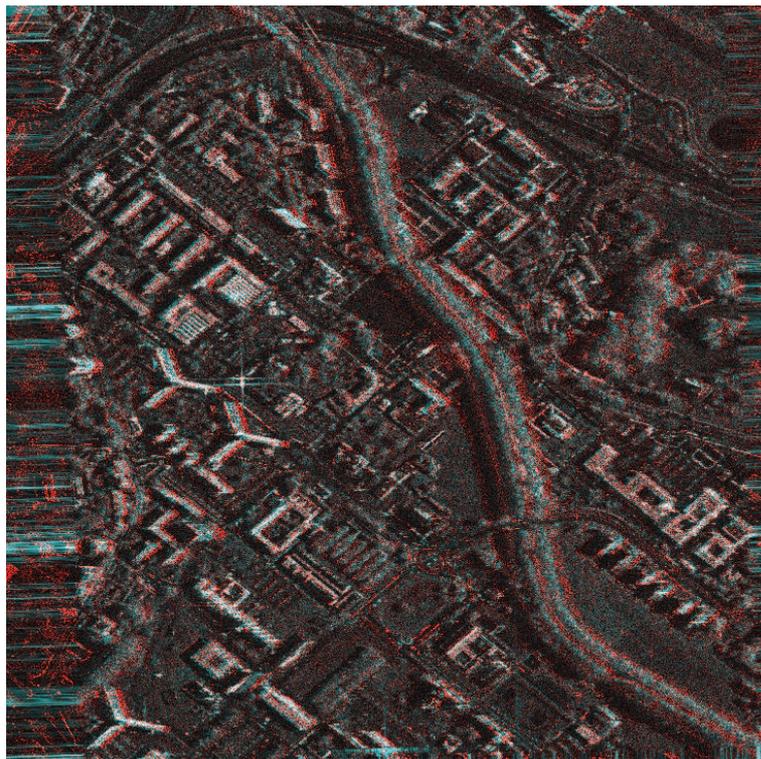


FIGURE 6.2 – Superposition de deux images recalées (différence d'angle de 14°) du Tableau 6.1 avec la méthode SAR-SIFT + AC-RANSAC. L'image maître est en rouge et l'image esclave en bleu.

Méthode de recalage	Recalage manuel	SAR-SIFT + AC-RANSAC
Erreur quadratique moyenne	2.46	2.03

TABLE 6.2 – Précision des deux méthodes de recalage pour les images du Tableau 6.1.

la méthode manuelle, en calculant l’erreur quadratique moyenne. Ce processus est répété 10000 fois et les moyennes des erreurs quadratiques sont présentées dans le Tableau 6.2. La superposition des deux images recalées par SAR-SIFT + AC-RANSAC est présentée en Fig. 6.2. Le décalage systématique autour des objets élevés (bâtiments, arbres) est dû aux effets de recouvrement (Section 1.2.5).

La méthode automatique, SAR-SIFT + AC-RANSAC, offre une bonne précision du recalage, comparable à la méthode manuelle. Sachant que l’extraction des points de contrôle est fastidieuse et sujette aux erreurs, cette nouvelle méthode est intéressante pour le recalage d’images radar acquises dans des conditions différentes.

6.2.2.2 Estimation de la translation résiduelle

Nous souhaitons ici évaluer la translation résiduelle entre des images géoréférencées. Nous n’estimons donc pas une transformation affine, mais juste un décalage. Les images étudiées sont décrites dans le Tableau 6.3 et observables en Figure 6.4. Elles ont été géoréférencées à l’aide des paramètres capteur et d’un MNT [NTF⁺12]. Toutes les images Cosmo-Skymed d’un côté et TerraSAR-X de l’autre sont donc recalées entre elle avec une précision inférieure au pixel. Nous souhaitons estimer le décalage résiduel (de l’ordre du pixel) entre toutes les paires Cosmo-Skymed/TerraSAR-X, soit 195 paires. Le décalage entre chaque paire d’images Cosmo-Skymed/TerraSAR-X doit donc être sensiblement le même. Manuellement nous avons mesuré le décalage suivant : **(7 , -1.5)**.

Pour chacune des paires étudiées, nous estimons une translation avec la méthode SAR-SIFT + AC-RANSAC. La répartition des décalages horizontaux et verticaux est représentée Figure 6.3. On obtient **(7.29 , -1.79)** comme valeurs moyennes et **(0.59 , 0.67)** comme écarts-types des décalages. Ces valeurs moyennes sont très proches du décalage estimé manuellement. Les écarts-types sont inférieurs au pixels, et comme on peut l’observer Figure 6.3 on ne trouve pas de décalages aberrants.

Capteur	Angle d’incidence	Résolution	Orbite	Taille	Nombre
Cosmo-Skymed®	34°	3 m	Ascendante	1024 × 1024	15
TerraSAR-X®	34°	3 m	Ascendante	1024 × 1024	13

TABLE 6.3 – Paire d’images utilisées pour évaluer la précision de la méthode de recalage SAR-SIFT + AC-RANSAC. Les images sont acquises sur la zone de Saint-Gervais, France.

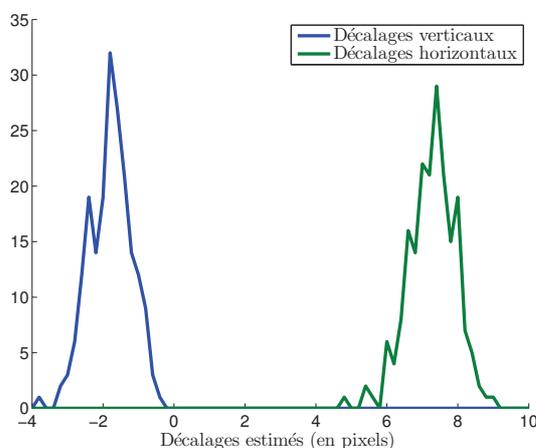


FIGURE 6.3 – Répartition des décalages horizontaux et verticaux estimés pour chaque paire d’images Cosmo-Skymed/TerraSAR-X à l’aide de la méthode SAR-SIFT + AC-RANSAC.

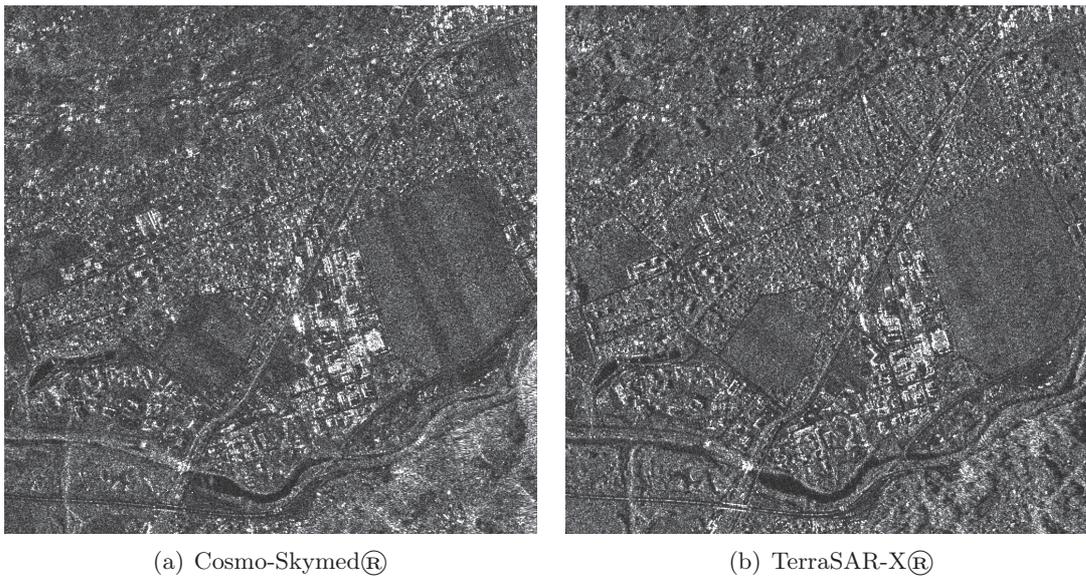


FIGURE 6.4 – Exemples de deux images du Tableau 6.3.

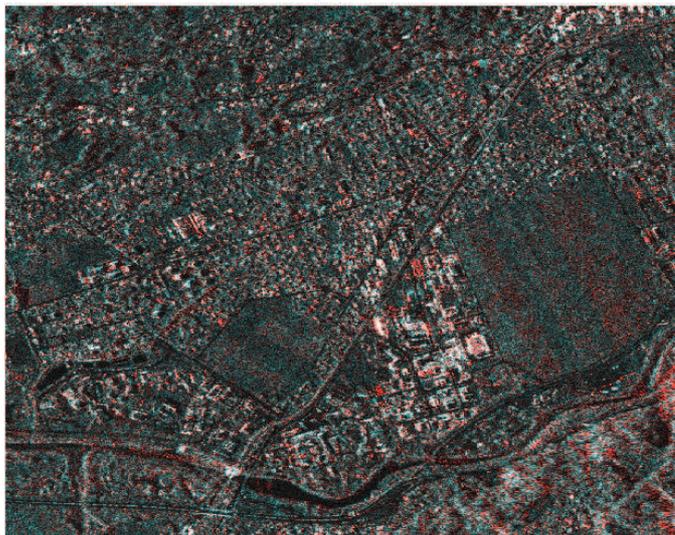


FIGURE 6.5 – Superposition des deux images de Figure 6.4 recalées avec la méthode SAR-SIFT + AC-RANSAC. Le décalage estimé est $(7.97, -1.77)$.

D'autre part, on mesure que 90% des décalages ont une erreur en distance inférieure à 1.5 pixels, par rapport au décalage estimé manuellement. La méthode SAR-SIFT + AC-RANSAC est donc efficace pour mesurer ces translations résiduelles. Les décalages sont mesurés avec une erreur de précision de l'ordre du pixel, ce qui est suffisant pour ce type de situations.

Afin d'illustrer cette expérience, nous présentons Figure 6.5 la superposition des images de Figure 6.4 après recalage.

6.3 Filtrage des faux appariements

Comme présenté dans la Section 5.2.2, l'étape de filtrage DR des appariements échoue dans des situations de changement d'angle d'incidence. Néanmoins le pourcentage d'ap-

Jeu	Capteur	Angle d'incidence	Résolution	Orbite	Nombre	Taille
<i>a</i>	Cosmo-SkyMed®	48°	1 m	Descendante	4	2048 × 2048
<i>b</i>	Cosmo-SkyMed®	34°	1 m	Descendante	2	2097 × 1914

TABLE 6.4 – Images utilisées : caractéristiques, nombre pour chaque jeu et taille en pixels. Toutes les images ont été acquises en mode SpotLight ou Stripmap et avec une visée à droite.

Situation étudiée	Etape NN	Filtrage DR	Filtrage AC-RANSAC
Bons appariements	2251	45	1979
Faux appariements	48747	11	104

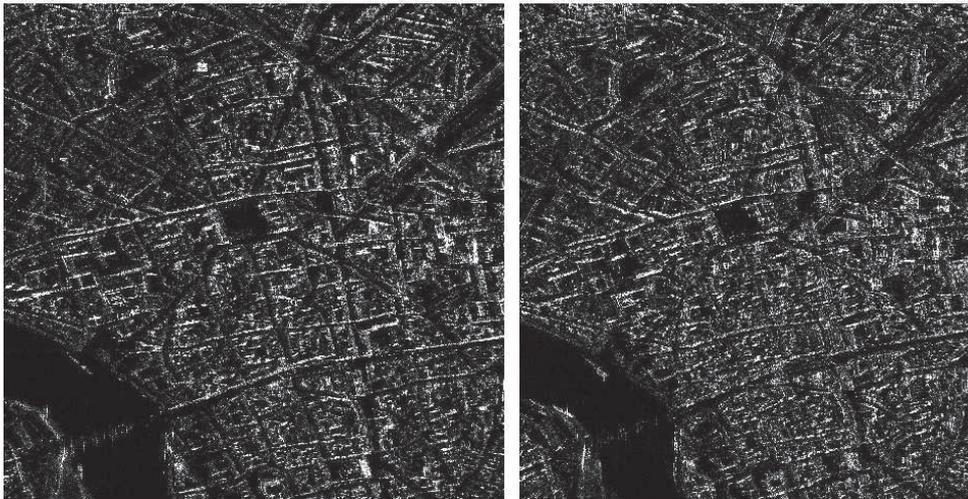
TABLE 6.5 – Nombre moyen d'appariements (bons et faux) par paire d'images. Trois situations sont étudiées : les appariements obtenus avec l'étape NN, ceux filtrés par l'étape DR ($th = 0.8$) et ceux par l'algorithme AC-RANSAC.

pariements corrects n'est pas nul (autour de 5 à 10%). Il peut être intéressant d'obtenir une plus grande quantité d'appariements, notamment pour des applications de détection de changements (voir Chapitre 7). Dans cette section, nous étudions la capacité de l'algorithme AC-RANSAC à filtrer les faux appariements dans ce type de situation.

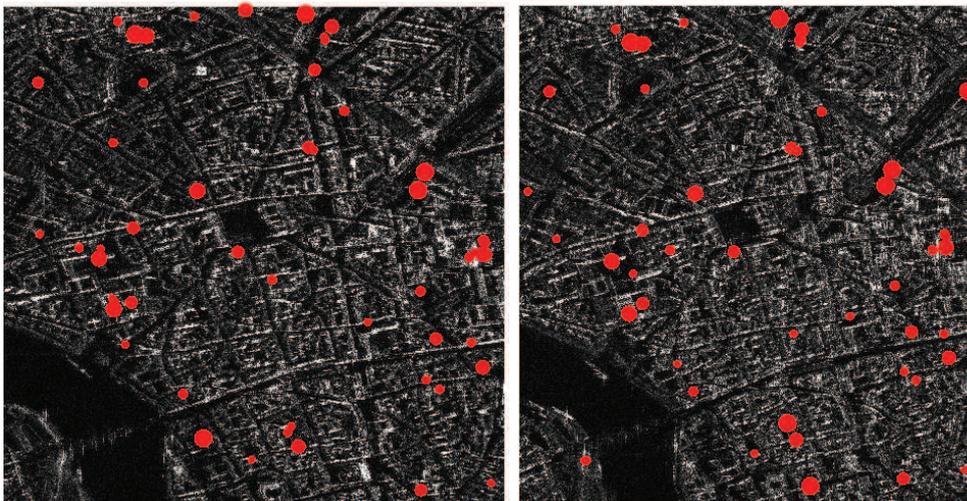
Nous reprenons les données de Section 5.4 et nous concentrons notre étude sur la comparaison des images des jeux *a* et *b* (Tableau 6.4). Cette situation offre la plus grande différence d'angle d'incidence (14°) et s'est montrée difficile pour l'algorithme SAR-SIFT. Pour chaque paire d'images considérée, nous filtrons les appariements avec la méthode proposée basée sur AC-RANSAC et nous calculons le nombre d'appariements corrects et faux obtenus. Les mises en correspondance sont évaluées de la même façon que dans la Section 5.4. Pour avoir un élément de comparaison, nous avons aussi calculé le nombre d'appariements corrects et faux obtenus avec l'étape NN puis avec le filtrage DR (seuil $th = 0.8$ sur le rapport des distances). Le Tableau 6.5 présente le nombre moyen d'appariements, corrects et faux, par paire d'images, pour les trois situations étudiées. L'utilisation de AC-RANSAC permet de supprimer la très grande majorité des faux appariements, tout en conservant 88% des bons. Le pourcentage d'appariement aberrants passe de 95% à 5%. Nous pouvons atteindre un grand nombre de bons appariements tout en conservant un taux de fausses alarmes raisonnable. En comparaison, le filtrage DR limite fortement le nombre d'erreurs mais le nombre d'appariements obtenus est beaucoup trop faible.

La Figure 6.6 montre les points-clés mis en correspondance entre une des paires précédemment étudiées. Deux filtrages sont considérés : DR avec un seuil $th = 0.8$ sur le rapport des distances, et AC-RANSAC avec une transformation affine. Cet exemple met bien en évidence une limitation du filtrage DR et l'intérêt d'utiliser AC-RANSAC. En effet, le nombre d'appariements obtenus avec le filtrage DR pour un tel seuil est faible (61) et le taux de fausses alarmes (18%) est assez important. Le filtrage avec l'algorithme AC-RANSAC conserve un nombre très important d'appariements (2234) tout en maintenant un faible taux de fausses alarmes (1%).

