THESE

présentée à

L'UNIVERSITE DE NICE - SOPHIA ANTIPOLIS
Ecole Doctorale des Sciences Pour l'Ingénieur

pour obtenir le titre de

Docteur en Sciences

spécialité

Sciences de l'Ingénieur

par

Marc MANGOLINI

Sujet :

APPORT DE LA FUSION D'IMAGES SATELLITAIRES MULTICAPTEURS AU NIVEAU PIXEL EN TELEDETECTION ET PHOTO-INTERPRETATION

Soutenue le 15 novembre 1994 devant le jury composé de :

MM.	**Albert PAPIERNIK**	Président
	Roger MANIERE	Rapporteurs
	Marcel RAFFY	
	Albert BIJAOUI	Examinateurs
	Jean-Marc NASR	
	Christian ROUCHOUZE	
	Jean-Paul RUDANT	
	Lucien WALD	

AVANT-PROPOS

Cette thèse a été réalisée sous contrat CIFRE entre Aérospatiale et l'Ecole des Mines de Paris. Elle s'est déroulée à Aérospatiale dans l'établissement de Cannes, au sein du service "Télédétection et Traitement d'Images" que dirige M. Jean-Marc NASR.

L'encadrement scientifique a été assuré par le laboratoire "Télédétection & Modélisation" de l'Ecole des Mines de Paris, situé à Sophia Antipolis et dirigé par M. Lucien WALD.

De plus, ce travail a été mené à bien dans le cadre universitaire de la formation doctorale "Propagation, Télécommunications et Télédétection" de l'Université de Nice - Sophia Antipolis, dirigée par M. Albert PAPIERNIK.

La proximité géographique de ces entités et la complémentarité des compétences qu'elles représentent ont grandement contribué au bon déroulement de la thèse.

Afin de respecter la confidentialité industrielle et militaire, ce mémoire a été soumis à l'approbation de la hiérarchie et du service de sécurité d'Aérospatiale Cannes, qui ont autorisé sa diffusion.

REMERCIEMENTS

Je voudrais tout d'abord remercier M. Albert PAPIERNIK d'avoir bien voulu présider le jury de cette thèse, après avoir accepté mon inscription dans sa formation doctorale. J'exprime également ma reconnaissance à MM. Roger MANIERE et Marcel RAFFY qui ont accepté d'exercer le rôle de rapporteurs, et à MM. Albert BIJAOUI, Christian ROUCHOUZE, et Jean-Paul RUDANT, pour m'avoir fait l'honneur de prendre part au jury.

Je remercie sincèrement M. Jean-Marc NASR, chef du service Télédétection et Traitement d'Images d'Aérospatiale Cannes, qui a été à l'origine de la thèse et m'a fait confiance pendant ces trois années. C'est grâce à son concours que j'ai pu goûter à la recherche, tout en baignant dans un milieu industriel de pointe. Je tiens à exprimer ma profonde reconnaissance à mon directeur de thèse, M. Lucien WALD, qui a suivi et orienté ce travail avec une disponibilité et une sympathie qui m'ont beaucoup touché. Il m'a fait découvrir ce qu'est un travail de recherche, et si j'ai parfois maudit la rigueur au nom de laquelle il me faisait recommencer ou compléter un travail, je l'en remercie *a posteriori*. Sa contribution à cette thèse a été fondamentale, et j'espère qu'à l'avenir nous aurons de nouveau l'occasion de travailler ensemble.

Je remercie tous mes collègues de service, en compagnie desquels j'ai passé d'excellents moments, et qui m'ont toujours apporté dans la bonne humeur leur soutien technique et moral. Je témoigne en particulier ma sympathie à Philippe PEREZ dont j'ai partagé le bureau, et qui m'a fait profiter de sa vision "philosophique" de la fusion de données. Je remercie Roberto ALOISI qui a assuré la valorisation industrielle des méthodes développées dans la thèse. Merci aussi à Pierre ARMAND et Alain FRIZON de m'avoir aidé à déboguer mes programmes dans les cas désespérés, et à Eric SAVARIA pour les discussions constructives concernant les outils de traitements d'images. Un grand merci à Laurent POLIDORI qui a spontanément proposé de relire cette thèse ; il pourra constater que ses remarques ont été prises en compte !

