

Fusion de données optique et radar à haute résolution en milieu urbain

Gabrielle Lehureau

- Florence Tupin Directrice de thèse
- Marine Campedel Co-Directrice de thèse
- Celine Tison Encadrant CNES
- Guillaume Oller Encadrant Magellium

Table des matières

1	Introduction					
2	Prés	Présentation des capteurs et de leurs données 1.				
	2.1	Les satellites	3			
		2.1.1 Systèmes radar	3			
		2.1.2 Systèmes optiques	4			
	2.2	Introduction aux images radar	5			
	2.3	Introduction aux images optiques	9			
	2.4	Géométrie des images	1			
	2.5	Présentation de la base de données	1			
3	Reca	lage 2	9			
	3.1	Introduction	9			
		3.1.1 Pourquoi faire du recalage?	0			
		3.1.2 Les types de déformations	0			
	3.2	État de l'art sur le recalage 3	4			
		3.2.1 Généralités sur le recalage	4			
		3.2.2 État de l'art sur le recalage en Optique/Radar	6			
	3.3	3.3 Recalage rigide par l'invariant de Fourier-Mellin				
		3.3.1 Initialisation et choix des primitives	0			
		3.3.2 Rotation et changement d'échelle	1			
		3.3.3 Translation	6			
		3.3.4 Analyse des résultats	2			
	3.4	Recalage polynômial	3			
		3.4.1 Extraction de points d'amer	4			
		3.4.2 Appariement des points	6			
		3.4.3 Choix des paramètres	7			

TABLE DES MATIÈRES

		3.4.4	Transformation polynomiale		
		3.4.5	Estimation du polynôme		
	3.5	3.5 Résultats			
		3.5.1	Image du CNES		
		3.5.2	Image de la prison		
		3.5.3	Image TerraSAR du complexe scientifique de Rangueil 74		
	3.6	Conclu	lsion		
4	Clas	sificatio	on 81		
	4.1	Problé	matique		
	4.2	Contex	te		
	4.3	Chaîne	de traitement		
		4.3.1	Partition de l'image optique		
		4.3.2	Extraction des primitives utiles à la classification		
		4.3.3	Paramètres extraits de l'image radar		
		4.3.4	Classificateurs		
		4.3.5	Evaluation		
	4.4	Expéri	mentations		
		4.4.1	Pertinence des primitives		
		4.4.2	Nombre de primitives		
		4.4.3	Choix du classificateur		
		4.4.4	Robustesse à la segmentation		
		4.4.5	Complémentarité optique/radar		
		4.4.6	Robustesse au changement de scène		
	4.5	Conclu	lsion		
5	Opti	misatio	n jointe de la classification et du recalage 119		
	5.1	Introdu	uction		
	5.2	Classification et recalage joints			
		5.2.1	Principe		
		5.2.2	Mise en oeuvre		
		5.2.3	Zone d'apprentissage		
		5.2.4	SVM probabiliste		
		5.2.5	Classification : bâtiments/non-bâtiments		
		5.2.6	Classification itérative		

TABLE DES MATIÈRES

		5.2.7	Conclusion	130
	5.3	Classif	fication et recalage successifs : par le contexte	131
		5.3.1	Principe	131
		5.3.2	Détection du coin réflecteur	132
		5.3.3	Détection des ombres	132
		5.3.4	Fusion	133
		5.3.5	Résultats	135
		5.3.6	Conclusion sur le recalage par le contexte	136
	5.4	Conclu	usion	139
6	Con	clusion	et Perspectives	141

TABLE DES MATIÈRES

Chapitre 1

Introduction

Problématique

L'une des problématiques actuelles de l'interprétation des images issues des archives satellitaires concerne la multiplication des satellites et des capteurs. La nouvelle génération de capteurs satellitaires radar et optique qui ont des résolutions de l'ordre du mètre soulève de nouveaux défis. Cette amélioration de la résolution nécessite la définition d'outils qui leur soient adaptés. Les différences de contenu pour des résolutions de 10m, 1m et 30cm sont drastiques non seulement pour les données optiques mais aussi pour les données radar.

De nombreux satellites ont été lancés récemment, et en particulier la famille des Pléiades (2009) à résolution sub-métrique et Cosmo-Skymed (2007), constituant le système Franco-Italien d'observation de la terre. La spécificité des capteurs optiques et radar fait que les images sont porteuses d'informations potentiellement complémentaires, d'où la volonté de mettre à disposition un système dual optique/radar. Dans le but de promouvoir et accompagner l'utilisation de ces données, le CNES a mis en place le programme d'accompagnement ORFEO¹. Les travaux présentés ici ont pour but d'aider à interpréter les données optiques et radar à haute résolution en milieu urbain.

La haute résolution permet désormais d'identifier des objets auparavant nonreconnaissables sur des images décamétriques, comme des bâtiments, des arbres ou des ponts, très présents en milieu urbain. "Le monde dans lequel nous vivons est devenu urbain. La population urbaine mondiale est passée de près de 47 millions vers 1700 à 75 millions vers 1800 et à 335 millions en 1910. En 1950, la population urbaine mon-

http://smsc.cnes.fr/PLEIADES/Fr/A_prog_accomp.htm

1. Introduction

diale regroupait environ 724 millions d'individus, en 1980, 1.806 milliards et 3.15 en 2005." extrait de l'Encyclopedie Universalis². L'étude de la forte croissance du milieu urbain, ces dernières années, a encouragé le développement d'outils de télédétection. Les méthodes d'identification utilisées pour des images basse résolution ne donnent pas de résultats satisfaisants appliquées aux images submétriques [Chini *et al.*, 2009]. En effet, ces méthodes ne tiennent pas compte de certains changements, présents dans les images haute résolution. Par exemple, sur une image optique, les ombres des bâtiments peuvent varier en fonction des saisons. Il faut également tenir compte de l'angle d'observation du capteur qui change la parallaxe, et donc la forme d'un objet, en fonction de sa hauteur.

Les systèmes d'imagerie optiques, passifs, mesurent directement le rayonnement réfléchi par la surface terrestre dans le domaine du visible ou de l'infra rouge. Bien que l'utilisation de tels imageurs soit très intéressante car les données fournies sont très proches de la vision qu'aurait un utilisateur, ils ont cependant quelques inconvénients. Le premier concerne leur dépendance vis à vis de la source lumineuse, le soleil, puisqu'il est alors impossible de faire des acquisitions de nuit. De plus, les longueurs d'ondes utilisées (400 à 800 nm) sont très sensibles à la teneur en eau de l'atmosphère, il est donc impossible d'obtenir une image du sol en cas de couverture nuageuse ou de brume.

L'utilisation de systèmes radar (actifs) permet de combler les lacunes de l'imagerie optique. De tels capteurs utilisent le domaine des micro-ondes (1 à 10 Ghz), qui les rend peu sensibles à l'humidité de l'atmosphère. La qualité des images ne dépend donc pas des conditions météorologiques. Ces systèmes sont actifs, c'est à dire qu'il disposent de leur propre source d'illumination et sont donc indépendants de l'illumination solaire, ils peuvent ainsi acquérir des images de jour comme de nuit.

Ces différents modes d'observations sont complémentaires et riches en information. Ils permettent d'appréhender la planète dans son ensemble. Par exemple dans le cas de gestions de catastrophes, comme un incendie ou une irruption volcanique, une évaluation des dégâts pourrait être obtenue très rapidement grâce à l'imagerie radar. Désormais les images satellites couvrent toute la planète et un satellite met environ une heure et demie pour en faire le tour (pour un capteur à 500 km). Les images radar peuvent nous donner des informations malgré les nuages de fumées ou de cendres et les données optiques peuvent nous permettre une interprétation facile. Dans ces cas particuliers, il est impératif de pouvoir traiter les données rapidement et donc de développer des méthodes automatisées d'interprétation.

²http://www.universalis.fr/encyclopedie/ville-le-fait-urbain-dans-le-monde/

Malheureusement cette diversité dans les données est également un facteur de confusion et d'erreur car les différences entre les images, qui font à la fois leur complémentarité rendent aussi leur exploitation plus difficile. L'interprétation des images fournies par ces capteurs peut être effectuée manuellement, mais devant l'augmentation croissante de la quantité de données à disposition, l'automatisation de l'interprétation est d'un intérêt majeur. Il est donc nécessaire de développer de nouveaux outils automatiques de traitements, adaptés à l'interprétation de ce type d'images. C'est dans cette optique que nous nous sommes intéressés à la fusion de données optiques et radar à des résolutions métriques en milieu urbain.

Notre approche

La fusion permet d'utiliser les différentes caractéristiques des données en jeu pour améliorer la précision de l'information extraite. Dans ces travaux, nous souhaitons interpréter une scène du milieu urbain et identifier certains objets, comme les bâtiments.

Bien sûr, pour pouvoir fusionner automatiquement des images, il est nécessaire qu'elles se correspondent. Aussi la première étape de la fusion est une étape de recalage qui permet de mettre en correspondance les pixels de chaque image. Ceci peut se faire à l'aide des paramètres capteurs donnés dans les méta-données fournies généralement avec les images. Seulement, l'incertitude sur certains paramètres peut entraîner de grandes erreurs dans le recalage [Tupin et Galland, 2004]. De plus, parfois les fournisseurs ne souhaitent pas délivrer ces paramètres qu'ils considèrent confidentiels. Nous avons donc choisi comme contrainte de développer une méthode de recalage qui se base sur le contenu de la scène et non sur les paramètres capteurs. Etant donné que nous traitons des données de milieu urbain, les éléments choisis pour guider le recalage sont les contours de bâtiments et de routes. Nous avons utilisé une approche hiérarchique pour déterminer la transformation rigide entre les images, puis nous avons affiné le résultat à l'aide d'une transformation polynomiale. Nous montrons expérimentalement dans ce travail, sur différentes données, la qualité des résultats obtenus par notre chaîne, ainsi que ses limites. En particulier, à cause de la géométrie des images radar, les objets de la scène en hauteur sont mal recalés. Nous envisageons donc de tenir compte d'une information supplémentaire sur la hauteur des éléments de la scène pour guider le recalage, et donc de classer les éléments de l'image suivant leur hauteur.

Etant donné que nous souhaitons ensuite développer une méthode d'interprétation

1. Introduction

et/ou d'identification des images, nous proposons une méthode de classification qui utilise les informations issues des deux images. Nous avons choisi de classer la scène par régions, par opposition aux classifications pixelliques, car nous souhaitons identifier des objets, comme les bâtiments. Nous proposons dans ces travaux, une chaîne de classification avec apprentissage, robuste à la segmentation, qui utilise les séparateurs à vaste marge (SVM). Nous proposons plusieurs descripteurs extraits des données optiques et radar et montrons expérimentalement l'intérêt d'utiliser simultanément les deux types de données pour la classification. Les classes choisies sont au nombre de cinq, fréquemment présentes dans le milieu urbain : bâtiments, routes, végétation (arbres, forêts), sol et ombres. Nous mettons en avant l'interêt de fusionner les deux images pour améliorer la classification, notamment pour les classes bâtiments et routes du milieu urbain, bien discriminés sur une image radar. Nous voyons aussi le besoin d'avoir un bon recalage, afin que les éléments de chaque image se correspondent, pour améliorer la classification.

Finalement, nous en concluons que les opérations de recalage et de classification sont dépendantes l'une de l'autre. Nous nous intéressons donc à des méthodes permettant d'optimiser l'une et l'autre simultannément ou successivement. Dans un premier temps, nous proposons une méthode originale qui permet de classer et recaler simultanément un objet. Cette méthode se base sur l'hypothèse qu'un classificateur peut reconnaître un objet bien recalé si nous le lui avons appris. Nous classons plusieurs fois un objet, avec des caractéristiques correspondant à différentes translations. Dans un deuxième temps, nous proposons de recaler nos objets conditionnellement à leur classe. Pour cela nous utilisons la connaissance a priori que nous avons des images, notamment l'information contextuelle. En conclusion, nous voyons qu'il est possible de classer et recaler simultanément, toutefois nous devons envisager l'intervention d'un utilisateur ainsi que le besoin d'ajouter des informations contextuelles.

Plan

Ce document est constitué de quatre chapitres principaux. Le premier chapitre suivant cette introduction est consacré aux différents capteurs optiques et radar existants et aux notions associées aux techniques du radar imageur. Nous y présenterons les différences entre les types d'images, les principes de l'imagerie SAR seront ensuite étudiés afin d'établir les notations. Nous verrons également les bases de données utilisées au cours de ces travaux pour valider les algorithmes.

Le deuxième chapitre portera sur le recalage ; une étape nécessaire à l'interprétation jointe de nos images. Nous verrons dans ce chapitre, dans quel cadre le recalage est utilisé ainsi que la méthodologie afin de recaler des images. Nous proposerons ensuite une chaîne de traitement adaptée aux images optiques et radar à haute résolution en milieu urbain. Cette chaîne se décompose en deux parties : tout d'abord un recalage rigide par l'invariant de Fourier-Mellin, puis un recalage polynômial.

Le troisième chapitre portera sur la classification des images. Tout d'abord nous introduirons la notion d'"objets" : la scène représentée dans les images sera décomposée en objets, que nous classerons. Nous proposerons une méthodologie pour la classification, basée tout d'abord sur une segmentation de la scène et sur différents descripteurs à la fois optiques et radar ainsi que sur un classificateur SVM à noyau gaussien. Nous verrons également l'intérêt d'utiliser conjointement les deux types de données pour aider la classification.

Dans le dernier chapitre, nous présenterons deux approches permettant d'optimiser le recalage ainsi que la classification. La première permet de classer et recaler simultanément, tandis que la seconde introduit des a priori contextuels pour guider le recalage. Chacune de ces approches sera illustrée par des expériences sur des objets de la classe bâtiments. Nous conluerons sur la possibilité de déterminer le recalage exact au moyen d'un classificateur SVM et le besoin d'utiliser l'information contextuelle. 1. Introduction

Chapitre 2

Présentation des capteurs et de leurs données

2.1 Les satellites

2.1.1 Systèmes radar

Le premier système radar à synthèse d'ouverture (RSO) civil, nommé SEASAT, fut lancé en 1978 par la NASA. Son objectif était de collecter des données océanographiques (vent, température, topographie ...) et de démontrer la possibilité d'une surveillance satellitaire radar des eaux. Ce système n'a fonctionné que quelques jours. Les systèmes suivants furent SIR-A, SIR-B en 1982 et 1984 sur navette spatiale, puis les radars européens ERS (ERS1 en 1991 et ERS2 en 1995), le radar japonais JERS lancé en 1992, RADAR-SAT en 1995. Ces derniers fournissaient des images de basse résolution de l'ordre d'une dizaine de mètres, adaptées à l'étude des reliefs.

Avec l'avènement des capteurs haute résolution, de l'ordre de quelques mètres, voire très haute résolution, on peut s'intéresser à de nouvelles applications comme la reconnaissance de cibles ou de bâtiments. Citons par exemple SAR-Lupe lancé en 2006, TerraSar-X en 2007 ou Cosmo-Skymed constitué de 4 satellites dont les deux premiers ont été lancés en 2007.

Une autre catégorie de systèmes radar sont les systèmes aéroportés. *Une campagne d'acquisition est plus simple à mettre en oeuvre et ils permettent d'obtenir des images à très haute résolution.

capteur	opérateur	pays	résolution (en m)	dates
ERS1	ESA	Europe	20	1991-2000
ERS2	ESA	Europe	20	1995-2003
JERS	NASDA	Japon	18	1992-1998
RadarSat 1	CSA	Canada	9 à 100	1995-
Radarsat 2	CSA+MDA	Canada	3 à 100	2007-
Envisat	ESA	Europe	30 à 150	2002-
ALOS	JAXA	Japon	7 à 100	2006-
COSMO SkyMed	ASI	Italie	1	2007-
TerraSAR-X	DLR+EADS+InfoTerra	Allemagne	1	2007-

TAB. 2.1 – Systèmes satellites radars.

capteur	opérateur	pays	résolution (en m)
AES1	InterMap Technologies	USA	0.5
AIRSAR	NASA+JPL	USA	2.5
DOSAR	EADS	Allemagne	0.5
ESAR	DLR	Allemagne	1.2
EMISAR	DCRS	Danemark	2
MEMPHIS	FGAN	Allemagne	
PHARUS	TNO+FEL	Pays-Bas	3
PISAR	NASDA+CLR	Japon	1.5
RAMSES	ONERA	France	<1
SAR580	Environnement Canada	Canada	<3

TAB. 2.2 – Systèmes aéroportés radars.

2.1.2 Systèmes optiques

Le premier système permettant de photographier la Terre depuis l'espace fut embarqué sur le satellite Explorer-6 lancé en 1959 par la NASA. Les premières applications furent pour la météo, la surveillance de l'agriculture, des forêts, des océans, des ressources en eau...

Le premier satellite de télédétection civile a été Landsat-1 (1972). A cette époque, plusieurs satellites de reconnaissance militaires avaient déjà été lancés par les USA et l'URSS [Nicolas, 2008]. Depuis les satellites n'ont cessé de se développer et de s'améliorer pour offrir une vision globale de plus en plus détaillée de la surface terrestre. Ils permettent d'exploiter les domaines du visible et proche infra rouge (0.4 à 15 μ m).

Visible	0.4 à $0.8~\mu m$
Proche Infra rouge	0.8 à 1.5 μm
Moyen Infra Rouge	1.5 à 5 μ m
Infra rouge thermique	8 à 15 μm

TAB. 2.3 – Longueurs d'onde.

capteur	pays	résolution (en m)	dates
Proba HRC	Europe	8	2001-
Landsat 1 à 7	USA	15	1972-
AVNIR	Japon	10	1996-
SPOT 1 à 4	France	10	1986-1996 (pour 1 à 3)
SPOT 5	France	2.5	2002-
Resurs DK	Russie	0.5	2006-
CBERS 1, 2 et 2B	Bresil	20	1999-2003 (pour le 1)
IRS (8 satellites)	Inde	5.8	1988-
Ikonos	USA	1	1999-
Quickbird	USA	0.6	2001-
WorldView 1	USA	0.5	2007-
EROS A	Israel	1.8	2000-
EROS B	Israel	0.7	2006-

TAB. 2.4 – Systèmes satellites optiques.

2.2 Introduction aux images radar

Le radar est un système actif de télédétection, il délivre ses propres ondes électromagnétiques et capte ces ondes rétro-diffusées par les objets illuminés. Les longueurs d'ondes utilisées sont de l'ordre du centimètre ou du mètre, ce qui correspond à des fréquences allant de 300 Mhz à 100 Ghz. Ces portions du spectre sont avantageuses car il y a peu d'absorption due à la vapeur d'eau ou l'atmosphère. Une grande longueur d'onde permet de pénétrer certains sols et végétation, alors qu'une petite longueur d'onde est plus sensible à la rugosité.

Le système radar se compose d'une antenne fonctionnant en émission et/ou réception. Le modèle simplifié de l'imagerie radar peut être décrit par une compression en distance et en azimut : l'antenne émet une onde électromagnétique qui est réfléchie par des cibles. La mesure du temps entre l'émission (t_E) et la réception (t_R) de cette onde permet de



FIG. 2.1 – Bandes spectrales et transmission dans l'atmosphère

déterminer la distance D entre la cible et le capteur :

$$D = \frac{t_R - t_E}{2c}$$

où c représente la vitesse de la lumière.

La trajectoire du porteur définit l'axe azimut de l'image. L'axe distance correspond à la distance de la cible par rapport au capteur. L'acquisition se fait en visée latérale avec un angle d'émission θ non nul. En effet, lorsque l'angle d'incidence est nul, deux zones situées de part et d'autre du capteur à égale distance de celui-ci, réfléchissent l'onde en même temps. Il est alors impossible de faire la distinction entre ces différentes zones.

Une cible peut répondre de différentes manières. Les mécanismes de rétro-diffusion élémentaires sont : la simple réflection, la double réflection ou la réflection volumique (Figure 2.3).

Lors d'une simple réflection sur une surface plane par rapport à la longueur d'onde, l'énergie de l'onde diffusée ne revient pas vers le capteur et la cible apparaît alors très sombre sur l'image, comme pour de l'eau ou des routes. On considère qu'une surface



FIG. 2.2 – Axes d'une image radar et positionnement du capteur



FIG. 2.3 – *Mécanismes de rétro-diffusions élémentaires :* simple réflexion, double réflexion et réflexions volumiques

est plane par rapport à sa longueur d'onde. Autrement dit, une surface qui possède des rugosités dont l'amplitude ne dépasse pas la longueur d'onde peut réfléchir de la même manière qu'une surface complètement lisse. Bien sûr, l'énergie rétro-diffusée dépend de l'inclinaison de la surface. Une surface lisse dirigée "face au capteur" renverra une grande quantité d'énergie.

Le mécanisme de double réflection est très fréquent dans les zones urbaines en raison de la présence de bâtiments. Comme le montre le second schéma de la Figure 2.3, l'onde émise se réfléchit une première fois sur le sol, puis une seconde fois sur le mur. Dans ce cas une grande partie de l'onde émise est rétro-diffusée vers le capteur. Par conséquent, un coin sol-mur d'un bâtiment, orienté vers le capteur apparaîtra sous forme d'une ligne brillante dans l'image radar.

Dans le cas de réflection volumique ou alors de réflection simple sur des surfaces rugueuses, il est difficile de prévoir dans quelles directions l'onde sera réfléchie, seule

une partie est renvoyée vers le capteur. Cela dépend de l'angle d'incidence, la fréquence, la polarisation de l'onde électromagnétique et des propriétés électromagnétiques de la cible.

Basly [Basly, 2000] rapporte dans ses travaux l'influence de l'orientation, l'inclinaison du toit et la dimension d'un bâtiment, sur le type de réflexion. Une différence de quelques degrés suffit à diviser par dix la puissance du signal réfléchi. D'autre part, plus la cible est grande devant la longueur d'onde, plus l'énergie sera réfléchie de façon directive et se comportera comme un miroir. La puissance réfléchie dépend alors de l'angle d'incidence de l'onde radar.

Nous avons vu qu'un capteur radar a une vision latérale et l'image acquise possède deux axes, dont un (l'axe range) qui correspond au temps d'aller-retour de l'onde émise, et donc à la distance entre le capteur et la cible. Cette mesure en distance implique que différentes cibles seront projetées sur le même point de l'image si elles sont à égales distance du capteur. La Figure 2.4 montre ce phénomène, les points A et B sont à égale distance du capteur, sur le même front d'onde, les signaux renvoyés sont reçus en même temps par le capteur, ce qui entraîne une superposition de ces points sur l'image radar. De la même manière, une partie du toit du bâtiment sera confondue avec le sol.

De plus, la visée latérale implique des déformation géométriques de la scène qui n'existent pas dans le cas des images optiques :

- la contraction : les pentes orientées vers le capteur apparaissent comprimées, à l'inverse les pentes orientées dans l'autre sens sont étirées. La compression est maximale lorsque la pente est parallèle au front d'onde.
- le recouvrement : si la pente est supérieure à l'angle de visée du capteur, comme c'est le cas pour un bâtiment avec un mur vertical, les signaux de la pente et du sol sont mélangés.
- l'inversion : en cas de trop forte pente, le sommet de l'objet est représenté avant sa base. Sur la Figure 2.4, le sommet du toit (point B) répond avant sa base (point C).
- l'ombre : lorsqu'une pente est très élevée, l'onde ne rencontre pas d'objet, c'est le cas sur le segment EF, la taille de l'ombre dépend alors de la hauteur de l'objet.



FIG. 2.4 – Déformations géométriques d'un bâtiment dans une image radar.

2.3 Introduction aux images optiques

Les systèmes de télédétection optique permettent de capter la lumière du soleil réfléchie sur la terre (visible et infra-rouge) ainsi que le rayonnement propre des sols (infra-rouge thermique), comme le montre la figure 2.5. Les divers satellites optiques sont dotés de capteurs technologiquement très variés :

- les chambres photographiques, nécessitant de ramener les films à terre,
- les capteurs solides comme les caméras, radiomètres ou matrice de CCD (Charged-Coupled Device).

Les capteurs solides permettent l'acquisition d'une image selon diverses techniques (Figure 2.6) :

 un unique radiomètre doté d'une optique balayant une ligne au sol dans le sens orthogonal à celui du déplacement du satellite. Le mouvement du satellite permet



FIG. 2.5 – Acquisition d'une image satellite optique d'un bâtiment.

alors d'acquérir les lignes suivantes nécessaires à la construction d'une image.

- le *whiskbroom* est un système identique au précédent, mais avec plusieurs capteurs, généralement dédiés à différentes bandes de fréquences.
- le *pushbroom* est un système possédant plusieurs capteurs alignés, permettant l'acquisition simultanée de toute une ligne. L'image est construite ligne à ligne en prenant en compte le mouvement du satellite.
- la matrice CCD 2D qui permet d'obtenir une image instantanée d'une zone sans utiliser le déplacement du satellite pour sa construction.

Les données envoyées par les satellites ne sont donc pas des images toutes faites. Elles doivent être composées à partir des informations fournies par leurs instruments.

Différents modes d'acquisition existent : le mode "multibandes" (multispectral) ou le mode panchromatique. Dans le cas multibandes, plusieurs capteurs mesurent la réflectance dans de nombreuses bandes spectrales allant de 0.3 à quelques micromètres, dans le cas panchromatique, les capteurs mesurent la réflectance dans une large bande électromagnétique.



FIG. 2.6 – Capteurs optiques, figure issue de [Schowengerdt, 2007]

2.4 Géométrie des images

La présence du sur-sol, bâtiments ou végétation dans le milieu urbain, peut modifier considérablement l'aspect des données en fonction du type de capteur. La visée latérale de l'imagerie radar implique des déformations géométriques des objets en sur-sol, figure 2.7. La projection de chaque point de la scène dans chacune des images peut être calculée à partir d'équations explicitées dans le paragraphe 3.2.2. Il existe toutefois des zones d'ombres dans les images, de plus, chaque capteur a ses propres zones d'ombres qui dépendent de la position du capteur pour l'imagerie radar ou de la position du soleil pour l'imagerie optique.

2.5 Présentation de la base de données

Le cadre de travail de cette thèse est le milieu urbain et la haute résolution, de l'ordre du mètre. Les jeux de données qui nous ont permis de tester nos algorithmes proviennent du CNES pour l'image optique, de l'ONERA, Infoterra et le DLR (l'agence spatiale Allemande) pour les images radar. Il s'agit d'images optiques et radar (RAMSES TerraSAR-X) de la zone de Toulouse : figure 2.8, 2.9 et 2.10. Les images Quickbird et RAMSES font partie de la base de données mise à disposition par le CNES pour le programme ORFEO¹.

L'image optique sur la zone de Toulouse a été acquise par le satellite Quickbird 02 le

http://smsc.cnes.fr/PLEIADES/Fr/A_prog_accomp.htm



FIG. 2.7 – **Projection d'un bâtiment,** (a) : sur une image radar, (b) : sur une image optique en visée verticale.