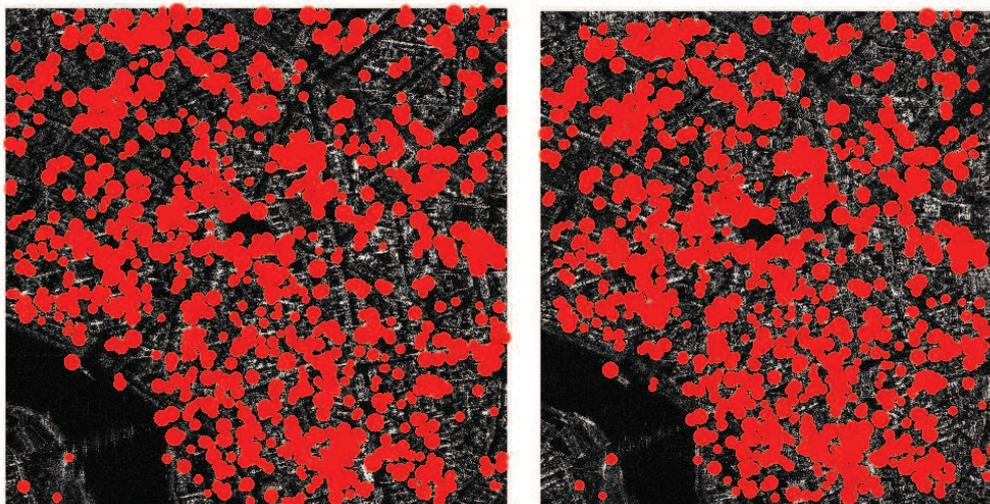
Un autre exemple de mise en correspondance est présenté à la Figure 6.7. La paire d'images utilisée offre une situation multi-capteurs multi-angles d'incidence. Les images ont été acquises avec des satellites différents (TerraSAR-X et Cosmo-SkyMed), en passe ascendante et avec la même résolution (3 m), mais avec des angles d'incidence différents (34° et 45°). Par ailleurs, les images représentent la ville de Saint-Gervais, France, située dans une zone très montagneuse. On est alors en présence d'un relief important. Bien que cette



(a) Paire d'image étudiée : Images Cosmo-SkyMed de même résolution (1 m) mais avec des angles d'incidence différents (34° et 48°).

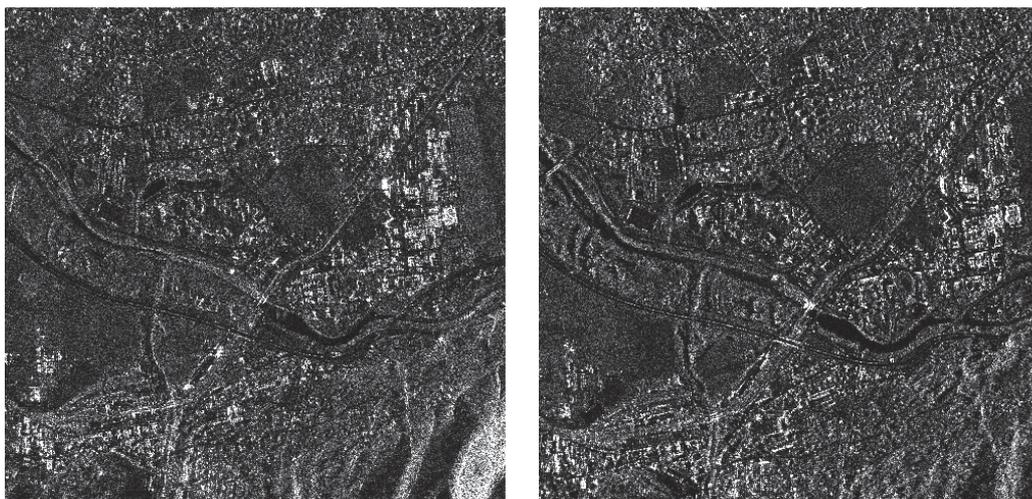


(b) Filtrage DR : 61 appariements dont 11 faux.

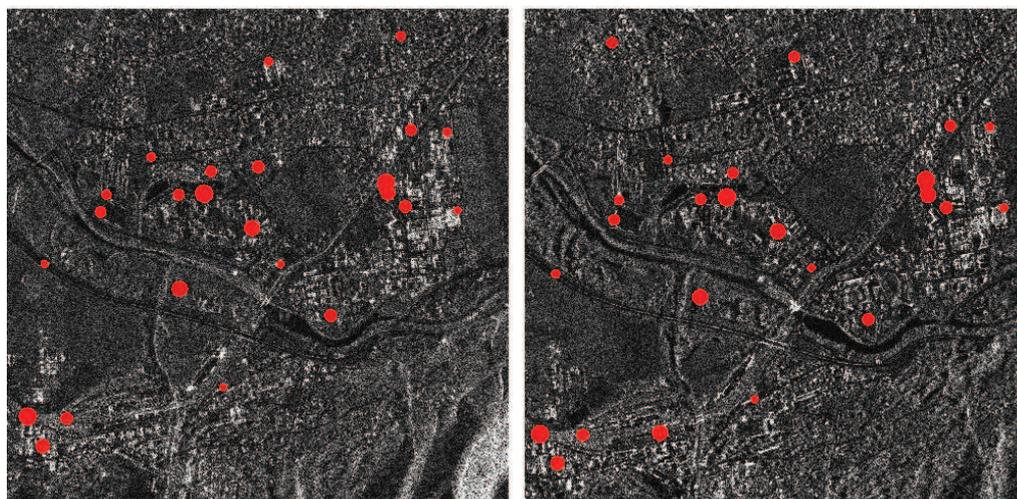


(c) Filtrage AC-RANSAC : 2234 appariements dont 23 faux.

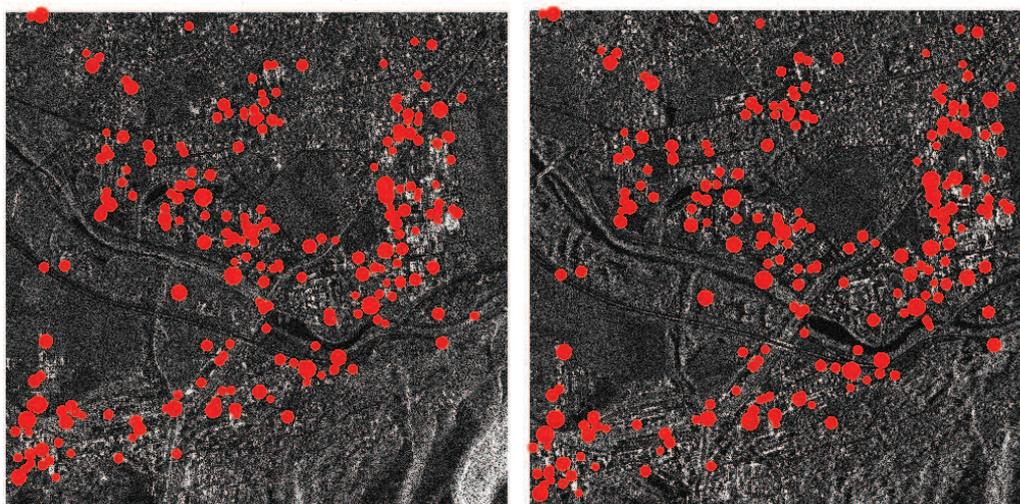
FIGURE 6.6 – Affichage de points-clés mis en correspondance. Deux méthodes de filtrage sont testées : DR avec le seuil $th = 0.8$ et AC-RANSAC avec une transformation affine.



(a) Image Cosmo-SkyMed, avec 3 m de résolution et 34° d'angle d'incidence. (b) Image TerraSAR-X, avec 3 m de résolution et 45° d'angle d'incidence.



(c) Filtrage DR : 31 appariements.



(d) Filtrage AC-RANSAC : 521 appariements.

FIGURE 6.7 – Affichage de points-clés mis en correspondance. Deux méthodes de filtrage sont testées : DR avec le seuil $th = 0.8$ et AC-RANSAC avec une transformation affine.

situation soit plus compliquée que la précédente, les résultats sont raisonnables. Le nombre d'appariements passe de 31 avec le filtrage DR à 521 avec le filtrage AC-RANSAC (taille des images : 1000×1000). L'absence de vérité-terrain ne permet pas d'évaluer correctement les appariements, mais une inspection visuelle donne 2 faux pour le filtrage DR et aucun pour le filtrage AC-RANSAC.

L'algorithme AC-RANSAC, associé à la recherche d'une transformation affine, est donc particulièrement efficace pour filtrer les faux appariements obtenus avec SAR-SIFT, même dans les zones non plates. Nous avons montré son intérêt dans des situations multi-angle d'incidence et multi-capteurs.

6.4 Conclusions

L'algorithme AC-RANSAC, associé à la recherche d'une transformation affine, est donc performant pour le filtrage de faux appariements et l'estimation de transformation globale entre images acquises avec des angles d'incidence différents. La méthode SAR-SIFT + AC-RANSAC peut être utilisée pour estimer la transformation affine entre deux images acquises avec des angles d'incidence différents, sous réserve que les images ne présentent pas un relief trop important, mais aussi pour évaluer le décalage résiduel entre deux images géoréférencées à l'aide de paramètres capteur et de MNT. Sur les zones de relief, une transformation affine ne permet pas de modéliser correctement la déformation. Cette méthode ne permettra pas d'obtenir un recalage précis mais elle est suffisante pour filtrer les faux appariements.

Le chapitre suivant présente une application de détection de changements à l'aide des algorithmes SIFT et SAR-SIFT pour les images de télédétection.

Chapitre 7

Détection de changements par descripteurs locaux

La détection de changements en télédétection consiste à repérer des variations dans une paire ou une série d'images satellites (optiques, radar, multispectrales...). Ces changements peuvent être structurels (construction ou destruction de bâtiments), mais aussi saisonniers (végétation, champs). Les images de télédétection, grâce à leur importante couverture spatiale et au large choix offert en résolution spatiale et spectrale, sont utilisées pour de nombreuses applications pratiques, comme le suivi de végétation, l'évaluation de dommages après catastrophes, la mesure de l'expansion urbaine ou le suivi d'occupation du sol.

Afin d'évaluer correctement les changements liés au contenu, il est nécessaire de les séparer de changements causés par d'autres effets, comme l'illumination, les conditions atmosphériques, la dynamique du capteur ou l'angle d'incidence. Ainsi, pour comparer deux zones, il faudrait idéalement deux images acquises à des dates anniversaires, par le même capteur, sous le même angle d'incidence et avec la même résolution, et il faudrait également pouvoir les recalculer avec une précision inférieure au pixel. Bien que cela soit possible pour des programmes au long terme comme le suivi d'occupation du sol, cela est rarement le cas dans les situations de catastrophe, où l'on souhaite avoir des données le plus rapidement possible. Ainsi les méthodes de détection de changements nécessitent un pré-traitement pour corriger au maximum les différences entre images non liées au contenu. Il est souvent essentiel d'effectuer des corrections radiométriques, prenant en compte l'illumination de la scène, l'atmosphère et les caractéristiques du capteur (données de calibration), ainsi que des corrections géométriques précises (géoréférencement, recalage, prise en compte de l'élévation du terrain). Les images corrigées sont ensuite comparées par des méthodes de détection de changements que l'on peut regrouper en deux familles : les méthodes basées sur la comparaison de pixels et celles basées sur la comparaison d'objets. Les méthodes de haut niveau comme celles comparant des objets peuvent permettre de s'affranchir d'un certain nombre de corrections radiométriques ou géométriques.

Ces dernières années ont vu une forte augmentation du volume de données acquises, allant de pair avec une hausse du nombre de capteurs et une amélioration de la résolution. Le suivi des changements peut ainsi se faire de façon plus précise et plus fréquente. Cependant à cause de la quantité de données, le traitement doit se faire de façon plus automatique. Conjointement, l'augmentation de la résolution et la multiplication des situations multi-capteurs et multi-incidences compliquent la comparaison des images (voir la Section 1.3). Il est ainsi plus difficile d'effectuer les corrections nécessaires à la comparaison des images. Il est donc nécessaire de développer des méthodes automatiques et invariantes permettant de détecter des changements dans des situations complexes.

Nous présentons ici une nouvelle méthode automatique de détection de changements

entre deux images optiques ou deux images radar, reposant sur les algorithmes SIFT ou SAR-SIFT (selon les modalités d'acquisition). Cette méthode ne nécessite pas de prétraitement, ni de recalage. Nous illustrons cette approche avec différents exemples d'applications (situations multi-capteurs ou multi-incidence).

Nous exposerons dans un premier temps les méthodes de détection de changements de la littérature, puis nous développerons notre approche.

7.1 Etat de l'art sur la détection de changements en télédétection

Les méthodes de détection de changements peuvent être regroupées en trois catégories [Dee98] selon le type de comparaison effectué : comparaison de pixels (*pixel-based*), de caractéristiques (*feature-based*) ou d'objets (*object-based*).

Les méthodes de comparaison de pixels ou de caractéristiques sont assez proches. Les deux se basent sur des comparaisons au niveau du pixel et ne prennent généralement pas en compte le contexte spatial. Néanmoins, les méthodes de comparaison de caractéristiques s'appuient sur des opérateurs statistiques appliqués en chaque pixel, et non sur les seules valeurs de pixel. Ces deux types de méthodes nécessitent des recalages sous-pixeliques ainsi que de possibles corrections radiométriques. Elles sont bien adaptées aux images de basse et moyenne résolution, mais présentent des limites pour les images de haute résolution.

Dans les méthodes orientées objet, l'unité n'est plus un pixel individuel mais un groupe de pixels (objet). Un objet est défini par un ensemble de pixels adjacents et homogènes. Le critère d'homogénéité peut reposer sur des informations spectrales, texturales ou géométriques. Ce type d'approche est apparenté à la vision humaine, qui identifie des objets homogènes plutôt que des pixels individuels. Ces méthodes permettent de mieux gérer les erreurs de recalage et de calibration. Elles sont particulièrement bien adaptées aux images de haute résolution.

Nous présentons un résumé des méthodes de détections de changement pour ces trois catégories. Une revue plus détaillée de ces techniques peut être trouvée dans [JHGQ08, HCC⁺13].

Comparaison de pixels Ces méthodes sont les plus anciennes. On peut citer les comparaisons directes de pixels, comme le calcul de la différence [CE96] ou le rapport [HW81] entre deux images. Les pixels sont ensuite détectés comme changement à l'aide de seuils adaptés.

Comparaison de caractéristiques Ces méthodes s'appuient sur la comparaison d'informations mesurées en chaque pixel. Elles nécessitent une transformation des propriétés spatiales ou spectrales de l'image. On peut ainsi citer l'analyse en composantes principales [DWDQ08] ou la comparaison de textures [TEK11]. L'algorithme GeoCDX [KCS⁺13] calcule des caractéristiques comme l'entropie, l'asymétrie et l'orientation de lignes, puis effectue une détection de changements par pixel avec une tolérance spatiale sur l'erreur de recalage. L'algorithme MIMOSA [QPPNL14] repose sur le calcul de deux types de moyennes temporelles sur une pile d'images. Les changements sont ensuite relevés à l'aide d'un seuillage automatique sur la densité de probabilité jointe de ces moyennes.

Comparaison d'objets Parmi les méthodes orientées objet, la comparaison post-classification [JMTU06, KSS⁺12] est très utilisée : les changements sont mesurés en comparant les classifications de deux images. Les algorithmes d'apprentissage automatique

(*machine learning*) de type réseaux de neurones [PPSA05] ou SVM (*Support Vector Machine*) [NY06,BBM08] ont prouvé leur efficacité pour ce type d'applications. On peut aussi citer les méthodes combinant des détections orientées pixels et objets. Gamba [GDL06] propose ainsi d'effectuer une comparaison de lignes extraites sur des images radar multi-temporelles pour valider des changements détectés par une méthode pixellique.

7.2 Détection de changements à l'aide d'algorithmes de type SIFT

Afin d'éviter les problèmes d'erreur de recalage ou de calibration radiométrique, nous proposons d'utiliser les points-clés détectés et mis en correspondance entre deux images à l'aide d'algorithmes de type SIFT. L'utilisation de ce type d'algorithme permet en effet la comparaison directe d'images, sans avoir recours à des pré-traitements (radiométriques ou géométriques) grâce aux propriétés d'invariance des points-clés et des descripteurs. L'idée est de considérer que les points-clés donnent une information sur les zones présentant des structures, tandis que les points appariés indiquent les emplacements qui n'ont pas changé. On souhaite alors repérer les changements en recherchant les zones où l'on trouve des points-clés mais peu d'appariements.