Je garderai également un très bon souvenir des moments passés au Groupe Télédétection & Modélisation de l'Ecole des Mines. Une pensée particulière pour Thierry RANCHIN avec qui j'ai mené un travail commun passionnant, et pour Michel ALBUISSON dont les conseils avisés tant en informatique qu'en télédétection ont été précieux.

Enfin, merci à tous ceux qui n'ont pas été cités ci-dessus, mais qui ont contribué d'une manière ou d'une autre au bon déroulement de la thèse.

TABLE DES MATIERES

Chapitre 1 :

INTRODUCTION 7

Chapitre 2 :

DONNEES UTILISEES ET ASPECTS DE FUSION ABORDES 11

2.1 INTRODUCTION 11

2.2 DONNEES UTILISEES 11

2.2.1 Un choix limité 11

2.2.2 Les capteurs retenus 12

2.2.3 Prétraitements nécessaires 15

2.3 DIFFERENTS NIVEAUX DE FUSION ET METHODES ASSOCIEES 17

2.4 FORMALISME DE FUSION ADOPTE 20

Chapitre 3 :

UNE SOLUTION A LA FUSION MULTIRESOLUTION : LA METHODE ARSIS 23

3.1 INTRODUCTION 23

3.2 DESCRIPTION DE LA METHODE ARSIS 25

3.2.1 Prétraitements et hypothèses 25

3.2.2 Description de la méthode 26

3.2.3 Influence de la fonction de transfert de l'instrument et des effets atmosphériques 33

3.3 APPLICATION A LA FUSION SPOT P+XS 35
3.3.1 Introduction 35
3.3.2 Résultats numériques 36
3.3.3 Résultats visuels 42
3.3.4 Conclusion 42
3.4 APPLICATION A L'ESTIMATION DE L'INDICE DE VEGETATION 49
3.4.1 Définition du NDVI 49
3.4.2 Mise en œuvre d'ARSIS 50
3.5 APPLICATION A D'AUTRES RAPPORTS DE RESOLUTION 52
3.6 VALIDITE DE L'ESTIMATION DE LA QUALITE 54
3.7 CONCLUSION 55

Chapitre 4 :

PRESENTATION VISUELLE DU VOLUME DE DONNEES 59
4.1 INTRODUCTION 59
4.2 CHANGEMENT D'ESPACE DE REPRESENTATION 60
4.3 UNE METHODE DE FUSION D'IMAGES OPTIQUE ET RADAR 79
4.4 CONCLUSION 82

Chapitre 5 :

CLASSIFICATION - ANALYSE DE DIFFERENTES ARCHITECTURES DE FUSION 89
5.1 INTRODUCTION 89
5.2 CLASSIFICATION D'IMAGES 90
5.2.1 Choix des classificateurs 90
5.2.2 Capteurs disponibles et classes choisies 91
5.2.3 Sélection des données d'apprentissage et des données test 94
5.2.4 Critères d'évaluation de la qualité des classifications 100

5.2.5 Prétraitements des images 104
5.2.5.1 Filtrage du speckle 104
5.2.5.2 Texture 104
5.3 DIFFERENTES ARCHITECTURES DE FUSION POUR LA CLASSIFICATION 107
5.3.1 Généralités 107
5.3.2 Architecture centralisée 107
5.3.3 Architecture décentralisée 108
5.3.4 Architecture mixte 111
5.4 ANALYSE DES RESULTATS 114
5.4.1 Introduction 114
5.4.2 Classification sur SPOT, ERS-1, et Landsat 115
5.4.2.1 Classifications monocapteurs 115
5.4.2.2 Architecture centralisée 117
5.4.2.3 Architecture décentralisée 117
5.4.2.4 Architecture mixte 119
5.4.3 Classification sur SPOT et ERS-1 122
5.4.4 Classification sur SPOT bruité et ERS-1 123
5.4.5 Synthèse des résultats 123
5.5 CONCLUSION 125