01/04/2002. La résolution de cette image est de 65.9 cm, l'acquisition est faite en mode panchromatique entre 0.45 et $0.9\mu m$. La première image radar est acquise par l'avion et le capteur Ramses de l'ONERA en décembre 2006, en bande X, avec un angle de visée de 50° au centre de la fauchée. Le pas est de 35 cm en géométrie radar et en azimut. La seconde image radar est acquise par le satellite TerraSAR-X le 19/12/2009.

Sur ces images nous nous sommes intéressés à plusieurs zones de test : le CNES et la prison Saint Michel sur l'image radar de l'ONERA, présentés Figure 2.11. Une autre zone a été extraite de l'image satellite TerraSAR : le complexe scientifique de Rangueil, présentée Figure 2.12.



FIG. 2.8 – *Image Quickbird de Toulouse.* Le rectangle rouge délimite la zone correspondant à l'image radar. ©DigitalGlobe



FIG. 2.9 – Image radar de Toulouse acquise par Ramses. ©ONERA



FIG. 2.10 - Image radar de Toulouse acquise par TerraSAR-X. ©InfoTerra



FIG. 2.11 – **Zones de test :** (*a*) le CNES, (*b*) la prison. A gauche se trouvent les images optiques et à droite les images radar.

2.5. Présentation de la base de données



FIG. 2.12 – **Zone de test :** complexe scientifique de Rangueil. A gauche se trouve l'image optique et à droite l'image radar.

Chapitre 3

Recalage

3.1 Introduction

En traitement d'images, le recalage consiste à établir une relation géométrique entre plusieurs images qui peuvent être acquises à différents moments, avec différents capteurs ou de différents points de vues. Le recalage est la mise en correspondance de deux ou plusieurs images d'une même scène. Le recalage des images est une étape cruciale lorsque l'on souhaite combiner des données obtenues par différents capteurs, comme en fusion d'image, faire des mosaïques, de la détection de changement, de la reconnaissance d'objets, créer un modèle numérique d'élévation ou bien extraire les informations sur une scène à partir de plusieurs observations.

Un recensement des méthodes de recalage est proposé dans [Brown, 1992] [Zitová et Flusser, 2003] pour tous types de données. Les méthodes les plus fréquemment utilisées pour le recalage optique/radar sont des méthodes iconiques (qui se basent sur les intensités des pixels), des méthodes statistiques ou des méthodes géométriques utilisant les paramètres capteurs. Le choix que nous avons fait est d'élaborer une chaîne de recalage à partir d'un minimum d'informations sur les images. En effet, les paramètres des capteurs peuvent être difficiles à obtenir car les constructeurs ne souhaitent pas les partager, ou parce que certaines images ont été archivées sans ces paramètres. La chaîne que nous proposons doit être adaptée aux problèmes induits par les nouvelles résolutions et plus particulièrement au milieu urbain où les effets de parallaxe dûs aux bâtiments sont plus importants.

Nous verrons ici la méthodologie générale pour recaler des images, puis nous nous intéresserons aux méthodes les plus couramment utilisées pour le recalage des images

optiques et radar, avant de proposer notre approche.

3.1.1 Pourquoi faire du recalage?

Le recalage est une étape fondamentale en traitement d'images lorsque l'on souhaite mettre en correspondance deux données [Brown, 1992]. Par exemple, lorsqu'on souhaite recaler des images prises à différents instants, par des capteurs différents ou donnant différents points de vues.

Typiquement, nous avons recours au recalage pour combiner deux ou plusieurs sources d'informations. C'est le cas en télédétection pour faire de la surveillance de terrain ou de la détection de changement [Chesnel, 2008]. En médecine, on l'utilise par exemple afin de combiner les images de tomographies utilisées pour établir des diagnostics.

On utilise également le recalage en vision, pour faire des mosaïques d'images ou panoramiques d'images aériennes [Liu *et al.*, 2008], du suivi d'objets ... Il est aussi utile afin de construire des modèles d'élévation 3D par stéréoscopie [Lucas et Kanade, 1981].

3.1.2 Les types de déformations

Il existe plusieurs méthodes de recalage adaptées à différents types de déformation. Il est nécessaire de connaître le type de déformation afin de déterminer la méthode de recalage la mieux adaptée. On distingue 3 types de déformations : les déformations spatiales (ou géométriques), les variations temporelles et les différences d'acquisitions. Bien sûr plusieurs types de déformations peuvent être présents dans les données à recaler.

3.1.2.1 Déformations spatiales

Les déformations spatiales peuvent être corrigées par une transformation géométrique. Elles peuvent être dues à une différence d'acquisition ou un angle de vue différent.

Une déformation spatiale peut être globale, c'est à dire appliquée à toute l'image, ou locale, c'est à dire qu'elle diffère sur l'ensemble de l'image, comme le montre la Figure 3.1.

Les déformations usuelles sont :

• les déformations rigides (ou similitudes) : elles résultent de la composition d'une

rotation R d'angle θ , d'une translation T et d'un changement d'échelle s.

$$m' = \begin{pmatrix} x' \\ y' \end{pmatrix} = sR.m + T = s \begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix} \begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} t_x \\ t_y \end{pmatrix}$$

• les déformations affines : elles sont composées, comme la similitude, d'une rotation R et d'une translation T mais possèdent deux facteurs d'échelles s_1 et s_2 .

$$m' = A.m + T = \begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos\theta \end{bmatrix} \cdot \begin{bmatrix} s_1 & 0 \\ 0 & s_2 \end{bmatrix} \cdot \begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} t_x \\ t_y \end{pmatrix}$$

• les déformations polynomiales (ou élastiques). Elles s'écrivent sous la forme :

$$\begin{cases} x' = \sum_{i=0}^{n} \sum_{j=0}^{n-i} a_{i,j} x^{i} y^{j} \\ y' = \sum_{i=0}^{n} \sum_{j=0}^{n-i} b_{i,j} x^{i} y^{j} \end{cases}$$

• les déformations projectives (ou homographiques). Elles permettent de décrire les effets du changement de perspective. Pour exprimer la relation entre les points m et m' sous forme matricielle, il faut utiliser les notations en coordonnées homogènes :

$$m' = \begin{pmatrix} X' \\ Y' \\ Z' \end{pmatrix} = H.m = \begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{bmatrix} \cdot \begin{pmatrix} x \\ y \\ 1 \end{pmatrix}$$

Les coordonnées cartésiennes de m' sont obtenues de la manière suivante :

$$\begin{cases} x' = \frac{X'}{Z'} = \frac{h_{11}x + h_{12}y + h_{13}}{h_{31}x + h_{32}y + h_{33}}\\ y' = \frac{Y'}{Z'} = \frac{h_{11}x + h_{22}y + h_{23}}{h_{21}x + h_{22}y + h_{23}} \end{cases}$$

Dans le cadre du recalage d'images optique et radar, les déformations que nous chercherons sont de type rigide et polynomiale d'ordre deux.



FIG. 3.1 – Exemples de déformations spatiales.

3.1.2.2 Variations temporelles

Des changements temporels peuvent également apparaître entre deux images prises à des instants différents et ne sont généralement pas corrigeables par des transformations géométriques. Comme par exemple, un changement d'intensité, un changement atmosphérique ou un objet en mouvement. Ces variations rendent le recalage plus difficile en particulier dans le cas du milieu urbain à haute résolution, les variations temporelles sont souvent présentes : des bâtiments sont détruits, d'autres se batissent et les véhicules sont en mouvement.

3.1.2.3 Différences d'acquisition

Le troisième type de déformation est celui dû aux différences de capteurs. Une même scène prise sous le même angle avec des capteurs différents ne répondra pas de la même manière. C'est le cas en imagerie médicale, lorsque l'on souhaite recaler une image IRM (Imagerie par Résonnance Magnétique), fournissant des informations anatomiques et une TEP (Tomographie par Emission de Positons) fournissant des informations fonctionnelles

[Sarrut, 2000]. Ces deux images correspondent à la même scène mais apportent des informations différentes, comme le montre la Figure 3.2.



FIG. 3.2 – Images médicales : (a) coupe d'un volume IRM, (b) coupe d'un volume TEP.

On observe les différences d'acquisition en imagerie satellitaire sur la Figure 3.3. Parmi les différences, les bâtiments répondent avec des lignes brillantes sur l'image radar et les objets sur le toit, présents dans l'image optique, sont peu visibles sur l'image radar.



FIG. 3.3 – Images satellites de Toulouse : (a) Quickbird, (b) Ramses.

Sur la partie gauche de la Figure 3.3, des variations temporelles sont visibles. L'image optique montre un chantier, alors qu'un parking apparaît sur l'image radar.

Les différences peuvent également venir de la géométrie particulière des images, comme par exemple, en imagerie radar, en raison de la visée latérale, comme le montre la Figure 2.7 qui représente la projection d'un bâtiment. Sur l'image radar, une partie du

sol et du bâtiment sont superposées, de plus une *ombre* apparaît, elle correspond à une zone où le capteur radar ne reçoit aucune information. Dans l'image optique, une zone d'ombre peut également apparaître, mais elle dépend de la position du soleil et n'a pas de correspondance avec l'ombre radar. Ce schéma montre l'ambiguïté du recalage, *faut-il recaler la projection du sol entre les deux images ou alors la projection du bâtiment* ? Ceci nous oblige donc à introduire une analyse sémantique de la scène pour identifier le sol du sur-sol.

3.2 État de l'art sur le recalage

Afin de recaler nos images, nous devons trouver la transformation qui doit être appliquée pour que chaque pixel d'une image puisse être mis en correspondance avec un pixel de l'autre.

3.2.1 Généralités sur le recalage

On ne peut définir une unique méthode de recalage, quelles que soient les images, néanmoins la plupart des méthodes se décomposent généralement en quatre étapes :

- Modalité de l'image : choisir des informations à mettre en correspondance (signaux des images, contours, points de contrôle...).
- Critère de similarité : choix d'une mesure de la qualité d'une mise en correspondance.
- *Espace de recherche* : choix d'un ensemble des transformations géométriques (globale, locale, rigide...).
- Stratégie : choix d'une méthode d'optimisation du critère.



FIG. 3.4 – Méthode de recalage

Le choix de chacun de ces paramètres dépend bien entendu du type de déformation existant entre les images ; la connaissance que l'on a de celui-ci nous permettra de choisir un modèle adapté à ce type de recalage.

La première étape est de choisir les éléments qui seront mis en correspondance entre les images, ce que l'on appelle primitives. Dans certains cas, il n'est pas possible de travailler directement sur les intensités des pixels, par exemple dans le cas d'images multimodales ; on choisit alors d'extraire certains éléments qui serviront pour la mise en correspondance. Le point délicat est d'utiliser des éléments significatifs, présents dans les deux images et qui représentent la même réalité physique [Shekhar *et al.*, 1999]. Les primitives choisies peuvent être des primitives intrinsèques (régions, intersections, coins...) ou des primitives extrinsèques, c'est à dire des marqueurs n'appartenant pas à l'image, mais dont on connaît la position. Par exemple, [Eugenio *et al.*, 2004] utilise les bords de mer ou de fleuves comme primitives ; [Ali et Clausi, 2002] choisit d'extraire des regions homogènes. Les primitives extrinsèques sont souvent utilisées en imagerie médicale où on pose sur le patient des capteurs qui apparaitront sur l'image et permettront le recalage [Van den Elsen *et al.*, 1993] [Roche, 2001]. En imagerie satellite, cela s'apparente à utiliser des points de controles comme [Cheng et Toutin, 1997].

La deuxième étape est le choix du critère de similarité. Ce critère doit être adapté aux images que l'on souhaite recaler ; il permet d'attribuer une note à la transformation et ainsi d'évaluer sa qualité. Il existe trois familles de critères de similarité : les critères dérivés de la corrélation, les critères statistiques et les critères spatiaux ou de forme. [Roche, 2001] les classe en fonction de la dépendance statistique entre les deux images. Pour des images dont les radiométries suivent une dépendance affine (conservation de l'intensité, changement de contraste ...), on utilisera plutôt des variantes de la corrélation : corrélation normalisée, critère de Woods... L'hypothèse la plus faible que l'on puisse faire sur les radiométries est celle d'une dépendance statistique, dans ce cas on utilise un critère qui se base sur les densités de probabilité calculées à partir de l'histogramme joint (entropie conjointe, information mutuelle...). La dernière famille de critères s'appuie sur des caractéristiques spatiales : des longueurs, des superficies, des formes...

Le troisième point est le choix de la transformation. Le transformation choisie doit rendre compte de la déformation existante entre les images. Celles essentiellement utilisées en recalage d'images satellitaires sont les transformations rigides (translation, rotation, homothétie), les transformations polynomiales et les transformations élastiques.

Le quatrième point concerne la stratégie; il va permettre de guider le recalage afin de diminuer la complexité algorithmique. En effet, on peut pour toutes les transformations possibles, effectuer une recherche exhaustive et garder celle qui donne la meilleure similarité. Cependant si la transformation recherchée possède plusieurs degrés de liberté le temps de calcul sera long, on peut le diminuer en tenant compte de la stratégie. Par exemple, choisir de travailler localement, passer dans un espace (fréquentiel par exemple) où les calculs seront simplifiés ou utiliser des algorithmes qui permettent de converger plus rapidement vers la solution (descente de gradient, Levenberg-Marquardt...).

3.2.2 État de l'art sur le recalage en Optique/Radar

Dans le cadre spécifique du recalage d'images optique/radar, les images n'ont pas la même radiométrie ni la même géométrie, comme le montre la Figure 3.5. En dessous du rond-point, nous voyons sur l'image optique un bâtiment et un parking, alors que sur l'image radar, des lignes brillantes forment un rectangle. Ces lignes correspondent en fait à une réflection sur une barrière métallique qui n'est pas visible sur l'image optique. Nous observons également les différences de radiométries sur les bâtiments. L'image optique montre des rectangles homogènes correspondants aux toits, alors que l'image radar présente de grandes lignes verticales brillantes correspondant aux coins réflecteurs du toit.



FIG. 3.5 – *Images de la zone de Toulouse : extraits de l'image de Toulouse (a) optique, (b) radar, (c) zoom sur un bâtiment de l'image optique, (d) zoom sur un bâtiment de l'image radar.*

Dans la littérature trois types de méthodes sont utilisées pour le recalage des images optique et radar : lorsqu'on connaît les paramètres d'acquisition des images [Goncalves et Dowman, 2002], lorsqu'on utilise des primitives ou des points d'appuis
(souvent utilisés pour le recalage d'images basse résolution) ou bien des méthodes iconiques, c'est à dire qui se basent sur les intensités des pixels.

3.2.2.1 Recalage avec paramètres capteurs

Ainsi, lorsqu'on dispose des paramètres des capteurs, pour géoréferencer une image optique, deux approches co-existent. D'une part, il existe **un modèle paramétrique** qui décrit le mode physique d'acquisition des images. Mais hélas, les données respectives à cette acquisition ne sont pas toujours accessibles. **Le modèle non-paramétrique** (ou modèle empirique) quant à lui, utilise une fonction polynomiale qui fait le lien entre un pixel et sa position au sol. Par exemple, pour une image Quickbird, une vingtaine de paramètres, appelés Rational Polynomial Coefficient (RPC), sont donnés. Ils correspondent aux coefficients des polynômes qui modélisent la fonction de transfert [Gianinetto *et al.*, 2004]. Un point de coordonnées au sol (X, Y, Z) aura pour coordonnées (r, c) dans l'image.

$$\begin{cases} r = \frac{p_1(X, Y, Z)}{p_2(X, Y, Z)} \\ c = \frac{p_3(X, Y, Z)}{p_4(X, Y, Z)} \end{cases}$$

où :

$$p_n = \sum_{i=0}^{m_1} \sum_{j=0}^{m_2} \sum_{k=0}^{m_3} a_{i,j,k} X^i Y^j Z^k$$

sont les fonctions polynomiales et $a_{i,j,k}$ les coefficients RPC.

Pour une image radar, dans le modèle simplifié, l'antenne émet des impulsions avec une certaine fréquence. L'ensemble des échos reçus après émission forme alors une ligne i de l'image. La position j dans la ligne nous renseigne donc sur la durée aller-retour de l'onde.

Autrement dit, un pixel situé sur la colonne j de l'image, correspondant au temps d'aller-retour t_c , est localisé sur une sphère de rayon r égal à la distance parcourue par l'onde pendant le temps d'aller $t_c/2$. Le centre de cette sphère est la position du satellite à l'instant t de la prise de vue.

Un pixel situé sur la ligne i est vu par le capteur à l'instant t de la prise de vue avec une fréquence Doppler f_d liée au dépointage de l'antenne.

Ceci nous amène à formuler deux équations :

$$(x - X)^{2} + (y - Y)^{2} + (z - Z)^{2} = r^{2}$$

$$(x-X)\dot{X} + (y-Y)\dot{Y} + (z-Z)\dot{Z} = \frac{\lambda r f_d}{2}$$

où (x, y, z) sont les coordonnées de la cible, (X, Y, Z) et $(\dot{X}, \dot{Y}, \dot{Z})$ sont les vecteurs positions et vitesse du capteur.

Connaissant le pas temporel en azimut P_a , le pas temporel en range P_r , les temps t_c^0 , t^0 et r^0 pour un point de référence, on peut faire le lien entre les coordonnées (i, j) d'un point de l'image et les grandeurs physiques :

$$t = P_a i + t^0$$

$$t_c = P_r j + t_c^0$$

$$r = c \frac{t_c}{2} = c P_r j + r^0$$

Nous disposons alors d'un système de deux équations à trois inconnues. Nous supposons généralement Z connu, par exemple donné par un modèle numérique de terrain (MNT) pour retrouver le point (X, Y, Z), qui sera ensuite projeté dans l'image optique avec le polynôme rationnel. La précision de ce recalage dépend donc de la précision des paramètres capteurs et du Modèle Numérique de Terrain utilisé.

3.2.2.2 Recalage sans paramètres capteurs

Dans le cas où l'on ne dispose pas de ces paramètres, plusieurs méthodes ont été proposées en se basant sur les intensités des pixels. La plupart des méthodes de recalage iconique utilisent le coefficient de corrélation comme critère de similarité entre deux images, seulement celui-ci suppose que les radiométries suivent une dépendance linéaire, ce qui n'est pas le cas pour des images issues de capteurs optiques et radars, comme le montre la Figure 3.6.

[Inglada et Giros, 2004] et [Shabou *et al.*, 2007] proposent plusieurs critères de similarité adaptés aux cas multi-capteurs comme la distance de Kolmogorov ou l'information mutuelle. Ils montrent tout d'abord les limites du coefficient de corrélation dans le cas d'images multi-modales (Spot 5 et ERS), puis l'intérêt d'utiliser un critère statistique. [Oller *et al.*, 2006] recale des images optiques/radar en utilisant comme critère l'information mutuelle et montre les difficultés de recalage dues aux effets de parallaxe.

D'autres méthodes de recalage qui ne se basent pas sur les intensités des pixels,



FIG. 3.6 – **Histogrammes joints** (a) entre deux images optiques avec un changement d'intensité, (b) entre une image optique et radar. Le niveau de gris de l'histogramme est inversement proportionnel à la probabilité d'apparition du couple (x,y).

utilisent des primitives, comme des points de contrôle, des régions, des contours ou des coins. [Ali et Clausi, 2002] recale des images basse résolution Radarsat et Landsat en utilisant les contours de régions homogènes, en particulier les lacs. De même, [Shabou, 2006] utilise des lignes extraites dans chacune des images, tandis que [Wong et Clausi, 2007] utilise des points particuliers. [Tupin et Galland, 2004] s'intéresse aux erreurs de recalage obtenues à cause de l'incertitude sur les paramètres capteurs.

3.3 Recalage rigide par l'invariant de Fourier-Mellin

Dans le cas du recalage d'images optique et radar, nous disposons de deux images avec des radiométries différentes et des distorsions complexes dues aux caractéristiques propres des capteurs. Dans un premier temps, nous avons fait l'hypothèse d'une déformation rigide unique sur l'image globale, composée d'une rotation, translation et changement d'échelle entre les deux images, dont nous allons déterminer les paramètres.

Dans le cadre de cette thèse, nous nous intéressons particulièrement au milieu urbain avec des données à des résolutions métriques. Les éléments que nous avons choisis de mettre en correspondance sont donc des contours de bâtiments ou des routes. En effet, nous avons vu dans 3.2.1 que les primitives choisies doivent être significatives, ce qui est le cas pour celles-ci.

Afin de diminuer l'espace de recherche, nous avons choisi de faire une analyse pyra-

midale multi-échelles, en utilisant la corrélation de phase, cela entre dans la phase optimisation. Nous déterminerons la transformation sur une image à basse résolution, puis nous affinerons les résultats en descendant dans les niveaux de la pyramide.

Le recalage rigide a été fait de la manière suivante :

- initialisation : création de la pyramide multi-échelles sur 3 niveaux et extraction des primitives sur chacun d'eux,
- détermination des paramètres de rotation et de changement d'échelle sur le premier niveau,
- application de la transformation au niveau inférieur et itération jusqu'à la pleine résolution,
- détermination de la translation.

3.3.1 Initialisation et choix des primitives

La stratégie que nous proposons passe par une analyse multi-échelles. Pour ce faire, nous avons construit une pyramide d'images sur 3 niveaux en utilisant un moyennage incohérent pour l'image radar. Ce choix se justifie car cela revient à obtenir une image multi-vues. Pour l'image optique, les niveaux ont été créés en convoluant l'image avec un filtre Gaussien, puis en sous-échantillonnant d'un facteur deux. On appellera "premier niveau" celui ayant la plus basse résolution.

Le critère choisi pour estimer la transformation est expliqué dans 3.3.2.1, il se base sur la corrélation. Les images optique et radar ayant des radiométries différentes, l'information issue des intensités des pixels n'est pas suffisante pour déterminer la transformation, nous avons donc choisi de travailler sur des primitives.

A pleine résolution, les contours des bâtiments permettent un meilleur recalage que le réseau routier, à l'inverse quand la résolution diminue, ces contours sont moins précis, le réseau routier donne par contre un bon alignement pour débuter le recalage. Nous avons donc choisi d'extraire les lignes des routes sur le premier niveau (basse résolution) et d'affiner le recalage sur les autres niveaux en recalant sur les contours de bâtiments.

Le premier niveau de la pyramide permet d'extraire des routes. Pour l'image optique, l'extraction peut se faire avec un opérateur comme celui proposé par [Duda et Hart, 1972] ou bien alors un détecteur de contour comme celui proposé par [Canny,]. Nous avons utilisé le détecteur de contour de Canny. Dans l'image radar, à cause de la longueur d'onde de quelques centimètres, on peut considérer les routes

comme des surfaces lisses. L'onde émise par le radar sera réfléchie dans une autre direction, les routes apparaîtront donc comme des lignes sombres. Nous avons donc pris comme primitives les lignes sombres de l'image radar en utilisant un détecteur fondé sur [Tupin *et al.*, 1998].

Sur les autres niveaux de la pyramide l'image est mieux résolue. Des contours plus fins sont visibles, comme les contours de bâtiments. Nous cherchons alors à utiliser cette information ; pour cela nous allons utiliser le détecteur de contour de Canny sur l'image optique en supposant que ceux obtenus représentent les contours des bâtiments et des routes. En ce qui concerne l'image radar, le coin mur-sol induit souvent une double réflexion, ce qui se traduit par une forte rétrodiffusion. Les contours de bâtiments vont donc apparaître comme des lignes brillantes et sont extraits avec un détecteur de lignes claires [Tupin *et al.*, 1998]. Les pyramides et les primitives utilisées pour les images sont illustrées Figure 3.7 et 3.8 pour le jeu de données du CNES.

La dernière étape pour initialiser le recalage est de mettre les images à même résolution, en général ce paramètre est connu de manière très précise pour les données satellitaires. La méthode de recalage proposée permet de déterminer la rotation entre [-90°,90°], aussi nous devons orienter les images de manière à ce que le paramètre de rotation soit compris entre ces valeurs. Nous n'avons pas eu recours à cette étape dans ces travaux, car les données dont nous disposons étaient déjà correctement orientées.

3.3.2 Rotation et changement d'échelle

3.3.2.1 Invariant de Fourier-Mellin

• Méthode

Afin de déterminer les paramètres de rotation et de changement d'échelle, nous utilisons l'invariant de Fourier-Mellin [Reddy et Chatterji, 1996]. Cette méthode est une extension de la corrélation de phase qui permet d'estimer la translation entre deux images.

Soit f_1 et f_2 deux images différant d'une translation $(\delta x, \delta y)$ et F_1 et F_2 leurs transformées de Fourier, on a :

$$f_2(x,y) = f_1(x - \delta x, y - \delta y)$$

$$F_2(u,v) = e^{-j2\pi(u\delta x + v\delta y)}F_1(u,v)$$



FIG. 3.7 – Pyramide et primitives de l'image optique.



FIG. 3.8 – Pyramide et primitives de l'image radar.

$$\frac{F_1(u,v)F_2'^*(u,v)}{|F_1(u,v)F_2'(u,v)|} = e^{j2\pi(u\delta x + v\delta y)}$$

En prenant la transformée de Fourier inverse, on obtient un pic de dirac en $(\delta x, \delta y)$, correspondant aux paramètres de translation.

L'invariant de Fourier-Mellin est une extension de cette méthode pour obtenir les paramètres de rotation et changement d'échelle. Elle utilise pour cela une transformée logpolaire du module de la transformée de Fourier des images.

Soient g_1 et g_2 deux images qui diffèrent par une rotation d'angle θ_0 et un changement d'échelle de α , soient G_1 et G_2 leurs transformées de Fourier :

$$g_2(x,y) = g_1(\alpha(x\cos\theta_0 + y\sin\theta_0), \alpha(-x\sin\theta_0 + y\cos\theta_0))$$

D'après les propriétés sur la transformée de Fourier, une rotation reste une rotation de même angle et un changement d'échelle de rapport α devient un changement d'échelle de rapport $1/\alpha$.

$$G_2(u,v) = \frac{1}{|\alpha|} G_1(\frac{u}{\alpha}\cos\theta_0 + \frac{v}{\alpha}\sin\theta_0, \frac{-u}{\alpha}\sin\theta_0 + \frac{v}{\alpha}\cos\theta_0)$$

Si on convertit G_1 et G_2 dans un système de coordonnées log-polaire, on obtient :

$$G_2(\log \rho, \theta) = \frac{1}{|a|} G_1(\log \rho - \log \alpha, \theta - \theta_0)$$

La rotation et le changement d'échelle deviennent des translations. Nous pouvons alors utiliser la corrélation de phase pour en déterminer les paramètres. Étant donnée la symétrie dans le module de Fourier, il n'est pas nécessaire de projeter toute l'image en coordonnées log-polaire, nous ne projetons qu'une moitié de l'image correspondant aux angles compris dans [-90°,90°]. Le paramètre de rotation trouvé se situera entre ces valeurs modulo 180°.