Dans ce chapitre, nous présentons une étude prospective sur l'utilisation des algorithmes de type SIFT pour la détection de changement, suivant l'idée exposée ci-dessus. Nous présentons dans un premier temps des observations préliminaires sur le comportement des points-clés et des appariements en zone de changement et de non-changement. Ces observations nous amènent à proposer une approche *a contrario* [DMM07] pour la détection de changements sur des paires optiques ou radar. Nous réalisons ensuite une étude sur les paramètres de cette méthode, suivie de quelques exemples d'applications sur des images réelles. Enfin nous présentons des exemples d'applications de cette méthode avec des paires d'images optiques et radar dans différentes situations (multi-capteurs ou multi-incidences).



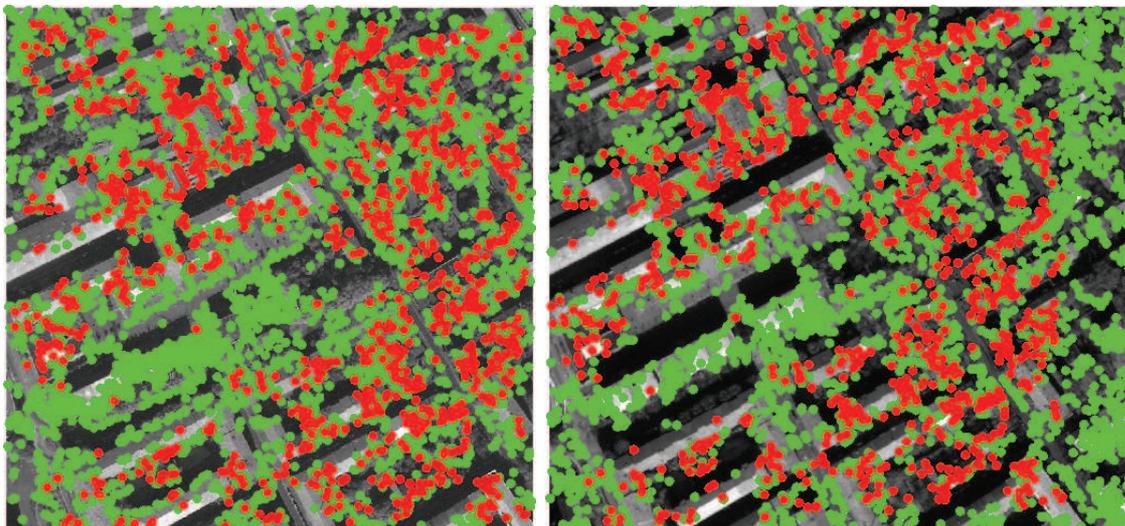
(a) Image Geoeye 18/10/09.

(b) Image Worldview 24/02/10

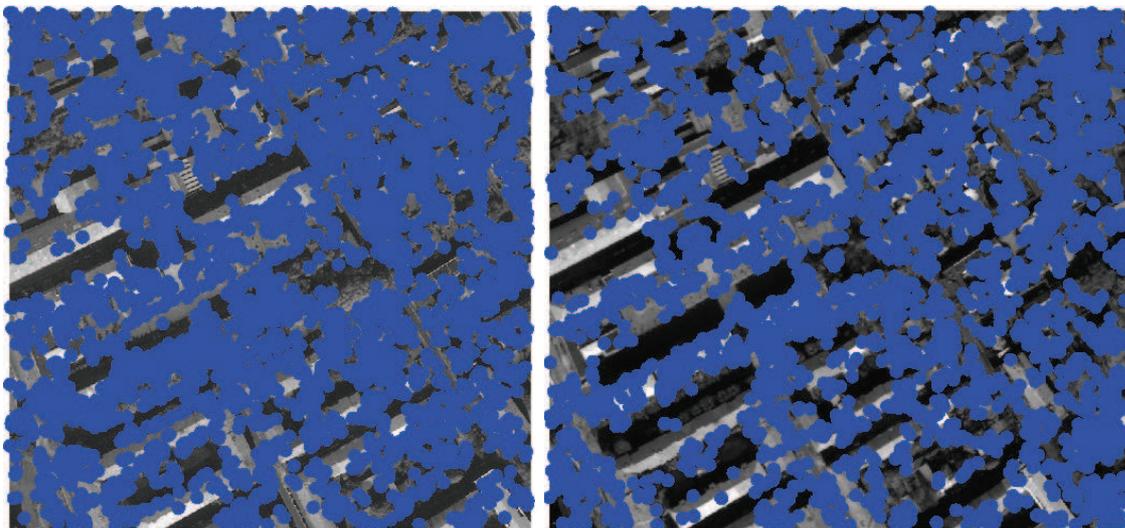
FIGURE 7.1 – Exemple de zone de changement. Les images ont été acquises par deux capteurs différents avec des incidences proches du nadir et une même résolution (50 cm). La zone de changement est entourée en vert.

7.2.1 Observations préliminaires

Les points-clés sont détectés et mis en correspondance à l'aide de l'algorithme SIFT pour les couples d'images optiques et SAR-SIFT (Section 5) pour les couples d'images radar. Les appariements sont ensuite filtrés par l'algorithme AC-RANSAC, associé à la recherche d'une transformation affine (voir la Section 6.1).



(a) points-clés détectés (en vert) et appariements (en rouge).



(b) points-clés non mis en correspondance (en bleu).

FIGURE 7.2 – Affichage des points-clés (en vert) et mises en correspondance (en rouge) du couple de Figure 7.1. On trouve 4388 points-clés et 1065 appariements sur l'image de gauche, et 3567 points-clés et 965 appariements sur l'image de droite. Les points-clés non mis en correspondance sont ensuite présentés en bleu.

Afin d'observer le comportement des points-clés et des appariements sur les zones de changement et de non changement, l'algorithme SIFT est appliqué au couple de la Figure 7.1, offrant un exemple de changement avec un bâtiment en cours de construction. La Figure 7.2(a) présente les points-clés détectés (en vert) et les appariements (en rouge) sur chacune des images. On remarque que des points-clés sont détectés sur l'ensemble de l'image, zone de changement ou non. Par contre on trouve très peu d'appariements sur la

zone de changement, à l'inverse du reste de l'image.

Sur les zones sans changement apparent, le nombre d'appariements est faible comparé au nombre de points-clés détectés. Le pourcentage de points-clés mis en correspondance après filtrage est de l'ordre de 25% sur ces images. Supprimer les points-clés appariés du reste des points ne permet pas de mettre en évidence les changements. L'affichage des points-clés non mis en correspondance (Fig. 7.2(b)) montre que ce n'est pas suffisant. On ne peut donc pas considérer les points-clés individuellement et il faut plutôt prendre en compte leur densité. On appellera densité la répartition des points-clés sur l'image par rapport au nombre total de points-clés.

Néanmoins la densité des points-clés n'est pas constante sur toute l'image. Certaines structures vont présenter plus de points-clés que d'autres de taille similaire. On ne peut pas non plus se contenter de soustraire la densité des appariements à la densité des points-clés. Il y aurait en effet un risque de détecter par erreur comme changement certaines zones avec de nombreux points-clés.

Il semble ainsi plus adéquat de comparer les distributions de points-clés et d'appariements de façon locale. Dans la suite, nous proposons une approche s'appuyant sur cette idée pour mettre en évidence les zones de changement.

7.2.2 Détection de changements par densité de points-clés et approche a contrario

Nous rappelons dans un premier temps le principe des approches a contrario [DMM07], puis proposons une approche originale de détection de changement.

7.2.2.1 Principe de l'approche a contrario

L'approche a contrario est basée sur le principe de Helmholtz [DMM07] : nous ne percevons une structure que s'il est peu probable qu'elle soit due au hasard. Cette approche repose sur deux concepts :

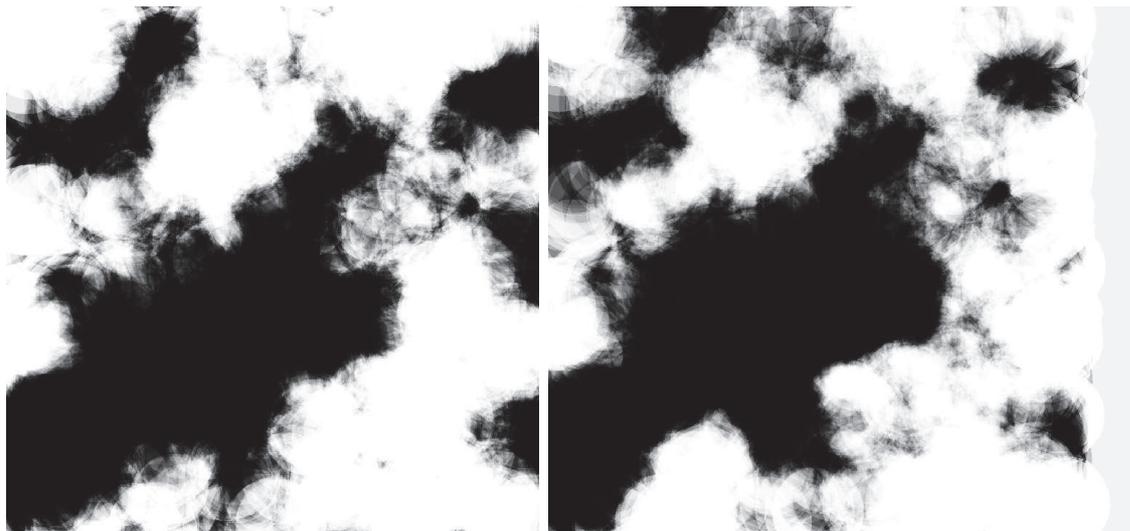
- un modèle de fond définissant les configurations pour lesquelles aucune structure ne devrait être détectée (hypothèse \mathcal{H}_0)
- une mesure x sur les structures à détecter.

A l'aide de ces concepts, on définit la quantité NFA (Nombre de Fausses Alarmes) qui est une borne sur l'espérance du nombre de structure générée par le modèle de fond. Une faible valeur de NFA correspondra à des structures ayant une faible probabilité d'apparaître et indiquera donc la présence de structures pertinentes. Le NFA est obtenu à partir de la probabilité $P(X \geq x)$, X étant la variable aléatoire correspondant au modèle de fond. Les structures sont détectées à l'aide d'un seuillage $NFA \leq \varepsilon$, avec de faibles valeurs de ε .

7.2.2.2 Approche proposée

On effectue la supposition suivante : la densité de points-clés sur une zone \mathcal{A} de l'image est la même que la densité des appariements. La densité de points-clés est la proportion de points dans \mathcal{A} par rapport à l'ensemble des points-clés. Connaissant la densité des appariements sur une zone, on s'étonnera si l'on trouve bien plus de points-clés qu'attendu. Cette situation caractérise bien les zones de changement : on y trouve, en proportion par rapport au reste de l'image, bien plus de points-clés que d'appariements.

Afin d'exprimer cette situation dans une approche a contrario, le modèle de fond \mathcal{H}_0 est défini de la façon suivante : les points-clés sont distribués selon un processus de Poisson de même densité que les appariements. La distribution des appariements est ainsi choisie comme référence.

(a) Valeurs de NFA .

(b) points-clés détectés comme changement.

FIGURE 7.3 – Exemples de détection de changements sur la paire d'image de 7.1.

On se donne une zone de l'image \mathcal{A} . On mesure la quantité x comme le nombre de points-clés réellement détectés dans la zone \mathcal{A} , parmi les N points-clés détectés sur toute l'image. On définit une variable X comme la distribution du nombre de points-clés sous l'hypothèse \mathcal{H}_0 . On appelle p l'intensité du processus de Poisson sur la zone \mathcal{A} (dont le calcul est explicité plus loin). Conditionnellement au fait que le domaine de l'image contient N points-clés et que les points-clés sont distribués selon un processus de Poisson, X suit une loi binomiale de paramètre p . La probabilité de trouver x points-clés dans \mathcal{A} sous l'hypothèse \mathcal{H}_0 est ainsi :

$$P_{\mathcal{H}_0}(X = x) = \binom{N}{x} p^x (1-p)^{N-x}. \quad (7.1)$$

Nous obtenons ensuite comme valeur de NFA :

$$NFA = N_{test} \cdot P_{\mathcal{H}_0}(X \geq x) = \sum_{i=x}^N \binom{N}{i} p^i (1-p)^{N-i}. \quad (7.2)$$

La quantité N_{test} correspond au nombre de points-clés N .

En pratique, on définit \mathcal{A} comme un voisinage circulaire de rayon r centré sur un point-clé. On définit également la quantité n de points-clés dans \mathcal{A} et m d'appariements. On en déduit les valeurs $x = n$ et $p = \frac{m}{M}$, où M est le nombre total d'appariements. La quantité NFA est ainsi calculée en chaque point-clé. Les points-clés sont ensuite classifiés comme changement à l'aide du seuil ε sur NFA . Les deux paramètres à fixer dans cette approche sont le rayon r et le seuil ε , dont nous allons étudier l'influence.

Les détections sont calculées indépendamment sur chaque image. La détection n'est donc pas symétrique. S'il y a apparition d'un objet sur une zone homogène, le changement ne sera détecté que sur la deuxième image. En effet, il n'y aura pas de détection de points-clés sur la zone correspondante dans la première image. Une comparaison des zones détectées comme changement pourrait permettre de mettre en évidence les apparitions ou disparition d'objets.

Un exemple d'application sur le couple d'images de la Figure 7.1 est présenté en Figure 7.3. Le choix des paramètres ($r = 60$ et $\varepsilon = 10^{-15}$) est empirique. De nombreux points sont correctement classifiés dans la zone de changement, et on ne trouve aucune fausse détection sur le reste de l'image. Cet exemple offre ainsi un premier résultat encourageant. Le principal problème réside maintenant dans le choix des paramètres r et ε . Les zones de changement étant de tailles quelconques, il n'est pas aisé de déterminer les valeurs optimales.

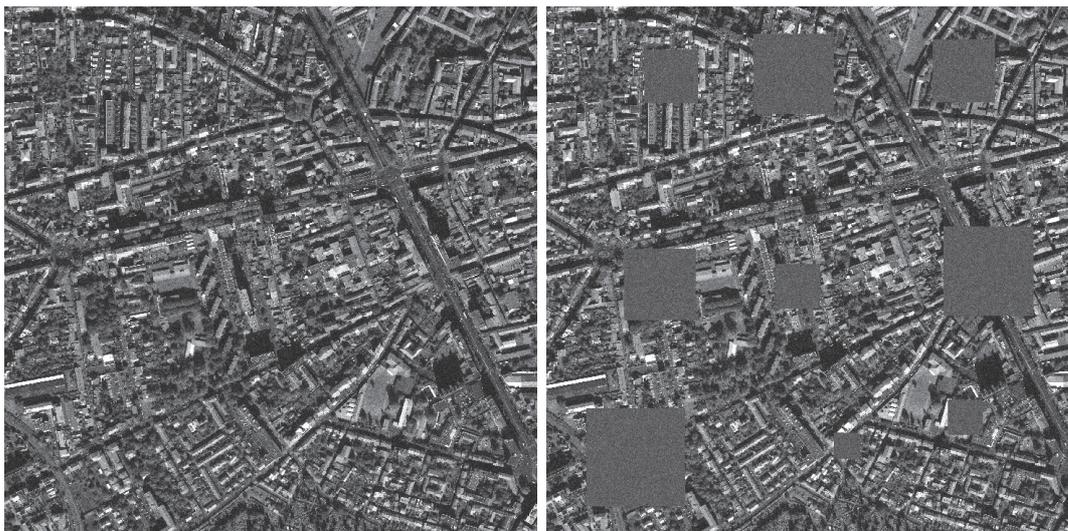
7.2.2.3 Etude de l'influence des paramètres

Nous présentons une étude sur des images simulées. Nous avons utilisé des images optiques et radar débruitées auxquelles ont été ajoutés de façon aléatoire des carrés homogènes de différentes tailles. Du bruit (gamma pour l'image radar et gaussien pour l'image optique) a été ajouté sur chacune des images afin d'éviter que les points-clés et descripteurs soient identiques. Les couples d'images étudiées sont présentées Figure 7.4. Neuf carrés sont rajoutés par image et leur taille varie de 100 à 380. Ces tailles correspondent aux changements observés sur des images réelles (construction/destruction de bâtiments). Le taux de points-clés appariés dans cette expérience est autour de 50%, tandis qu'il varie de 10 à 30% sur les images réelles. Nous avons choisi de supprimer de façon aléatoire des appariements pour atteindre un taux de 10%.