Chapitre 6 :

CONCLUSION 127

REFERENCES BIBLIOGRAPHIQUES 129

ANNEXES 137

Annexe A :

METHODES DE FUSION D'IMAGES MULTISPECTRALES MULTIRESOLUTIONS 139

A.1 Méthodes combinant les radiométries 139

A.1.1 Méthode P+XS du CNES 140

A.1.2 Autres méthodes 147

A.2 Méthodes dissociant information spectrale et information spatiale 149

Annexe B :

CLASSIFICATEUR K-PLUS PROCHES VOISINS 153

Annexe C :

RESEAUX DE NEURONES APPLIQUES A LA CLASSIFICATION 155

C.1 Introduction 155

C.2 Le perceptron multicouches 155

C.3 Classification multicapteurs 159

Annexe D :

RESULTATS COMPLEMENTAIRES SUR LES ARCHITECTURES DE FUSION 163

D.1 SITE DE FOS 164

D.1.1 CLASSIFICATEUR RESEAU DE NEURONES 164

D.1.1.1 Classifications monocapteurs 164

D.1.1.2 Architectures appliquées à SPOT, ERS, et Landsat 165

D.1.2 CLASSIFICATEUR K-PPV 166

D.1.2.1 Classifications monocapteurs 166

D.1.2.2 Architectures appliquées à SPOT, ERS, et Landsat 167

D.2 SITE DE MARIGNANE 168

D.2.1 CLASSIFICATEUR RESEAU DE NEURONES 168

D.2.1.1 Classifications monocapteurs 168
D.2.1.2 Architectures appliquées à SPOT, ERS, et Landsat 169
D.2.2 CLASSIFICATEUR K-PPV 170
D.2.2.1 Classifications monocapteurs 170
D.2.2.2 Architectures appliquées à SPOT, ERS, et Landsat 171

LISTE DES FIGURES 173

Chapitre 1

INTRODUCTION

On dispose aujourd'hui d'images de la Terre provenant de nombreux capteurs satellitaires de type optique (visible ou infrarouge) ou radar, représentant des informations variées et complémentaires. La mise au point de produits optimaux de télédétection et d'aide à la photo-interprétation passe par une utilisation conjointe de ces différentes images, en leur appliquant des méthodes dites de **fusion de données**. Ce terme, quoique récent, désigne une démarche adoptée spontanément dans de nombreux domaines. Une définition générale de la fusion de données peut être : **ensemble de méthodes et de moyens utilisant les données issues de plusieurs sources de natures différentes, en vue d'accroître la qualité (au sens très large) de l'information.** La fusion de données est exploitée par de nombreux systèmes biologiques, dont un exemple éloquent est le système humain qui fait appel à ses différents "sens". Dans le domaine scientifique, son utilisation à des fins d'observation et de description d'une scène s'étend en particulier à la robotique et à la médecine (vision multicapteurs) ainsi qu'à la télédétection (classification multisources). Nous développons dans la thèse les applications de la fusion de données spécifiques aux images de **télédétection spatiale**, mais les méthodes peuvent être largement étendues à des capteurs aéroportés, voire à d'autres domaines d'observation.