Validation sur des images simulées

Nous avons utilisé la corrélation de phase sur deux images translatées manuellement de $(\delta_x, \delta_y) = (107, 27)$ pixels. Afin de montrer que cette méthode est robuste au bruit, nous avons également rajouté un bruit Gaussien de moyenne nulle et d'écart-type égal à 30. Les images sont représentées Figure 3.9. La réponse obtenue après avoir appliquée la transformée de Fourier inverse ressemble bien à un pic de Dirac. Elle atteint son



FIG. 3.9 – Corrélation de phase pour déterminer une translation.

maximum pour les coordonnées (106,27), ce qui correspond, à un pixel près, aux paramètres de translation. Cette méthode permet effectivement de retrouver les paramètres de translation, même en présence d'un bruit gaussien. En effet, dans le domaine spectral, un bruit gaussien devient un bruit uniforme, le fait de normaliser nous permet de nous affranchir de ce type de bruit.

• Validation sur des images de télédétection

La Figure 3.10 illustre l'invariant de Fourier-Mellin sur des images aériennes. Les primitives utilisées sont superposées aux images (a) et (b). Les transformées de Fourier des images montrent deux lignes importantes, correspondants aux deux directions privilégiées des images. Une rotation entre deux images induit une rotation des directions privilégiées, ce qui se répercute dans la transformée de Fourier, comme le montrent les Figures 3.10 (c) et (d). La corrélation de phase s'applique à des translations, or dans un système logpolaire, une rotation correspond à une translation sur l'axe polaire et un zoom à une translation sur l'axe radial. La projection en coordonnées log-polaire des transformées de Fourier est montrée Figure 3.10 (e) et (f). Comme cela a été dit précédemment, seule la moitié de l'image entre [-90°,90°] a été projetée en coordonnées log-polaire. Un rectangle rouge indique les directions qui doivent être mises en correspondance pour trouver l'angle de rotation.

Théoriquement, les "lignes verticales" observées devraient être absolument identiques entre les deux images, ce qui n'est pas le cas sur les Figures 3.10 (e) et (f). En pratique, les primitives utilisées sur les images ne sont pas strictement les mêmes, c'est pourquoi on observe ces différences. Cependant, **ces directions sont suffisantes pour déterminer l'angle de rotation entre les images.**

On applique la corrélation de phase sur les Figures 3.10 (e) et (f) et on obtient un maximum pour un angle de rotation de -78.6° et un zoom égal à 0.94. La Figure 3.11 montre les résultats après correction de la rotation et du changement d'échelle.

3.3.2.2 Application

L'invariant de Fourier-Mellin est une méthode très sensible au choix des primitives qui doivent être mises en correspondance. Afin d'augmenter la robustesse, nous avons choisi d'affiner le recalage au fur et à mesure que nous avançons dans la pyramide. De plus, afin d'éviter des erreurs dues à un mauvais angle et un mauvais facteur d'échelle, nous avons limité le facteur d'échelle. Etant donné que dans l'étape d'initialisation nous avons mis les images à même résolution, au lieu de chercher le pic de corrélation dans toute l'image, nous cherchons le pic dans une zone correspondant à un facteur d'échelle compris dans [0.95,1.05].

Au premier niveau de la pyramide, l'espace de recherche du maximum de similarité se fait sur $[-90^{\circ},90^{\circ}]$ pour l'angle de rotation et [0.95,1.05] pour le facteur d'échelle. La transformation trouvée est appliquée au niveau suivant, nous affinons la recherche autour de ces paramètres sur $[-10^{\circ},10^{\circ}]$ et [0.95,1.05], jusqu'au dernier niveau.

Nous avons alors trouvé les paramètres de rotation et changement d'échelle entre les deux images optique et radar, comme le montre la Figure 3.13. Pour terminer le recalage rigide, il nous reste à déterminer les paramètres de translation.

3.3.3 Translation

Pour déterminer les paramètres de translation, la méthode de Fourier-Mellin n'est pas satisfaisante. Le critère de similarité utilisé par l'invariant de Fourier-Mellin est la



FIG. 3.10 – Corrélation de phase pour déterminer une rotation et un changement d'échelle (a)image optique et ses primitives en rouge, (b) image radar et ses primitives en vert, (c) transformée de Fourier des primitives de l'optique, (d) transformée de Fourier des primitives du radar, (e) projection de l'optique en coordonnées log-polaire, (f) projection du radar en coordonnées log-polaire.



FIG. 3.11 – Corrélation de phase pour déterminer une rotation et un changement d'échelle (a)image optique, (b) image radar recalée.



FIG. 3.12 – Schéma synoptique pour déterminer la rotation.

corrélation. Précédemment, nous cherchions la rotation qui maximisait la corrélation entre les directions principales des images, ce critère de similarité était donc pertinent. Par contre, utiliser l'invariant de Fourier-Mellin pour déterminer la translation équivaut à chercher la translation qui maximise la corrélation entre nos primitives. Cependant, les primitives que nous avons extraites ne sont pas exactement les mêmes, comme le montre les Figures 3.7 et 3.8. Un décalage de quelques pixels entre les primitives induit une erreur sur la translation obtenue.

Nous devons donc mettre au point une étape complémentaire, c'est à dire déterminer de nouvelles primitives et un nouveau critère de similarité.

3.3.3.1 Critère de similarité

Plusieurs critères de similarité existent, adaptés à différents types de primitives et d'images à recaler. [Roche, 2001] classe les différents critères suivant les dépendances statistiques des radiométries des images. Ainsi, deux images ayant une conservation de l'intensité utiliseront comme critère la norme L1 ou la norme L2. Des images dont les



FIG. 3.13 – Correction de la rotation (a) image optique, (b) image radar.

intensités suivent une dépendance affine utiliseront des critères dérivés de la corrélation. Pour des images dont l'intensité ne peut pas s'exprimer de manière fonctionnelle, on utilise des critères statistiques, dérivés de l'entropie conjointe, comme le coefficient de corrélation entropique ou l'information mutuelle.

Dans le cas d'images multimodales, on ne peut établir de dépendance fonctionnelle entre les radiométries des images. Nous avons donc fait le choix d'utiliser un critère statistique : l'information mutuelle [Inglada et Giros, 2004].

3.3.3.2 Pré-traitement

Le choix de la fenêtre de 400x400 est un compromis entre un nombre suffisant de points pour avoir une bonne appréciation de l'histogramme joint et le temps de calcul. Les images radar ont pour caractéristique d'avoir quelques points très brillants ; ainsi si l'on veut avoir une bonne représentation de l'image quantifiée par son histogramme, il est important d'appliquer une transformation sur ce dernier, comme le montre la Figure 3.14(a). Une méthode couramment utilisée consiste à seuiller cette image [Lambers *et al.*, 2008]. Nous avons choisi comme seuil : $m + 3\sigma$ où m est la moyenne de l'image et σ son écart-type.

Nous avons également quantifié l'image radar sur 10 niveaux de gris, ce qui permet de diminuer les temps de calcul sans dégrader la qualité des résultats, ainsi que cela est montré à la section 3.4.3. La figure 3.15 illustre les effets visuels de la quantification



FIG. 3.14 – **Dynamique d'une image radar quantifiée** (a) avant seuillage, les valeurs les plus faibles sont les plus fréquentes, (b) après seuillage à $m + 3\sigma$, ce qui correspond à un zoom sur la partie proche de 0 de (a).

scalaire uniforme et du seuillage sur l'image radar ainsi que sur son histogramme.

3.3.3.3 Mise en oeuvre

Nous cherchons la translation qui maximise un critère de similarité adapté à notre recalage. Une translation ne possèdant que deux degrés de liberté, il est possible de faire une recherche exhaustive sans que le temps de calcul soit trop long.

L'information mutuelle est une mesure de similarité statistique basée sur l'histogramme joint des images. Soient deux variables aléatoires X et Y, on a :

$$MI(X,Y) = H(Y) - H(Y|X) = H(X) - H(X|Y)$$

= $H(X) + H(Y) - H(X,Y)$

où $H(X) = -E_X(\log(P(X)))$ représente l'entropie de X, P(X) la densité de probabilité et E_X l'espérance.

La méthode utilisée pour déterminer la translation est donnée Figure 3.16. Nous extrayons une zone de taille 400x400 au centre de l'image optique et nous calculons l'information mutuelle entre la zone extraite de l'image optique et une zone de même taille sur l'image radar pour toutes les translations possibles. Un pré-traitement de l'image radar est fait avec un seuillage et une quantification. La translation désirée correspond à un maximum de l'information mutuelle, comme le montre la Figure 3.17.



(c)

FIG. 3.15 – *Image radar seuillée :* (*a*) non quantifiée, (*b*) quantifiée sur 10 niveaux, (*c*) histogramme de l'image radar quantifiée.

3.3. Recalage rigide par l'invariant de Fourier-Mellin



FIG. 3.16 – Schéma synoptique pour déterminer la translation.



FIG. 3.17 – Information mutuelle calculée sur un décalage de ± 200 pixels autour de la position d'origine. Le pic obtenu nous indique une translation de (128,5) pixels.

Afin de diminuer les temps de calcul, nous n'avons fait ici qu'une recherche dans un voisinage de taille 200×200 pixels autour du centre de l'image. L'espace de recherche est important et doit être plus grand que la translation entre les données, sans a priori sur celle-ci, il faut étendre l'espace de recherche à toute l'image.

3.3.4 Analyse des résultats

Les résultats obtenus sur la zone du CNES à Toulouse sont donnés Figure 3.18. Nous ne disposons pas de vérité terrain afin de calculer l'erreur de recalage, nous l'avons donc évaluée manuellement. Pour cela, nous avons choisi des points de contrôle au sol sur chacune des images et mesuré l'erreur moyenne de recalage. Nous obtenons **une erreur moyenne de 30 pixels en valeur absolue**.

La figure 3.19 nous montre la répartition des erreurs de recalage en deux dimensions. Les erreurs se répartissent selon une droite, cela signifie que les déformations restantes sont dirigées selon cet axe. En effet, nous nous sommes intéressés aux translations locales



FIG. 3.18 – *Images recalées d'après une transformation rigide :* (*a*) *image optique,* (*b*) *image radar.*

en quelques points représentés figure 3.20. Les flèches rouges indiquent les directions des erreurs résiduelles de recalage, elles ont été agrandies pour une meilleure visibilité. Nous voyons qu'il reste une déformation semblable à un cisaillement qui doit encore être corrigée afin d'améliorer le recalage.

3.4 Recalage polynômial

Les modèles polynômiaux ont été parmi les premiers modèles employés à l'ordre un ou deux pour corriger des effets de translation, rotation, échelle, torsion et convexité.

Dans le paragraphe 3.3 nous avons fait l'hypothèse d'une déformation rigide entre les deux images. C'est cette déformation que nous avons cherché à corriger, mais cette hypothèse n'est pas entièrement vérifiée. Il reste notamment une erreur moyenne de 30 pixels. Afin d'améliorer le recalage, nous allons rechercher et appliquer une déformation polynomiale entre les images, qui permet de corriger l'effet de cisaillement que nous avons observé dans la section précédente.

Le schéma de recalage proposé est le suivant :

- extraction de points d'amer sur l'image optique.
- utilisation de l'information mutuelle pour trouver dans l'image radar, les points correspondants à ceux extraits.
- détermination des coefficients du polynôme par minimisation de l'erreur entre les paires de points.



FIG. 3.19 – *histogramme des erreurs de recalage données en pixel.* Ce graphe montre la répartition des erreurs de recalage verticales et horizontales de la base étiquetée. Les zones rouges correspondent aux distorsions résiduelles souvent présentes dans l'image et les zones bleues, celles moins présentes.

3.4.1 Extraction de points d'amer

Les points d'amer ou d'appui, sont les points utilisés pour estimer la transformation globale entre les images. L'utilisation d'un détecteur de coins permet de s'assurer que les points d'amers ne seront pas pris dans une zone homogène ou une zone texturée comme une forêt. En effet, il est difficile d'apparier des zones homogènes, le critère de similarité donnant une valeur constante. Le coin détecté nous assure que dans la zone que l'on souhaite recaler, il se trouve une structure sur laquelle s'appuyer.

Le détecteur de coin utilisé est le détecteur de Harris [Harris et Stephens, 1988]. Ce détecteur consiste en la recherche de points où les courbes (c'est à dire les lignes de niveaux de gris de l'image) présentent une très forte courbure localement. En effet, un coin est généralement caractérisé par une large variation des gradients suivant deux directions. Cela se caractérise par l'analyse des valeurs propres de la matrice d'autocorrélation A.



FIG. 3.20 – *Déformations résiduelles :* les flèches rouges indiquent la direction des erreurs de recalage mais ne sont pas à l'échelle pour une meilleure visibilité.

$$A = \begin{bmatrix} \langle I_x^2 \rangle & \langle I_x I_y \rangle \\ \langle I_x I_y \rangle & \langle I_y^2 \rangle \end{bmatrix}$$

où I_x et I_y sont les gradients de l'image I suivant les axes x et y et $\langle . \rangle$ représente la moyenne sur un voisinage.

Trois cas se distinguent suivant les amplitudes des valeurs propres λ_1 et λ_2 de A :

- $\lambda_1 \simeq 0$ et $\lambda_2 \simeq 0$, il n'y a pas de point d'intérêt.
- $-\lambda_1 \simeq 0$ et λ_2 est une valeur positive de grande amplitude, il y a un bord.
- $-\lambda_1$ et λ_2 sont deux valeurs positives de grandes amplitudes, il y a un coin.

3.4. Recalage polynômial

Une fois le détecteur de Harris appliqué à l'image optique, nous ne gardons qu'un certains nombre de points équi-répartis sur l'image, qui serviront d'amers. Il est nécessaire de prendre suffisemment de points d'amers pour pouvoir calculer le polynôme, soit six points au minimum pour un polynôme d'ordre deux. Ici, nous avons fait le choix de sélectionner 25 points. Pour cela, l'image est virtuellement découpée en 25 zones égales. Pour chacune des 25 zones de la grille, le point donnant la plus forte réponse au détecteur de Harris est conservé. Nous avons ainsi extrait 25 points particuliers répartis sur l'image optique, comme le montre la figure 3.21. Parmi eux, certains points seront "en dehors" ou au bord de l'image, ces points seront alors rejetés.



FIG. 3.21 – Extraction des points sur l'image optique.

3.4.2 Appariement des points

Les images ont déjà été pré-recalées à l'aide d'un recalage rigide. Nous avons donc supposé que la déformation restante est locale, c'est à dire que pour un point donné, son correspondant doit se trouver dans un voisinage. De plus nous avons supposé que dans ce voisinage restreint, la déformation est une translation. Pour trouver les correspondants des points extraits sur l'image optique, l'information mutuelle est à nouveau utilisée.

L'hypothèse de déformation locale ainsi que le temps de calcul nécessaire obligent à

réduire au maximum la taille de la fenêtre. Une étude sur la taille de fenêtre à utiliser pour calculer l'information mutuelle a été faite, ainsi que sur un critère permettant de déterminer si les appariements sont corrects. En effet, il peut arriver que l'information mutuelle donne un maximum pour un appariement qui ne soit pas le bon.

3.4.3 Choix des paramètres

Afin de déterminer si l'appariement donné par l'information mutuelle est valide, nous avons calculé des probabilités de fausses alarmes pour différentes tailles de fenêtre et niveaux de quantification.

8000 valeurs d'informations mutuelles entre deux images optiques et radar recalées ont été calculées, de même entre deux images non recalées. Cette opération a été faite avec différentes tailles de fenêtre et différentes quantifications. Pour chaque cas, nous avons déterminé le seuil qui permet de minimiser les mauvaises détections et les fausses alarmes.

Nous appellons **faux négatif** les valeurs de l'information mutuelle correspondant à un bon recalage, mais qui sont inférieures au seuil de détection.

Les **fausses alarmes** sont les valeurs de l'information mutuelle obtenues entre deux fenêtres non recalées qui sont supérieures au seuil de détection et donc considérées comme bonnes.

La Figure 3.22 représente les histogrammes des réponses obtenues par l'information mutuelle pour différentes tailles de fenêtre. La courbe rouge correspond aux réponses obtenues entre deux images optiques et radar quelconques et la courbe bleue entre deux images recalées. Notons que pour une fenêtre de 400x400 les deux courbes sont distinctes. Il est donc possible de déterminer un seuil s_{min} tel que si le maximum d'information mutuelle lui est inférieur, nous avons un mauvais recalage. A l'inverse, les réponses supérieures à ce seuil s_{min} seront attribuées à un bon recalage. Le seuil choisi correspond à la valeur qui minimise le taux de faux négatifs et de fausses alarmes. La courbe, figure 3.23, représente le pourcentage d'erreur en fonction du seuil. Finalement le seuil choisi est celui qui permet de minimiser cette courbe. Ce seuil se détermine également graphiquement sur une courbe ROC (*Receiver Operating Characteristic* ou Caractéristique de fonctionnement du récepteur). Il s'agit d'une courbe qui représente le taux de faux positif en fonction du taux de faux négatif pour plusieurs valeurs de seuil. Dans ce cas, le seuil s_{min} que nous choisissons est l'intersection entre notre courbe et la droite de pente -1

qui a une ordonnée à l'origine minimale, nous voyons un exemple graphique sur la figure 3.6(b).

$$s_{min} = argmin(FA(s) + FP(s))$$

 $FA(s) = taux de fausses alarmes > s$
 $FP(s) = taux de faux négatifs > s$

Nous observons également que plus la taille de fenêtre diminue, plus les courbes se superposent. Pour calculer l'information mutuelle, il ne faut donc pas prendre une taille de fenêtre trop petite. La Figure 3.22 nous montre la limite, en terme de taille de fenêtre, qui nous permet de séparer les bonnes réponses des mauvaises. Dans cet exemple, prendre une taille de fenêtre plus petite que 250x250 ne nous permet pas d'avoir suffisamment d'information pour avoir une bonne estimation de l'information mutuelle. Cette constatation montre que l'étape de détection des points par le détecteur de Harris n'est, dans le cas général, pas cruciale. En effet, nous pouvons supposer qu'une fenêtre de taille 250x250 pixels contient suffisamment de structures.

Un autre paramètre qui peut influencer les résultats est la quantification appliquée aux images pour calculer l'information mutuelle. Nous avons donc étudié l'influence de la taille de fenêtre et de la quantification sur les faux négatifs et les fausses alarmes. A cause de sa radiométrie, il est possible de modifier l'histogramme avant la quantification. Les méthodes couramment utilisées pour cela sont le seuillage de l'image, le calcul du logarithme ou l'application d'une fonction Gamma [Lambers *et al.*, 2008]. Plusieurs méthodes de quantification sont possibles, nous nous sommes intéressés à deux en particulier : la quantification scalaire uniforme et une quantification non uniforme obtenue par filtrage.

• Quantification scalaire uniforme

Comme nous l'avons vu dans le paragraphe 3.3.3.2, les images radar ont une dynamique telle qu'il est nécessaire de les seuiller avant d'appliquer une quantification scalaire uniforme, sous peine de perdre beaucoup d'information. Nous avons donc seuillé notre image à $m + 3\sigma$ où m est la moyenne de l'image et σ son écart-type. Nous avons ensuite fait une étude sur le nombre de niveaux de quantification et la taille de fenêtre. Les résultats sont donnés dans les tableaux 3.1 à 3.5, dans lesquels pour chaque niveau de



FIG. $3.22 - Distribution des valeurs prises par l'information mutuelle. Les courbes rouges représentent les valeurs prises par deux zones distinctes, en bleu celles obtenues pour deux zones identiques et recalées : (a) pour une fenêtre de taille <math>400 \times 400$, (b) 300×300 , (c) 250×250 , (d) 200×200 . Plus la taille de la fenêtre d'analyse est grande, plus les courbes sont distinctes, nous pouvons alors séparer les valeurs de l'information mutuelle correspondant à un bon recalage des autres.



FIG. $3.23 - Pourcentage d'erreur en fonction du seuil pour une fenêtre de taille <math>250 \times 250$. L'erreur correspond à la somme des faux négatifs et des faux positifs.

quantification, nous étudions l'influence de la taille de fenêtre. Ces résultats sont aussi représentés sous forme de courbes ROC, figure 3.6, où pour une taille de fenêtre donnée nous voyons l'influence de la quantification ou du filtrage.

Dans les tableaux 3.1 à 3.3, l'image radar a été quantifiée en 255, 100 et 10 niveaux de gris. Nous observons que la quantification de l'image radar a eu un effet positif. Pour une même taille de fenêtre, quantifier fortement l'image radar permet de réduire les faux négatifs et fausses alarmes. Autrement dit, **pour un taux équivalent de fausses alarmes, la quantification de l'image radar permet de réduire la taille de fenêtre.**

Nous allons étudier l'influence de la quantification sur l'image optique dans les tableaux 3.4 et 3.5. Pour cela, nous avons calculé les taux de mauvaises détections et fausses alarmes de l'information mutuelle, pour différentes quantifications de l'image optique en 100 et 10 niveaux. L'image radar est quantifiée en 10 niveaux.

Les tableaux 3.3 et 3.4 montrent que la quantification de l'image optique de 255 à 100 niveaux de gris n'a pas beaucoup d'impact sur le calcul de l'information mutuelle. En effet, pour une même taille de fenêtre, les taux de mauvaise détection et de fausse alarme sont du même ordre. Par contre, quantifier trop fortement l'image optique, à 10 niveaux de gris, fait augmenter ces taux, comme le montre le tableau 3.5. Nous trouvons les mêmes conclusions lorsque nous regardons les courbes ROC, figure 3.6. Ces graphes représentent le taux de faux positif en fonction du taux de faux négatif pour plusieurs tailles de fenêtres (200x200, 250x250, 300x300 et 400x400 pixels) et différents

Taille de fenêtre	s_{min}	% faux négatif	% fausse alarme
400x400	0.105	2.09	0.28
300x300	0.171	14.87	4.72
250x250	0.231	34.26	16.77
200x200	0.23	10.77	83.29

TAB. 3.1 – Seuil pour l'information mutuelle lorsque les images sont quantifiées sur 256 niveaux.

Taille de fenêtre	s_{min}	% faux négatif	% fausse alarme
400x400	0.054	0	0
300x300	0.087	1.36	0.31
250x250	0.125	15.14	4.06

TAB. 3.2 – Seuil pour l'information mutuelle avec quantification de l'image radar en 100 niveaux et 256 niveaux pour l'image optique.

Taille de fenêtre	s_{min}	% faux négatif	% fausse alarme
400x400	0.016	0.0	0.0
300x300	0.024	0.01	0.01
250x250	0.031	0.65	1.05
200x200	0.047	8.82	10.75

TAB. 3.3 – Seuil pour l'information mutuelle avec quantification : 10 niveaux pour l'image radar et 256 pour l'image optique.

Taille de fenêtre	s_{min}	% faux négatif	% fausse alarme
400x400	0.013	0	0
300x300	0.017	0.11	0.11
250x250	0.023	0.68	1.15
200x200	0.033	7.32	11.03

TAB. 3.4 – Seuil pour l'information mutuelle avec quantification de l'image optique en 100 niveaux et de l'image radar en 10.

Taille de fenêtre	s_{min}	% faux négatif	% fausse alarme
400x400	0.007	0.17	0.11
300x300	0.008	0.76	0.91
250x250	0.01	3.19	3.89
200x200	0.016	16.8	14.36

TAB. 3.5 – Seuil pour l'information mutuelle avec quantification des images optique et radar en 10 niveaux.

niveaux de quantification des images optiques et radar. La courbe ayant les meilleures performances est celle qui a le moins d'erreur, autrement dit, celle qui est inférieure aux autres ou la plus proche de 0. Nous obtenons donc de meilleures performances lorsque nous quantifions l'image radar sur 10 niveaux et que nous laissons l'image optique sur 255.

• Quantification et Filtrage

De nombreuses approches ont été proposées pour filtrer les images corrompues par un bruit de chatoiement, comme les images radar. La modélisation par champs de Markov (CdM) fournit un cadre adapté pour prendre en compte la nature du bruit et imposer la régularité de la solution. Dans ce contexte, la minimisation de la variation totale a été abondamment utilisée [Rudin *et al.*, 1992] afin de limiter les oscillations dans l'image régularisée tout en préservant les bords.

Le bruit de chatoiement suit une distribution de probabilité à queue lourde et la formulation par CdM conduit à un problème de minimisation mettant en jeu des attaches aux données non-convexes ainsi qu'une régularisation non lisse. Une telle minimisation peut être réalisée par une approche d'optimisation combinatoire en calculant des coupures minimales de graphes (graph-cuts). Bien que cette optimisation puisse être parfois menée de façon exacte en théorie, ce type d'approche ne peut être appliqué en pratique sur les images de grande taille rencontrées dans les applications de télédétection à cause de leur grande consommation de mémoire. Le temps de calcul des algorithmes de minimisation approchée (en particulier α -extension) est également généralement trop élevé quand la régularisation conjointe de plusieurs images est considérée.

[Denis *et al.*, 2009] propose une méthode rapide et efficace pour régulariser des images radar par coupures minimales successives. L'image du CNES filtrée sur 256 niveaux est représentée figure 3.24. Le tableau 3.7 nous donne les taux de mauvaise détection et de fausse alarme pour différentes tailles de fenêtre. De même que précédemment, nous voyons l'importance de bien choisir la taille de fenêtre, plus elle est petite et plus notre erreur est grande. Les résultats obtenus en filtrant l'image radar sont moins bons, pour une même taille de fenêtre, qu'en effectuant une quantification scalaire uniforme sur 10 niveaux. Toutefois, une étude plus approfondie sur le nombre de niveaux peut peut-être améliorer ces résultats.

• Conclusion

Nous avons supposé que la déformation résiduelle entre nos images était une transla-



TAB. 3.6 - Courbe ROC: taux de faux positif en fonction du taux de faux négatif pour différentes taille de fenêtres et niveaux de quantification des images optiques et radar. Pour une même taille de fenêtre, la quantification de l'image radar améliore le calcul de l'information mutuelle, tandis que ce n'est pas nécessairement le cas pour l'image optique. Nous cherchons les paramètres de quantification qui permettent de minimiser les taux d'erreurs sur une fenêtre d'analyse la plus petite possible. Les paramètres choisis sont une fenêtre de 250×250 et une quantification sur 256 niveaux de l'image optique et 10 niveaux de l'image radar.



FIG. 3.24 – Image du CNES de Toulouse quantifiée et filtrée par CdM avec optimisation par graph-cuts.

tion locale. Dans ce cas, il vaut mieux choisir les paramètres de manière à ce que la taille de fenêtre soit la plus petite possible, car dans ces conditions notre hypothèse peut être vérifiée. De plus, minimiser la taille de la fenêtre, dans laquelle on calcule l'information mutuelle, permet de diminuer les temps de calcul.

D'après l'étude que nous avons faite sur la quantification, nous avons choisi d'utiliser une quantification uniforme associée à un seuillage de l'image radar. Nous avons vu qu'il est intéressant de quantifier l'image radar, mais pas nécessairement l'image optique. Nous avons donc choisi une quantification sur 10 niveaux pour l'image radar et laissé l'image optique sur 256 niveaux de gris. Les taux de mauvaise détection et fausse alarme pour ces paramètres sont donnés dans le tableau 3.3. La Figure 3.22 montre les histogrammes des

Taille de fenêtre	s_{min}	% faux négatif	% fausse alarme
400x400	0.161	0	0
300x300	0.239	2.87	0.59
250x250	0.325	9.42	1.35
200x200	0.4	13.97	32.01

TAB. 3.7 – Seuil pour l'information mutuelle avec filtrage de l'image radar.

valeurs de l'information mutuelle.