La méthode a été appliquée pour différentes valeurs de r et ε et deux données sont mesurées : le taux de détection et le taux de fausses alarmes. Pour le taux de détection nous avons choisi de compter le nombre de carrés présentant au moins 30 points-clés classifiés comme changement. L'idée est en effet qu'une zone de changement doit être repérée par un nombre suffisant de points mais qu'il n'est pas nécessaire que tous les points-clés de la zone soient détectés comme changement. Les changements ne seront observés que sur les images originales («images sans carrés»). En effet, on ne devrait pas trouver de changement sur les images auxquelles ont été ajoutés des carrés («images avec carrés») car elles ne présentent aucune nouvelle structure. La Figure 7.5(a) présente le nombre de carrés détectés en fonction de r et ε . Le nombre maximal de carrés est dix-huit.

Le taux de fausses alarmes est calculé comme le pourcentage de points-clés classifiés comme changements dans les zones de non-changement. Ce taux a été calculé sur les «images sans carrés» (Fig. 7.5(b)) et sur les «images avec carrés» (Fig. 7.5(c)).

En premier lieu, on peut remarquer que le nombre maximal de détections est atteint pour de nombreuses combinaisons de r et ε . Le taux de fausses alarmes pour les «images avec carrés» dépend peu de r et atteint des valeurs proches de zéro rapidement quand ε diminue (dès $\varepsilon = 10^{-5}$). En conséquence, la valeur de r ne semble pas influencer le taux



(a) Images optiques.

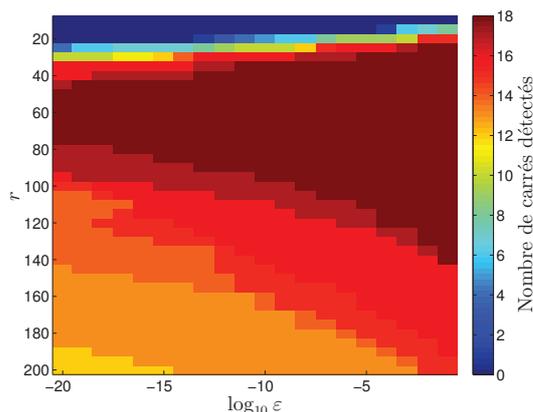


(b) Images radar.

FIGURE 7.4 – Couples d’images simulées utilisées pour évaluer les paramètres de l’approche a contrario. A gauche les images originales («images sans carrés»), à droite les «images avec carrés», où sont simulées des zones de changement.

de fausses alarmes sur les zones sans changement. Ce taux dépend seulement de ε . Cette observation n’est pas valable pour le taux de fausses alarmes sur les «images sans carrés». Pour les grandes valeurs de ε ($\varepsilon > 10^{-5}$), ce taux est autour de 1% pour $r < 60$ mais il devient vite très grand quand r est supérieur à 60. On observe également une augmentation de ce taux pour ces valeurs de r quand, conjointement, r augmente et ε diminue. On suppose que de trop fortes valeurs de r entraînent de nombreuses fausses détections dans le voisinage des zones de changements. Toutefois, les taux de fausses alarmes sont proches de zéro pour $10^{-20} < \varepsilon < 10^{-10}$ et $40 < r < 80$. Par ailleurs, ces valeurs correspondent à de très bons taux de détection. Les valeurs de r et ε doivent donc être choisies dans cette plage de valeurs.

Nous choisissons ici $r = 60$ et $\varepsilon = 10^{-10}$ pour obtenir un compromis entre un bon taux de détection et un faible taux de fausses alarmes. D’autres études sont néanmoins nécessaires pour préciser ces paramètres.



(a) Nombre de carrés détectés.

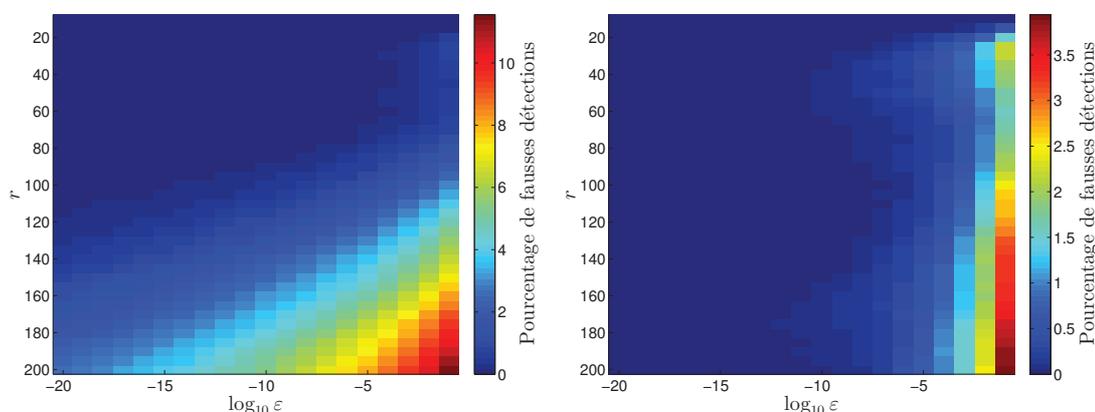
(b) Taux de fausses alarmes sur les images sans car- (c) Taux de fausses alarmes sur les images avec car-
rés. rés.

FIGURE 7.5 – Nombre de carrés détectés et taux de fausses alarmes pour les couples de Fig. 7.4 et pour différentes valeurs de r et ε . Pour faciliter la lecture, l'échelle définit $\log_{10} \varepsilon$ plutôt que ε .

7.2.2.4 Exemples de résultats

Nous présentons dans la suite quatre exemples de détection de changements : deux paires radar (Fig. 7.6 et 7.7) et deux paires optiques (Fig. 7.8 et 7.9). Pour chaque cas, la zone de changement est entourée en vert sur les premières images. Les points-clés détectés comme changements sont ensuite affichés en rose. Les paramètres ont été fixés à $r = 60$ et $\varepsilon = 10^{-10}$ pour ces expériences.

Les Figures 7.6, 7.7 et 7.8 représentent une même zone subissant un changement récurrent : l'occupation régulière d'un parking. Dans les trois situations, le parking est vide ou faiblement occupé sur une image, tandis qu'il est majoritairement occupé sur l'autre. On souhaite alors détecter ces changements, soit apparition et disparition d'objets. Il y a aussi un changement supplémentaire sur la Figure 7.8 avec une zone en construction.

Néanmoins, les trois situations n'offrent pas la même complexité. La Figure 7.6 est un cas plutôt simple puisqu'elle présente deux images radar acquises dans les mêmes conditions et avec des dates proches. Elles sont donc facilement comparables. Le cas de la Figure 7.7 est plus compliqué : ce sont deux images radar acquises avec des incidences différentes. Les images de la Figure 7.8 sont des images optiques acquises par des capteurs différents mais dans des conditions proches.

L'approche proposée est efficace pour détecter les zones de changement dans les trois



(a) Paires d'images Cosmo-SkyMed.



(b) Points-clés détectés comme changement.

FIGURE 7.6 – Exemple de détection de changements sur une paire d'images radar Cosmo-SkyMed. Les images ont été acquises dans les mêmes conditions (48° d'angle d'incidence et 1 m de résolution) et à des dates proches (16 jours d'écart).

cas. Un grand nombre des points-clés détectés dans ces zones sont correctement classifiés comme changement. On trouve cependant quelques erreurs liées aux différences de géométrie des bâtiments pour la Figure 7.7. Sur la Figure 7.8, d'autres points-clés sont détectés comme changements, mais ils sont liés à des différences de végétation ou des déplacements de voitures. Ce ne sont donc pas à proprement parler des erreurs.

La Figure 7.9 présente un dernier exemple, avec une paire d'images optiques acquises dans des conditions différentes et avec un écart de plusieurs années. Cette paire présente une grande zone de changement qui est correctement repérée sur l'image de droite, mais moins bien sur l'image de gauche. On remarque aussi d'autres détections liées à de la végétation ou des voitures. On trouve cependant des fausses détections sur plusieurs bâtiments (en haut de l'image). Ces bâtiments offrent des différences de luminosité très importantes et l'algorithme SIFT échoue à les mettre en correspondance.



(a) Images CosmoSkyMed 34° (à gauche) et 48° d'incidence (à droite).



(b) Points-clés détectés comme changement.

FIGURE 7.7 – Exemple de détection de changements sur une paire d'images radar CosmoSky-Med. Les images ont été acquises avec la même résolution (1 m) mais des incidences différentes. L'acquisition s'est faite avec quelques mois d'écart (04/11 et 09/11), il y a donc des variations pour la végétation.

7.2.2.5 Conclusions

Nous avons proposé une nouvelle méthode de détection de changement, basée sur l'algorithme SIFT et une approche a contrario. Cette méthode présente des résultats encourageants sur différents types de situations : paires d'images optiques et radar, en situation multi-capteurs et/ou multi-incidences. Il semble cependant nécessaire d'effectuer des études plus poussées pour déterminer les paramètres optimaux. Une approche multi-échelle peut être envisagée afin de considérer les différentes tailles de changements possibles. Une détection annexe de la végétation et des voitures pourrait être effectuée afin de supprimer les détections liées à ces perturbations.



(a) Image GeosEye 10/09 (à gauche) et WorldView 02/10 (à droite).



(b) Points-clés détectés comme changement.

FIGURE 7.8 – Exemple de détection de changements sur une paire d'images optiques. Les images ont été acquises par des capteurs différents (GeosEye et WorldView), avec une même résolution (50 cm) et une même incidence (proche du nadir). L'acquisition s'est faite avec quelques mois d'écart (10/09 et 02/10), il y a donc des variations pour la végétation.



(a) Image Geoeye (à gauche) et Worldview (à droite).



(b) Points-clés détectés comme changement.

FIGURE 7.9 – Exemple de détection de changements sur une paire d'images optiques. Les images ont été acquises par des capteurs différents (Quickbird et Geoeye), avec des résolutions différentes (60 et 50 cm) et des incidences proches (proche du nadir). L'acquisition s'est faite à plusieurs années d'intervalle (04/02 et 10/09), il y a donc des variations au niveau de la végétation mais aussi des bâtiments et du sol.

Conclusions et perspectives

Conclusions

L'objectif de ce travail doctoral était de proposer des descripteurs locaux pour les images de télédétection, puis d'étudier leur intérêt pour différentes applications. Devant la multiplication des données et des types d'acquisition, les descripteurs locaux apportent un intérêt certain pour des applications en télédétection, compte tenu de leur efficacité pour mettre en correspondance des images acquises dans des conditions différentes. Nous avons choisi d'adopter l'algorithme SIFT, déjà couramment utilisé en imagerie optique en raison de ses invariances et de son efficacité sur les images naturelles. Cet algorithme associe une détection de points d'intérêt à l'extraction de descripteurs locaux, codant la géométrie locale.

A la différence des images optiques, les images radar présentent cependant un fort bruit multiplicatif, dit de speckle, qui complique les traitements. Nous avons mis en évidence les limites de l'algorithme SIFT sur ce type d'images puis identifié les étapes à améliorer. Nous avons constaté qu'il était nécessaire de développer un nouveau calcul du gradient adapté à ce type de bruit. En nous inspirant de travaux sur la détection de contour, nous avons proposé le Gradient par Ratio, qui permet de calculer, à différentes échelles, une norme et une orientation du gradient robustes face au bruit de speckle. Nous nous sommes appuyés sur cet opérateur pour adapter les étapes de l'algorithme SIFT aux statistiques des images radar et ainsi proposer un nouvel algorithme, appelé SAR-SIFT, pour la comparaison d'images radar. Cet algorithme représente la contribution principale de nos travaux. Nous avons montré que les performances de l'algorithme SAR-SIFT sont supérieures, dans ce contexte, à celle de l'algorithme SIFT original. Nous avons ensuite évalué son comportement pour des paires d'images acquises dans des conditions différentes. L'algorithme montre ses limites dans des situations où l'angle d'incidence varie.

Nous avons proposé des approches pour des applications en télédétection, basées sur les descripteurs locaux. Nous avons démontré l'utilité de l'algorithme SAR-SIFT pour l'estimation d'une transformation globale entre deux images radar. L'utilisation des points-clés mis en correspondance comme points homologues, associée à l'algorithme AC-RANSAC, nous a permis d'envisager deux applications de recalage : l'estimation d'une transformation affine entre deux images pour les zones plates et l'estimation de la translation résiduelle entre deux images géoréférencées. Nous avons également dégagé un autre avantage de l'utilisation de l'algorithme AC-RANSAC : le filtrage des faux appariements, en particulier pour les paires d'images avec des changements d'angle d'incidence.

Enfin, nous avons mené une étude prospective sur l'utilisation des descripteurs locaux (SIFT en imagerie optique et SAR-SIFT en imagerie radar) pour la détection de changements en télédétection. Nous avons proposé de comparer, à l'aide d'une approche a contrario, les densités de points-clés mis en correspondance aux densités de points-clés détectés. Nous avons mis en évidence l'intérêt de cette méthode à l'aide de résultats préliminaires sur des paires d'images radars et des paires d'images optiques.

Perspectives

A partir des travaux que nous avons développés, différentes pistes peuvent être envisagées.

Adaptation de l’algorithme SAR-SIFT aux transformations affines L’algorithme proposé est invariant aux transformations de type similitude (translation, rotation et changement d’échelle). Cependant, on rencontre souvent des situations de transformations affines, notamment pour les paires d’images acquises avec des angles d’incidence différents. La transformation globale des points au sol peut effectivement être modélisée par une transformation affine (Section 6.2.2). D’autre part, les changements d’incidence entraînent des distorsions géométriques pour les objets en hauteur (Fig. 1.18) que l’on peut rapprocher des effets de parallaxe en imagerie optique. Nous avons remarqué que l’algorithme SAR-SIFT offre des performances limitées dans ces situations. Il semble intéressant de développer l’invariance du descripteur Ratio aux transformations affines. Nous pourrions nous inspirer du détecteur Harris-Affine [MS04a], qui estime une région elliptique autour de chaque point-clé détecté à l’aide de la matrice d’auto-corrélation.

Détection de changements basée sur les descripteurs locaux Nous avons présenté une étude prospective de l’utilisation de descripteurs locaux pour une application de détection de changements en télédétection. Cette étude mérite d’être prolongée à l’aide de cas réels notamment. Par ailleurs, nous avons choisi de calculer le *NFA* (Nombre de Fausses alarmes) de façon discrète (en chaque point-clé) mais il pourrait aussi s’effectuer de façon plus dense en l’évaluant en chaque pixel. L’information d’échelle des points-clés pourrait également être prise en compte, par exemple en donnant plus de poids aux mises en correspondance des points avec de grandes échelles. Afin d’améliorer la précision de cette approche de détection de changements, il est aussi nécessaire d’augmenter le pourcentage de points-clés mis en correspondance. Cela peut s’effectuer en utilisant des descripteurs locaux invariants aux transformations affines. Enfin, une détection annexe des zones de végétation peut être envisagée pour gérer les détections liées aux changements saisonniers. L’utilisation de la bande proche-infrarouge pour les images optiques, notamment, pourrait être utile.

Reconnaissance d’objets à l’aide des descripteurs locaux En vision par ordinateur, les descripteurs locaux sont très utilisés pour la reconnaissance d’objets. Des applications de reconnaissance de bâtiments à l’aide de l’algorithme SIFT sont déjà proposées en imagerie optique [SU09]. On pourrait envisager de la même façon l’utilisation de l’algorithme SAR-SIFT pour la reconnaissance de bâtiments en imagerie radar, avec des approches comme les sacs de mots ou les graphes.

On peut aussi s’appuyer sur les descripteurs SIFT et SAR-SIFT pour définir un modèle générique de représentation des objets à l’aide de graphes et de méthodes d’apprentissage. L’appariements de graphes permet ensuite de comparer des paires d’images optiques/radar ou des séries multi-temporelles. Cela peut être utile pour de la reconnaissance d’objets, mais aussi de la détection de changements, de la reconstruction 3D ou de la segmentation.