L'élaboration de méthodes de fusion performantes implique une prise en compte des spécificités des images de télédétection spatiale. La première particularité de telles images est leur contenu thématique, constitué de paysages de la surface terrestre, en général beaucoup plus complexes que les scènes artificielles traitées en robotique. Par ailleurs la plate-forme satellitaire, qui supporte les instruments de prise de vue, impose des contraintes opérationnelles liées à l'orbite (dates de prise de vue par exemple), et influe sur les caractéristiques géométriques des images (résolution spatiale liée notamment à l'altitude, distortions résultant de l'attitude du satellite, etc.). Enfin une contrainte élémentaire liée au domaine spatial est la prise en compte des milieux physiques traversés qui déterminent le choix du rayonnement électromagnétique utilisé. Dans le spectre optique par exemple, le rôle de l'atmosphère est essentiel, au point que

certaines longueurs d'onde non transmises sont inutilisables. Le type de rayonnement détermine aussi la grandeur physique mesurée au sol puisque, de manière générale, l'onde électromagnétique interagit avec des objets dont la taille est de l'ordre de la longueur d'onde. La télédétection spatiale exploite principalement trois types de rayonnements :

- les ondes émises par des sources naturelles et réfléchies par les objets au sol, qui sont captées par des instruments passifs → visible et proche infrarouge,
- les ondes émises par les corps eux-mêmes également reçues par des instruments passifs → infrarouge thermique,
- les ondes émises en direction de la Terre à l'aide d'un instrument actif, qui mesure ensuite la partie réfléchie par le sol → radar microondes.

Les images issues du spectre visible présentent l'intérêt d'être aisément interprétables, puisqu'elles correspondent à la perception visuelle humaine. Le proche infrarouge est très utilisé pour l'étude de la végétation, alors que l'infrarouge thermique traduit la température des corps et présente l'avantage de fonctionner de jour comme de nuit. Enfin le radar à ouverture synthétique (ROS ou SAR en anglais) mesure une grandeur physique (surface équivalente radar ou SER) plus complexe à interpréter, qui dépend notamment de la pente du terrain, de sa rugosité, et des caractéristiques diélectriques du sol. Son intérêt tient à sa complémentarité thématique avec la mesure optique, et à ses capacités toute heure et tout temps.

Les méthodes de fusion de données mises en œuvre dans la thèse s'appliquent à des données images, c'est-à-dire bidimensionnelles. Ces données pourront être hétérogènes, ce terme désignant une différence de support électromagnétique et par conséquent de thématique. Nous utilisons des images provenant des spectres visible, proche infrarouge, et microondes. L'infrarouge thermique est écarté car nous disposons uniquement d'images Landsat TM6, dont la résolution spatiale (120 m) est très grossière comparativement à celles des autres images utilisées (résolutions décamétriques). La plupart des méthodes présentées dans la thèse sont toutefois applicables aux images infrarouges thermiques, et une illustration en est donnée dans Mangolini *et al.* (1992).

Il existe de nombreuses méthodes de fusion d'images qui sont généralement classées d'après le niveau auquel elles opèrent la fusion. La fusion de "bas niveau" combine les données brutes, proches du signal physique mesuré. On peut alternativement extraire des informations d'un plus "haut niveau", au sens sémantique

du terme, avant de les fusionner. Plus la fusion intervient à un niveau sémantique élevé, plus le volume de données est réduit, ce qui peut être rendu nécessaire par des contraintes opérationnelles telles que la limitation des débits de transmission, en cas de délocalisation de l'information. Mais l'extraction d'attributs ou primitives dans les images impose une sélection de l'information, donc des pertes qui peuvent dégrader les performances du système. Face à la grande diversité des techniques de fusion existantes, nous avons dû restreindre notre champ d'investigation à un niveau de fusion particulier, et notre choix s'est porté sur la fusion de "bas niveau". Une des motivations de ce choix tient au vaste éventail d'applications envisageables en utilisant les mêmes méthodes de fusion ; l'extraction d'attributs sémantiques impose au contraire de s'attacher à une application thématique spécifique. Lors de l'étude des architectures décentralisées de fusion, nous serons néanmoins amenés à utiliser des attributs images, mais ceux-ci restent de bas niveau sémantique dans le sens où ils sont localisés au niveau du pixel : on exclut donc les attributs de type formes ou régions. Plus précisément, la thèse développe donc des **méthodes de fusion d'images satellitaires au niveau pixel**.