Ainsi, pour une fenêtre de taille 400x400, on peut déterminer un seuil qui sépare les bons recalages des mauvais. Par contre, plus la taille de fenêtre diminue, plus l'erreur augmente. Pour une fenêtre de taille 250x250, nous avons une erreur de l'ordre de 1% ce qui est un bon compromis entre l'erreur autorisée et le temps de calcul. Par contre pour une plus petite fenêtre, comme 200x200, l'erreur monte à 10%. On retrouve ce résultat sur la Figure 3.22. Ces courbes montrent la limite, en terme de taille de fenêtre, qui permet de séparer les bonnes réponses des mauvaises. Prendre une taille de fenêtre plus petite que 250x250 ne nous permet pas d'avoir suffisamment d'information pour avoir une bonne estimation de l'information mutuelle. Pour une fenêtre de cette taille, le seuil que nous choisissons pour déterminer si le critère est correct ou non est de 0.031. Cette valeur a été choisie en tenant compte du seuillage et de la quantification appliqués à l'image radar.

En ce qui concerne le voisinage dans lequel nous recherchons la correspondance, nous avons choisi une taille de 60 pixels. En effet, nous avons remarqué que l'erreur moyenne obtenue après le recalage rigide était de l'ordre d'une trentaine de pixels, nous cherchons donc sur un voisinage supérieur à 30 pixels. Ce choix peut être modifié, cependant une taille de voisinage plus grande demandera des calculs plus longs et à l'inverse, prendre une taille de voisinage trop faible pourra donner de mauvais résultats si le décalage est plus grand.

L'information mutuelle est donc calculée sur une **fenêtre de 250x250** centrée sur les points extraits précédemment dans l'image optique. Une étape de quantification de l'image radar en 10 niveaux est faite au préalable. Nous allons ensuite chercher sur un **voisinage de 60x60 pixels** le maximum de similarité avec l'image radar. Ceci nous permet, pour chacun des points extraits précédemment dans l'image optique, de trouver son correspondant dans l'image radar. Tous les points dont l'information mutuelle est inférieure

au seuil déterminé (0,031 dans notre cas) sont rejetés.

3.4.4 Transformation polynomiale

Parmi les 25 points sélectionnés dans l'image optique, beaucoup ont été rejetés, soit parce qu'ils étaient en bordure de l'image et ne donnaient pas suffisamment d'information pour calculer l'information mutuelle, soit parce qu'aucun correspondant donnant une bonne valeur de similarité n'a été trouvé. Les points restants (environ une dizaine), vont servir à identifier la déformation polynomiale entre les images.

Nous recherchons un polynôme de degré deux, car prendre un degré égal à un revient à rechercher une translation, rotation et changement d'échelle ce qui a déjà été fait. A l'inverse, prendre un polynôme de degré élevé (ordre 3 ou 4) coûterait cher en temps de calcul et nécessiterait plus de points d'amers.

Le polynôme d'ordre deux recherché est de la forme :

$$\begin{cases} x' = \sum_{i=0}^{n} \sum_{j=0}^{n-i} a_{i,j} x^{i} y^{j} \\ y' = \sum_{i=0}^{n} \sum_{j=0}^{n-i} b_{i,j} x^{i} y^{j} \end{cases}$$

où (x, y) et (x', y') sont les coordonnées des points dans les images optique et radar, n = 2 le degré du polynôme et $(a_{i,j}, b_{i,j})$ les coefficients.

Le problème à résoudre est de déterminer les coefficients du polynôme d'ordre deux.

3.4.5 Estimation du polynôme

Plusieurs méthodes sont envisageables pour déterminer les paramètres de la transformation polynomiale à partir de points d'amers.

• La méthode RANSAC (Random Sample Consensus) [Fischler et Bolles, 1981] est une méthode de vote probabiliste qui permet de ne pas tenir compte des paramètres aberrants dans le calcul du modèle. L'algorithme consiste à évaluer un modèle à partir d'échantillons pris au hasard parmi ceux dont on dispose. Puis, on compte le nombre total d'échantillons qui correspondent à ce modèle. Cette opération est itérée un certain nombre de fois, on garde au final le modèle qui donne la plus petite erreur résiduelle. • La méthode utilisée qui a été proposée par Gauss et Legendre [Rousseeuw et Leroy, 1987], est la méthode des **moindres carrés**. On cherche les paramètres qui vont minimiser l'erreur quadratique. Nous disposons d'un modèle des signaux $s_k(\theta)$ dépendants de paramètres θ ainsi que d'une mesure de ces signaux s_k^* . On cherche le jeu de paramètres θ^* qui vérifie $s_k^* = s_k(\theta^*)$. On définit le résidu comme étant $r_k(\theta) = s_k^* - s_k(\theta)$.

La méthode des moindres carrés consiste à minimiser la fonction de coût :

$$C(\theta) = \sum_k r_k^2(\theta)$$

La solution de ce problème s'obtient pour grad $C(\theta) = 0$. La détermination des paramètres optimaux se ramène à la résolution d'un système d'équation linéaire.

Nous avons choisi d'appliquer la méthode des moindres carrés pour déterminer les coefficients $(a_{i,j}, b_{i,j})$ de notre polynôme. L'inconvénient de cette méthode est qu'elle est très sensible aux valeurs déviantes. Seulement, afin d'être robuste, nous avons sélectionnés au préalable les points d'amer en fonction de la valeur de l'information mutuelle. Cette méthode nous est apparue mieux adaptée que RANSAC car nous minimisons l'erreur de recalage pour tous les points de l'image ayant une information mutuelle significative, tandis que RANSAC ne considère que les points nécessaires à l'évaluation du polynôme (7 points pour un polynôme d'ordre 2). Or nous ne connaissons pas la répartition de ces points d'amer dans l'image.

3.5 Résultats

3.5.1 Image du CNES

Ceci nous permet de calculer au mieux la déformation restante entre nos deux images. Les résultats obtenus sur les images du CNES de Toulouse sont donnés Figure 3.25. De la même manière que pour le recalage rigide et sur les mêmes points de contrôles, nous avons évalué manuellement l'erreur. Le recalage polynômial nous permet de passer d'une erreur moyenne de 30 pixels à une distorsion de 10 pixels. Sur la Figure 3.26 les primitives de chaque image ont été superposées.

Afin d'évaluer la qualité du recalage, nous avons déterminé manuellement la déformation restante pour quelques points. Ces déformations sont montrées sur la figure



(b)

FIG. 3.25 – *Images recalées :* (a) image optique, (b) image radar.



FIG. 3.26 – Superposition des primitives des images optique et radar après recalage. Les lignes vertes représentent les primitives de l'image radar et les lignes rouges les primitives de l'image optique. Les primitives sont bien alignées, en particulier les contours des routes et des zones d'herbes, ce qui correspond à un bon recalage au sol.

3.28. On remarque que la zone en bas à droite de l'image est moins bien recalée, c'est à cause du manque de points d'amers dans cette zone. La transformation calculée n'a pas pris en compte la transformation locale dans cette zone. La répartition des erreurs de recalage est donnée par l'histogramme 2D, figure 3.27. Les erreurs semblent équi-réparties dans toutes les directions et nous voyons que, pour la majorité des points, l'erreur est inférieure à une dizaine de pixels, soit en tenant compte de la résolution des images, inférieure à 7 m.



FIG. 3.27 – Histogramme des erreurs de recalage après la transformation polynomiale.

3.5.2 Image de la prison

Nous observons Figure 3.29 le recalage obtenu sur une autre zone : celle de la prison de Toulouse. Cette image représente une zone urbaine dense. Nous avons appliqué le même processus de recalage que pour la zone du CNES, à savoir extraire les contours de routes et de bâtiments sur chaque image, chercher la transformation rigide entre les deux données, puis, améliorer le recalage à l'aide d'une transformation polynomiale. Pour cela, les mêmes paramètres que pour la zone du CNES ont été utilisés : même taille de fenêtre et de seuil pour l'information mutuelle. La Figure 3.30 représente la superposition des primitives de chaque image après la chaîne de recalage, soit après le recalage rigide et polynomial.

Au cours du processus de recalage, nous avons vu l'amélioration apportée par la transformation polynomiale dans le recalage d'une zone dense comme celle-ci. L'erreur moyenne que nous avons mesurée manuellement sur quelques points (une vingtaine) passe de 33 à 15 pixels, soit environ 10 m. Le recalage d'une zone urbaine dense est plus complexe à cause de la présence d'un grand nombre de bâtiments qui entraînent



FIG. 3.28 – **Qualité du recalage :** les points verts représentent les erreurs inférieures à 10 pixels, en orange les erreurs comprises entre 10 et 20 pixels et les points rouges les erreurs supérieures à 20 pixels.



(b)

FIG. 3.29 – *Images recalées :* (a) image optique, (b) image radar.


FIG. 3.30 – Superposition des primitives des images optique et radar après recalage. Les lignes vertes représentent les primitives de l'image radar et les lignes rouges les primitives de l'image optique.

des différences géométriques entre les images. Toutefois, nous obtenons un recalage de bonne qualité.

3.5.3 Image TerraSAR du complexe scientifique de Rangueil

Nous avons également testé notre méthode de recalage sur un autre type de données, acquis par le satellite radar TerraSAR-X. La figure 3.31 montre les résultats du recalage. Nous avons évalué le recalage manuellement, à l'aide d'une vingtaine de points d'amers et nous obtenons une erreur moyenne d'une dizaine de pixels après le recalage rigide, soit inférieure à 10 m.

Dans notre méthode, nous proposons d'améliorer le recalage rigide avec une transformation polynomiale. Les paramètres que nous avons déterminés, section 3.4.3, sont adaptés pour le couple d'image Quickbird/Ramses, mais pas pour le couple Quickbird/TerraSar. Aussi nous devons mener une autre étude sur les paramètres de quantification et la taille de fenêtre adaptée. Nous avons choisi de filtrer l'image radar à l'aide du filtre proposé par Denis [Denis *et al.*, 2009] et de quantifier l'image sur 10 niveaux. La fenêtre d'analyse utilisée est de 400x400 pixels. Finalement, nous obtenons une légère amélioration du recalage, nous passons d'une erreur moyenne de 9.7 à 9.5 pixels. Pour évaluer visuellement le résultat, nous avons superposé les images optiques et radar par morceaux, figure 3.32.

Afin de comparer ces résultats à un recalage s'appuyant ur les paramètres capteurs, nous avons utilisé les programmes disponibles dans l'Orfeo Toolbox¹ (OTB). Il s'agit d'une librairie C++ de traitement d'images, initié par le CNES dans laquelle se trouve des fonctions de projections de données satellitaires à partir de paramètres capteurs. Pour l'image optique, le modèle non-paramétrique est utilisé, basé sur le modèle RPC. Pour le cas radar, OTB fournit un modèle se basant sur les vrais paramètres de prise de vue stockés dans les méta-données. Nous utiliserons aussi un modèle numérique de terrain (MNT) pour fournir une information sur la hauteur ainsi que des points d'amer correspondants aux quatre coins de notre image initiale (figure 2.10). Le résultat de la projection de l'image optique sur l'image radar est représenté figure 3.33. Nous avons superposé des morceaux de l'image optique sur l'image radar.

Les résultats obtenus par la méthode décrite dans ce chapitre sont plus satisfaisants que ceux obtenus par la projection à l'aide des paramètres capteurs, sans autre point

http://smsc.cnes.fr/PLEIADES/Fr/A_prog_accomp.htm



(b)

FIG. 3.31 – *Images recalées du complexe scientifique de Rangueil :* (*a*) *image optique,* (*b*) *image radar.*



FIG. 3.32 – *Alternance des images optiques et radar après recalage par notre méthode. L'erreur moyenne de recalage est inférieur à une dizaine de pixels.*



FIG. 3.33 – Alternance des données optiques et radar, après projection à l'aide des paramètres capteurs et de la librairie de traitement d'images OTB. L'erreur de recalage au centre de l'image est de 126 pixels.

d'appui que ceux fournis avec les données. Les erreurs que nous observons lors de la projection des données peuvent venir de l'incertitude ou l'imprécision de certains paramètres, en particulier la hauteur, puisque le MNT utilisé à une précision de 90 m en planimétrie et une incertitude inférieure à 16 m² sur la hauteur.

Une perspective intéressante serait de combiner ces deux méthodes de recalage. La projection des données à l'aide des paramètres capteurs permet d'avoir une bonne initialisation pour la rotation et l'échelle, ce qui permet de restreindre l'espace de recherche de ces paramètres, même au niveau 1 de la pyramide multi-échelle. Alors qu'auparavant, nous initialisions manuellement l'échelle ainsi que la rotation entre [-90°,90°].

3.6 Conclusion

Dans ce chapitre, nous avons présenté une méthode de recalage en deux étapes, tout d'abord un recalage rigide (rotation, changement d'échelle et translation), puis un recalage polynomial d'ordre deux, basé sur des points d'amers détectés automatiquement par un filtre de Harris. Nous proposons une méthode hiérarchique utilisant l'information contenue dans l'image et l'invariant de Fourier-Mellin pour déterminer les paramètres de rotation et de zoom. Une fois ceux-ci corrigés, nous déterminons la translation globale grâce à l'information mutuelle. L'erreur moyenne de recalage, évaluée manuellement pour chaque zone, est rappelée dans le tableau 3.8.

	Erreur moyenne de recalage en pixels					
	CNES	Prison	Rangeuil			
Recalage Rigide	30	32	9.7			
Recalage Polynomiale	11	14	9.5			

TAB. 3.8 – *Erreur moyenne de recalage* exprimée en pixels, pour chaque méthode de recalage et chaque zone de test.

Afin d'affiner le recalage, nous avons cherché la déformation polynomiale globale de l'image, que nous avons calculée à partir de points d'amer. Nous avons sélectionné des points significatifs répartis sur toute l'image optique et nous avons déterminé leurs correspondants dans l'image radar à l'aide de l'information mutuelle. La transformation

²http://www2.jpl.nasa.gov/srtm/SRTM_paper.pdf

polynomiale s'obtient alors en minimisant l'erreur de recalage aux moindres carrés. Nous obtenons finalement une erreur moyenne d'une dizaine de pixels.

L'intérêt de cette méthode est qu'elle ne requiert pas de paramètres capteurs, c'est dans cette optique que nous l'avons utilisée : recaler des données dont on ne connaît que la résolution. Finalement, cette méthode permet d'obtenir des résultats satisfaisants. En effet, nous avons comparé nos résultats avec ceux obtenus par projection à l'aide des paramètres capteurs, sans prise de point d'appui manuelle et nous avons observé un meilleur recalage avec notre méthode. En effet, une petite incertitude sur les paramètres capteurs peut entraîner de grandes distorsions [Tupin et Galland, 2004]. C'est pourquoi il est avantageux de recourir à des méthodes qui se passent des paramètres et utilisent le contenu de l'image.

Le fait de recaler les images de manière globale induit des limites. En particulier, à cause de la géométrie des images radar, les objets en hauteur sont superposés à ceux du sol. Or dans les zones urbaines, on trouve beaucoup d'objets en hauteur qu'il s'agisse des bâtiments ou de la végétation. Il nous faut donc, pour améliorer ces résultats envisager de recaler différemment les objets appartenant au sol de ceux du sur-sol.

3.6. Conclusion

Chapitre 4

Classification

4.1 Problématique

Dans le chapitre précédent, nous avons mis au point une méthodologie pour le recalage des images optiques et radar à des résolutions métriques en milieu urbain. Le recalage obtenu nous donne une erreur moyenne d'une dizaine de pixels (soit environ 8 m) sur les différentes bases de données de test. L'une des problématiques du recalage d'images optiques et radar à haute résolution est la déformation géométrique des objets qui sont en hauteur dans la scène, comme par exemple des bâtiments [Chesnel, 2008]. La figure 2.7 illustre cela, les projections du sol et du bâtiment sur chacune des images ne correspondent pas. Une question se pose alors : *"recaler correctement l'image au sol ou recaler correctement le bâtiment ?"* Il devient donc essentiel de faire la différence entre les objets au sol et ceux en sur-sol. Nous avons choisi pour cela d'introduire une classification de la scène qui nous permettra de guider ensuite le recalage.

Nous voyons que la hauteur des objets nécessite de définir un recalage objet par objet. Étant donné le type de scène semi-urbaine, que nous étudions, nous proposons de classer chaque objet de l'image optique suivant cinq classes : bâtiment, route, végétation haute ou forêt, sol et ombre. Dans tout ce qui suit la classe ombre fera référence à l'ombre optique et non à l'ombre radar. Nous avons fait le choix de ces classes car une simple classification en sol et sur-sol peut s'avérer insuffisante. De manière intuitive, nous estimons que le recalage d'un bâtiment ne se fait pas de la même manière que le recalage d'une forêt, c'est pourquoi nous avons choisi de séparer ces classes présentes dans nos scènes. Cette décomposition nous permet également de mettre de côté l'ombre optique qui n'a pas de correspondance dans l'image radar et d'identifier les éléments du sol (route et sol) et ceux



FIG. 4.1 – Schéma de classification.

du sur-sol (forêt et bâtiment).

Nous présentons dans ce chapitre une chaîne de traitement permettant de classer les données. Tout d'abord, nous verrons différentes méthodes pour partitionner l'image optique en régions ou objets. Puis nous nous intéresserons à la classification de ces objets à partir d'informations extraites de l'image optique et de l'image radar. Enfin, nous étudierons la pertinence des différentes informations apportées par ces deux capteurs dans un objectif de classification et la robustesse de notre chaîne face à un changement de scène.

4.2 Contexte

sont intéressés à la fusion données Plusieurs travaux se de optique et radar [Brown et al., 1996], que ce soit pour extraire des objets parti-[Tupin et Roux, 2003] [Thiele et al., 2007], culiers faire une segmentation [Lombardo et al., 2003], de la détection de changement [Chini et al., 2009] ou une classification.

Certains travaux ont porté sur la fusion de différentes classifications obtenues séparément avec chaque images [Briem *et al.*, 2002], c'est ce que l'on appelle la prise de décision. Pour cela [Waske et Benediktsson, 2007] utilisent des machines à vecteurs de supports (ou Séparateur à Vaste Marges, SVM), d'autres approches sont basées sur la théorie de l'évidence de Dempster-Schafer [Le Hegarat-Mascle *et al.*, 1997]. Une autre famille de méthodes consiste à concaténer directement les signaux (les niveaux de gris des images) issus des différentes données en un vecteur à traiter. Ainsi, [Serpico et Roli, 1995] utilisent des données optiques et polarimétriques pour faire une classification de parcelles agricoles, tandis que [Benediktsson *et al.*, 1990] classe des zones montagneuses. D'autres méthodes fusionnent des primitives extraites à partir des différentes données, comme le NDVI (valeur combinatoire de la réponse du rouge et de l'infrarouge) [Liénou *et al.*, 2006], des primitives texturales ou statistiques [Solberg, 1999] [Bogdanov *et al.*, 2005]. En particulier, [Solberg et Jain, 1997] a montré l'interêt d'utiliser des caractéristiques statistiques (comme les log-moments) et texturales, dérivées de la matrice de co-occurence.

Dans ces travaux, nous souhaitons utiliser les données optique et radar conjointement, dans le but d'établir une classification des zones urbaines. L'imagerie optique à haute résolution permet de mieux distinguer les formes des objets d'une scène, cependant la présence d'ombres et la variation de l'illumination solaire peut rendre plus difficile une interprétation automatique. Les images radar, quant à elles, peuvent être acquises indépendamment de la position du soleil ou des conditions climatiques, mais le bruit et les déformations géométriques rendent leur interprétation plus difficile. Nous envisageons donc d'utiliser des informations extraites des différentes images pour établir une classification.

La classification automatique regroupe un ensemble de méthodes et théories d'analyse des données, en particulier concernant l'apprentissage qui peut être soit supervisé, soit non-supervisé. Dans la classification supervisée, le but est de prédire la classe en se basant sur un certain nombre de données en entrées. Le terme "supervisé" est justifié par la présence d'un expert qui doit étiqueter certaines des valeurs d'entrées. A partir de ces échantillons étiquetés, on peut concevoir un modèle qui est ensuite appliqué aux autres échantillons pour prédire leurs classes. Dans la classification non supervisée, on ne dispose que de données non étiquetées. Le but est donc de décrire la manière dont les données sont organisées pour en déterminer la classe. Plusieurs algorithmes sont utilisés pour estimer les paramètres d'un modèle à partir d'échantillons. Les échantillons sont une représentation de nos données d'entrées selon un vecteur de caractéristiques. Parmi ces algorithmes, on distingue entre autres, les réseaux de neurones, la méthode des k-plus-proches-voisins, les k-Moyennes, les arbres de décision [Breiman *et al.*, 1984], les méthodes bayésiennes et plus récemment les SVM [Boser *et al.*, 1992].

Plusieurs travaux se sont intéressés à comparer les SVM avec d'autres méthodes de classification [Melgani et Bruzzone, 2004] [Zammit *et al.*, 2007]. Et comme Melgani et Bruzzone l'ont conclu, les SVM s'avèrent être une alternative valide et efficace aux approches conventionnelles de reconnaissance des formes pour la classification de données de télédétection hyperspectrale. Nous avons choisi pour classer nos images d'utiliser une méthode avec apprentissage qui bien que plus contraignante puisqu'elle nécessite des données d'entrée étiquetées, donne généralement de meilleurs résultats.

4.3 Chaîne de traitement

4.3.1 Partition de l'image optique

Nous avons vu qu'une des difficultés du recalage d'images optiques et radar à haute résolution est la déformation géométrique des objets qui sont en hauteur dans la scène. Nous avons fait le choix de recaler l'image au sol, dès lors tous les objets au-dessus du sol induisent une distorsion. Pour la suite, nous ne souhaitons plus considérer l'image dans son ensemble mais comme une combinaison de différents objets que nous classerons. Il existe deux types d'approches pour la classification : celles basées sur le pixel et celles basées sur des objets ou régions préalablement détectés. Ces dernières considèrent l'image comme un ensemble de régions (généralement obtenues par un partitionnement de l'image) et classent ainsi chaque région plutôt que chaque pixel pris indépendamment ce qui a pour avantage de pouvoir tenir compte de leur agencement spatial.

Cette décomposition en objets est d'autant plus utile, qu'à haute résolution, les paramètres texturaux habituellement utilisés ne sont plus suffisants pour obtenir une bonne classification [Bellens *et al.*, 2008]. L'approche par objet utilisée ici, nous permet de rajouter des caractéristiques géométriques afin d'améliorer la classification. Bien sûr, pour déterminer ce que sont les *objets* de notre image, nous devons utiliser une partition de l'image en régions. Étant donné que nous recalons l'image optique sur l'image radar, nous avons choisi de segmenter l'image optique. Celle-ci est également plus facile à segmenter (pas de chatoiement, moins de réponses ponctuelles ou linéaires dues à des petits objets dihédriques). Nous pouvons ainsi classer et recaler indépendamment chaque objet de l'image optique sur l'image radar.

Plusieurs méthodes de segmentation sont présentées dans ce paragraphe. Les composantes connexes obtenues seront utilisées par la suite en tant qu'*objets* bien qu'elles ne correspondent pas toujours à des objets au sens sémantique.

4.3.1.1 Ligne de partage des eaux

En géologie, la ligne de partage des eaux désigne la limite géographique naturelle entre deux bassins versants : de chaque côté de cette ligne, les eaux de pluie s'écoulent dans des directions différentes. En traitement d'images, la ligne de partage des eaux est une méthode de segmentation qui utilise la description des images comme un relief géographique [Beucher et Lantuéjoul, 1979], le niveau de gris de chaque pixel étant assimilé à une altitude. De la même manière qu'en géologie, on définit alors la ligne de partage des eaux comme la crête limitant deux bassins versants. De l'eau coule sur le relief et tombe par gravité dans les bassins. La ligne de partage des eaux correspond aux lignes où deux lacs disjoints se rejoignent au cours de l'immersion.

En pratique, pour segmenter une image, on applique la ligne de partage des eaux sur une image de contours. L'inconvénient de cette méthode est qu'elle fait apparaître tous les minima locaux, c'est-à-dire beaucoup de petites régions et donc une sur-segmentation comme le montre la figure 4.2(c). Afin d'éviter cela, il est conseillé d'utiliser un filtre sur l'image de gradient. La figure 4.2(e) nous montre le résultat de la segmentation après un seuillage sur l'image du gradient, nous avons éliminé les petites régions créées par le bruit de l'image.



FIG. 4.2 – Segmentation d'une courbe par la ligne de partage des eaux : (a) image originale, (b) gradient, (c) contour de l'image segmentée, (d) image de gradient seuillée et (e) contour des segments obtenus sur l'image de gradient seuillée.

4.3.1.2 Segmentation de Guigues

[Guigues, 2004] propose une approche multi-échelle de la segmentation. Considérant une distance D sur les couples (image, partition), il s'agit de trouver la partition P qui

minimise D(I, P) où I est notre image. Bien sûr le meilleur modèle au sens de cette distance, par exemple la somme des erreurs quadratiques (norme L_2), sera l'image elle même divisée en autant de régions qu'elle possède de pixels.

Guigues introduit alors une énergie C(P) sur la partition et un réel positif λ . Trouver la bonne partition P revient à résoudre un problème sous contrainte :

Trouver P qui minimise D(I,P) sous la contrainte C(P)
$$\leq \lambda$$
.

L'énergie C peut être vue comme une complexité, on peut par exemple prendre pour C le nombre de régions de la partition ou la longueur de ses frontières. Le paramètre λ possède alors le sens d'une échelle. Quand $\lambda = 0$, l'image est partitionnée en autant de régions qu'il y a de pixels, autrement dit chaque pixel est une région de l'image. A l'inverse quand λ est très grand l'image sera découpée en une unique région. La Figure 4.3 montre les segmentations obtenues pour différentes valeurs de λ . Les sections sont réalisées selon une suite géométrique de raison 1/2 par rapport à la valeur de λ , notée $\lambda^+(I)$, qui correspond à un découpage de l'image en une seule région.

4.3.1.3 Méthode de Luo

Les travaux de [Luo, 2007] utilisent la *fast level set transform* (FLST) [Monasse, 2000]. C'est une transformation bijective permettant de passer d'une image numérique à l'ensemble de ses lignes de niveaux. Nous montrons les résultats de la FLST sur une image synthétique Figure 4.4.