Quatrième partie

Annexes

Annexe A

Bases de données à disposition

A.1 Données radar

Les images à disposition proviennent de deux satellites, TerraSAR-X (TSX) et Cosmo-SkyMed (CSK), et représentent la ville de Toulouse, France. Elles ont été acquises en mode Stripmap (SM) ou Spotlight (SL), avec une visée à droite et en orbite ascendante (A) ou descendante (D). Les images sont représentées en projection slant range (SLC). Ici nous n'utilisons que les images en amplitude (ou en intensité).

Le tableau suivant détaille les cinq configurations d'images utilisées. Pour chaque configuration, une ou plusieurs images étaient disponibles à différentes dates. Des exemples d'images sont représentés par la suite.

Données	1	2	3	4	5
Satellite	TSX	TSX	CSK	CSK	CSK
Acquisition	SL	SL	SL	SL	SM
Angle d'incidence	34°	34°	48°	34°	43°
Range	1 m	2 m	1 m	1 m	3 m
Azimuth	1 m	1 m	1 m	1 m	3 m
Orbite	A	A	D	D	D
Taille	7x5 km	10x5 km	10x10 km	10x10 km	40x40 km
Date	27/11/07	17/08/08 09/02/09	02/09/11 18/09/11 04/10/2011 05/10/2011	17/04/11 18/04/11	20/08/11 28/08/11 05/09/11 13/09/11 21/09/11 29/09/11
Figure	A.1	A.2	A.3	A.4	A.5,A.6



FIGURE A.1 – TerraSAR-X Spotlight 34° (27/11/2007)



FIGURE A.2 – TerraSAR-X Spotlight 34° (17/08/2008)



FIGURE A.3 – Cosmo-SkyMed Spotlight 48° (02/09/2011)



FIGURE A.4 – Cosmo-SkyMed Spotlight 34° (18/04/2011)

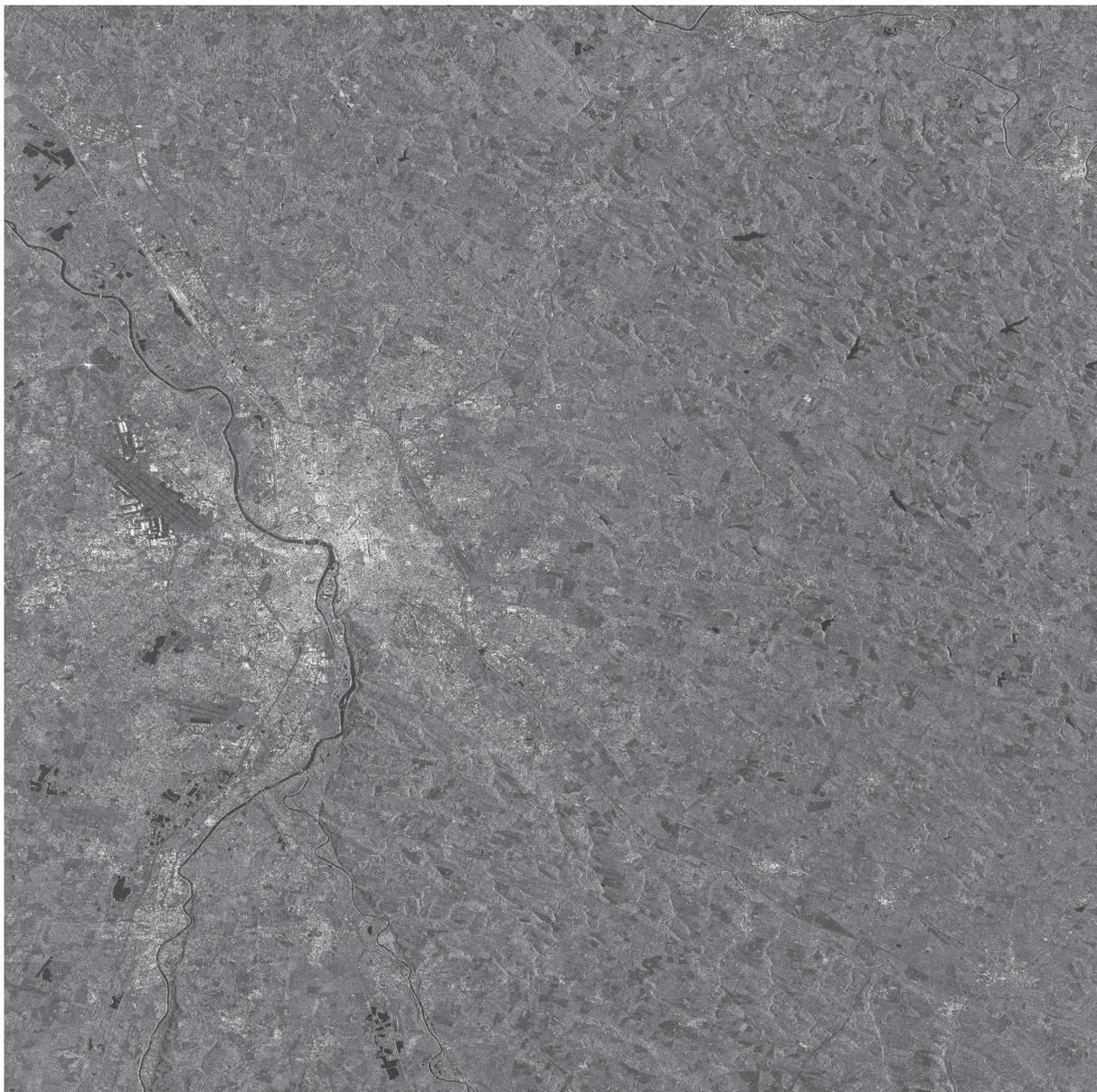


FIGURE A.5 – Cosmo-SkyMed Stripmap 43° (20/08/2011)



FIGURE A.6 – Cosmo-SkyMed Stripmap 43° (08/20/2011) avec sélection du milieu urbain

A.2 Données optiques

Nous avons à disposition une série temporelle avec des images acquises par des satellites différents pour des résolutions et angles différents. Cinq images sont mises à disposition. Elles ont toutes été acquises au dessus de la ville de Toulouse, France et proviennent de quatre satellites différents. Le tableau suivant détaille les caractéristiques de chaque image. Pour certaines images, l'angle de prise de vue n'est pas connu. Nous n'utilisons ici que les images panchromatiques. Ces images sont représentées par la suite.

Satellite	Quickbird	Quickbird	Geoeye	Worldview	Pleiades
Angle de visée				2°	23°
Résolution	60 cm	60 cm	50 cm	50 cm	70 cm
Date	14/03/02	01/04/02	18/10/09	24/02/10	07/05/12
Figure	A.7	A.8	A.9	A.10	A.11



FIGURE A.7 – Quickbird 60cm (14/03/2002)

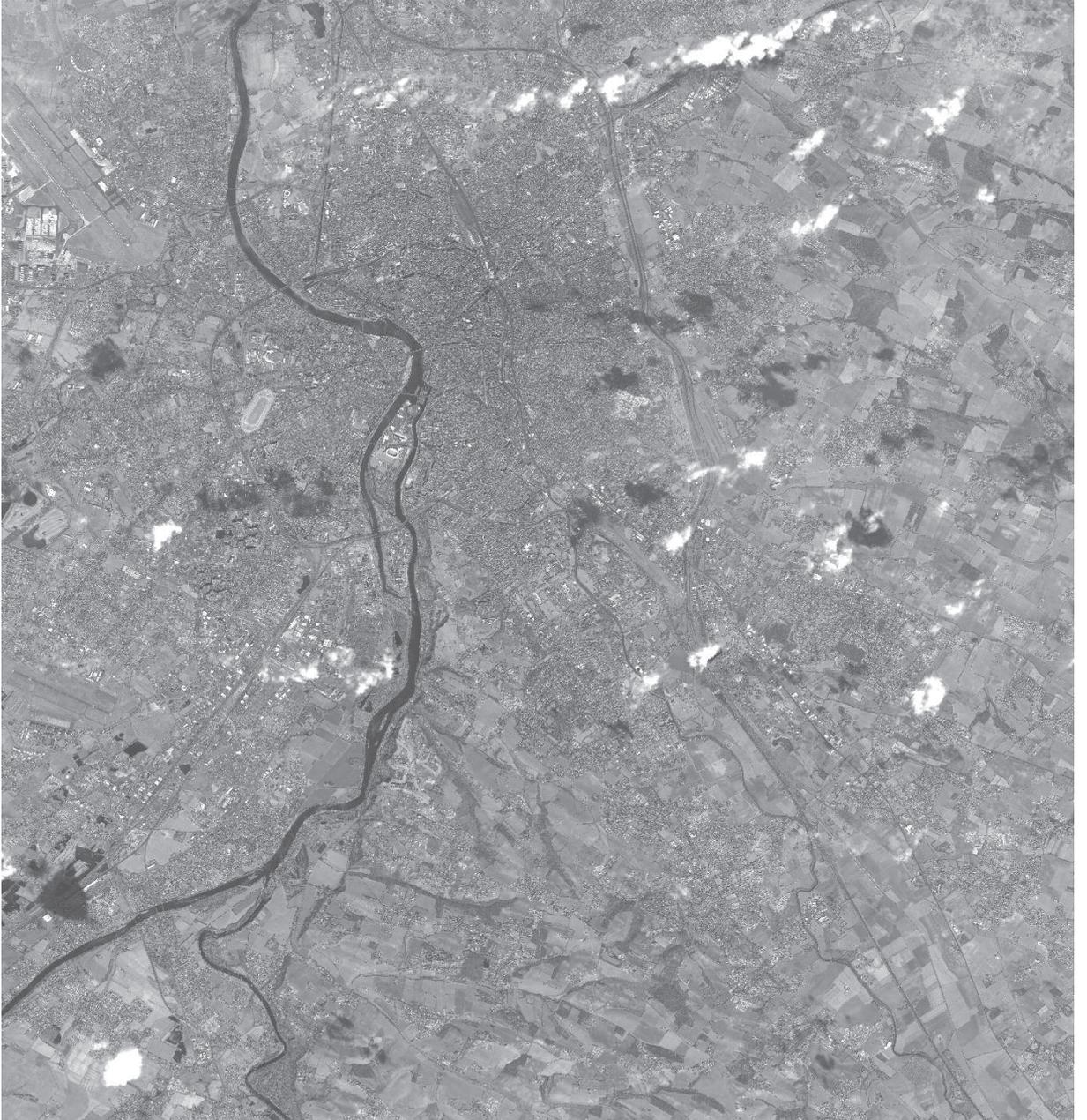


FIGURE A.8 – Quickbird 60cm (01/04/2002)



FIGURE A.9 – Geoeye 50cm (18/10/2009)

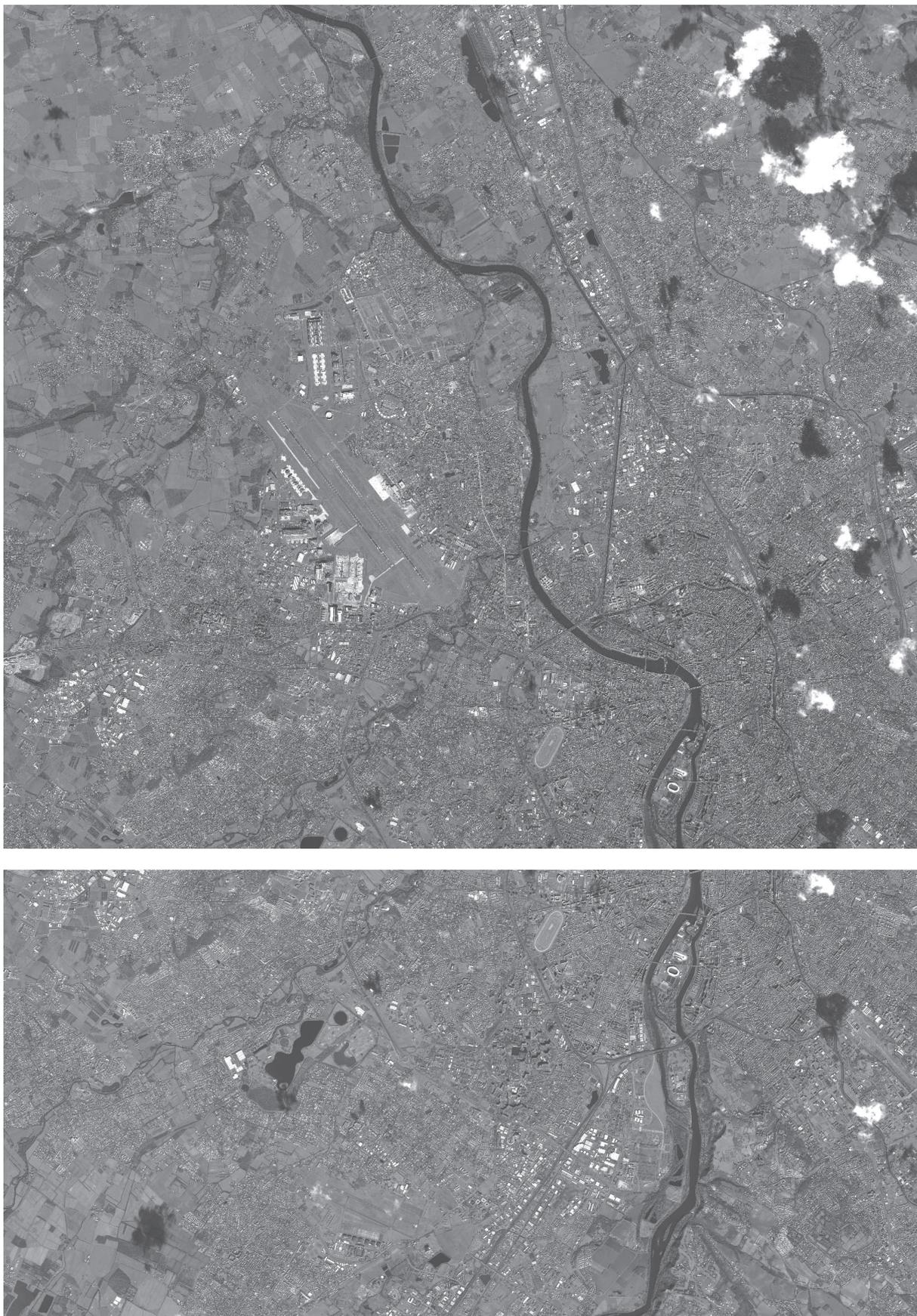


FIGURE A.10 – Worldview 50cm 2° (24/02/2010). Les données ont été fournies en deux parties.