Les méthodes de fusion développées concernent la constitution du volume de données, les pré-traitements, la combinaison des données proprement dite, ainsi que la présentation visuelle des résultats. Les applications ont porté sur l'aide à la photo-interprétation et sur la classification multicapteurs. La **photo-interprétation** est une discipline qui vise à reconstituer une réalité terrain à partir d'images de la Terre (aéroportées ou satellitaires). Du fait de la multitude d'informations contenues dans des images de la surface terrestre, la photo-interprétation touche un grand nombre de domaines thématiques tels que l'agriculture, la géologie, l'urbanisme, ou encore le renseignement militaire. Pour reconstituer un modèle de la réalité, le photo-interprète dispose souvent, parallèlement aux images, d'informations annexes comme des cartes ou des études de terrain préalables. A partir de ces données, il effectue une analyse de la scène, en faisant appel à ses connaissances et surtout à son expérience dans le domaine où il est spécialisé, qui lui permet d'établir des relations complexes entre les objets observés. Le fruit de ce travail peut se présenter sous des formes diverses : plan renseigné, estimation de paramètres physiques du terrain, dénombrement d'un type d'objets, changements relativement à une situation antérieure, etc. La classification est un aspect particulier de la photo-interprétation, qui vise à réaliser une carte d'occupation des sols, généralement sur des zones étendues.

Suite au présent chapitre d'introduction, le chapitre 2 détaillera les aspects de fusion envisagés dans la thèse, et présentera les données utilisées. Nous préciserons les contraintes associées à la fusion au niveau pixel, pour qu'elle puisse être appliquée à des images multicapteurs multidates, initialement non superposables, et de résolutions spatiales variées.

Le chapitre 3 proposera une méthode originale de fusion d'images multispectrales multirésolutions, nommée ARSIS. Les images obtenues par ARSIS pourront être soit considérées comme un produit final, soit utilisées en entrée d'autres traitements nécessitant des images de même résolution.

Nous examinerons ensuite au chapitre 4 comment présenter visuellement le volume de données au photo-interprète, afin qu'il puisse en extraire au mieux l'information qu'il recherche. Nous montrerons l'intérêt de méthodes spécifiques à la fusion de données hétérogènes optiques et radar, par rapport aux méthodes usuelles de présentation de données homogènes de type optique.

Le chapitre 5 sera consacré à la comparaison de différentes architectures de fusion, communément désignées sous le nom d'architectures centralisée et décentralisée. L'application thématique retenue est la classification multicapteurs qui permet, à l'aide d'une vérité terrain, d'effectuer une comparaison quantitative des performances des deux architectures. Nous proposerons également une architecture "mixte" (ou hybride), qui s'inspire des deux précédentes.

Enfin les annexes présenteront certaines méthodes de fusion issues de la littérature auxquelles il est fait référence, décriront des algorithmes codés pour les besoins de la thèse, ou encore apporteront des résultats complémentaires pouvant intéresser le lecteur.

Chapitre 2

DONNEES UTILISEES ET ASPECTS DE FUSION ABORDES

2.1 INTRODUCTION

Nous présentons tout d'abord dans ce chapitre les capteurs utilisés pour la mise au point et l'évaluation des méthodes de fusion développées dans la thèse. Les différents niveaux de fusion possibles sont alors détaillés, et nous dressons un rapide panorama des méthodes associées. Le cas de la fusion au niveau pixel ayant été retenu, on abordera les contraintes et prétraitements nécessaires à la mise en œuvre de ce type de fusion. On exposera ensuite le formalisme adopté, utilisé pour décrire les architectures des méthodes de fusion sous forme de schémas.