Luo utilise une fonction de coût afin de cumuler les contrastes. Dans une image numérique, les contours sont toujours lissés en raison des différents filtrages dus à la formation de l'image (fonction de transfert du capteur, interpolation...). L'image étant quantifiée, ceci se manifeste sous la forme de paquets de formes emboîtées ayant une différence de niveaux de gris très faible. Or cette notion de forme n'est satisfaisante que si l'on suppose qu'à chaque objet n'est attribué qu'une unique forme. L'idée principale de [Luo, 2007] est de cumuler les contours des formes suffisamment proches qui appartiennent au même objet. Ainsi pour un pixel x de l'image, notons $f_i(x)$ l'ensemble des formes contenant x, telles que $f_i(x) \subset f_{i+1}(x)$, $S(f_i)$ sa surface et $P(f_i)$ son périmètre. Les contours seront cumulés si :

$$S(f_{i+1}) - S(f_i) < \lambda P(f_i)$$



(a) Image d'origine



(g) $\lambda = \lambda^+(I)/2^6$



(b) $\lambda = \lambda^+(I)/2$



(e) $\lambda = \lambda^+(I)/2^4$



(c) $\lambda = \lambda^+(I)/2^2$



(f) $\lambda = \lambda^+(I)/2^5$



(i) $\lambda = \lambda^+(I)/2^8$

FIG. 4.3 – Segmentations de parcelles cultivées (Exemple extrait de [Guigues et al., 2006]).

(h) $\lambda = \lambda^+(I)/2^7$

4.3. Chaîne de traitement



FIG. 4.4 – *Exemple de FLST*: (a) image synthétique, (b) arbre d'inclusion obtenu par la FLST. (Exemple extrait de la thèse de [Luo, 2007])

avec λ une constante que nous avons prise égale à 1. Cette segmentation a pour avantage d'être adaptée aux images satellites optiques dont les contours d'objets apparaissent souvent flous et mal définis. De plus, cette méthode ne nécessite pas de paramètres en entrée, contrairement à la segmentation par ligne de partage des eaux. La Figure 4.5 illustre la segmentation obtenue sur l'image du CNES, les niveaux de gris associés à chaque segments sont proportionnels à leurs périmètres.

4.3.2 Extraction des primitives utiles à la classification

Après la segmentation vient l'étape d'extraction d'attributs. Pour chaque segment, nous extrayons plusieurs attributs qui vont le caractériser et être utilisés pour la classification.

4.3.2.1 Paramètres extraits de l'image optique

Les paramètres que nous avons utilisés sont des paramètres texturaux et géométriques. Les paramètres géométriques sont extraits à partir de la forme des objets : leur surface, leur périmètre et un paramètre d'échelle défini dans [Luo, 2007] correspondant à la surface sur le périmètre. Les paramètres texturaux sont la moyenne et la variance de chaque objet ainsi que la moyenne et la variance après avoir filtré les images par des filtres de Gabor [Manjunath et Ma, 1996].

Pour extraire la variance de chaque objet, nous avons tout d'abord calculé la variance de l'image initiale en tous points sur une fenêtre de taille 11x11. Ensuite pour chaque



FIG. 4.5 – *Segmentation de Luo.* Les niveaux de gris associés à chaque forme sont fonction de leurs périmètres.

objet nous avons gardé la variance minimale et la variance maximale. Il ne s'agit donc pas à proprement parler de la variance de l'objet mais d'une variance associée à l'objet.

En général, les paramètres texturaux sont extraits à l'aide de bancs de filtre ou d'ondelettes. Nous avons choisi d'utiliser les ondelettes de Gabor qui sont d'ailleurs utilisés comme descripteurs de textures dans la norme MPEG7. On peut les considérer comme un ensemble de canaux de filtrage dont la fonction mère est la suivante :

$$h(x, y; u) = exp\left(-\frac{1}{2}\left[\frac{x^2}{\sigma_x^2} + \frac{y^2}{\sigma_y^2}\right]\right)\cos(2\pi ux)$$

où *u* représente la fréquence centrale du filtre et σ_x et σ_y les largeurs de bande. On définit à partir de là d'autres fonctions g_{mn} qui dérivent de celle-ci à l'aide de *n* rotations et *m* changements d'échelle :

$$g_{mn}(x,y) = a^{-m}h(x',y')$$
$$a > 1$$
$$x' = a^{-m}(x\cos\theta + y\sin\theta)$$
$$y' = a^{-m}(-x\sin\theta + y\cos\theta)$$

où $\theta = k\pi/n$, k = 0, 1, ..., n - 1. Le terme a^{-m} est le facteur de la m-ième échelle. Le spectre de ces filtres pour m = 3 dimensions et n = 4 orientations est montré sur la Figure 4.6.

4.3.3 Paramètres extraits de l'image radar

Nous souhaitons dans ce paragraphe étudier l'apport des informations contenues dans l'image radar pour la classification de l'image optique. Sur la partition de l'image optique recalée sur l'image radar, avec le recalage obtenu précédemment (Figure 4.5), nous allons extraire des paramètres statistiques de l'image radar : les trois premiers log-cumulants.

[Nicolas, 2003] montre que les images radars peuvent être très bien modélisées par une loi de Fisher. Cette distribution, qui entre dans la classe des "lois à queue lourde", est un outil statistique peu utilisé jusqu'à présent car les méthodes traditionnelles d'estimation des paramètres sont souvent inadaptées. Les "statistiques de deuxième espèce" (ou Log-statistiques) permettent en revanche d'en estimer aisément les paramètres. Aussi son utilisation pour l'analyse des images de Radar à Synthèse d'Ouverture est perfor-



FIG. 4.6 – *Filtres de Gabor dans le domaine fréquentiel : en ligne les 4 orientations et en colonne les 3 dimensions.*

mante, tant sur le plan de la modélisation que sur le plan du traitement. [Tison *et al.*, 2004] montre l'efficacité des log-cumulants pour la classification d'images SAR de zones urbaines à haute résolution.

Les relations entre les statistiques usuelles et la transformée de Fourier sont bien connues : la fonction caractéristique d'une fonction f est la transformée de Fourier de sa densité de probabilité p_x . Le moment d'ordre n est la dérivée d'ordre n de cette fonction caractéristique et les cumulants sont les dérivées $n^{ième}$ de la fonction caractéristique du logarithme de f.

On définit les statistiques de seconde espèce de manière similaire [Nicolas, 2003] [Tison, 2004], en introduisant la transformée de Mellin d'une fonction f, définie sur \mathbb{R}^+ :

$$TM[f](s) = \int_0^\infty u^{s-1} f(u) du$$

Par analogie, on définit la première fonction caractéristique de seconde espèce par :

$$\phi_x(s) = TM[p_x] = \int_0^\infty u^{s-1} p_x(u) du$$

où $p_x(u)$ est une densité de probabilité.

Les moments de deuxième espèce d'ordre n ou log-moments m_n , sont donnés par la relation suivante :

$$m_n = \left. \frac{d^n \phi_x(s)}{ds^n} \right|_{s=1} = \int_0^\infty (\log u)^n p_x(u) du$$

Les cumulants de deuxième espèce d'ordre n, ou log-cumulants k_n , sont donnés par :

$$k_n = \left. \frac{d^n (\log \phi_x(s))}{ds^n} \right|_{s=1}$$

Après calcul, on obtient :

$$k_1 = m_1$$

$$k_2 = m_2 - m_1^2$$

$$k_3 = m_3 - 3m_1m_2 + 2m_1^3$$

Finalement les paramètres extraits des images radar choisis sont les trois premiers logcumulants k_1, k_2 et k_3 . La segmentation de l'image optique est projetée sur l'image radar et pour chaque région, les trois premiers log-cumulants sont calculés sur tous les pixels de la région.

4.3.4 Classificateurs

Dans ce paragraphe, nous présentons plusieurs méthodes de classifications appliquées aux images satellitaires.

4.3.4.1 Classification à l'aide d'un Modèle Numérique d'Elevation

Lorsque l'on dispose d'un modèle numérique d'élévation 3D (MNE), il est aisé de faire une classification sol/sur-sol. [Cord *et al.*, 2002] calcule, à partir de deux images optiques de 10 cm de résolution, un Modèle Numérique d'Elevation. Il évalue ensuite un modèle numérique de terrain (MNT) en partant de l'hypothèse que le MNE est en fait un MNT perturbé par des points aberrants : le sur-sol. Il utilise donc des méthodes d'estimation robuste pour les éliminer. A partir de là il obtient une classification sol/sur-sol.

[Baillard et Maitre, 1999] utilisent également un MNE calculé à partir de deux images optiques (résolution de 40 cm) parfaitement recalées. Ils utilisent ensuite un modèle markovien sur les informations du MNE pour classifier en sol/sur-sol. Une nouvelle classification est faite sur le sur-sol en bâtiment/végétation à partir d'un seuillage de l'entropie des directions du gradient.

Le même principe est utilisé par Sohn et Dowman [Sohn et Dowman, 2007] avec des images LiDAR. Ils calculent un MNT à partir du MNE obtenu par l'image, en utilisant un filtre RTF (Recursive Terrain Fragmentation) détaillé dans [Sohn et Dowman, 2002]. Ils en déduisent une classification sol/sur-sol, puis ils font une classification bâtiment/végétation à l'aide d'un seuillage sur le NDVI.

4.3.4.2 Méthodes avec apprentissage

4.3.4.2.1 Méthode Bayesienne

[Liénou *et al.*, 2006] classe des images optiques en effectuan une classification Bayésienne sur des critères spectraux (NDVI, IB, ISU, QMF) ainsi qu'un apprentissage sur la base de données CORINE.

[Amberg, 2005] fait une classification bayésienne des images radar haute résolution (métrique) : pré-traitement pour enlever le speckle (filtre de Lopes), apprentissage avec des lois de Fisher, classification au sens du maximum de vraisemblance, post-traitement pour linéariser les contours par règle majoritaire.

[Tison, 2004] utilise l'approche bayésienne globale pour la classification de données radar submétriques. Le résultat de la classification est ensuite fusionné avec une information de cohérence et de phase interférométrique afin de réduire certaines confusions. Cette classification est ensuite utilisée dans un processus de reconstruction 3D par interférométrie.

4.3.4.2.2 Réseaux de neurones

Benediktsson [Benediktsson *et al.*, 2003] classe les images optiques de résolutions métriques. Tout d'abord il crée une pyramide morphologique en effectuant des opérations d'ouverture et de fermeture. Il compare ensuite des primitives pour chaque pixel (DAFE, DBFE, taille des regions) qui permettent une bonne classification par réseaux de neurones.

Dell'Acqua [Dell'Acqua et Gamba, 2003] utilise des réseaux de neurones pour classer des images SAR à une résolution décamétrique. Il se base sur des paramètres texturaux (contraste, corrélation, entropie, moyenne, variance, second moment, matrice de cooccurence...).

4.3.4.2.3 Compression

Cerra [Cerra et Datcu, 2008] propose une méthode de classification par compression. Un dictionnaire de chaque classe est obtenu en compressant une base de données. Chaque image sera compressée avec un dictionnaire associé à une classe, l'image appartenant à la classe du dictionnaire qui a donné la compression la plus forte.

4.3.4.2.4 Support Vector Machines

Les SVM (ou Séparateurs à Vastes Marges en français) sont des classificateurs qui permettent de traiter des problèmes de discrimination non-linéaire, et de reformuler le problème de classement comme un problème d'optimisation quadratique. Ils ont été proposés par Boser, Guyon et Vapnik en 1992 [Boser *et al.*, 1992].

La première idée clé est la notion de marge maximale. La marge est la distance entre la frontière de séparation et les échantillons les plus proches. Dans les SVM, la frontière de séparation est choisie comme celle qui maximise la marge. La problématique est de trouver cette frontière séparatrice, à partir d'un ensemble d'apprentissage. Ceci est fait en formulant le problème comme un problème d'optimisation quadratique, pour lequel il existe des algorithmes connus.

Afin de pouvoir traiter des cas où les données ne sont pas linéairement séparables, la deuxième idée clé des SVM est de transformer l'espace de représentation des données d'entrées en un espace de plus grande dimension, dans lequel il est probable qu'il existe une séparatrice linéaire. Ceci est réalisé grâce à une fonction noyau.

L'INRIA [Lafarge *et al.*, 2004] propose une classification des zones urbaines/nonurbaines d'images optiques de 5 m de résolution par SVM. Les paramètres texturaux utilisés sont les variances conditionnelles associées à une direction.

[Waske et Benediktsson, 2007] fusionne des images optiques et radar de 30 m de résolution. Pour cela, il effectue au préalable une classification de chacune des images à l'aide de SVM, puis il compare différentes méthodes pour fusionner les images : SVM, MLC, DT et DT-based.

4.3.4.2.5 Fisher

L'analyse discriminatoire linéaire de Fisher [Hastie *et al.*, 2005] n'est pas seulement un classificateur à proprement parler, mais peut aussi être utilisée pour la sélection de primitives. Pour cela, les données sont projetées sur un hyperplan, comme illustré sur la figure 4.7 dans le cas de deux classes. L'objectif de la discrimination linéaire de Fisher est de trouver une droite telle que la projection des données sur celle-ci les sépare au mieux, c'est-à-dire en maximisant le rapport des variances des projections inter et intra classes. On peut alors classer les données en définissant un seuil sur cette droite : si un point est inférieur au seuil, il appartient à une classe sinon il appartient à l'autre classe. Le schéma de gauche montre un mauvais classificateur, les données projetées se chevauchent, il n'est pas possible de définir un seuil qui permette de séparer les classes. Le schéma de droite montre le bon discriminant de Fisher qui permet de **maximiser la distance** entre les barycentres de chaque classe et de **minimiser la compacité** de chaque classe.



FIG. 4.7 – *Exemple de la discrimination linéaire de Fisher :* à gauche la projection des échantillons sur la droite (\mathfrak{D}) ne permet pas de séparer les classes (cercles et croix), à droite les classes sont séparées.

L'analyse de Fisher est également utilisée pour la sélection de primitives. Le coefficient directeur du discriminant de Fisher donne des poids sur chaque primitive. Sur la figure 4.7, le discriminant de Fisher peut s'écrire de la forme (α, β) dans la base (x_1, x_2) avec $\beta \gg \alpha$. Les valeurs α et β peuvent être considérées comme des poids associés à chaque attribut. Ceci nous permet de ranger les primitives suivant leurs pertinences.

4.3.4.2.6 K-Plus-Proches-Voisins

L'algorithme des k-plus-proches-voisins (kppv) est un algorithme de la famille des algorithmes dits "paresseux" : à l'inverse de beaucoup d'autres méthodes d'apprentissage automatique aucun apprentissage ne prend réellement place, c'est-à-dire qu'il n'y a pas de phase de détermination de paramètres d'une fonction par le biais d'une optimisation mathématique. Le principe est le suivant : étant donnée une base d'apprentissage étiquetée et un entier k, le classificateur k-ppv détermine la classe d'un nouvel objet en lui attribuant la classe majoritaire des k objets lui ressemblant le plus dans la base d'apprentissage. La notion de voisinage est caractérisée par une mesure de distance ou de similarité (norme L1, distance Euclidienne...).

4.3.4.3 Méthodes non supervisées

[Viveros-Cancino, 2003] choisit dans sa thèse de faire une classification des zones urbaines d'une image optique en utilisant les K-moyennes.

[Oller *et al.*, 2006] fait une classification zone urbaine/non-urbaine d'images optiques hautes résolution, en faisant un seuillage sur des paramètres texturaux (variance dans plusieurs directions).

[Luo, 2007] propose une partition des images optiques haute résolution en fonction de l'échelle des objets présents dans l'image. Il définit l'échelle comme étant la surface sur le périmètre d'un objet.

4.3.4.4 Conclusion

Nous ne disposons pas de modèle numérique de terrain, ni d'images interférométriques ou stéréoscopiques qui permettraient d'en déduire un, nous n'utiliserons donc pas de modèle numérique d'élévation pour faire une classification sol/sur-sol.

Nous avons choisi d'utiliser une méthode de classification avec apprentissage qui donne généralement de meilleurs résultats qu'une méthode sans apprentissage. Parmi celles-ci, les SVM sont une méthode récente et efficace de classification. Waske [Waske et Benediktsson, 2007] montre que les SVM permettent d'obtenir des résultats de qualité identique à, ou meilleure, que d'autres méthodes, ce que nous montrerons dans le paragraphe 4.4.3. De plus, ils ne nécessitent pas d'a priori sur les distributions, comme les méthodes Bayésiennes, mais nécessitent le choix d'un noyau.

4.3.5 Evaluation

La qualité de la classification sera évaluée de plusieurs manières :

- une évaluation visuelle des résultats,
- les performances en prédiction de modèle grâce à la validation croisée,
- les performances en test par rapport au nombre d'objets, c'est-à-dire le pourcentage d'objets bien classés,
- les performances en test par rapport à la surface, c'est-à-dire le pourcentage de pixels bien classés.

4.3.5.1 Validation croisée

Concernant les méthodes d'évaluation des modèles, nous utilisons la validation croisée. La validation croisée est l'une des méthodes les plus utilisées pour prédire les performances d'un modèle [Kohavi, 1995]. Le principe est le suivant : l'ensemble des données est divisé en deux, une partie pour l'apprentissage et une autre pour le test. En général, l'ensemble des données est divisé aléatoirement en K parties égales. Ensuite, de manière itérative, K - 1 sous-ensembles sont utilisés pour l'apprentissage du modèle et l'erreur de prédiction du modèle est calculée sur le sous-ensemble restant. On peut ainsi prédire l'erreur moyenne du modèle.

La validation croisée est appliquée sur un ensemble *train* de 50 points par classe pris aléatoirement dans notre base d'objets étiquetés. Pour K = 5 itérations, on apprend sur 200 points et on teste sur les 50 restants, soit $1/5^{\text{ème}}$ de la base d'apprentissage. Les performances en prédiction sont alors obtenues en calculant la moyenne et la variance des performances de chaque itération.

4.3.5.2 Vérité terrain

Dans le but d'estimer le modèle, il faut au préalable avoir une base d'apprentissage dont on connaît la vérité terrain. Dans le cas des images satellitaires que nous utilisons, nous disposons d'une vérité terrain partielle représentée Figure 4.8. Cette vérité terrain a été faite pixel à pixel manuellement. Bien sûr, il faut alors faire attention au choix des objets à étiqueter. Nous avons fait en sorte que les objets étiquetés soient les plus représentatifs possibles de chaque classe dans leurs natures et leurs tailles. Par exemple, la classe végétation haute est représentée par des zones de forêts et des arbres individuels, la classe sol est représentée par des grandes zones de végétation comme des parcs ou

label	bâtiment	route	forêt	sol	ombre	non-classé	classe
objet 3	21	51	79	59326	0	7817	sol
objet 13	2433	0	72	0	9	17412	non-classé
objet 26	0	9849	0	296	6	2580	route
objet 148	1227	0	0	1046	0	115	bâtiment

TAB. 4.1 – **Exemple de vote majoritaire.** Le tableau est un exemple de nos résultats. Pour chaque objet $(1^{ere} \text{ colonne})$, il présente le nombre de pixels par classe donné par la vérité terrain. La colonne la plus à droite indique le résultat du vote majoritaire, ce qui corrrespond à la classe la plus représentée en nombre de pixels.

des champs, mais aussi par des petits ronds-points et des parkings. La zone n'a pas été entièrement étiquetée à cause de la petite taille et l'incertitude sur certains objets, par exemple, dans quelle classe mettre le canal ?

La vérité terrain dont nous disposons est pixellique, il nous faut donc la projeter sur nos images segmentées de manière à avoir une classe pour chaque objet. En effet, il est possible qu'une région segmentée avec une des méthodes décrites ci-dessus contienne plusieurs classes différentes. Afin de régulariser cela, nous avons procédé à un vote majoritaire. Pour chaque objet segmenté, on lui attribue la classe majoritairement représenté par la vérité terrain (Tableau 4.1). En outre, un objet dont majoritairement les pixels n'ont pas de vérité terrain sera considéré comme non classé. Ce type de projection dépend énormément de la qualité de la segmentation. Dans le cas idéal, chaque segment ne contiendra qu'un unique label issu de la vérité terrain. Or nous voyons dans le tableau 4.1 que ce n'est pas toujours le cas. Le choix d'attribuer la classe majoritaire est évident pour certains objets comme l'objet 3 ou 26, seulement nous voyons l'ambiguïté pour d'autres, notamment l'objet 148 dont 51% de sa surface est labelisée en bâtiment contre 44% en sol. Nous pouvons effectivement nous interroger sur la validité de cette projection qui servira ensuite de vérité terrain associée à la segmentation. La figure 4.9 montre la projection de la vérité terrain sur différentes segmentations de la scène. L'influence de la segmentation et sa projection sera discutée dans la section 4.4.4.



FIG. 4.8 – Vérité terrain superposée à l'image satellitaire de la zone du CNES de Toulouse : les bâtiments sont représentés en bleu, les routes en beige, la végétation haute en vert, la végétation basse en rouge et les ombres en rose.



FIG. 4.9 – **Projection de la vérité terrain sur différentes segmentations :** (a) ligne de partage des eaux, (b) méthode de Guigues et (c) méthode de Luo. Les bâtiments sont représentés en bleu, les routes en beige, la végétation en vert, le sol en rouge et les ombres en rose.

nombre	PRIMITIVES OPTIQUES					
3	géometriques	- périmètre				
		- surface				
		- compacité				
3	texturales	- moyenne				
		- minimum et maximum de la variance calculée sur une				
		fenêtre 11x11				
24		- moyenne et variance calculée sur les objets filtrés par les				
		ondelettes de Gabor sur 4 orientations et 3 dimensions				
nombre	PRIMITIVES RADAR					
3	statistiques	trois premiers log-cumulants				

TAB. 4.2 – Primitives extraites des images optiques et radar pour caractériser chaque objet.

4.4 Expérimentations

4.4.1 Pertinence des primitives

Après la procédure d'extraction des primitives, chaque objet est désormais représenté par un vecteur de 33 caractéristiques, recensées dans le tableau 4.2. Il est important de rappeler que nous avons calculé les log-cumulants sur les objets extraits de l'image optique recalée. Cependant les erreurs résiduelles dans le recalage, notamment pour les objets du sur-sol comme les bâtiments, vont se propager lorsque nous estimons les paramètres texturaux de ces objets dans l'image radar.

Une fois les primitives extraites, nous les avons normalisées de manière à avoir une moyenne nulle et une variance de un.

Nous souhaitons mettre en évidence l'intérêt de l'imagerie radar pour la classification. Pour cela, nous étudions l'influence de chaque type de primitive pour chaque classe (bâtiment, route, forêt, sol et ombre). Nous utilisons la discrimination linéaire de Fisher présentée dans le paragraphe 4.3.4.2.5 pour étudier les primitives les plus pertinentes, c'est à dire celles qui permettent au mieux de séparer les classes. Pour chaque classe les primitives les plus pertinentes sont données dans le tableau 4.3.

Nous constatons que les paramètres radar n'interviennent pas pour améliorer la classification de l'ombre. Cela semble cohérent, puisque nous ne nous attendons pas à ce qu'il y ait une signification physique à l'ombre optique dans l'image radar (et réciproquement).

Comme attendu, les caractéristiques texturales sont très utiles pour la discrimina-

bâtiment	paramètres statistiques radar, moyenne et variance ex-
	traites du filtrage de Gabor à l'échelle 2, moyenne extraite
	du filtrage de Gabor à l'échelle 0
route	paramètres statistiques radar, compacité, Gabor échelle
	0 : variance, Gabor échelle 1 et 2 : moyenne et variance
forêt	paramètres statistiques radar, compacité, variance de
	l'image, Gabor échelle 0 : variance, Gabor échelle 1 : va-
	riance, Gabor échelle 2 : moyenne
sol	moyenne de l'image, Gabor échelle 0 : moyenne, Gabor
	échelle 1 : moyenne et variance
ombre	moyenne et variance de l'image, compacité, Gabor échelle
	0 : moyenne, Gabor échelle 1 : moyenne et variance

TAB. 4.3 – Primitives les plus discriminantes de chaque classe.

tion de chaque classe. Il est intéressant de voir que l'image radar est utile à la discrimination des bâtiments et des forêts, c'est à dire aux "classes hors-sol". Les primitives géométriques ont également leur importance, en particulier pour discriminer les routes, les ombres et les forêts.

4.4.2 Nombre de primitives

Chaque type de primitives extraites, que ce soit géométriques, radar ou texturales, a donc son importance dans la classification. La figure 4.10 nous montre le résultat de la classification en fonction du nombre de primitives pour différents classificateurs, ce graphe est obtenu à l'aide de la boîte à outils spider¹. Nous observons sur cette figure que 10 primitives sont suffisantes pour avoir un résultat satisfaisant en terme de pourcentage de bonne classification, pourtant dans la suite nous faisons le choix de garder les 33 primitives proposées. En effet, nous voyons dans le tableau 4.3 que chacune à son importance. Les primitives radar permettent de discriminer les classe bâtiments, route et forêt ; les primitives géométriques sont utiles aux classes forêt et ombre. Parmi les primitives texturales utilisant les filtres de Gabor, certaines sont plus pertinentes que d'autres, en particulier certaines orientations ; seulement nous ne souhaitons pas faire d'a priori sur l'orientation de nos images. Dans le but de proposer une méthode générale, nous préférons garder toutes les orientations des filtres de Gabor et donc toutes les primitives.

http://www.kyb.mpg.de/bs/people/spider/

Chaque objet est désormais représenté par les 33 primitives décrites dans le tableau 4.2, donc par un vecteur de taille 33.



FIG. 4.10 – Performances de la classification prédites en validation croisée en fonction du nombre de primitives sélectionnées par Fisher (implantation de spider).

4.4.3 Choix du classificateur

Dans le paragraphe 4.3.4.2, nous avons présenté plusieurs méthodes de classification. Nous avons testé les suivantes pour la classification d'image optique à l'aide des primitives proposées ci-dessus :

- k plus proches voisins (kppv) avec k=3,
- classificateur linéaire de Fisher,
- SVM à noyau linéaire,
- SVM à noyau Gaussien dont le paramètre est estimé en validation croisée sur la base d'apprentissage.

Pour chacune de ces méthodes, les résultats sont donnés dans le tableau 4.4, sous la forme d'un pourcentage de bonne classification calculées en validation croisée. La performance moyenne ainsi que la variance sont calculées sur plusieurs itérations. Ce

tableau montre que les SVM donnent des résultats bien plus élevés que la méthode des k plus proches voisins ou la méthode de Fisher.

kppv	Fisher	SVM Linéaire	SVM Gaussien		
$59.6 \% \pm 4.77$	$59.2~\% \pm 4.82$	$76.4~\% \pm 8.88$	$76.8~\% \pm 5.22$		

TAB. 4.4 – Résultats des différents classificateurs en validation croisée sur la base d'apprentissage.

Ici, les 33 primitives ont été utilisées. La figure 4.10 montre les pourcentages de bonne classification de chaque méthode calculées en validation croisée, en fonction du nombre de primitives. Assymptotiquement les SVM donnent de meilleurs résultats que les méthodes kppv ou Fisher. Pour un grand nombre de primitives, le choix du noyau des SVM semble ne pas avoir d'importance puisque les résultats sont similaires, mais dans le cas de peu de primitives (moins de quinze) le SVM à noyau Gaussien converge plus rapidement vers son assymptote. Nous avons donc fait le choix pour la suite d'utiliser comme classificateur les **SVM avec un noyau Gaussien**. Les paramètres du SVM sont estimés en validation croisée sur la base d'apprentissage.