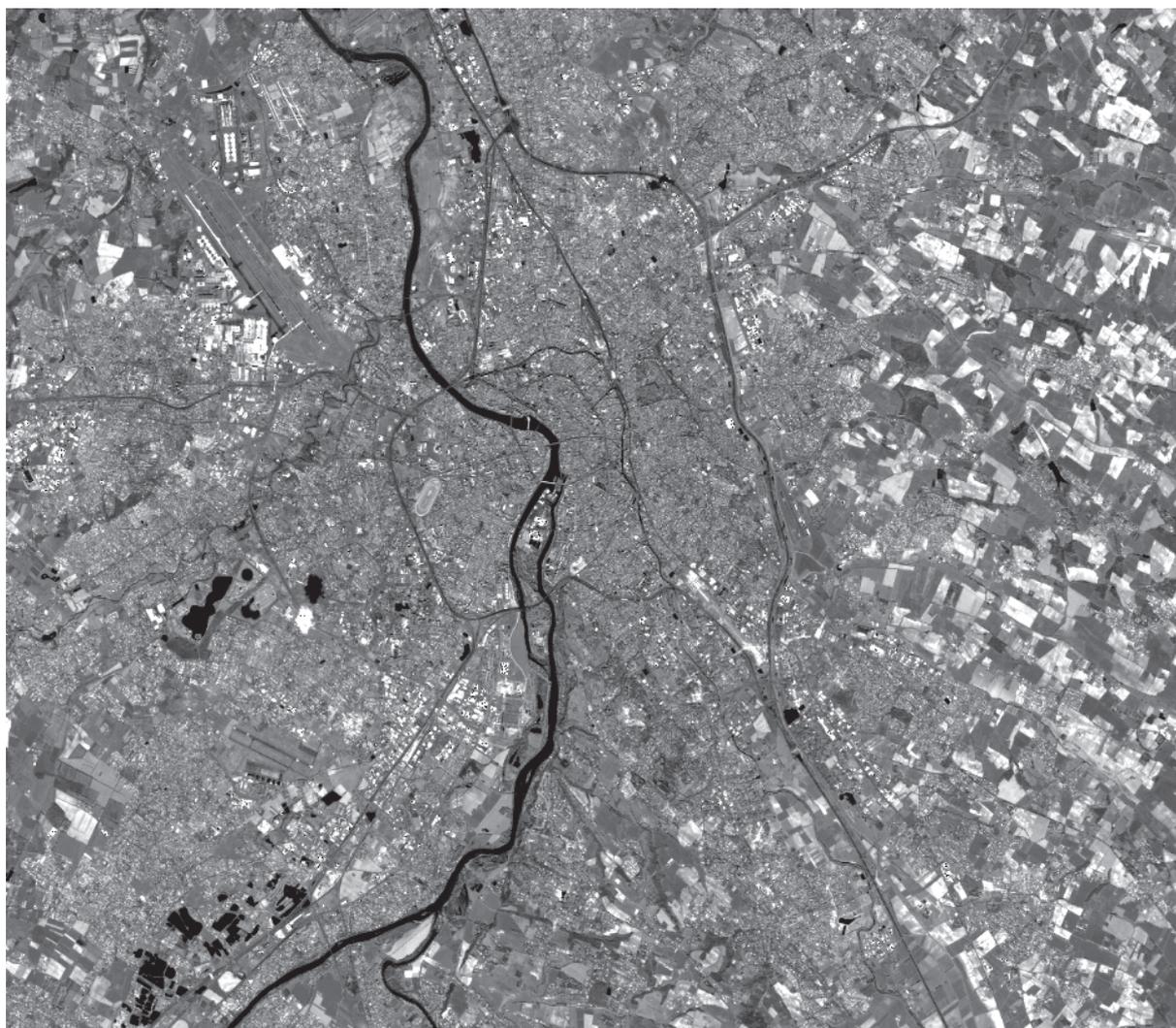


FIGURE A.11 – Pleïades 70cm 23° (07/05/2012)

Annexe B

Algorithme de type SIFT utilisé en imagerie optique

Ici nous utiliserons une variante de l'algorithme SIFT, développée par Rabin [Rab09]. Cette méthode sera utilisée pour l'imagerie optique, mais aussi pour comparer les performances de l'algorithme SAR-SIFT, développé pour les images radar. Les étapes utilisées ainsi que les paramètres sont présentés dans ce qui suit.

B.1 Détection de points-clés

La détection des points-clés se fait à l'aide de la méthode LoG-Harris (Section 2.1.2.4) : sélection des extrema en 3D du LoG, avec filtrage sur les petites valeurs du LoG et seuil sur le critère de Harris multi-échelle.

Les paramètres utilisés sont :

- $k_{max} = 13$: nombre d'échelles pour la construction de l'espace-échelle du LoG
- $\sigma_0 = 0.63$: valeur de la première échelle
- $r = 2^{1/3} = 1.26$: rapport entre deux échelles successives
- $t_l = 1$: seuil sur les petites valeurs du LoG (mais avec l'image I non normalisée)
- $a_1 = \sqrt{2}$: le facteur de proportionnalité de l'écart-type de la pondération exponentielle G du critère de Harris multi-échelle
- $k_1 = 0.04$: le paramètre de calcul du critère de Harris.

Le seuil d_h sur le critère de Harris, nécessaire pour la suppression des fausses détections, est souvent fixé à 1500 [MS01] ou 2000 [Rab09] pour les images numériques codées sur 8 bits. Cependant les images satellites optiques présentent généralement une radiométrie plus importante et spécifique au capteur (Section 1.1.5). Le critère de Harris multi-échelle étant proportionnel au contraste de l'image élevé à la puissance quatre, il est nécessaire de l'adapter à chaque image étudiée. Pour chaque image, on observe la distribution des valeurs des pixels et on mesure les valeurs c_1 et c_2 correspondant respectivement au premier et au dernier centile. On considère que le contraste de l'image étudiée est la différence $c_2 - c_1$ et que le contraste moyen c_m d'une image numérique sur 8 bits est égal à 200. On définit alors le seuil sur le critère de Harris comme $d_h = d_{h_0} \left(\frac{c_2 - c_1}{c_m} \right)^4$, avec $d_{h_0} = 2000$.

B.2 Sélection des orientations

Les orientations sont choisies à l'aide de la méthode de sélection *a contrario* sur les histogrammes d'orientation (Section 2.2.2). Les paramètres utilisés sont :

- $f_{orient} = 6$: facteur appliqué à l'échelle σ d'un point-clé pour définir le rayon de la zone de calcul des orientations.
- $nb_{orient} = 36$: nombre d'intervalles dans l'histogramme.

B.3 Extraction des descripteurs

Le type de descripteur utilisé est le descripteur SIFT circulaire présenté en section 2.3.5. Les paramètres utilisés sont :

- $r_{descp} = 12$, avec $r_{descp}\sigma$ le rayon du descripteur
- $r_1 = 3$, avec $r_1\sigma$ le rayon du premier anneau du masque
- $r_2 = 8.75$, avec $r_2\sigma$ le rayon du deuxième anneau du masque
- $nb_{descp} = 12$: nombre d'intervalles pour les histogrammes de chaque région

On obtient alors des descripteurs de taille $12 \times 9 = 128$.

Annexe C

Calculs et démonstrations

C.1 Calcul du bruit résiduel d'une image bruitée après ajout de flou gaussien

Soit une image I_0 constante et non bruitée, à laquelle est ajouté un bruit blanc gaussien $B \sim \mathcal{N}(0, \zeta^2)$. Cette image est convolée avec le filtre gaussien $G_\sigma(i, j)$ afin d'obtenir :

$$I_\sigma = G_\sigma * I = G_\sigma * I_0 + G_\sigma * B.$$

L'image I_0 étant non bruitée, $G * I_0$ est donc l'image convolée de I_0 , soit une image constante non bruitée. Le bruit résiduel de l'image I_σ est $B_2 = G_\sigma * B$. Expérimentalement, on a vérifié que B_2 est un bruit gaussien. On souhaite alors déterminer sa moyenne μ et sa variance ν .

L'image B_2 est la convolée de B :

$$B_2(k, l) = \sum_i \sum_j G_\sigma(i, j) B(k - i, l - j)$$

Les variables $B(i, j)$ sont indépendantes et identiquement distribuées puisque B est un bruit gaussien. Comme $B_2(k, l)$ est une combinaison linéaire des variables $B(i, j)$, on peut écrire :

$$\begin{aligned} \mu &= \mathbb{E}(B_2) = \frac{1}{2\pi\sigma^2} \sum_i \sum_j G_\sigma(i, j) \mathbb{E}(B) \\ \nu &= \text{var}(B_2) = \frac{1}{2\pi\sigma^2} \sum_i \sum_j G_\sigma(i, j)^2 \text{var}(B) \end{aligned}$$

L'espérance de B est nulle et sa variance est égale à ζ^2 . On obtient alors pour la moyenne $\mu = 0$ et pour la variance :

$$\begin{aligned} \nu &= \zeta^2 \sum_i \sum_j G(i, j)^2 \\ \nu &= \frac{\zeta^2}{4\pi\sigma^2} \end{aligned}$$

Le bruit résiduel d'une image bruitée avec un bruit gaussien de variance ζ^2 et après ajout d'un flou gaussien de variance σ^2 est un bruit blanc gaussien de variance $\nu^2 = \frac{\zeta^2}{4\pi\sigma^2}$.

C.2 Influence du bruit gaussien sur l'orientation du gradient d'un bord vertical ou horizontal

On effectue le calcul pour un bord horizontal. Soit une image I de dimension $N \times M$ avec :

$$I(x, y) = \begin{cases} m & \text{pour } y > M/2 \\ 0 & \text{pour } y \leq M/2. \end{cases}$$

On ajoute à I un bruit blanc gaussien $B \sim \mathcal{N}(0, \varsigma^2)$. On choisit de calculer le gradient de la façon suivante :

$$g_x(x, y) = \frac{I(x+1, y) - I(x-1, y)}{2}$$

$$g_y(x, y) = \frac{I(x, y+1) - I(x, y-1)}{2}.$$

Les valeurs de gradient $g_x(x, y)$ et $g_y(x, y)$ sont indépendantes et identiquement distribuées puisqu'elles ne font pas intervenir les mêmes pixels. Pour $y = \{M/2, M/2 + 1\}$, et pour tout x , les valeurs de $g_x(x, y)$ et $g_y(x, y)$ suivent les lois normales respectives $\mathcal{N}(0, \frac{\varsigma^2}{2})$ et $\mathcal{N}(\frac{m}{2}, \frac{\varsigma^2}{2})$.

Comme l'orientation du gradient est obtenue avec la formule $\theta(x, y) = \arctan\left(\frac{g_y(x, y)}{g_x(x, y)}\right)$, les valeurs de $\theta(x, y)$ satisfont l'égalité : $g_y = g_x \tan \theta$. On souhaite calculer la fonction de répartition de la variable θ : $F_\theta(t) = \mathbb{P}[\theta \leq t]$. A l'aide des symétries, on n'étudie que le cas $0 \leq \theta \leq t$ avec $0 \leq t \leq \frac{\pi}{2}$. La fonction de répartition est alors :

$$F_\theta(t) = \mathbb{P}[0 \leq \theta \leq t, 0 \leq t \leq \frac{\pi}{2}]$$

$$F_\theta(t) = \mathbb{P}[0 \leq g_x, 0 \leq g_y \leq g_x \tan t]$$

$$F_\theta(t) = \int_{a=0}^{a=+\infty} \int_{b=0}^{b=a \tan t} f_{g_y}(b) f_{g_x}(a) db da$$

$$F_\theta(t) = \frac{1}{\pi \varsigma^2} \int_{a=0}^{a=+\infty} \int_{b=0}^{b=a \tan t} \exp\left(-\frac{(b - \frac{m}{2})^2}{\varsigma^2}\right) \exp\left(-\frac{a^2}{\varsigma^2}\right) db da.$$

Après un changement de variables, on obtient :

$$F_\theta(t) = \frac{1}{\pi \varsigma^2} \int_{u=0}^{u=t} \int_{a=0}^{a=+\infty} a(1 + \tan^2 u) \exp\left(-\frac{a^2}{\varsigma^2}\right) \exp\left(-\frac{(a \tan u - \frac{m}{2})^2}{\varsigma^2}\right) da du.$$

On calcule alors la densité de probabilité de θ avec $f_\theta(t) = F'_\theta(t)$:

$$F'_\theta(t) = \frac{1}{\pi \varsigma^2} \int_{a=0}^{a=+\infty} a(1 + \tan^2 t) \exp\left(-\frac{a^2(1 + \tan^2 t) - am \tan t + \frac{m^2}{4}}{\varsigma^2}\right) da.$$

On effectue une intégration par parties :

$$f_\theta(t) = \exp\left(-\frac{m^2}{4\varsigma^2}\right) \left(\frac{1}{2\pi} + \frac{m \tan t}{2\pi \varsigma^2} \int_{a=0}^{a=+\infty} \exp\left(-\frac{a^2(1 + \tan^2 t) - am \tan t}{\varsigma^2}\right) da\right)$$

puis une factorisation :

$$f_\theta(t) = \exp\left(-\frac{m^2}{4\varsigma^2}\right) \left(\frac{1}{2\pi} + \frac{m \tan t}{2\pi \varsigma^2} \exp\left(\frac{m^2 \sin^2 t}{4\varsigma^2}\right) \int_{a=0}^{a=+\infty} \exp\left(-\frac{(a - \frac{m \sin t \cos t}{2})^2}{\varsigma^2 \cos^2 t}\right) da\right)$$

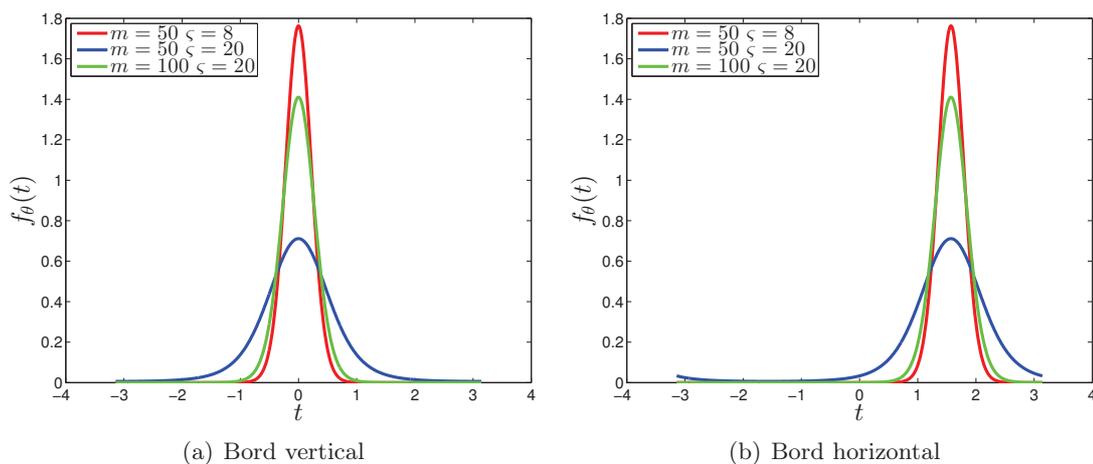


FIGURE C.1 – Fonction f_θ sur un bord vertical ou horizontal pour différentes valeurs de m et ζ .

et enfin un changement de variables :

$$f_\theta(t) = \exp\left(-\frac{m^2}{4\zeta^2}\right) \left(\frac{1}{2\pi} + \frac{m \sin t}{2\pi\zeta} \exp\left(\frac{m^2 \sin^2 t}{4\zeta^2}\right) \int_{a=-\frac{m \sin t}{2\zeta}}^{a=+\infty} \exp(-u^2) du \right).$$

En utilisant la fonction d'erreur $\operatorname{erf}(x) = \frac{2}{\sqrt{\pi}} \int_{u=0}^{u=x} \exp(-u^2) du$, on obtient la formule suivante pour la densité de probabilité des valeurs d'orientation :

$$f_\theta(t) = \frac{\exp\left(-\frac{m^2}{4\zeta^2}\right)}{2\pi} + \frac{m \sin t}{4\sqrt{\pi}\zeta} \exp\left(\frac{-m^2 \cos^2 t}{4\zeta^2}\right) \left(1 + \operatorname{erf}\left(\frac{m \sin t}{2\zeta}\right) \right).$$

Avec un changement de variable $u = t - \pi/2$, on obtient la densité de probabilité des valeurs d'orientations sur un bord vertical :

$$f_\theta(t) = \frac{\exp\left(-\frac{m^2}{4\zeta^2}\right)}{2\pi} + \frac{m \cos t}{4\sqrt{\pi}\zeta} \exp\left(\frac{-m^2 \sin^2 t}{4\zeta^2}\right) \left(1 + \operatorname{erf}\left(\frac{m \cos t}{2\zeta}\right) \right). \quad (\text{C.1})$$

La Figure C.1 affiche les valeurs de $f_\theta(t)$ en fonction de t sur un bord vertical et horizontal pour différentes valeurs de m et ζ . Les valeurs de θ sont centrées autour de 0 pour le bord vertical et $\frac{\pi}{2}$ pour le bord horizontal. Dans les deux cas, les valeurs sont plus resserrées quand m est important et/ou quand ζ est faible.