2.2 DONNEES UTILISEES

2.2.1 Un choix limité

Comme il a été précisé en introduction, les capteurs imageurs spatiaux opérationnels ou disponibles dans un futur proche opèrent dans les domaines électromagnétiques optique - visible et infrarouge - et microondes pour le radar. Nous avons souhaité expérimenter les méthodes de fusion sur des données aussi variées que possible, appartenant aux différents domaines précédents. Un grand nombre de contraintes d'ordres divers ont cependant limité le choix des images. Tout d'abord, le nombre de satellites opérationnels dont on peut commander des images est fort limité. Dans le domaine optique, les familles de satellites SPOT et Landsat sont pratiquement les seules à disposer d'une large distribution commerciale de leurs images. En ce qui concerne les images radar à ouverture synthétique (SAR en anglais), ERS-1 bénéficie de la meilleure diffusion en Europe, et il est toujours opérationnel contrairement aux satellites ou missions américains Seasat et SIR-A ou SIR-B. On peut également se

procurer des images optiques ou SAR en provenance de satellites russes, japonais, ou indiens, mais la disponibilité de données répondant à nos besoins reste aléatoire. Les contraintes sur les données, liées à leur exploitation en fusion d'images, sont les suivantes :

- on doit pouvoir trouver des images d'un même site géographique en provenance de chaque capteur d'intérêt,
- les résolutions spatiales de ces images doivent être du même ordre de grandeur, pour que puissent leur être appliquées des méthodes de fusion au niveau pixel. On retiendra des capteurs de résolution décamétrique, majoritairement utilisés en photo-interprétation, ce qui exclut par exemple les images NOAA ou Météosat,
- les dates de prise de vue doivent être telles que la composition de la scène imagée par les divers capteurs soit comparable entre les images. Il faut bien noter que cette composition peut varier considérablement en quelques jours, ne serait-ce que par les conditions météorologiques ou l'emplacement des objets mobiles.

En pratique, la sélection des images a été effectuée en recherchant l'intersection des catalogues images des différents capteurs, ce qui a imposé un écart temporel minimal de plusieurs mois. Il est possible de mieux répondre aux contraintes précédentes par une commande, qui consiste à programmer les satellites retenus en fonction du besoin de l'utilisateur. Ce mode d'approvisionnement allonge les temps de livraison, mais il est surtout beaucoup plus onéreux qu'une commande sur catalogue. Il va de soi que cette solution n'a pu être adoptée pour les besoins de la thèse, pour laquelle il a fallu utiliser des images préalablement acquises par Aérospatiale ou le laboratoire d'accueil.

2.2.2 Les capteurs retenus

Les capteurs retenus sont SPOT HRV, Landsat TM, et l'instrument SAR de ERS-1. Nous rappelons brièvement leurs caractéristiques pour mieux en cerner les différences qui vont intervenir sur les traitements de fusion.

Les satellites SPOT fournissent des images dans le visible et le très proche infrarouge. Ils sont munis de deux instruments HRV identiques qui fonctionnent au choix en mode panchromatique ou multispectral. Le mode panchromatique est très

utilisé en photo-interprétation, en raison de sa résolution spatiale de 10 m et de l'excellente qualité de ses images. Les trois bandes du mode multispectral complètent thématiquement le panchromatique, en particulier pour l'étude de la végétation, mais leur résolution spatiale est limitée à 20 m. Les caractéristiques des bandes spectrales ainsi que certains paramètres liés à l'orbite sont résumés par la figure 2.1. Une caractéristique intéressante de SPOT est sa capacité de dépointage, qui permet d'augmenter la fréquence d'observation d'un site spécifique, et de d'obtenir des couples stéréoscopiques utilisés pour réaliser des modèles numériques de terrain (MNT). L'ensemble des caractéristiques de SPOT sont rassemblées dans Anonyme (1986).