4.4.4 Robustesse à la segmentation

Dans cette partie, nous souhaitons étudier l'influence de la segmentation sur la classification. Nous avons utilisé en apprentissage, pour le classificateur SVM, 250 objets, soit 50 de chaque classe, pris aléatoirement dans la vérité terrain, ce qui correspond à ce que ferait un utilisateur qui indiquerait des points pour initialiser la classification. Plusieurs tirages de la base d'apprentissage ont été effectués afin d'obtenir une performance moyenne, indépendante des objets d'apprentissage donnés par un utilisateur.

Nous avons alors estimé la performance de la classification sur l'ensemble d'apprentissage et la performance en test exprimée en nombre d'objets (c'est-à-dire en se référant aux segments étiquetés, soit 828 segments) et en nombre de pixels bien classés (c'est-à-dire en se référant à la vérité terrain manuelle). Les performances en nombre d'objets correspondent au nombre d'objets bien classés, indépendamment de leur taille ; cependant un objet ayant une grande surface (jusqu'à 10 % de l'image dans nos zones de tests) n'a pas la même importance qu'un objet mal classé ayant une petite surface (de quelques pixels). Nous calculons donc également les performances en nombre de pixels, ce qui correspond à la surface bien classée de la scène. Ceci est effectué pour chacune des segmentations présentées précédemment : la ligne de partage des eaux, la segmentation proposée par Guigues et celle proposée par Luo.

Le tableau 4.5 nous donne les résultats de la classification obtenus en validation croisée (XVal) et en test. La classification est faite avec un SVM Gaussien sur 50 objets d'apprentissage pour chaque classe, les performances sont données pour plusieurs bases d'apprentissage.

	partage des eaux			Guigues			Luo		
tirage	XVal	objets	pixels	XVal	objets	pixels	XVal	objets	pixels
1	76.8	78.7	71.53	77.6	82.84	84.33	71.6	79.93	79.1
2	77.6	77.47	70.89	77.6	83.2	84.06	76.4	83.08	84.86
3	72.8	81.45	79.87	69.6	82.45	82	74.6	84.3	85.91
4	74.4	76.87	75.2	79.6	84.05	81.49	77.6	81.29	83.41
5	69.2	79.26	72.3	78.8	82.72	83.2	73.2	82.19	83.29
moyenne	74.16	78.75	73.96	76.64	83.05	83.02	74.68	82.16	83.31

TAB. 4.5 – *Résultats de la classification pour différentes segmentations.* Pour chacune on donne le résultat obtenu en validation croisée (XVal) sur la base d'apprentissage et les résultats en test. Le résultat en pixels correspond au pourcentage de pixels bien classés par rapport à la vérité terrain dont nous disposons, le résultat en objets correspond au pourcentage d'objets bien classés par rapport à la vérité terrain gropet à la vérité terrain projetée sur la segmentation.

En validation croisée, les résultats sont équivalents, autour de 75 % de bonne classification. Ceci montre l'**indépendance de la segmentation** vis-à-vis de la méthode de classification utilisée. Nous avons également testé cette méthode sur l'ensemble de l'image, puis nous avons évalué le résultat en comptant le nombre de pixels bien classés par rapport à la vérité terrain (Figure 4.8) et en comptant le nombre d'objets bien classés par rapport à la vérité terrain projetée sur chaque segmentation (Figure 4.9). Nous voyons que les performances en tests diffèrent un peu suivant la segmentation utilisée.

Tout d'abord nous pouvons remarquer que les résultats de classification sont à peu près équivalents d'une segmentation à une autre. Cependant, lorsqu'on comptabilise les résultats en fonction de la surface bien classée, on remarque que la segmentation obtenue par partage des eaux est nettement inférieure aux autres. Afin de comprendre cette différence, nous nous sommes intéressés à la projection de la vérité terrain sur les segmentations. En effet, dans un cas nous comparons une classification par rapport à une

4.4. Expérimentations

vérité terrain projetée et nous voyons que les résultats sont du même ordre et dans l'autre cas nous comparons la classification avec la vérité terrain pixellique, c'est alors que des différences interviennent. La figure 4.11 nous montre les projections d'une scène par vote majoritaire sur chacune des segmentations. Certains objets, comme des routes ou des arbres, ont disparu de la figure 4.11 (a) lors de la projection. Ceci explique pourquoi la segmentation par partage des eaux donne de moins bons résultats rapportés aux nombre de pixels.



FIG. 4.11 – *Exemple de projection de la vérité terrain pixellique sur différentes segmentation par vote majoritaire :* (a) vérité terrain pixellique, (b) ligne de partage des eaux, (c) méthode de Guigues et (d) méthode de Luo.

Afin d'évaluer au mieux l'influence de la segmentation dans notre chaîne de traitement, une étude quantitative n'est pas suffisante. Il faut également évaluer visuellement les résultats. Aussi la figure 4.12 montre les résultats de la classification obtenue pour chacune des segmentations superposées à l'image optique originale. La classification a été faite de la manière décrite dans le paragraphe précédent, c'est-à-dire avec un SVM à noyau Gaussien sur 50 points d'apprentissage pris aléatoirement dans l'image. Chaque segmentation donne 33307 objets pour la ligne de partage des eaux, 231975 pour la segmentation de Guigues et 168539 pour la segmentation de Luo, aussi nous n'avons classé qu'une partie de ces objets. Les régions non labellisées de l'image (en noir) sont les régions qui n'ont pas été classées manuellement.

Si on regarde attentivement la figure 4.12, il est très difficile de privilégier un résultat plutôt qu'un autre. Visuellement, ces résultats sont tous satisfaisants lorsque nous les comparons à la vérité terrain associée à chaque segmentation (Figure 4.9). Rappelons que les vérités terrain sont obtenues à partir de l'originale (Figure 4.8) en faisant un vote majoritaire pour chaque objet. La classe d'un objet est donc dépendante de sa segmentation ; une zone mal segmentée qui contient plusieurs classes se verra attribué le label majoritairement représenté. Les vérités terrain associées à chaque segmentation sont représentées



FIG. 4.12 – *Résultat de la classification par un SVM Gaussien pour différentes segmentations obtenues par :* (*a*) la ligne de partage des eaux, (*b*) la méthode de Guigues et (*c*) la méthode de Luo. Les bâtiments sont représentés en bleu, les routes en beige, la végétation en vert, le sol en rouge et les ombres en rose.



FIG. 4.13 – Évaluation de la classification pour chacune des cinq classes et des trois segmentations. Les notes vont de 1 à 5, 5 étant la meilleure note.

figure 4.9. Ces cartes montrent l'importance de la segmentation, les objets mal segmentés (comme certaines routes sur la Figure 4.9(a)) ne sont pas représentés, on obtient alors une vérité terrain non conforme avec celle qu'un interprète ferait. Finalement, **juger visuelle-ment la classification revient à juger la pertinence de la segmentation**.

Nous avons demandé à une dizaine de personnes de noter les classifications obtenues avec chaque segmentation. Le graphe 4.13 résume les notes attribuées pour chaque classe et chaque segmentation. Celle nous donnant les meilleurs performances toutes classes confondues est la segmentation de Luo, bien que pour la classe Bâtiment, la ligne de partage des eaux puisse sembler meilleure ou la segmentation de Guigues pour la classe sol.

La ligne de partage des eaux est un algorithme de segmentation *généraliste*, conçu pour tout type d'images. En utilisant des techniques plus adaptées aux images que l'on traite, dans notre cas les images satellitaires, nous pouvons obtenir de meilleurs résultats. Un autre inconvénient est que cette méthode effectue une analyse mono-échelle, ce qui signifie qu'il faut choisir au préalable le bon niveau de détail. En pratique, il est très difficile de déterminer le bon niveau de détail, trop petit nous obtenons une sur-segmentation, trop grand certains objets n'apparaissent pas.

L'algorithme de Guigues a pour avantage de permettre une analyse multi-échelle, grâce au paramètre λ . Cependant le résultat de cet algorithme est une hiérarchie de régions emboîtées difficile à utiliser dans notre cas. Nous retrouvons alors le problème similaire à savoir : **Quels paramètres ou échelles l'utilisateur doit-il donner ?**

L'algorithme de Luo propose une analyse multi-échelle qui tient compte du modèle
d'acquisition d'une image satellite. Il utilise la FLST pour décomposer l'image en arbre d'inclusions de ses composantes connexes, ce qui fournit une décomposition multi-échelle de l'image. Il cumule ensuite les différentes lignes de niveaux de manière à faire ressortir la forme la plus contrastée. Cette méthode est très intéressante car elle est adaptée aux images que nous utilisons et elle permet également d'extraire des régions de différentes échelles. De plus, elle a pour avantage de ne pas nécessiter de paramètres.

Pour notre chaîne de traitement, nous avons choisi d'utiliser la segmentation de Luo. D'une part, nous avons vu que la classification était robuste au choix de la segmentation, les performances étant similaires. D'autre part, cette méthode ne requiert pas de paramètres adaptés à chaque image, du coup notre méthode est quasiment automatique, l'utilisateur n'intervient que pour donner quelques points d'apprentissage.

4.4.5 Complémentarité optique/radar

Dans ce paragraphe, nous étudions l'apport des informations issues de l'image radar dans la classification. Pour cela, nous avons classé les images une première fois en utilisant uniquement les paramètres extraits de l'image optique, puis une seconde fois en introduisant les paramètres extraits de l'image radar.

Il est aujourd'hui reconnu que l'information issue d'autres types de données, comme les images multispectrales [Kusaka *et al.*, 1990] [Ünsalan et Boyer, 2004], aide à la classification des zones urbaines. Pour classer nos images satellites optiques, nous avons fait le choix d'utiliser les informations contenues dans des données radar correspondantes. Bien sûr, l'apport de ces nouvelles informations est dépendant du recalage fait entre les données. Dans le chapitre 3, nous avons proposé une méthode de recalage avec un erreur résiduelle d'une dizaine de pixels. Cette erreur peut en induire une autre dans la classification, puisque les informations extraites peuvent correspondre à deux objets différents si le recalage est mal fait. Néanmoins dans ce paragraphe nous allons étudier **l'influence pour la classification des informations extraites des images radar**.

Les tableaux 4.6 à 4.8 donnent respectivement les résultats moyens sur 5 itérations d'une classification à partir des données optiques seulement, radar seulement et les deux réunies. Les données d'apprentissage sont de 50 objets par classe et sont les mêmes pour le calcul des performances optiques, radar ou mixtes. Le tableau 4.6 donne les résultats

4.4. Expérimentations

r		1				
	bâtiment	route	forêt	sol	ombre	total
validation croisée	71 %	56 %	58 %	53 %	94 %	66 %
performance par objets	76.76 %	75.73 %	78.14 %	78.67 %	96.62 %	79.98 %
performance par pixels	72.63 %	69.45 %	70.32 %	72.68 %	95.27 %	76.07 %

TAB. 4.6 – Résultats de la classification sur les données optiques seules. Les chiffres représentent les taux de bonne classification.

	bâtiment	route	forêt	sol	ombre	total
validation croisée	46 %	45 %	32 %	31 %	23 %	35 %
performance par objets	77.57 %	80 %	9.29 %	24.14 %	6.48 %	39.5 %
performance par pixels	77.49 %	78.14 %	9.82 %	22.16 %	7.4 %	39 %

TAB. 4.7 – Résultats de la classification sur les données radar. Les chiffres représentent les taux de bonne classification.

	bâtiment	route	forêt	sol	ombre	total
validation croisée	86 %	71 %	65 %	67 %	86 %	75 %
performance par objets	77.57 %	85.84 %	75.76 %	74.58 %	97.74 %	82.3 %
performance par pixels	95.63 %	78.41 %	73.19 %	90.21 %	96.32 %	86.75 %

TAB. 4.8 – Résultats de la classification sur les données optiques et radar. Les chiffres représentent les taux de bonne classification.

de la classification, obtenus à partir de 30 primitives optiques (géométriques et texturales). Le tableau 4.7 donne les résultats obtenus à partir des trois primitives radar, les log-cumulants, et le tableau 4.8, les résultats obtenus avec les 33 primitives optique et radar. En utilisant uniquement les caractéristiques optiques, nous avons environ 76 % de bonne classification, les caractéristiques radar donnent 39 % de bonne classification, tandis qu'en utilisant les deux types de primitives, nous obtenons 86 % de bonne classification. Ces résultats appuient l'hypothèse que nous avons faite à savoir que l'image radar contenait des informations utiles à la classification. En effet, mise à part la classe *ombre*, nous obtenons de meilleures performances et de meilleurs résultats lorsque nous utilisons les deux types de primitives et ceci malgré les erreurs éventuelles de recalage. En ce qui concerne la classe *ombre*, qui correspond à l'ombre de l'image optique, il paraît évident que l'image radar n'apporte pas d'amélioration puisque cette classe n'a pas de raison d'être dans cette dernière.

Ces résultats nous ouvrent une nouvelle voie pour la classification d'images satel-

litaires en milieu urbain. Nous voyons ici que **l'imagerie radar peut être une source d'information complémentaire à l'imagerie optique** dans le cadre de la classification des données haute résolution en milieu urbain, même avec un recalage imparfait.

4.4.6 Robustesse au changement de scène

Nous avons proposé jusqu'ici une chaîne de traitement pour la classification. Tout d'abord, nous avons choisi de segmenter l'image optique, ceci nous permet d'une part de découper des *objets* que nous pourrons recaler indépendamment, d'autre part d'extraire des informations géométriques. Ensuite, nous avons vu l'intérêt d'utiliser l'information contenue dans l'image radar, en plus de celle contenue dans l'image optique pour guider la classification. Finalement, nous proposons une méthode de classification avec apprentissage.

Une question peut alors se poser : L'apprentissage utilisé peut-il être utilisé sur une autre scène ? Autrement dit, les primitives choisies sont-elles robustes à un changement de scène ?

4.4.6.1 Image de la prison de Toulouse

La scène que nous avons utilisée pour l'apprentissage est une zone semi-urbaine. Dans ce cas-là, les classes choisies (bâtiment, route, végétation-forêt, sol et ombre) étaient bien représentées dans la scène et distinctes entre elles. La zone de la prison de Toulouse appartient à la catégorie zone urbaine dense. Dans ce type de zones, les classes que nous avons choisies ne sont plus distinctes. Pour un interprète il ne serait pas facile de classer cette image selon les cinq classes utilisées précédemment. En effet, cette zone urbaine dense est constituée de quartiers pavillonnaires. Les maisons sont rattachées à des jardins avec des arbres, les classes *végétation-forêt* et *sol* sont donc mélangées. Les bâtiments sont pour la plupart des maisons avec des toits à deux pentes avec des expositions différentes. Les rues sont étroites et entourées de maisons qui projettent leur ombre. Il est donc difficile de séparer la classe *route* de la classe *ombre*.

Nous voyons Figure 4.14 le résultat de la classification de l'image de la prison de Toulouse avec des données apprises sur la zone du CNES.

Nous ne disposons pas de vérité terrain sur cette zone pour évaluer numériquement la classification. Aussi nous proposons une évaluation visuelle dans le tableau 4.9.



FIG. 4.14 – *Classification de la zone de la prison de Toulouse à partir de données apprises sur la zone du CNES : les bâtiments sont représentés en bleu, les routes en beige, la végétation haute en vert, le sol en rouge et l'ombre en rose.*

hôtiment	Il y a pau d'arraurs sur las hâtiments détactés par contra la
Datiment	If y a peu u effeurs sur les bauments détectes, par contre la
	plupart ne le sont pas. Les gros bâtiments sont mis dans la
	classe sol.
route	très peu de routes sont détectées, la plupart sont classées
	comme des ombres.
végétation	difficile à évaluer car on trouve de la végétation dans toute
	l'image de manière parcimonieuse : autour des routes, dans
	les jardins
sol	mauvaise classification : cette classe contient beaucoup de
	bâtiment.
ombre	bonne classification.

TAB. 4.9 – Résultat de la classification de la prison de Toulouse avec une base d'apprentissage extraite de la zone du CNES.

Comme on s'y attendait, les classes routes et ombres sont mélangées. La végétation est mélangée avec d'autres classes telles que le sol ou les routes. Les bâtiments de notre image sont pour la plupart des pavillons avec des toits à deux pentes alors que la base d'apprentissage contient de grands bâtiments aux toits plats, ce qui peut expliquer la mauvaise détection que nous avons pour cette classe.

Le problème que nous rencontrons ici concerne le choix de nos classes. Le résultat de l'apprentissage que nous utilisons sur la zone du CNES ne peut s'exporter à n'importe quel type de scène urbaine. Dans le cas de zones urbaines denses, comme pour l'image de la prison de Toulouse, il est nécessaire définir de nouveaux labels adaptés à ce type de scène, par exemple : routes, grands bâtiments et zones résidentielles (comprenant maison individuelles et jardins).

Bien que le modèle de classification ne soit pas robuste d'un type de zone à un autre, nous allons montrer la robustesse de la chaîne, c'est-à-dire de la méthode de traitement que nous proposons, à partir de la segmentation jusqu'à la classification. Cela suppose de donner manuellement quelques points d'apprentissage, extraits de notre image, afin de calculer un modèle de classification mieux adapté à notre image.

La figure 4.15 montre les résultats de notre chaîne de traitement lorsque nous faisons intervenir un utilisateur pour donner des informations en apprentissage. L'utilisateur donne pour chaque classe une cinquantaine de points d'apprentissage. Pour la classe qui nous intéresse particulièrement, nous voyons que les bâtiments sont bien détectés. Les routes, elles aussi, sont mieux détectées bien qu'il en manque toujours beaucoup. Les ombres sont encore bien détectées, mais la confusion avec les routes subsiste.

La zone de la prison est une zone difficile à labelliser. De plus le choix des classes que nous avons fait peut être discuté sur ce type de zones urbaines denses. Nous voyons néanmoins que la méthode de classification que nous proposons donne des résultats visuellement satisfaisants quel que soit le type de scène (semi-urbain ou urbain dense).

4.4.6.2 Complexe scientifique de Rangeuil

La chaîne de traitement a également été appliqué aux images TerraSar sur la zone industrielle de Rangueil. Nous avons donné en entrée du classificateur 50 objets d'apprentissage par classe. Nous observons le résultat de la classification sur la figure 4.16.

Une analyse visuelle du résultat a été faite. Les bâtiments à toits plats sont très bien détectés, tandis que ceux avec un toit à plusieurs pentes ne le sont pas. Cette classe a tout de même plusieurs fausses alarmes, en particulier, une zone de travaux et certains arbres isolés. En ce qui concerne la classe route, le réseau routier apparaît, mais nous voyons beaucoup de fausses alarmes au centre de l'image. Il s'agit en fait d'un seul objet mal classé, représenté figure 4.17, qui devrait appartenir à la classe sol. La végétation semble bien identifiée, bien qu'il soit diffile de l'évaluer. Les ombres des bâtiments sont bien classées, par contre les ombres des arbres peuvent être confondues avec ceux-ci.

Nous obtenons une classification satisfaisante des données satellitaires sur le complexe scientifique de Rangeuil, à l'exception de la classe route, où un objet appartenant à la classe sol a été détecté.

4.5 Conclusion

Dans ce chapitre, nous avons proposé une méthodologie pour classer des données sattelitaires optiques et radar. Nous souhaitions distinguer les objets de la scène en hauteur, qui se projettent différemment, par rapport à ceux au sol, sur chacune des images. Pour cela nous avons choisi de décomposer notre scène en objets à l'aide d'une segmentation de l'image optique et de les classer selon 5 classes : bâtiments, routes, sol, végétation, ombre optique.

La chaîne de traitement que nous proposons se décompose en trois étapes : segmentation, extraction des primitives et classification.

Nous avons vu que le choix de la méthode de segmentation a peu d'influence sur



FIG. 4.15 – *Classification de la zone de la prison de Toulouse à partir d'informations données manuellement : les bâtiments sont représentés en bleu, les routes en beige, la végétation haute en vert, le sol en rouge et l'ombre en rose.*



FIG. 4.16 – *Classification de la zone industrielle de Rangueil en 5 classes : bâtiments, routes, végétation, sol et ombre.*



FIG. 4.17 – Objet de la classe sol classé en tant que route dans la figure 4.16.

le résultat de la classification. Nous avons choisi la segmentation proposé par Luo [Luo, 2007] qui a pour avantage d'être automatique et robuste au bruit d'acquisition des images optiques.

Le fait d'utiliser une segmentation de la scène nous permet d'en extraire des caractéristiques géométriques (périmètre, surface et compacité), en plus des texturales (moyenne, variance et filtrage de Gabor). Bien que nos images ne soient pas parfaitement recalées, nous avons ajouté des caractéristiques statistiques radar : les log-cumulants. Nous avons montré l'information que portait chaque type de primitives et en particulier l'amélioration apportée par l'utilisation conjointe des données optiques et radar lorsqu'on utilise un classificateur SVM avec un noyau Gaussien.

Le but de ce chapitre était de proposer une méthode de classification qui permette de guider le recalage suivant que les objets soient au sol ou en sur-sol. Notre méthode permet

d'obtenir cela avec des performances de l'ordre de 80 % de bonne classification malgré l'erreur moyenne de recalage d'une dizaine de pixels. Aussi nous pouvons améliorer ces résultats avec des données parfaitement recalées.

Chapitre 5

Optimisation jointe de la classification et du recalage

5.1 Introduction

Nous avons vu dans les chapitres précédents, une méthodologie pour recaler des données optiques et radar de scènes urbaines. Ce recalage peut être affiné en tenant compte de la nature des objets, nous avons donc proposé une méthode de classification. Nous avons alors montré l'interêt de l'information contenue dans l'image radar pour améliorer la classification. Cependant, si les données ne sont pas parfaitement recalées, les caractéristiques radar utilisées par le classificateur peuvent ne pas correspondre précisément aux objets d'interêts, les résultats de la classification dépendent donc du recalage. Autrement dit, **ces deux opérations sont dépendantes l'une de l'autre.**

Dans le cadre de cette thèse, deux tâches dépendantes l'une de l'autre sont à optimiser : le recalage et la classification. Dans ce chapitre nous proposons deux pistes pour optimiser ces critères : la première consiste à effectuer les deux tâches simultanément, la seconde consiste à les faire successivement. Etant donné les déformations entre les images optiques et radar, en particulier pour les objets en hauteur, il paraît évident que nous devons développer des méthodes propres à chaque classe. Les bâtiments sont parmi les objets les plus complexes des zones urbaines, l'urbanisme change constamment (démoliton et nouvelle construction) aussi il est important de pouvoir traiter ces données de manière automatique. De plus, identifier et recaler précisément les bâtiments permet d'extraire une information sur leur hauteur et donc sur le type de scène. Nous avons donc choisi de concentrer notre étude sur les bâtiments qui sont les plus difficiles à recaler car en hauteur.



FIG. 5.1 – Classification et recalage joints.

5.2 Classification et recalage joints

Dans un problème d'optimisation multi-critères, l'idée la plus générale consiste à combiner les critères en une unique fonction. Dans le cas du recalage et de la classification, nous ne pouvons pas définir une fonction de coût qui puisse être optimisée globalement. Le moyen que nous avons choisi afin de classer et recaler simultanément nos objets consiste à introduire dans la classification plusieurs recalages possibles, comme le montre le schema 5.1. La méthode que nous proposons consiste à translater un objet et le classer comme dans le chapitre 4, dans ce cas les seuls primitives qui changent sont les primitives statistiques de l'image radar, la forme et la texture extraites de l'image optique restent identiques. Cela est effectué pour plusieurs translations, une décision est ensuite prise pour déterminer la meilleure translation et la classe associée.

Notre hypothèse est la suivante : "Ayant effectué notre apprentissage sur des données recalées, pour chaque segment, la sortie du SVM ayant la plus forte probabilité, en testant plusieurs couples de translation $(\delta x, \delta y)$ correspond à la bonne translation et la bonne classe".

5.2.1 Principe

Pour recaler et classer simultanément nos images, nous avons choisi de classer chaque objet en considérant différents recalages possibles. Pour cela, quelques objets sont choisis au préalable et éventuellement recalés afin de produire une base d'apprentissage d'objets recalés, la mise en oeuvre est expliquée dans le paragraphe 5.2.3. Nous utilisons ensuite la procédure de classification décrite dans le chapitre précédent.

Nous supposons que le classificateur appris sur des données recalées permettra d'identifier la nature ainsi que la position des objets de test. Pour cela, nous nous intéressons à *la distance à la marge* du classificateur SVM. Plus cette distance est grande, plus la probabilité que notre objet appartienne à la classe est grande.

5.2.2 Mise en oeuvre

Pour chaque objet, nous devons extraire les paramètres correspondants aux différentes translations. Les paramètres géométriques et texturaux de l'image optique restent inchangés quel que soit le recalage entre les objets. Ce sont les paramètres statistiques de l'image radar qui changent en fonction du recalage, ils sont donc recalculés pour différentes translations.

La classification s'effectue sur notre image recalée grossièrement dans le chapitre 3. L'erreur moyenne obtenue était d'une dizaine de pixels sur les différents jeux de données, aussi l'espace de recherche est limité à une translation comprise entre -20 et 10 pixels horizontalement et entre -20 et 20 pixels verticalement en utilisant un pas de 5 pixels. Nous nous sommes intéressés de manière plus précise (avec un pas de 1 pixel) au déplacement horizontal car il nous permet d'avoir une indication sur la hauteur des objets. Nous avons également étendu la zone de recherche sur la gauche car nous savons que les objets en hauteur sont projetés dans cette direction. Le capteur étant situé conventionnellement à gauche de l'image.

5.2.3 Zone d'apprentissage

Les informations données en apprentissage au classificateur doivent être extraites sur des objets recalés entre l'image radar et l'image optique. Pour les classes routes, bâtiments et végétation basse, nous avons recalé manuellement une dizaine d'objets, par contre la classe ombre n'a pas de correspondance dans l'image radar, nous avons donc laissé les objets tels quels sans chercher à les recaler. En ce qui concerne la classe arbre, nous

voyons qu'elle est souvent sur-segmentée, il est difficile pour un utilisateur de recaler ces objets, nous avons donc veillé à ce que les objets étiquetés de cette classe soient bien des arbres dans chacune des images. Toutefois, il n'est pas certain que le recalage donné soit très précis, en particulier dans les zones de forêts très texturées. La figure 5.2 présente les objets servant de base d'apprentissage.



FIG. 5.2 – Zones d'apprentissage superposées à l'image optique.

5.2.4 SVM probabiliste

Dans le cas multi-classes, la stratégie employée est le "*un contre tous*". Nous considérons autant de classificateurs binaires qu'il y a de classes et pour chacun d'entre eux, nous calculons la distance à la marge du SVM. Afin de pouvoir comparer les sorties des SVM, nous avons besoin de les représenter dans un espace normalisé, aussi nous avons choisi de probabiliser les sorties. Bien peu de points sont disponibles en appren-

tissage (une dizaine par classe), ce qui peut rendre difficile la détermination de la bonne séparation entre les classes ou la densité de probabilité. Pour transformer cette distance à la marge en une probabilité, nous utilisons la méthode proposé par Platt [Platt, 1999]. Ce dernier décrit la probabilité d'appartenance à la classe fonction de la distance à la marge comme une sigmoïde, soit :

$$P(y|f) = \frac{1}{1 + exp(Af + B)}$$

où y est la classe et f la distance à la marge. A et B sont des paramètres à estimer à partir des données d'apprentissage.