C.3 Estimateur MV du logarithme du contraste d'une rupture de réflectivité

Une intensité I sur une zone homogène de réflectivité R bruitée avec du speckle monovue suit la distribution :

$$p_I(I|R) = \frac{e^{-\frac{I}{R}}}{R}$$

Soit \mathbf{I}_1 et \mathbf{I}_2 des vecteurs constitués respectivement des N_1 intensités $I_1(k)$, $k = 1, 2, \dots, N_1$, et des N_2 intensités $I_2(l)$, $l = 1, 2, \dots, N_2$, observées sur des zones de réflectivités respectives R_1 et R_2 . La distribution conjointe de ces intensités conditionnellement

aux réflectivités R_1 et R_2 est :

$$\begin{aligned} p_{\mathbf{I}_1, \mathbf{I}_2}(\mathbf{I}_1, \mathbf{I}_2 | R_1, R_2) &= p_{\mathbf{I}_1}(\mathbf{I}_1 | R_1) \cdot p_{\mathbf{I}_2}(\mathbf{I}_2 | R_2) \\ &= \prod_{k=1}^{N_1} p_{I_1}(I_1(k) | R_1) \cdot \prod_{l=1}^{N_2} p_{I_2}(I_2(l) | R_2) \\ &= \prod_{k=1}^{N_1} \frac{e^{-\frac{I_1(k)}{R_1}}}{R_1} \cdot \prod_{l=1}^{N_2} \frac{e^{-\frac{I_2(l)}{R_2}}}{R_2} \end{aligned}$$

On cherche à estimer le logarithme du rapport des intensités $C_l = \log \frac{R_1}{R_2}$ avec comme observations \mathbf{I}_1 et \mathbf{I}_2 . L'estimateur MV de C_l maximise la densité $p(C_l, R_2) = p_{\mathbf{I}_1, \mathbf{I}_2}(\mathbf{I}_1, \mathbf{I}_2 | R_2 e^{C_l}, R_2)$ [Fjo99] ou aussi $\log p(C_l, R_2)$:

$$\log p(C_l, R_2) = -N_1 C_l - (N_1 + N_2) \log R_2 - \frac{1}{R_2} (N_1 \bar{I}_1 e^{-C_l} + N_2 \bar{I}_2)$$

Avec comme notation :

$$\bar{I}_1 = \frac{1}{N_1} \sum_{k=1}^{N_1} I_1(k) \quad \text{et} \quad \bar{I}_2 = \frac{1}{N_2} \sum_{l=1}^{N_2} I_2(l).$$

L'estimateur C_l doit donc vérifier :

$$\frac{\partial \log p(C_l, R_2)}{\partial C_l} = 0 \quad \text{et} \quad \frac{\partial \log p(C_l, R_2)}{\partial R_2} = 0.$$

Soit :

$$\begin{aligned} \frac{\partial \log p(C_l, R_2)}{\partial C_l} &= -N_1 + \frac{N_1 \bar{I}_1 e^{-C_l}}{R_2} = 0 \\ \frac{\partial \log p(C_l, R_2)}{\partial R_2} &= -\frac{N_1 + N_2}{R_2} + \frac{1}{R_2^2} (N_1 \bar{I}_1 e^{-C_l} + N_2 \bar{I}_2) = 0. \end{aligned}$$

On obtient $\hat{R}_2 = \bar{I}_2$ et $\hat{C}_l = \log \frac{\bar{I}_1}{\bar{I}_2}$. L'estimateur MV de C_l est bien le logarithme du rapport des intensités moyennes.

C.4 Log-cumulants de la loi de Fisher

La loi de Fisher est obtenue avec la convolution entre une loi Gamma \mathcal{G} et une loi Gamma Inverse \mathcal{IG} [NTD10] :

$$\mathcal{F}[\mu, L, M](x) = \mathcal{G}[1, L] * \mathcal{IG}[\mu, M],$$

avec

$$\begin{aligned} \mathcal{G}[m_1, L_1](x) &= \frac{L_1^{L_1}}{m_1^{L_1} \Gamma(L_1)} x^{L_1-1} \exp\left(-\frac{L_1 x}{m_1}\right) \\ \mathcal{IG}[m_2, L_2](x) &= \frac{(m_2 L_2)^{L_2}}{\Gamma(L_2)} x^{-L_2-1} \exp\left(-\frac{L_2 m_2}{x}\right). \end{aligned}$$

Soit deux lois p_A et p_B et leurs log-cumulants d'ordre n respectifs $\tilde{\kappa}_{A,n}$ et $\tilde{\kappa}_{B,n}$. Les log-cumulants de la convolution $p_A * p_B$ sont $\tilde{\kappa}_{A,n} + \tilde{\kappa}_{B,n}$ [NTD10].

Les log-cumulants aux deux premiers ordres sont pour la loi Gamma $\mathcal{G}(m_1, L_1)$:

$$\begin{aligned}\tilde{\kappa}_{\mathcal{G},1} &= \log(m_1) + \Psi(L_1) - \log(L_1) \\ \tilde{\kappa}_{\mathcal{G},2} &= \Psi(1, L_1)\end{aligned}$$

et pour la loi Gamma Inverse :

$$\begin{aligned}\tilde{\kappa}_{\mathcal{IG},1} &= \log(m_2) - \Psi(L_2) + \log(L_2) \\ \tilde{\kappa}_{\mathcal{IG},2} &= \Psi(1, L_2),\end{aligned}$$

avec Ψ la fonction Digamma et $\Psi(r, L)$ la fonction Polygamma d'ordre r .

On obtient alors comme log-cumulants pour la loi de Fisher :

$$\tilde{\kappa}_{\mathcal{F},n} = \tilde{\kappa}_{\mathcal{G},n}(1, L) + \tilde{\kappa}_{\mathcal{IG},n}(\mu, M).$$

Soit pour les deux premiers ordres :

$$\begin{aligned}\tilde{\kappa}_{\mathcal{F},1} &= \log(\mu) + \Psi(L_1) - \Psi(L_2) - \log(L_1) + \log(L_2) \\ \tilde{\kappa}_{\mathcal{F},2} &= \Psi(1, L_1) + \Psi(1, L_2).\end{aligned}$$

C.5 Invariance aux changements d'échelle de la matrice SAR-Harris

Soit deux images A et B telles que $A(a, b) = B(\frac{a}{m}, \frac{b}{m})$. On définit $A_{x,\alpha}$, $A_{y,\alpha}$, $B_{x,\alpha}$ et $B_{y,\alpha}$ les composantes du GR de respectivement A et B avec comme paramètre de pondération α . Afin de définir les notations, prenons l'exemple de la composante $A_{x,\alpha}$:

$$A_{x,\alpha} = \log \frac{\overline{A_{1x,\alpha}}}{\overline{A_{2x,\alpha}}}$$

avec :

$$\begin{aligned}\overline{A_{1x,\alpha}}(a, b) &= \int_{y=\mathbb{R}} \int_{x=\mathbb{R}^+} A(a+x, b+y) \times e^{-\frac{|x|+|y|}{\alpha}} dx dy \\ \overline{A_{2x,\alpha}}(a, b) &= \int_{y=\mathbb{R}} \int_{x=\mathbb{R}^-} A(a+x, b+y) \times e^{-\frac{|x|+|y|}{\alpha}} dx dy.\end{aligned}$$

On peut écrire :

$$\overline{A_{1x,\alpha}}(a, b) = \int_{y=\mathbb{R}} \int_{x=\mathbb{R}^+} B\left(\frac{a+x}{m}, \frac{b+y}{m}\right) \times e^{-\frac{|x|+|y|}{\alpha}} dx dy.$$

Après un changement de variable, on obtient :

$$\begin{aligned}\overline{A_{1x,\alpha}}(a, b) &= \int_{v=\mathbb{R}} \int_{u=\mathbb{R}^+} B\left(\frac{a}{m} + u, \frac{b}{m} + v\right) \times e^{-\frac{m(|u|+|v|)}{\alpha}} m^2 du dv \\ \overline{A_{1x,\alpha}}(a, b) &= m^2 \overline{B_{1x,\frac{\alpha}{m}}}\left(\frac{a}{m}, \frac{b}{m}\right).\end{aligned}$$

Soit pour la composante $A_{x,\alpha}$:

$$\begin{aligned}A_{x,\alpha}(a, b) &= \log \frac{m^2 \overline{B_{1x,\frac{\alpha}{m}}}\left(\frac{a}{m}, \frac{b}{m}\right)}{m^2 \overline{B_{2x,\frac{\alpha}{m}}}\left(\frac{a}{m}, \frac{b}{m}\right)} \\ A_{x,\alpha}(a, b) &= B_{x,\frac{\alpha}{m}}\left(\frac{a}{m}, \frac{b}{m}\right).\end{aligned}$$

On développe ensuite :

$$\begin{aligned} \mathcal{G}_s * A_{x,\alpha}^2(a,b) &= \iint_{(x,y) \in \mathbb{R}^2} \mathcal{G}_s(x,y) A_{x,\alpha}^2(a-x, b-y) dx dy \\ &= \iint_{(x,y) \in \mathbb{R}^2} \mathcal{G}_s(x,y) B_{x,\frac{\alpha}{m}}^2\left(\frac{a-x}{m}, \frac{b-y}{m}\right) dx dy \end{aligned}$$

puis on effectue un changement de variable :

$$\mathcal{G}_s * A_{x,\alpha}^2(a,b) = \iint_{(u,v) \in \mathbb{R}^2} \mathcal{G}_s(mu, mv) B_{x,\frac{\alpha}{m}}^2\left(\frac{a}{m} - u, \frac{b}{m} - v\right) m^2 du dv.$$

Or on peut réduire :

$$m^2 \mathcal{G}_s(mu, mv) = \frac{m^2}{2\pi s^2} e^{-\frac{m^2(u^2+v^2)}{2s^2}} = \mathcal{G}_{\frac{s}{m}}(u, v).$$

On obtient alors l'égalité :

$$\begin{aligned} \mathcal{G}_s * A_{x,\alpha}^2(a,b) &= \iint_{(u,v) \in \mathbb{R}^2} \mathcal{G}_{\frac{s}{m}}(u, v) B_{x,\frac{\alpha}{m}}^2\left(\frac{a}{m} - u, \frac{b}{m} - v\right) du dv \\ \mathcal{G}_s * A_{x,\alpha}^2(a,b) &= \mathcal{G}_{\frac{s}{m}} * B_{x,\frac{\alpha}{m}}^2\left(\frac{a}{m}, \frac{b}{m}\right). \end{aligned}$$

Afin d'être invariant aux changements d'échelle, le paramètre s doit donc être proportionnel à l'échelle α . On définit la matrice SAR-Harris de la façon suivante, avec $a_2 > 1$:

$$M_{SH}(a, b, \alpha) = \mathcal{G}_{a_2\alpha} \star \begin{bmatrix} (A_{x,\alpha})^2 & A_{x,\alpha} \cdot A_{y,\alpha} \\ A_{x,\alpha} \cdot A_{y,\alpha}^2 & (A_{y,\alpha})^2 \end{bmatrix}.$$

Publications

Journal

F. Dellinger, J. Delon, Y. Gousseau, J. Michel et F. Tupin, "SAR-SIFT : A SIFT-like algorithm for SAR images," *IEEE Transactions on Geoscience and Remote Sensing* (accepté pour publication).

Conférences

F. Dellinger, J. Delon, Y. Gousseau, J. Michel et F. Tupin, "SAR-SIFT : A SIFT-like algorithm for applications on SAR images," *IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, 2012 (Présentation orale).

F. Dellinger, J. Delon, Y. Gousseau, J. Michel et F. Tupin, "Change detection for High Resolution satellite images, based on SIFT algorithm and a contrario approach," *IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, 2014 (Présentation orale en 07/2014).

Bibliographie

- [ALAB13] F. ARGENTI, A. LAPINI, L. ALPARONE et T. BIANCHI : A tutorial on speckle reduction in synthetic aperture radar images. *IEEE Geosci. Remote Sens. Mag.*, pages 6–35, 2013.
- [BBM08] F. BOVOLO, L. BRUZZONE et M. MARCONCINI : A novel approach to unsupervised change detection based on a semisupervised SVM and a similarity measure. *Geoscience and Remote Sensing, IEEE Transactions on*, 46(7):2070–2082, July 2008.
- [BMP02] S. BELONGIE, J. MALIK et J. PUZICHA : Shape matching and object recognition using shape contexts. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 24(4):509–522, 2002.
- [Bov88] A. C. BOVIK : On detecting edges in speckle imagery. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 36(10):1618–1627, October 1988.
- [BTG06] H. BAY, T. TUYTELAARS et L. Van GOOL : SURF : Speeded up robust features. *In European Conference on Computer Vision*, pages 404–417, 2006.
- [CCS14] Tianze CHEN, Limin CHEN et Yi SU : A SAR image registration method based on pixel migration of edge-point feature. *Geoscience and Remote Sensing Letters, IEEE*, 11(5):906–910, May 2014.
- [CE96] P. R. COPPIN et Bauer M. E. : Digital change detection in forest ecosystems with remote sensing imagery. *Remote Sensing Reviews*, 1(3):207–234, 1996.
- [DDT⁺13] C.-A. DELEDALLE, L. DENIS, F. TUPIN, A. REIGBERS et M. JÄGER : NL-SAR : a unified non-local framework for resolution-preserving PolInSAR denoising. Preprint HAL – hal-00844118, 2013.
- [Dee98] P DEER : *Digital change detection in remotely sensed imagery using fuzzy set theory*. Thèse de doctorat, Department of Geography and Department of Computer Science, University of Adelaide, Australia, 1998.
- [DGL04] F. DELL’ACQUA, P. GAMBA et G. LISINI : Coregistration of multiangle fine spatial resolution SAR images. *Geoscience and Remote Sensing Letters, IEEE*, 1(4):237–241, Oct 2004.
- [DMM03] A. DESOLNEUX, L. MOISAN et J.-M. MOREL : A grouping principle and four applications. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 25(4):508–513, 2003.
- [DMM07] A. DESOLNEUX, L. MOISAN et J.-M. MOREL : *From Gestalt Theory to Image Analysis : A Probabilistic Approach*. Springer Publishing Company, Incorporated, 1st édition, 2007.
- [DSH00] Y. DUFOURNAUD, C. SCHMID et R. P. HORAUD : Matching images with different resolutions. *In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 612–618, 2000.

- [DT05] N. DALAL et B. TRIGGS : Histograms of oriented gradients for human detection. In *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, volume 1, pages 886–893 vol. 1, 2005.
- [DWDQ08] J. S. DENG, K. WANG, Y. H. DENG et G. J. QI : PCA-based land-use change detection and analysis using multitemporal and multisensor satellite data. *International Journal of Remote Sensing*, 29(16):4823–4838, 2008.
- [FB81] M. A. FISCHLER et R. C. BOLLES : Random sample consensus : a paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, 24(6):381–395, juin 1981.
- [FHPK13] B. FAN, C. HUO, C. PAN et Q. KONG : Registration of optical and SAR satellite images by exploring the spatial relationship of the improved SIFT. *IEEE Geoscience and Remote Sensing Letters*, 10(4):657–661, July 2013.
- [Fjo99] R. FJORTOFT : *Segmentation d’images radar par détection de contours*. Thèse de doctorat, Institut National Polytechnique de Toulouse, 1999.
- [FLMCC98] R. FJORTOFT, A. LOPES, P. MARTON et E. CUBERO-CASTAN : An optimal multiedge detector for SAR image segmentation. *IEEE Transactions on Geoscience and Remote Sensing*, 36(3):793–802, May 1998.
- [Gao08] Jay GAO : *Digital Analysis of Remotely Sensed Imagery*. McGraw-Hill Professional, 1st édition, 2008.
- [GDL06] P. GAMBA, F. DELL’ACQUA et G. LISINI : Change detection of multitemporal SAR data in urban areas combining feature-based and pixel-based techniques. *Geoscience and Remote Sensing, IEEE Transactions on*, 44(10):2820–2827, Oct 2006.
- [HCC⁺13] Masroor HUSSAIN, Dongmei CHEN, Angela CHENG, Hui WEI et David STANLEY : Change detection from remotely sensed images : From pixel-based to object-based approaches. *{ISPRS} Journal of Photogrammetry and Remote Sensing*, 80(0):91 – 106, 2013.
- [HPHZ12] C. HUO, C. PAN, L. HUO et Z. ZHOU : Multilevel SIFT matching for large-size VHR image registration. *IEEE Geoscience and Remote Sensing Letters*, 9(2):171–175, March 2012.
- [HS88] C. HARRIS et M. STEPHENS : A combined corner and edge detector. In *Proceedings of the 4th Alvey Vision Conference*, 1988.
- [HW81] P. J. HOWARTH et G. M. WICKWARE : Procedures for change detection using landsat digital data. *International Journal of Remote Sensing*, 2(3):277–291, 1981.
- [JGKM14] J. JOGLEKAR, S.S. GEDAM et B. KRISHNA MOHAN : Image matching using SIFT features and relaxation labeling technique - a constraint initializing method for dense stereo matching. *Geoscience and Remote Sensing, IEEE Transactions on*, PP(99):1–10, 2014.
- [JHGQ08] Gong JIANYA, Sui HAIGANG, Ma GUORUI et Zhou QIMING : A review of multi-temporal remote sensing data change detection algorithms. In *ISPRS Congress Beijing 2008*, volume 27, pages 757–762, 2008.
- [JMTU06] Wei Ji, Jia MA, Rima Wahab TWIBELL et Karen UNDERHILL : Characterizing urban sprawl using multi-stage remote sensing images and landscape metrics. *Computers, Environment and Urban Systems*, 30(6):861–879, 2006.
- [KA06] Y. KELLER et A. AVERBUCH : Multisensor image registration via implicit similarity. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 28(5):794–801, May 2006.
-