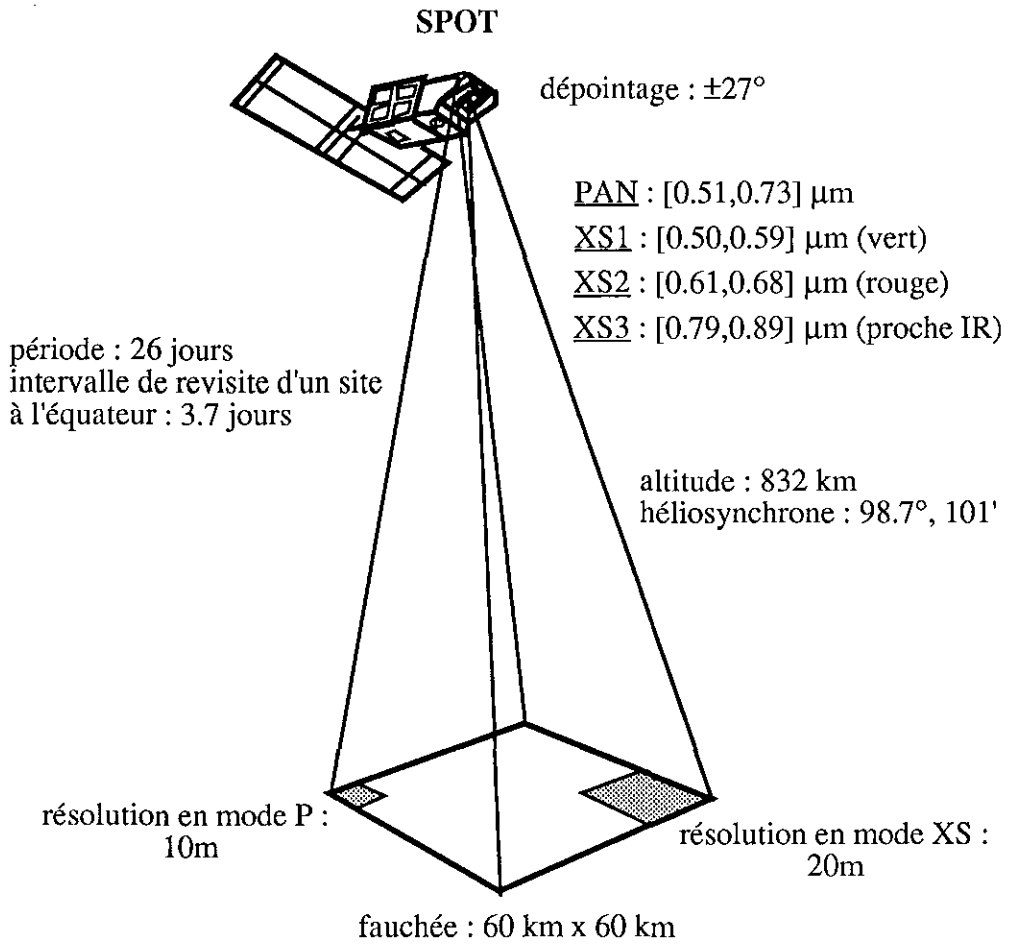


Figure 2.1 : caractéristiques synthétiques de SPOT 1, 2, et 3

Les satellites américains Landsat disposent du capteur TM (Thematic Mapper) qui possède six bandes visible et proche infrarouge à une résolution de 30 m, plus une bande infrarouge thermique à 120 m. Cette dernière sera délaissée pour les applications de fusion traitées ici en raison de sa médiocre résolution spatiale. Les avantages de Landsat sur SPOT tiennent à un plus grand nombre de bandes spectrales, qui s'étendent

du bleu au proche infrarouge, et à une couverture au sol dix fois plus importante. Par contre, SPOT garde l'avantage pour la résolution spatiale tant en mode panchromatique qu'en multispectral. La figure 2.2 résume les caractéristiques du capteur TM de Landsat-5.