Dans les exemples qui suivent, nous nous intéressons plus particulièrement aux bâtiments de la scène. Nous ne considérons donc que deux classes : bâtiments ou nonbâtiments. Dans ce cas, il n'est pas nécessaire de passer par les probabilités, nous comparons directement les distances à la marge.

5.2.5 Classification : bâtiments/non-bâtiments

Pour déterminer la classe *bâtiments*, nous devons donner en apprentissage des points de la classe *bâtiments* et de la classe *non-bâtiments*. Des travaux récents ont montré que les résultats obtenus avec les SVM, lorsque le nombre de données d'apprentissage est nettement plus important pour une classe, sont plus favorables pour la classe majoritaire [Zeng et Gao, 2009]. La méthode courante pour corriger cela est d'introduire une pénalité associée à la classe majoritaire [Veropoulos *et al.*, 1999]. Dans nos travaux, afin d'éviter ce problème de *balancement des données*, nous avons donné en apprentissage autant de points pour chaque classe. Etant donné que nous avons déjà recalé une dizaine d'objets par classe (bâtiments, routes, végétation, sol et ombre optique), nous choisissons donc une dizaine d'objets *non-bâtiments* et autant de la classe *bâtiments*. Les objets donnés en apprentissage sont représentés figure 5.3. Une fois cette base d'apprentissage constituée d'objets bien recalés, nous calculons la séparation entre ces données, puis pour chaque objet et plusieurs translations, la distance à la marge.

5.2.5.1 Evaluation de la classification

Pour évaluer le résultat de classification, nous avons calculé les taux de rappel et de précision. Le rappel correspond au taux de bâtiments détectés et la précision au taux de



FIG. 5.3 – Zones d'apprentissage pour la classification des bâtiments, superposées à l'image optique : les objets de la classe bâtiments sont représentés en bleu et ceux de la classe non-bâtiments en jaune.

vrais bâtiments parmis ceux détectés. Soit M, la matrice de confusion (Tableau 5.1), on a :

- VP = nombre de vrais positifs
- VN = nombre de vrais négatifs
- FP = nombre de faux positifs
- FN = nombre de faux négatifs

On définit alors le taux de rappel comme étant $\frac{VP}{VP+FN}$ et le taux de précision par $\frac{VP}{VP+FP}$.

$$Rappel = 64\%$$
 $Precision = 26\%$

Ces résultats sont moins bons que ceux obtenus dans le chapitre précédent, lorsque nous effectuons une classification en 5 classes. Ceci est dû entre autres au nombre de

		Vérité		
		Bâtiment	Non-Bâtiment	
Issif	Bâtiment	VP	FP	
Cla	Non-Bâtiment	FN	VN	

TAB. 5.1 – Résultats de la classification des bâtiments.

points utilisés pour l'apprentissage. Nous passons de 50 points d'apprentissage par classe à 10, nous avons fait ce choix pour limiter le temps nécessaire à un utilisateur pour recaler manuellement les objets d'apprentissage.

5.2.5.2 Evaluation du recalage

Dans ces travaux, nous cherchons à vérifier la validité de notre hypothèse, c'est-à-dire valider si la translation qui a une réponse maximale correspond à un bon recalage. Pour vérifier cela, nous nous sommes intéressés à quelques objets représentés figure 5.4. Nous avons choisi des objets représentatifs des types de bâtiments que nous voyons dans la scène, c'est à dire, des immeubles, des bâtiments isolés et non-isolés. Nous avons sélectionné des bâtiments bien segmentés qui n'appartiennent pas aux données d'apprentissage. Nous voyons les objets à leur position initiale, les mêmes une fois recalés selon la réponse maximale du classificateur, ainsi que les réponses : classe et score (= distance à la marge).

Au vu de la figure 5.4, nous constatons que le score de classification porte bien l'empreinte du recalage. En effet, la distance à la marge du classificateur est plus grande autour du bon recalage. Les scores de l'objet 1 forment un pallier plus élevé autour de la bonne translation, sans pour autant que le maximum corresponde à la translation exacte. L'objet 4 est une partie d'un bâtiment, cette partie isolée peut être recalée tout le long du bâtiment sur l'image radar. Nous observons également cela sur le graphe des scores, les déplacements horizontaux forment une fonction concave qui admet un maximum, alors que les déplacements verticaux le long du bâtiment adoptent une forme constante, il en va de même pour l'objet 2. Le troisième objet représente un bâtiment juxtaposé à plusieurs autres. Visuellement il est difficile de déterminer la bonne position de notre objet, il en est de même pour le classificateur. Nous voyons sur le graphe que les scores sont quasi-équivalents pour toutes les positions. Certes le maximum nous donne une position



FIG. 5.4 – **Recalage par classification.** La première colonne montre la position des objets d'origine, la deuxième leur position une fois recalés dans l'image radar, la troisième colonne montre la classe renvoyée par le classificateur en fonction de la position de l'objet et la dernière, les distances à la marge en valeur absolue, pour différentes positions.



FIG. 5.5 – **Recalage par classification.** La première colonne montre la position des objets d'origine, la deuxième leur position une fois recalés dans l'image radar, la troisième colonne montre la classe renvoyée par le classificateur en fonction de la position de l'objet et la dernière, les distances à la marge en valeur absolue, pour différentes positions.

qui semble juste, mais nous voyons tout de même que le classificateur a une forte sensibilité autour de celle-ci, toute les positions nous donnent une distance à la marge autour de 1.

La figure 5.5 représente des objets qui ont été mal classés ou mal recalés en suivant notre hypothèse. L'objet 5 n'a pas été reconnu comme bâtiment, sur l'image radar plusieurs bâtiments juxtaposés se superposent ce qui rend le recalage complexe. Lorsque nous regardons la distance à la marge obtenue par le SVM pour plusieurs positions, nous observons deux maxima aux positions [-20,20] et [-19,-15], seulement aucune de ces positions ne correspond à un bon recalage. A cause des multiples réflexions des bâtiments voisins, ni le bon recalage, ni la bonne classe n'ont été trouvés. L'objet 6 est un bâtiment isolé qui a été recalé sur l'ombre radar. La radiométrie de ce bâtiment est la même que celle du sol, les caractéristiques statistiques radar que nous utilisons (log-cumulants) ne sont donc pas suffisantes pour discriminer la bonne position. Etant donné que le capteur est à gauche de l'image, l'ombre radar se trouve forcément à droite du bâtiment, nous voyons donc que le recalage obtenu n'est pas correct, mais le bâtiment a tout de même été détecté. Les objets 7 et 8 ne sont pas des bâtiments qui ont pourtant été détectés comme tels. Le graphe de classification en fonction de la translation montre la sensibilité de la classification lorsqu'on déplace l'objet. L'objet 7 possède deux maxima : à la position [-20,15] avec une amplitude de 3.12 et à [2,-20] avec une amplitude de 2.88. Le premier maximum est associé à la classe bâtiment et le second à la classe non-bâtiment. Dans le premier cas, l'objet est recalé sur le bâtiment et classé come tel, dans le deuxième cas il est positionné sur du non-bâtiment et classé comme tel. Les deux maxima ont une amplitude assez proche et c'est le second qui correspond à la vraie nature de l'objet. En ce qui concerne l'objet 8, nous observons le même phénomène : superposé à des bâtiments, l'arbre est classé comme tel mais lorsqu'on déplace l'objet sur des zones non-bâtiments, la classification change.

Dans le cas des objets 7 et 8, on ne peut pas parler de bon recalage, car ces objets n'existent pas dans l'image radar, l'ombre optique n'a pas de correspondance dans l'image radar et l'arbre a laissé place à des bâtiments. Cependant, nous voyons que le résultat de la classification change en fonction de la position de l'objet, ce qui montre **l'influence des primitives radar pour la détection des bâtiments**.

5.2.6 Classification itérative

Parmis les résultats présentés au paragraphe précédent, nous constatons la forte sensibilité du classificateur autour de la bonne translation. La distance à la marge est effectivement plus élevée autour du bon recalage, mais pas sufisamment discriminante, c'est pourquoi prendre la translation ayant un score maximum ne donne pas toujours un recalage exact. Cela s'explique par le peu de points donnés en apprentissage au classificateur. Celui-ci calcule une marge de séparation à partir de seulement dix objets appris par classe. Nous pouvons diminuer la sensibilité du classificateur en augmentant le nombre d'objets d'apprentissage de manière itérative.

Plusieurs travaux existent sur le mode de sélection d'objets pertinents pour le SVM en télédétection [Ferecatu et Boujemaa, 2007]. Plusieurs questions se posent alors, concernant la stratégie à employer pour sélectionner des objets à ajouter à la base d'apprentissage.

En ce qui concerne la classe *bâtiments* :

- nous pouvons sélectionner automatiquement des objets ayant un score élevé et leur associer la translation associée. Cela nécessite tout de même l'intervention d'un utilisateur pour vérifier la classification et le recalage. Seulement rajouter des points loin de la marge, ne modifiera pas celle-ci et ne permet pas d'obtenir un score plus discriminant.
- nous pouvons aussi sélectionner des bâtiments ayant un score faible et faire intervenir un utilisateur pour déterminer le recalage.

En ce qui concerne la classe non-bâtiments :

- nous pouvons sélectionner des objets non-bâtiments ayant un score faible, c'est-àdire des objets ambigus, que le classificateur a du mal à classer. Seulement, quelle translation leur attribuer ? Devons-nous leur attribuer la translation ayant un score élevé ou faible ou faire intervenir un utilisateur qui recale l'objet ?
- nous pouvons sélectionner des objets de la classe non-bâtiments ayant un score élevé, ceux qui sont sans ambiguïtés pour le classificateur, mais à nouveau quelle translation leur attibuer ? De plus ces objets, loin de la marge, ne la modifieront pas et ne permettront pas de rendre les scores plus discriminants.
- Notons que le choix de la translation est important, car l'objet non-bâtiment ne doit pas être recalé sur un bâtiment.
- nous pouvons également sélectionner des bâtiments, correctement classés, mais mal recalés.

5.2.7 Conclusion

En étudiant les réponses du classificateur pour plusieurs positions, nous avons vu que celles-ci reflètent la position des bâtiments. En effet, **le score est plus élevé au niveau du bon recalage du bâtiment**, comme le montre les objets 1, 2 et 4. Toutefois, choisir le score maximum n'est pas le seul critère à prendre en compte, c'est ce que nous voyons pour l'objet 1 : le maximum ne correspond pas exactement au bon recalage. Nous donnons deux explications à ce problème : tout d'abord entre deux positions, les différences données au classificateur sont les primitives radar, cela signifie que les caractéristiques statistiques que nous extrayons ne sont pas suffisantes pour recaler de manière exacte nos objets. L'autre explication concerne le peu de points que nous utilisons pour l'apprentissage, nous évaluons un modèle de classification et recalage des bâtiments à partir de seulement dix données en apprentissage, aussi nous pouvons envisager que la barrière de sélection calculée par le SVM puisse être mal définie.

En même temps, nous voyons **l'importance des primitives radar pour la détection des bâtiments**, car nous arrivons à détecter et recaler approximativement des objets, ayant pourtant des caractéristiques autres, sur des bâtiments, grâce aux primitives radar, comme les objets 7 et 8.

La méthode que nous proposons ici pour classer et recaler des bâtiments a ses limites, en particulier elle dépend de la segmentation des objets. L'objet 4 est une sous-partie d'un bâtiment, considérée indépendamment; nous ne pouvons pas déterminer s'il correspond à la partie haute ou basse du bâtiment, il a donc plusieurs possibilités de recalage et c'est ce que nous observons dans la réponse du SVM. Dans le cas où plusieurs objets se superposent, comme les objets 3 ou 5, il est difficile d'évaluer visuellement le recalage, nous voyons qu'il en est de même pour le classificateur.

Les perspectives que nous pouvons envisager pour améliorer ce système {Classification + Recalage} sont :

- introduire une information contextuelle qui permettra de guider le recalage d'objets qui se superposent, c'est à dire considérer un ensemble d'objets,
- augmenter le nombre de points d'apprentissage, ce qui peut permettre d'affiner le score du bon recalage. Toutefois plusieurs questions se posent concernant la stratégie à suivre pour sélectionner ces points.

5.3 Classification et recalage successifs : par le contexte

Nous présentons dans cette partie une seconde méthode pour optimiser le recalage et la classification de données satellitaires en milieu urbain. Chacune de ces opérations est dépendante l'une de l'autre, nous proposons donc de les optimiser de manière successive.

Nous avons vu que la chaîne de classification établie au chapitre 4 est conditionnée au recalage. Nous développons ici une méthode de recalage conditionné à la classification. Pour cela nous utilisons les informations contextuelles liées à chaque classe. Dans cette partie nous nous intéresserons uniquement à la classe bâtiment.

5.3.1 Principe

Pour améliorer le recalage, nous utilisons les résultats de classification au chapitre précédent. De plus, de par les propriétés géométriques et radiométriques des images radar, nous savons qu'un bâtiment possède sur l'image radar un coin brillant dans le sens du capteur et une zone d'ombre du côté opposé, comme le montre la figure 5.6



FIG. 5.6 – Projection d'un bâtiment sur des données optiques et radar.

Nous proposons d'utiliser l'hypothèse de la classe obtenue précedemment et l'information contextuelle (un coin reflecteur brillant et une zone d'ombre de l'autre coté) pour recaler les bâtiments. Pour cela, nous définissons un détecteur de coin réflecteur ainsi qu'un détecteur des ombres, puis nous fusionnons ces deux informations.

Notre a priori pour recaler les bâtiments est donc le suivant : "Un bâtiment a sur son contour gauche, dans l'image radar, une ligne brillante et à sa droite une zone d'ombre."

5.3.2 Détection du coin réflecteur

Dans cette section, nous souhaitons calculer une carte de détection des coins réflecteurs sur l'image radar, propre à chaque objet. La méthode utilisée est la suivante :

- extraire le contour gauche de l'onjet, qui correspond sur l'image radar à un coin réflecteur,
- faire une corrélation avec l'image radar.

Le sens de déplacement du satellite radar ainsi que la visée sont des paramètres connus lors de l'acquisition d'une donnée radar. En ce qui concerne l'image RAMSES sur Toulouse, figure 2.9, nous savons que le capteur se déplace verticalement et vise vers la droite, les coins réflecteurs vont donc apparaître du côté du capteur, c'est à dire à la gauche des bâtiments. Nous pouvons donc utiliser cette information pour guider le recalage. La première étape consiste à extraire le contour gauche de l'objet correspondant au coin réflecteur (figure 5.7 (b)), où doit apparaître sur l'image radar une ligne brillante. Pour cela, un gradient est appliqué sur le masque de l'objet.

Une fois le coin réflecteur extrait, nous calculons, pour plusieurs translations, une probabilité de ligne brillante sur l'image radar, ce qui revient à calculer la radiométrie moyenne le long du coin réflecteur. Pour cela, nous effectuons une corrélation entre le coin réflecteur et la radiométrie de l'image radar. La figure 5.7 montre l'exemple d'un bâtiment, l'extraction de son contour gauche et le résultat de la corrélation en se déplaçant sur l'image radar.

5.3.3 Détection des ombres

L'autre information utilisée pour le recalage des bâtiments est la zone d'*ombre radar* qui se trouve du côté opposé au capteur. Nous savons qu'un bâtiment se trouvera juxtaposé à cette ombre, nous souhaitons donc calculer une carte de détection du bord des ombres radar. Le principe est le suivant :



FIG. 5.7 – *Carte des coins réflecteurs :* (a) objet superposé à l'image radar avant recalage, (b) extraction du contour gauche correspondant à l'emplacement d'un coin réflecteur, (c) corrélation entre le contour gauche et l'image radar pour plusieurs translations.

- extraire le bord droit du bâtiment contre lequel se trouve l'ombre radar,
- calculer un masque des ombres sur l'image radar,
- chercher les discontinuités des ombres, c'est-à-dire les contours,
- faire une corrélation entre le bord du bâtiment et les contours des ombres radar.

L'étape d'extraction du bord du bâtiment où se trouve l'ombre radar, consiste à extraire le bord droit du bâtiment, comme le montre la figure 5.8. Pour calculer un masque des ombres, il est possible d'utiliser une classification [Tison, 2004], mais nous avons choisi d'effectuer un seuillage de l'image, car les ombres radar comme optique sont caractérisées par une radiométrie faible. Nous avons donc seuillé l'image radar brut.

masque des ombres = image radar
$$< m/3$$

où m est la moyenne calculée sur toute l'image radar.

De fait, nous cherchons à recaler le bord d'un bâtiment, non pas sur les zones d'ombres, mais sur le bord de l'ombre, nous cherchons donc les contours des zones d'ombres. Pour cela, nous avons calculé les contours des zones d'ombres. La corrélation entre ces derniers et le contour droit de notre bâtiment, nous donne une carte de probabilité des emplacements du bâtiment par rapport aux ombres de la scène.

5.3.4 Fusion

Nous disposons alors de deux détecteurs : un pour les coins réflecteurs et un pour les gradients des ombres radar en fonction de la position du bâtiment. Nous souhaitons



FIG. 5.8 – *Carte des ombres :* (a) contour droit du bâtiment, (b) masque des ombres calculés sur l'image radar, (c) gradient des zones d'ombres, (d) corrélation entre le bord du bâtiment et le gradient des ombres pour plusieurs positions.



FIG. 5.9 – *Fusion des deux détecteurs* : (a) détecteur de coin réflecteur, (b) détecteur de gradient des ombres, (c) produit des deux détecteurs.

fusionner ces informations. [Bloch, 2003] propose plusieurs méthodes pour fusionner des informations : Dempster-Schafer, moyenne géométrique, pondération, produit ... Dans le cas présent, les informations portées par chacun des détecteurs sont aussi importantes l'une que l'autre, nous avons donc choisi de **normaliser** les détecteurs entre 0 et 1, puis de **prendre le produit**. Le bon recalage du bâtiment aura une position ayant une forte réponse avec chacun des détecteurs, nous choisissons donc le maximum pour déterminer cette position.

5.3.5 Résultats

Les figures 5.10 et 5.11 présentent les résultats de recalage obtenus pour les mêmes objets que précédemment.

L'objet 1 est très bien recalé, c'est un bâtiment isolé qui correspond parfaitement à notre a priori : une ligne brillante à gauche et un gradient d'ombre sur le contour droit.

Les objets 2 et 4 sont des immeubles de plusieurs étages, nous voyons sur l'image radar, plusieurs lignes brillantes correspondant à des réflexions sur plusieurs étages, notre détecteur de coins réflecteurs donne plusieurs possibilités de translations, et le maximum après fusion correspond à un mauvais recalage. Dans ce cas, il est intéressant de privilégier l'information apportée par le détecteur d'ombre sur celui de lignes brillantes. Pour l'objet 2, nous voyons apparaître deux maxima après la fusion, le premier correspond à une translation de [0,4], la seconde à [-5,-15], le second maximum correspond à la bonne translation.

Les objets 3 et 5 sont des bâtiments pour lesquels notre a priori n'est pas entièrement validé. Le bâtiment 3 est bien recalé horizontalement, mais plusieurs translations verticales répondent fortement à notre détecteur. C'est un objet difficile à recaler manuellement car les coins réflecteurs et les zones d'ombres des bâtiments voisins viennent interférer dans le recalage du bâtiment ; c'est aussi ce que l'on voit pour l'objet 5. Les coins réflecteurs des bâtiments voisins et parallèles répondent fortement sur notre détecteur, il est donc difficile de l'utiliser pour trouver le bon recalage. De plus, cet objet n'a pas de zone d'ombre visible car elle est cachée par un bâtiment voisin à sa droite. Cet objet n'a pas été correctement recalé car il ne répond pas à notre a priori.

L'objet 6 est un bâtiment dont la radiométrie du toit est identique à celle du sol, il est donc difficile d'identifier à l'oeil son recalage. Notre méthode a recalé ce bâtiment sur le coin réflecteur qui correspond en fait au coin sol-mur et non au bord du toit, le recalage n'est donc pas correct. Toutefois, si on s'intéresse à la réponse donnée par le détecteur d'ombre, nous obtenons un maximum pour une translation de [-9,-7] pixels, ce qui correspond à un bon recalage. Ceci montre **l'importance de l'information contextuelle** puisque les log-statistiques ne permettaient pas de déterminer le bon recalage à l'aide de la méthode précédente. Cela nous amène également à nous interroger sur le choix de notre critère de fusion : *les informations portées par chaque détecteur sont-elles aussi importantes l'une que l'autre ?*

L'objet 7 montre l'inconvénient d'utiliser des méthodes successives. En effet, l'arbre que nous voyons a été classé dans le chapitre 4 comme un bâtiment. Notre méthode de recalage ne remet pas en cause la classification, les erreurs se propagent donc.

5.3.6 Conclusion sur le recalage par le contexte

Nous avons proposé ici une méthode de recalage propre aux bâtiments. Cette méthode se base sur plusieurs a priori, celui de la classe ainsi que des informations contextuelles, en particulier sur le fait qu'un bâtiment a une ligne brillante sur un bord et une zone d'ombre de l'autre côté.

Nous avons développé deux descripteurs permettant de détecter les lignes brillantes et les zones d'ombre de part et d'autre d'un bâtiment, puis nous les avons fusionnés. Cette méthode permet d'obtenir de très bons résultats, lorsque nos a priori sont vérifiés, comme sur le bâtiment 1 de la figure 5.10.

Toutefois, nous voyons que l'information contextuelle que nous utilisons n'est pas toujours suffisante. En effet un bâtiment peut posséder plusieurs lignes brillantes (le coin sol-mur, le bord du toit et différentes réflexions qui peuvent apparaître sur les étages). De plus, la zone d'ombre n'est pas toujours bien marquée, comme sur les bâtiments 2 ou 4.



FIG. 5.10 – **Recalage des bâtiments :** (a) bâtiment avant recalage, (b) bâtiment après recalage, (c) produit des deux descripteurs contextuels.



FIG. 5.11 – **Recalage des bâtiments :** (a) bâtiment avant recalage, (b) bâtiment après recalage, (c) produit des deux descripteurs contextuels.

Nous avons supposé ces critères d'égales importances lors de la fusion, or pour certains bâtiments, la zone d'ombre ou le coin réflecteur n'est pas visible. Nous pouvons améliorer cela en pondérant les critères en fonction des bâtiments.

Le fait d'utiliser les résultats de la classification obtenue au chapitre précédent introduit un autre problème, celui de la propagation des erreurs. Les objets mal classés seront recalés comme des bâtiments, comme l'objet 7 de la figure 5.11 qui correspond à un arbre dans l'image optique.

5.4 Conclusion

Dans ce chapitre nous avons cherché à optimiser le recalage et la classification d'objets présents dans le milieu urbain, en nous intéressant principalement aux bâtiments. Ces deux opérations sont dépendantes l'une de l'autre, comme nous l'avons vu dans les chapitre précédents.

Nous avons donc proposé deux approches pour optimiser ces opérations. La première consiste à les faire simultanément. Pour cela nous supposons qu'un classificateur appris sur des données recalées sera capable de les reconnaître parmi des données mal recalées. La seconde méthode s'appuie sur une classification déjà obtenue et nous recalons les objets conditionnellement à leur classe par a priori contextuels.

La première méthode offre de bons résultats, illustrés par les scores obtenus sur les bâtiments. Ces résultats montrent notamment l'importance des données radar, nous constatons aussi certaines difficultés quant à la classification d'objets superposés, ces difficultés étant corroborées par le fait qu'il est aussi difficile pour un interprète de déterminer visuellement le bon recalage. Cette conclusion nous amène à l'idée d'introduire une information contextuelle supplémentaire afin d'outrepasser ces obstacles, ou considérer plus de points d'apprentissage. Cependant cette dernière solution soulève beaucoup de questions concernant la stratégie à utiliser.

Nous avons également mis en oeuvre une méthode de recalage basée sur différents a priori, tels que, par exemple, la présence de lignes brillante sur l'image radar au niveau du côté d'un bâtiment. L'utilisation de cette information a permis d'améliorer sensiblement les résultats lorsque les hypothèses utilisées étaient effectivement vérifiées. Toutefois, le cas contraire a démontré que cette technique présentait le défaut de propager certaines erreurs, car parmi nos a priori il y a celui de la classe. Or celui-ci n'est pas remis en cause, une erreur de classification va donc entraîner une erreur dans le recalage,

comme lorsque nous cherchons à recaler un arbre comme s'il s'agissait d'un bâtiment. Dans cette méthode, nous proposons deux détecteurs (coins réflecteurs et ombres) ainsi qu'une méthode de fusion qui accorde autant d'importance à chacun des détecteurs. Nous avons vu que dans certains cas, il vaut mieux privilégier un détecteur par rapport à l'autre et donc utiliser une méthode de fusion par pondération plutôt que le produit des détecteurs.

Dans ce chapitre, nous avons donc exploré des perspectives intéressantes : d'une part, nous avons mis en évidence la possibilité de recaler à l'aide d'un classificateur SVM à partir d'une base d'apprentissage d'objets recalés. D'autre part, nous avons montré l'importance d'utiliser l'information contextuelle pour le recalage. Ces travaux restent toutefois préliminaires et méritent d'être approfondis sur une plus grande base de test.

Chapitre 6

Conclusion et Perspectives

Conclusion

Nous avons abordé, dans cette thèse, la problématique de la fusion d'images optique et radar de milieu urbain à des résolutions métriques. Nous reprenons ici les conclusions des différents travaux abordés : le recalage global de données optiques et radar, la classification de l'image optique à partir des données conjointes et l'optimisation de chacune de ces opérations. Nous présentons également des perspectives.

Recalage global de données optiques et radar en milieu urbain

Tout d'abord nous avons présenté une méthode de recalage originale des données optiques et radar basée sur l'information contenue dans l'image (Chapitre 3), ce qui est une étape essentielle pour pouvoir analyser conjointement les données. La méthode proposée se décompose en deux parties et ne requiert aucun paramètre capteur. Dans la première partie, nous avons choisi d'utiliser une approche hiérarchique pour déterminer les paramètres d'une transformation rigide et globale. Nous mettons en correspondance des éléments de chaque image correspondant aux contours de routes et bâtiments par l'invariant de Fourier-Mellin. L'approche hiérarchique nous permet d'avoir une première estimation des paramètres de transformation sur des éléments visibles à basse résolution, comme les routes. Nous affinons ensuite la recherche en tenant compte des éléments de plus petite échelle. La translation est ensuite calculée, elle correspond à la translation qui maximise l'information mutuelle entre les données.

La seconde partie de notre chaîne de recalage consiste à affiner le recalage en déterminant la transformation polynomiale résiduelle, pour cela des points d'amers sont

extraits de l'image optique. Nous avons utilisé pour extraire ces points, le détecteur de coin de Harris, car nous recherchons des points particuliers avec une structure sur laquelle s'appuyer pour guider le recalage. Nous avons ensuite recherché les translations locales résiduelles en ces points à l'aide de l'information mutuelle. Cette seconde étape permet de corriger les déformations non rigides (cisaillement) qui peuvent apparaître entre les données optiques et radar.