-
- [KB01] T. KADIR et M. BRADY : Saliency, scale and image description. *Int. J. Comput. Vision*, 45(2):83–105, novembre 2001.
- [KCS⁺13] M.N. KLARIC, B.C. CLAYWELL, G.J. SCOTT, N.J. HUDSON, O. SJAHPUTERA, Y. LI, S.T. BARRATT, J.M. KELLER et C.H. DAVIS : GeoCDX : An automated change detection and exploitation system for high-resolution satellite imagery. *Geoscience and Remote Sensing, IEEE Transactions on*, 51(4):2067–2086, April 2013.
- [KS04] Y. KE et R. SUKTHANKAR : PCA-SIFT : a more distinctive representation for local image descriptors. In *Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, volume 2, pages II–506–II–513 Vol.2, 2004.
- [KSS⁺12] P. KEMPENEERS, F. SEDANO, P. STROBL, D.O. MCINERNEY et J. SANMIGUEL-AYANZ : Increasing robustness of postclassification change detection using time series of land cover maps. *Geoscience and Remote Sensing, IEEE Transactions on*, 50(9):3327–3339, Sept 2012.
- [Lin98] T. LINDBERG : Feature detection with automatic scale selection. *International Journal of Computer Vision*, 30:79–116, 1998.
- [LL04] W. LI et H. LEUNG : A maximum likelihood approach for image registration using control point and intensity. *Image Processing, IEEE Transactions on*, 13(8):1115–1127, Aug 2004.
- [Low99] D. G. LOWE : Object recognition from local scale-invariant features. In *Proceedings of the International Conference on Computer Vision*, 1999.
- [Low04] D. G. LOWE : Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60:91–110, 2004.
- [LWGZ09] J. LU, B. WANG, H. M. GAO et Z.Q. ZHOU : SAR images matching based on local shape descriptors. In *IET International Radar Conference*, pages 1–4, 2009.
- [LWLC09] Q. LI, G. WANG, J. LIU et S. CHEN : Robust scale-invariant feature matching for remote sensing image registration. *IEEE Geoscience and Remote Sensing Letters*, 6(2):287–291, April 2009.
- [LWW08] L. LIU, Y. WANG et Y. WANG : SIFT based automatic tie-point extraction for multitemporal SAR images. In *International Workshop on Education Technology and Training and on Geoscience and Remote Sensing*, volume 1, pages 499–503, 2008.
- [LXWH06] W. LI, X.LI, Y. WU et Z. HU : A novel framework for urban change detection using VHR satellite images. In *Pattern Recognition, 2006. ICPR 2006. 18th International Conference on*, volume 2, pages 312–315, 2006.
- [LY08] J. Z. LIU et X.C. YU : Research on SAR image matching technology based on SIFT. In *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 2008.
- [LYWW11] W. LV, W. YU, J. WANG et K. WANG : SAR image matching based on SIFT keypoints and multi-subregions information. In *International Asia-Pacific Conference on Synthetic Aperture Radar*, pages 1–4, 2011.
- [LZbPQ10] Zhen LI, Quo ZHANG, Hong bo PAN et Qiang QIANG : RPC-based adjustment model for COSMO-SkyMed stereo slant/ground-range images. In *Progress in Informatics and Computing (PIC), 2010 IEEE International Conference on*, volume 1, pages 593–597, Dec 2010.
-

- [MFFL10] M. MAKITALO, A. FOI, D. FEVRALEV et V. LUKIN : Denoising of single-look SAR images based on variance stabilization and nonlocal filters. *In Mathematical Methods in Electromagnetic Theory (MMET), 2010 International Conference on*, pages 1–4. IEEE, 2010.
- [MS01] K. MIKOLAJCZYK et C. SCHMID : Indexing based on scale invariant interest points. *In International Conference on Computer Vision*, volume 1, pages 525–531, 2001.
- [MS02] K. MIKOLAJCZYK et C. SCHMID : An affine invariant interest point detector. *In Proceedings of the 7th European Conference on Computer Vision*, pages 128–142, 2002.
- [MS04a] K. MIKOLAJCZYK et C. SCHMID : Scale and affine invariant interest point detectors. *International Journal of Computer Vision*, 60(1):63–86, 2004.
- [MS04b] Lionel MOISAN et Bérenger 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.
- [MS05] K. MIKOLAJCZYK et C. SCHMID : A performance evaluation of local descriptors. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(10):1615–1630, Oct. 2005.
- [MTS⁺05] K. MIKOLAJCZYK, T. TUYTELAARS, C. SCHMID, A. ZISSERMAN, J. MATAS, F. SCHAFFALITZKY, T. KADIR et L. Van GOOL : A comparison of affine region detectors. *Int. J. Comput. Vision*, 65(1-2):43–72, novembre 2005.
- [MY09] J.-M. MOREL et G. YU : ASIFT : A new framework for fully affine invariant image comparison. *SIAM Journal on Imaging Sciences*, 2(2):438–469, avril 2009.
- [Nic02] J.-M. NICOLAS : Introduction aux statistiques de deuxième espèce : applications des logs-moments et des logs-cumulants à l’analyse des lois d’images radar. *TS. Traitement du signal*, 19(3):139–167, 2002.
- [NTD10] J.-M. NICOLAS, F. TUPIN et M.-P. DOIN : Traitement des images SAR. Rapport technique, Télécom Paristech, 2010.
- [NTF⁺12] J. M NICOLAS, E. TROUVE, R. FALLOURD, F. VERNIER, F. TUPIN, O. HARRANT, M. GAY et L. MOREAU : A first comparison of Cosmo-SkyMed and TerraSAR-X data over Chamonix Mont-Blanc test-site. *In Geoscience and Remote Sensing Symposium (IGARSS), 2012 IEEE International*, pages 5586–5589, July 2012.
- [NY06] H. NEMMOUR et Chibani Y. : Multiple support vector machines for land cover change detection : An application for mapping urban extensions. *ISPRS Journal of Photogrammetry and Remote Sensing*, 61(2):125–133, 2006.
- [OBW96] C. J. OLIVER, D. BLACKNELL et R. G. WHITE : Optimum edge detection in SAR. *IEE Proceedings on Radar, Sonar and Navigation*, 143(1):31–40, February 1996.
- [PPAV12] S. PARRILLI, M. PODERICO, C.V. ANGELINO et L. VERDOLIVA : A nonlocal SAR image denoising algorithm based on LLMMSE wavelet shrinkage. *Geoscience and Remote Sensing, IEEE Transactions on*, 50(2):606–616, Feb 2012.
- [PPSA05] Bryan C. PIJANOWSKI, Snehal PITHADIA, Bradley A. SHELLITO et Konstantinos ALEXANDRIDIS : Calibrating a neural network-based urban change model for two metropolitan areas of the upper midwest of the united states. *International Journal of Geographical Information Science*, 19(2):197–215, 2005.
-

-
- [QPPNL14] G. QUIN, B. PINEL-PUYSSÉGUR, J.-M. NICOLAS et P. LOREAU : MIMOSA : An automatic change detection method for sar time series. *Geoscience and Remote Sensing, IEEE Transactions on*, 52(9):5349–5363, Sept 2014.
- [Rab09] J. RABIN : *Approches robustes pour la comparaison d'images et la reconnaissance d'objets*. Thèse de doctorat, Telecom ParisTech, 2009.
- [RB13] V. RISOJEVIC et Z. BABIC : Fusion of global and local descriptors for remote sensing image classification. *IEEE Geoscience and Remote Sensing Letters*, 10(4):836–840, 2013.
- [RD05] E. ROSTEN et T. DRUMMOND : Fusing points and lines for high performance tracking. In *Computer Vision, 2005. ICCV 2005. Tenth IEEE International Conference on*, volume 2, pages 1508–1515 Vol. 2, 2005.
- [RDG09] J. RABIN, J. DELON et Y. GOUSSEAU : A statistical approach to the matching of local features. *SIAM Journal on Imaging Sciences*, 2:931–958, September 2009.
- [RDGM10] J. RABIN, J. DELON, Y. GOUSSEAU et L. MOISAN : MAC-RANSAC : a robust algorithm for the recognition of multiple objects. In *Proceedings of 3D Data Processing, Visualization and Transmission*, 2010.
- [SB97] S. M. SMITH et J. M. BRADY : SUSAN : A new approach to low level image processing. *Proc. of International Journal of Computer Vision*, 23(1):45–78, mai 1997.
- [SBM⁺06] E. SANSOSTI, P. BERARDINO, M. MANUNTA, F. SERAFINO et G. FORNARO : Geometrical SAR image registration. *Geoscience and Remote Sensing, IEEE Transactions on*, 44(10):2861–2870, Oct 2006.
- [SMB00] C. SCHMID, R. MOHR et C. BAUCKHAGE : Evaluation of interest point detectors. *International Journal of Computer Vision*, 37(2):151–172, juin 2000.
- [SME11] A. SEDAGHAT, M. MOKHTARZADE et H. EBADI : Uniform robust scale-invariant feature matching for optical remote sensing images. *IEEE Transactions on Geoscience and Remote Sensing*, 49(11):4516–4527, November 2011.
- [Spo10] H. SPORTOUCHE : *Extraction et reconstruction de bâtiments en milieu urbain à partir d'images satellitaires optiques et radar à haute résolution*. Thèse de doctorat, Télécom ParisTech, 2010.
- [SR10] S. SURI et P. REINARTZ : Mutual-information-based registration of TerraSAR-X and IKONOS imagery in urban areas. *Geoscience and Remote Sensing, IEEE Transactions on*, 48(2):939–949, Feb 2010.
- [SSRS10] P. SCHWIND, S. SURI, P. REINARTZ et A. SIEBERT : Applicability of the SIFT operator to geometric SAR image registration. *International Journal of Remote Science*, 31(8):1959–1980, mars 2010.
- [SSRU09] S. SURI, P. SCHWIND, P. REINARTZ et J. UHL : Combining mutual information and scale invariant feature transform for fast and robust multisensor SAR image registration. In *75th Annual American Society for Photogrammetry and Remote Sensing Conference*, 2009.
- [SSUR10] S. SURI, P. SCHWIND, J. UHL et P. REINARTZ : Modifications in the SIFT operator for effective SAR image matching. *International Journal of Image and Data Fusion*, 1(3):243–256, 2010.
- [SU09] B. SIRMACEK et C. UNSALAN : Urban-area and building detection using SIFT keypoints and graph theory. *IEEE Transactions on Geoscience and Remote Sensing*, 47(4):1156–1167, 2009.
-

- [TEK11] D. TOMOWSKI, M. EHLERS et S. KLONUS : Colour and texture based change detection for urban disaster analysis. *In Urban Remote Sensing Event (JURSE), 2011 Joint*, pages 329–332, April 2011.
- [TLB88] R. TOUZI, A. LOPES et P. BOUSQUET : A statistical and geometrical edge detector for SAR images. *IEEE Transactions on Geoscience and Remote Sensing*, 26(6):764–773, November 1988.
- [TNI14] F. TUPIN, J.-M. NICOLAS et J. INGLADA : *Imagerie de télédétection*. Editions Hermes, Lavoisier, 2014.
- [TTCT11] C. TAO, Y. TAN, H. CAI et J. TIAN : Airport detection from large IKONOS images using clustered SIFT keypoints and region information. *IEEE Geoscience and Remote Sensing Letters*, 8(1):128–132, 2011.
- [WC07] A. WONG et D.A. CLAUSI : ARRSI : Automatic registration of remote-sensing images. *Geoscience and Remote Sensing, IEEE Transactions on*, 45(5):1483–1493, May 2007.
- [WHR07] B. WESSEL, M. HUBER et A. ROTH : Registration of near real-time SAR images by image-to-image matching. *In Photogrammetric Image Analysis*, page 179, 2007.
- [WY11] Y. WU et M. YANG : A multi-sensor remote sensing image matching method based on SIFT operator and CRA similarity measure. *In International Conference on Intelligence Science and Information Engineering*, pages 115–118, 2011.
- [WYF12] S. WANG, H. YOU et K. FU : BFSIFT : A novel method to find feature matches for SAR image registration. *IEEE Geoscience and Remote Sensing Letters*, 9(4):649–653, July 2012.
- [XTPG11] Y. XIAOPING, L. TONG, L. PINGXIANG et H. GUOMAN : The application of improved SIFT algorithm in high resolution SAR image matching in mountain areas. *In International Symposium on Image and Data Fusion*, pages 1–4, 2011.
- [YM09] G. YU et J.-M MOREL : A fully affine invariant image comparison method. *In Acoustics, Speech and Signal Processing, 2009. ICASSP 2009. IEEE International Conference on*, pages 1597–1600, 2009.
- [YN13] Y. YANG et C. NEWSAM : Geographic image retrieval using local invariant features. *IEEE Transactions on Geoscience and Remote Sensing*, 51(2):818–832, 2013.
- [ZF03] B. ZITOVA et J. FLUSSER : Image registration methods : a survey. *Image and Vision Computing*, 21:977–1000, 2003.
- [ZFL⁺10] Guo ZHANG, WB FEI, Zhen LI, Xiaoyong ZHU et DR LI : Evaluation of the RPC model for spaceborne SAR imagery. *Photogrammetric Engineering & Remote Sensing*, 76(6):727–733, 2010.
-

Descripteurs locaux pour l'imagerie radar et applications

Flora Dellinger

Résumé : Dans cette thèse, nous étudions l'intérêt des descripteurs locaux pour les images satellites optiques et radar. Ces dernières années, la multiplication des satellites de télédétection haute résolution conduit à une forte augmentation des situations multi-capteurs, multi-résolutions et multi-angles d'incidence. Les descripteurs locaux, du fait de leurs invariances et de leur représentation compacte, présentent un intérêt certain pour la comparaison d'images satellitaires acquises dans des conditions différentes. Bien que facilement applicables aux images optiques, ils offrent des performances très limitées sur les images radar. Effectivement ces images sont très fortement bruitées par un bruit multiplicatif qui perturbe le traitement et l'interprétation des données.

Nous proposons ainsi un descripteur local original pour la comparaison d'images radar. Cet algorithme, appelé SAR-SIFT, repose sur la même structure que l'algorithme SIFT, à savoir une détection de points d'intérêts associée à l'extraction de descripteurs locaux. Afin d'adapter ces étapes aux statistiques des images radar, nous avons au préalable développé un nouvel opérateur différentiel, appelé Gradient par Ratio. Cet opérateur permet de calculer, de façon multi-échelles, une norme et une orientation du gradient robustes au bruit multiplicatif des images radar. Nous nous sommes ensuite appuyés sur ce gradient pour modifier les étapes de détection de points-clés et d'extraction de descripteurs. Les performances de l'algorithme SAR-SIFT proposé sont supérieures à celle de l'algorithme SIFT original pour des données radar. Nous avons ensuite testé le comportement de l'algorithme SAR-SIFT dans différentes situations d'acquisition.

Nous présentons par la suite deux applications pour la télédétection, basées sur les descripteurs locaux. En premier lieu, nous estimons une transformation globale entre deux images radar à l'aide de l'algorithme SAR-SIFT. L'estimation est réalisée à l'aide d'un algorithme de type RANSAC et en utilisant comme points homologues les points-clés mis en correspondance. Nous proposons d'utiliser cette méthode pour estimer deux types de transformation : une transformation affine entre deux images pour les zones plates, ou une translation résiduelle entre deux images géoréférencées. Enfin nous avons mené une étude prospective sur l'utilisation des descripteurs locaux (SIFT en imagerie optique et SAR-SIFT en imagerie radar) pour la détection de changements en télédétection. La méthode proposée consiste à comparer à l'aide d'une approche a contrario les densités de points-clés mis en correspondance aux densités de points-clés détectés pour mettre en évidence les zones de changement. L'intérêt de cette méthode est illustré sur des cas réels (paires d'images optiques et radar).

Mots-clés : imagerie radar, images satellites, SIFT, détection de changements, recalage d'images radar