Finalement, nous obtenons un recalage avec une erreur moyenne d'une dizaine de pixels au sol, soit moins d'une dizaine de mètres, sur plusieurs jeux de données. Afin d'améliorer encore le recalage, nous sommes arrivés à la conclusion, qu'il nous fallait passer à un recalage local par objets et tenir compte de leur contexte spatial et de leur nature car un objet en hauteur sera projeté différemment d'un objet au sol, sur l'image radar.

Notons que la chaîne de recalage proposée a été validée sur différents jeux de données issues de plusieurs capteurs aéroportés et satellitaires et a montré son efficacité. De plus, la méthode automatique fournit des résultats plus satisfaisants que la projection par paramètres capteurs. Dans ces travaux, nous nous sommes intéressés à l'influence de certains paramètres utiles pour le calcul de l'information mutuelle (taille de fenêtre de calcul et quantification) et nous avons montré expérimentalement l'influence que ceux-ci peuvent avoir dans le résultat de l'information mutuelle.

Classification de l'image optique à partir de données optiques et radar conjointes

Nous nous sommes ensuite intéressés à la classification de l'image optique à partir des données optiques et radar conjointes et avons proposé une méthode pour cela. Dans un premier temps, nous avons segmenté l'image optique, dans un deuxième temps, il nous faut déterminer des caractéristiques propres à nos objets qui serviront de primitives à la classification. Aussi nous avons extrait des caractéristiques de l'image optique ainsi que de l'image radar et nous avons étudié leur pertinence. Un des interêts d'utiliser des objets est la possibilité d'en extraire des caractéristiques géométriques, dans notre cas : périmètre, surface et compacité. Nous avons également utilisé les filtres de Gabor, reconnus pour simuler le système visuel humain, afin d'extraire des primitives texturales de l'image optique. Nous avons en outre extrait des images radar, des primitives statistiques : les log-Cumulants, calculés sur les objets obtenus lors de la segmentation de l'image optique projetée sur l'image radar. L'apport des différents types de primitives a été évaluée par l'analyse discriminatoire de Fisher.

Enfin, nous avons testé le comportement de plusieurs classificateur et nous avons porté notre choix sur les SVM qui présentent les meilleures performances en terme d'erreur de classification. D'un point de vue expérimental, nous ne disposons pas de base d'apprentissage sur les images de télédétection que nous avons à traiter, aussi nous avons besoin de l'intervention de l'utilisateur pour indiquer une cinquantaine d'objets de chaque classe nécessaires à la classification. Etant donné le milieu urbain que nous étudions, nous avons considéré cinq classes : bâtiments, routes, végétation (arbres, forêts), sol et ombre optique. Ce choix nous permet d'identifier, entre autres, les objets de la scène en hauteur, que sont les bâtiments et la végétation, utiles pour améliorer le recalage.

La chaîne de classification que nous avons proposée a été testée sur plusieurs couples de données. En ce qui concerne la zone du CNES, nous avons mis au point une procédure de validation quantitative et qualitative des résultats. Nous avons montré expérimentalement la robustesse de la chaîne de classification face à la segmentation, ainsi que l'amélioration conséquente qu'apportaient les primitives radar (log-Cumulants) pour la classification de l'image optique, malgré l'erreur résiduelle de recalage. Nous pouvons donc supposer qu'un recalage exact améliorerait les résultats de classification.

Optimisation du recalage et de la classification

Ceci nous a amenés à réfléchir sur la dépendance entre les opérations de recalage et de classification. En effet, le recalage peut être amélioré, si nous connaissons la nature de l'objet que nous souhaitons recaler. Par ailleurs, pour déterminer avec le plus de précision possible la nature d'un objet lors de la classification, il est utile d'avoir des données bien recalées. Nous avons donc cherché à optimiser ces opérations, pour cela nous proposons deux approches innovantes.

La première approche consiste à recaler et classer chaque objet simultanément, elle se base sur l'hypothèse "qu'ayant donné en apprentissage à un classificateur SVM des données bien recalées, si nous testons plusieurs translations pour un segment, la sortie du classificateur ayant le score maximum correspond à la bonne translation et à la bonne classe." Cette approche a été testée pour recaler et détecter des bâtiments, en utilisant des primitives géométriques, texturales issues de l'image optique ainsi que des primitives statistiques radar.

Nous avons montré expérimentalement l'interêt de cette approche, en particulier la possibilité pour un classificateur SVM de discriminer des bâtiments bien recalés.

Toutefois, cette approche mérite d'être approfondie, en effet lors du bon recalage le classificateur renvoie un score elevé, mais qui ne correspond pas toujours au maximum. Notre étude expérimentale montre également l'importance des log-Cumulants dans le recalage et la classification des bâtiments, car ce sont eux qui font la différence d'une translation à l'autre.

La seconde approche a consisté à effectuer les opérations de recalage et de classification successivement et à introduire des informations contextuelles. Nous avons cherché à recaler un objet, connaissant sa classe ainsi que des a priori contextuels sur le recalage. Nous avons appliqué cela à la classe bâtiment, où notre a priori est "Un bâtiment possède une ligne brillante sur un bord et une zone d'ombre à l'opposé." Nous avons donc proposé deux détecteurs, correspondant à chacun de nos a priori, que nous avons fusionné.

Cette approche ne donne de bons résultats que lorsque notre a priori est parfaitement vérifié. Nous rencontrons des limites dans le cas de plusieurs bâtiments juxtaposés, où il est également difficile pour un interprète de déterminer le bon recalage. Nous voyons également le besoin d'approfondir la méthode, notamment dans le critère de fusion, car suivant les bâtiments, nos deux a priori n'ont pas la même pertinence.

Perspectives

La méthode de recalage proposée est intéressante, de part sa qualité de résultat, mais aussi le fait qu'elle soit automatique. Nous avons étudié l'influence de la taille de fenêtre dans le calcul de l'information mutuelle et nous avons vu que nous obtenons de bons résultats avec une grande taille de fenêtre. Cependant, plus la taille de fenêtre est grande, plus elle intègre de déformations (relief, bâtiments...) et donc ne vérifie pas notre hypothèse d'ue translation locale. De plus, évaluer l'information mutuelle sur une grande fenêtre est coûteux en temps de calcul, aussi nous souhaitons déterminer de manière optimale les paramètres, propres à chaque image, permettant de réduire au maximum la fenêtre d'analyse.

Nous avons montré que la chaîne de classification développée était indépendante de la segmentation. Cependant elle est importante pour l'étape de recalage, nous avons vu qu'il était difficile de recaler et classer conjointement des bâtiments mal segmentés, il en va de même lorsque l'on souhaite recaler les bâtiments à l'aide d'a priori sur l'environnement. Une meilleure segmentation apporterait de meilleur résultats lors du recalage et de la
classification conjoints. En particulier, nous avons montré l'importance de l'information portée par l'image radar et les log-Cumulants pour la détection des bâtiments, il serait intéressant d'utiliser cette information, en plus de l'image optique pour la segmentation des bâtiments.

Lors de l'étape d'optimisation du recalage et de la classification, nous proposons deux approches prometteuses. Lorsque nous classons et recalons simultannément, le peu de points d'apprentissage explique la marge d'erreur autour de la bonne translation. Des méthodes itératives, faisant intervenir l'utilisateur, peuvent améliorer cela (méthode d'apprentissage actif, par exemple par boucle de pertinence). En ce qui concerne l'approche contextuelle, d'autres pistes restent à explorer, notamment ajouter l'information contextuelle dans le classificateur SVM. Par exemple, pour la classe bâtiment, ajouter les cartes de détection de coins réflecteurs et d'ombres. Nous pouvons également envisager, un recalage contextuel qui tienne compte des objets voisins et de leurs positions.

D'autre part, le recalage local des bâtiments nous permet d'extraire une information sur leur hauteur. En géométrie radar, plus le bâtiment est décalé par rapport au sol, plus il est élevé. Ainsi, connaissant l'angle de visée du capteur radar et la résolution, nous pouvons donc envisager une reconstruction 3D de la scène à partir de la translation détecté entre le sol et le bâtiment.

6. Conclusion et Perspectives

Publications

Conférences

Registration of metric resolution SAR and Optical images in urban areas, G. Lehureau, F.Tupin, C. Tison, G. Oller et D. Petit, EUSAR, 2008.

Combining SAR and Optical features in a SVM classifier for man-made structures detection, G. Lehureau, M. Campedel, F. Tupin, C. Tison et G. Oller, IGARSS, 2009.

6. Conclusion et Perspectives

Bibliographie

- [Ali et Clausi, 2002] ALI, M. et CLAUSI, D. (2002). Automatic registration of SAR and visible band remote sensing images. *In IEEE International Geoscience and Remote Sensing Symposium, Toronto, Canada*, volume 3.
- [Amberg, 2005] AMBERG, V. (2005). Analyse de scènes péri-urbaines à partir d'images radar haute résolution, application à l'extraction semi-automatique du réseau routier.
 Thèse de doctorat, PhD thesis, Institut National Polytechnique de Toulouse.
- [Baillard et Maitre, 1999] BAILLARD, C. et MAITRE, H. (1999). 3-D reconstruction of urban scenes from aerial stereo imagery : A focusing strategy. *Computer Vision and Image Understanding*, 76(3):244–258.
- [Basly, 2000] BASLY, L. (2000). Télédétection pour la qualité de l'air en milieu urbain. *Thèse de doctorat, Université de Nice Sophia-Antipolis.*
- [Bellens et al., 2008] BELLENS, R., GAUTAMA, S., MARTINEZ-FONTE, L., PHILIPS, W., CHAN, J. et CANTERS, F. (2008). Improved classification of VHR images of urban areas using directional morphological profiles. *IEEE Transactions on Geoscience and Remote Sensing*, 46(10 Part 1):2803–2813.
- [Benediktsson et al., 2003] BENEDIKTSSON, J., PESARESI, M. et ARNASON, K. (2003). Classification and feature extraction for remote sensing images from urban areas based on morphological transformations. *IEEE Transactions on Geoscience and Remote Sensing*, 41(9):1940–1949.
- [Benediktsson *et al.*, 1990] BENEDIKTSSON, J., SWAIN, P. et ERSOY, O. (1990). Neural network approaches versus statistical methods in classification of multisource remote sensing data. *IEEE Transactions on Geoscience and Remote Sensing*, 28(4):540–552.
- [Beucher et Lantuéjoul, 1979] BEUCHER, S. et LANTUÉJOUL, C. (1979). Use of watersheds in contour detection. *In International Workshop on image processing, real-time edge and motion detection/estimation, Rennes*, pages 17–21.

- [Bloch, 2003] BLOCH, I. (2003). Fusion d'informations en traitement du signal et des images. *Hermes Science Publication, Paris*.
- [Bogdanov et al., 2005] BOGDANOV, A., SANDVEN, S., JOHANNESSEN, O., ALEXAN-DROV, V. et BOBYLEV, L. (2005). Multisensor approach to automated classification of sea ice image data. Geoscience and Remote Sensing, IEEE Transactions on, 43(7):1648–1664.
- [Boser *et al.*, 1992] BOSER, B., GUYON, I. et VAPNIK, V. (1992). A training algorithm for optimal margin classifiers. *In Proceedings of the fifth annual workshop on Computational learning theory*, pages 144–152. ACM New York, NY, USA.
- [Breiman et al., 1984] BREIMAN, L., FRIEDMAN, J. et OLSHEN, R. (1984). Classification and regression trees. WADSWORTH ADV. BOOK PROG., 10 DAVIS DR., BEL-MONT, CA 94002, USA.
- [Briem *et al.*, 2002] BRIEM, G., BENEDIKTSSON, J. et SVEINSSON, J. (2002). Multiple classifiers applied to multisource remote sensing data. *IEEE Transactions on Geoscience and Remote Sensing*, 40(10):2291–2299.
- [Brown, 1992] BROWN, L. G. (1992). A survey of image registration techniques. ACM Comput. Surv., 24(4):325–376.
- [Brown et al., 1996] BROWN, R., WOODING, M., BATTS, A., FOOK, L., NOH, K. et NORDIN, L. (1996). Complementary Use of Ers/sar and Optical Data for Land Cover Mapping in Johor, Malaysia. In ERS Applications, Proceedings of the Second International Workshop held 6-8 December, London, volume 383.
- [Canny,] CANNY, J. A computational approach to edge detection. *Readings in Computer Vision : Issues, Problems, Principles and Paradigms*, pages 184–203.
- [Cerra et Datcu, 2008] CERRA, D. et DATCU, M. (2008). A Model Conditioned Data Compression Based Similarity Measure. In Proceedings of the Data Compression Conference. IEEE Computer Society Washington, DC, USA.
- [Cheng et Toutin, 1997] CHENG, P. et TOUTIN, T. (1997). Urban planning using data fusion of satellite and aerial photoimages. *1997 IEEE International Geoscience and Remote Sensing*, *1997. IGARSS'97.*, 2.
- [Chesnel, 2008] CHESNEL, A. (2008). Quantification de dégâts sur le bâti liés aux catastrophes majeures par images satellite multimodales très haute résolution. *Thèse de doctorat, Ecole Nationale Supérieure des Mines de Paris*.

- [Chini *et al.*, 2009] CHINI, M., PIERDICCA, N. et EMERY, W. (2009). Exploiting sar and vhr optical images to quantify damage caused by the 2003 bam earthquake. *Geoscience and Remote Sensing, IEEE Transactions on*, 47(1):145–152.
- [Cord *et al.*, 2002] CORD, M., JORDAN, M., BELLI, T. et VIEIRA, M. (2002). Analyse d'images aériennes haute resolution pour la reconstruction de scènes urbaines. *Bulletin de la Société Francaise de Photogrammetrie et de Télédetection*, pages 34–43.
- [Dell'Acqua et Gamba, 2003] DELL'ACQUA, F. et GAMBA, P. (2003). Texture-based characterization of urban environments on satellite SAR images. *IEEE Transactions on Geoscience and Remote Sensing*, 41(1):153–159.
- [Denis *et al.*, 2009] DENIS, L., TUPIN, F., DARBON, J. et SIGELLE, M. (2009). Sar image regularization with fast approximate discrete minimization. *Image Processing, IEEE Transactions on*, 18(7):1588–1600.
- [Duda et Hart, 1972] DUDA, R. O. et HART, P. E. (1972). Use of the hough transformation to detect lines and curves in pictures. *Commun. ACM*, 15(1):11–15.
- [Eugenio et al., 2004] EUGENIO, F., MARCELLO, J. et MARQUES, F. (2004). An automated multisensor satellite imagery registration technique based on the optimization of contour features. In IEEE International Geoscience and Remote Sensing Symposium, 2004. IGARSS'04., volume 2.
- [Ferecatu et Boujemaa, 2007] FERECATU, M. et BOUJEMAA, N. (2007). Interactive remote-sensing image retrieval using active relevance feedback. *Geoscience and Remote Sensing, IEEE Transactions on*, 45(4):818–826.
- [Fischler et Bolles, 1981] FISCHLER, M. et BOLLES, R. (1981). Random sample consensus : A paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, 24(6):381–395.
- [Gianinetto et al., 2004] GIANINETTO, M., GIUSSANI, A., LECHI, G. et SCAIONI, M. (2004). "Fast Mapping" from High Resolution Satellite Images : A Sustainable Approach to Provide Maps for Developing Countries. In XXth ISPRS Congress, Commission VI, Istanbul, volume 6, pages 12–23.
- [Goncalves et Dowman, 2002] GONCALVES, J. et DOWMAN, I. (2002). Precise orientation of SPOT panchromatic images with tie points to a SAR image. *International archives of Photogrammetry Remote Sensing ans Spatial Information Sciences*, 34(3/A):125–130.

- [Guigues, 2004] GUIGUES, L. (2004). Modeles multi-échelles pour la segmentation d'images. *Thèse de doctorat, Ecole doctorale Sciences et Ingénierie de l'Université de Cergy-Pontoise, Cergy-Pontoise.*
- [Guigues et al., 2006] GUIGUES, L., TRIAS-SANZ, R., CHEHATA, N., TAILLANDIER, F. et DEVEAU, M. (2006). Segmentation multi-échelles d'images : théorie et applications. Bulletin d'Information Scientifique & Technique de l'IGN, 75(1):41.
- [Harris et Stephens, 1988] HARRIS, C. et STEPHENS, M. (1988). A combined corner and edge detector. *In Alvey vision conference*, volume 15, page 50. Manchester, UK.
- [Hastie *et al.*, 2005] HASTIE, T., TIBSHIRANI, R., FRIEDMAN, J. et FRANKLIN, J. (2005). The elements of statistical learning : data mining, inference and prediction. *The Mathematical Intelligencer*, 27(2):83–85.
- [Inglada et Giros, 2004] INGLADA, J. et GIROS, A. (2004). On the possibility of automatic multisensor image registration. *Geoscience and Remote Sensing, IEEE Transactions on*, 42(10):2104–2120.
- [Kohavi, 1995] KOHAVI, R. (1995). A study of cross-validation and bootstrap for accuracy estimation and model selection. *In International joint Conference on artificial intelligence*, volume 14, pages 1137–1145. Citeseer.
- [Kusaka et al., 1990] KUSAKA, T., EGAWA, H. et KAWATA, Y. (1990). Classification of the SPOT image using spectral and spatial features of primitive regions that have nearly uniform color. *IEEE Transactions on Geoscience and Remote Sensing*, 28(4):749–752.
- [Lafarge *et al.*, 2004] LAFARGE, F., DESCOMBES, X. et ZERUBIA, J. (2004). Noyaux texturaux pour les problèmes de classification par SVM en télédétection.
- [Lambers *et al.*, 2008] LAMBERS, M., NIES, H. et KOLB, A. (2008). Interactive dynamic range reduction for sar images. *Geoscience and Remote Sensing Letters, IEEE*, 5(3): 507–511.
- [Le Hegarat-Mascle et al., 1997] LE HEGARAT-MASCLE, S., BLOCH, I. et VIDAL-MADJAR, D. (1997). Application of Dempster-Shafer evidence theory to unsupervised classification in multisource remote sensing. *IEEE Transactions on Geoscience* and Remote Sensing, 35(4):1018–1031.
- [Liénou *et al.*, 2006] LIÉNOU, M., MAITRE, H. et DATCU, M. (2006). Is it possible to automatically produce a CORINE land cover map from a single SPOT image. *ESA EUSC, Madrid, Espagne.*

- [Liu et al., 2008] LIU, J., CHEN, Z. et GUO, R. (2008). A mosaic method for aerial image sequence by r/c model. International Conference on Computer Science and Software Engineering, Los Alamitos, CA, USA, 6:58–61.
- [Lombardo et al., 2003] LOMBARDO, P., OLIVER, C., MACRI PELLIZZERI, T. et ME-LONI, M. (2003). A new maximum-likelihood joint segmentation technique for multitemporal sar and multiband optical images. *Geoscience and Remote Sensing, IEEE Transactions on*, 41(11):2500–2518.
- [Lucas et Kanade, 1981] LUCAS, B. et KANADE, T. (1981). An iterative image registration technique with an application to stereo vision. *In International joint conference on artificial intelligence, University of British Columbia Vancouver, Canada*, volume 3, page 3.
- [Luo, 2007] LUO, B. (2007). *Echelle et résolution en imagerie de télédétection*. Thèse de doctorat, Télécom ParisTech, 2007.
- [Manjunath et Ma, 1996] MANJUNATH, B. et MA, W. (1996). Texture features for browsing and retrieval of image data. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18(8):837–842.
- [Melgani et Bruzzone, 2004] MELGANI, F. et BRUZZONE, L. (2004). Classification of hyperspectral remote sensing images with support vector machines. *IEEE Transactions on Geoscience and Remote Sensing*, 42(8):1778–1790.
- [Monasse, 2000] MONASSE, P. (2000). Morphological representation of digital images and application to registration. *Thèse de doctorat, Paris IX*.
- [Nicolas, 2003] NICOLAS, J. (2003). La loi de Fisher et ses applications aux images de Radar à Synthèse d'Ouverture. *GRETSI, Groupe d'Etudes du Traitement du Signal et des Images*.
- [Nicolas, 2008] NICOLAS, J. (2008). Les bases de l'imagerie satellitaire. Rapport technique, Technical report, Télécom Paris : polycopié de cours, 2008.
- [Oller *et al.*, 2006] OLLER, G., PETIT, D. et INGLADA, J. (2006). On the use of SAR and optical images combination for scene interpretation. *Revue française de photo-grammétrie et de télédétection*, (182):87–91.
- [Platt, 1999] PLATT, J. (1999). Probabilistic outputs for support vector machines and comparison to regularized likelihood methods. *Advances in Large Margin Classifiers*, pages 61–74.

- [Reddy et Chatterji, 1996] REDDY, B. et CHATTERJI, B. (1996). An FFT-based technique for translation, rotation, and scale-invariant image registration. *IEEE Transactions on Image Processing*, 5(8):1266–1271.
- [Roche, 2001] ROCHE, A. (2001). Recalage d'images médicales par inférence statistique. *Thèse de doctorat, Université de Nice Sophia-Antipolis*.
- [Rousseeuw et Leroy, 1987] ROUSSEEUW, P. et LEROY, A. (1987). Robust regression and outlier detection. *John Wiley & Sons, Inc. New York, NY, USA*, page 329.
- [Rudin *et al.*, 1992] RUDIN, L., OSHER, S. et FATEMI, E. (1992). Nonlinear total variation based noise removal algorithms. *Physica D*, 60(1-4):259–268.
- [Sarrut, 2000] SARRUT, D. (2000). Recalage multimodal et plate-forme d'imagerie médicale à acces distant. *These de doctorat, Université Lumiere, Lyon.*
- [Schowengerdt, 2007] SCHOWENGERDT, R. (2007). Remote sensing : models and methods for image processing.
- [Serpico et Roli, 1995] SERPICO, S. et ROLI, F. (1995). Classification of multisensor remote-sensing images by structured neural networks. *IEEE Transactions on Geoscience and Remote Sensing*, 33(3):562–578.
- [Shabou, 2006] SHABOU, A. (2006). Recalage rigide et non-rigide entre images multicapteurs radar/optique.
- [Shabou et al., 2007] SHABOU, A., TUPIN, F. et CHAABANE, F. (2007). Similarity measures between sar and optic data. In IEEE International Geoscience and Remote Sensing Symposium, Barcelone, Espagne., pages 4858–4861.
- [Shekhar *et al.*, 1999] SHEKHAR, C., GOVINDU, V. et CHELLAPPA, R. (1999). Multisensor image registration by feature consensus. *Pattern Recognition*, 32(1):39–52.
- [Sohn et Dowman, 2002] SOHN, G. et DOWMAN, I. (2002). Terrain surface reconstruction by the use of tetrahedron model with the MDL criterion. *International Archives of Photogrammetry Remote Sensing and Spatial Information Sciences*, 34(3/A):336–344.
- [Sohn et Dowman, 2007] SOHN, G. et DOWMAN, I. (2007). Data fusion of highresolution satellite imagery and LiDAR data for automatic building extraction. *ISPRS Journal of Photogrammetry and Remote Sensing*, 62(1):43–63.
- [Solberg, 1999] SOLBERG, A. (1999). Contextual data fusion applied to forest map revision. *Geoscience and Remote Sensing, IEEE Transactions on*, 37(3):1234–1243.

- [Solberg et Jain, 1997] SOLBERG, S. et JAIN, A. (1997). Texture fusion and feature selection applied to SAR imagery. *IEEE Transactions on Geoscience and Remote Sen*sing, 35(2):475–479.
- [Thiele et al., 2007] THIELE, A., CADARIO, E., SCHULZ, K., THONNESSEN, U. et SOERGEL, U. (2007). Building recognition from multi-aspect high-resolution insar data in urban areas. *IEEE Transactions on Geoscience and Remote Sensing*, 45(11):3583.
- [Tison, 2004] TISON, C. (2004). *Interférométrie RSO à haute résolution en milieu urbain : application au calcul de MNS urbain.* Thèse de doctorat, Ecole Nationale Supérieure des Télécommunications, Paris.
- [Tison *et al.*, 2004] TISON, C., NICOLAS, J., TUPIN, F. et MAÎTRE, H. (2004). A new statistical model for Markovian classification of urban areas in high-resolution SAR images. *IEEE transactions on Geoscience and Remote Sensing*, 42(10):2046–2057.
- [Tupin et Galland, 2004] TUPIN, F. et GALLAND, F. (2004). Recalage d'images optique et radar par extraction de primitives, rapport technique. *École Nationale Supérieure des Télécommunications, rapport de l'étude CNES DCT/SI/AP*, pages 04–259.
- [Tupin *et al.*, 1998] TUPIN, F., MAITRE, H., MANGIN, J., NICOLAS, J. et PECHERSKY,
 E. (1998). Detection of linear features in SAR images : application to road network extraction. *IEEE Transactions on Geoscience and Remote Sensing*, 36(2):434–453.
- [Tupin et Roux, 2003] TUPIN, F. et ROUX, M. (2003). Detection of building outlines based on the fusion of SAR and optical features. *ISPRS Journal of Photogrammetry and Remote Sensing*, 58(1-2):71–82.
- [Ünsalan et Boyer, 2004] ÜNSALAN, C. et BOYER, K. (2004). Classifying land development in high resolution satellite imagery using hybrid structural–multispectral features. *IEEE Trans. Geosci. Remote Sens*, 42(12):2840–2850.
- [Van den Elsen et al., 1993] Van den ELSEN, P., POL, E. et VIERGEVER, M. (1993). Medical image matching-A review with classification. *IEEE Engineering in Medicine and Biology Magazine*, 12(1):26–39.
- [Veropoulos et al., 1999] VEROPOULOS, K., CAMPBELL, C. et CRISTIANINI, N. (1999). Controlling the sensitivity of support vector machines. In Proceedings of the international joint conference on Artificial Intelligence, volume 60. Citeseer.
- [Viveros-Cancino, 2003] VIVEROS-CANCINO, O. (2003). Analyse du milieu urbain par une approche de fusion de donnees satellitaires optiques et radar.

- [Waske et Benediktsson, 2007] WASKE, B. et BENEDIKTSSON, J. (2007). Fusion of support vector machines for classification of multisensor data. *IEEE transactions on Geoscience and Remote Sensing*, 45(12):3858.
- [Wong et Clausi, 2007] WONG, A. et CLAUSI, D. (2007). ARRSI : Automatic registration of remote-sensing images. *IEEE Transactions on Geoscience and Remote Sensing*, 45(5):1483.
- [Zammit et al., 2007] ZAMMIT, O., DESCOMBES, X. et ZERUBIA, J. (2007). Assessment of different classification algorithms for burnt land discrimination. In IEEE International Geoscience and Remote Sensing Symposium, 2007. IGARSS 2007, pages 3000–3003.
- [Zeng et Gao, 2009] ZENG, Z. et GAO, J. (2009). Improving SVM Classification with Imbalance Data Set. *In Neural Information Processing*, pages 389–398.
- [Zitová et Flusser, 2003] ZITOVÁ, B. et FLUSSER, J. (2003). Image registration methods : a survey. *Image and Vision Computing*, 21(11):977 – 1000.