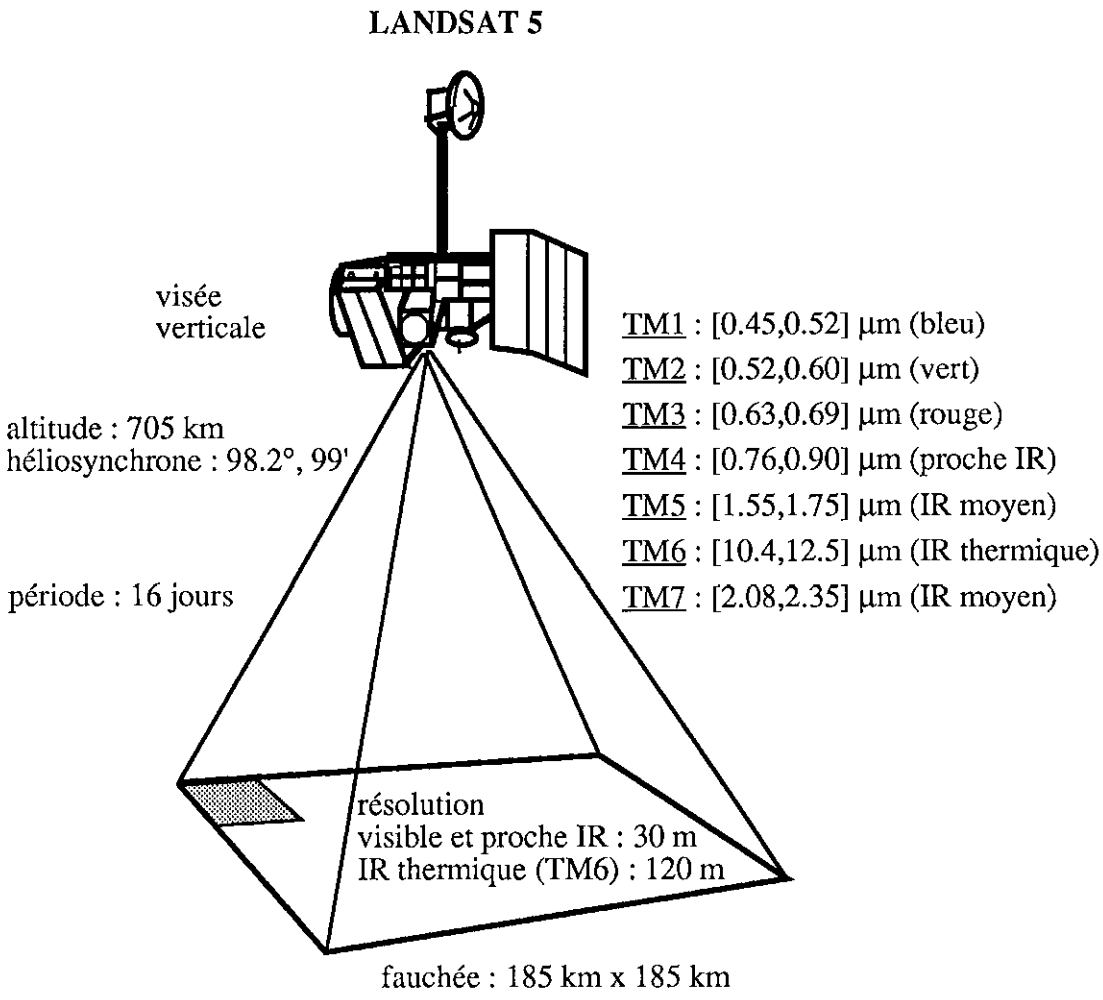


Figure 2.2 : caractéristiques synthétiques de Landsat 5

Enfin le capteur SAR de ERS-1 était initialement conçu pour l'étude des glaces et des océans, mais les besoins de la communauté scientifique ont élargi son succès à l'étude des paysages terrestres. Les applications prévues ont déterminé les caractéristiques du système radar, qui sont plus variables et plus restrictives que celles des systèmes optiques : le choix de la bande de fréquence et de la polarisation par exemple influe radicalement sur les paramètres mesurés. Nous ne développerons pas ici les principes de l'imagerie SAR qui sont supposés connus ; le lecteur peut se reporter

par exemple à Elachi (1988), ou Ulaby (1982) pour une description exhaustive. Les principales caractéristiques de ERS-1 sont données par la figure 2.3.

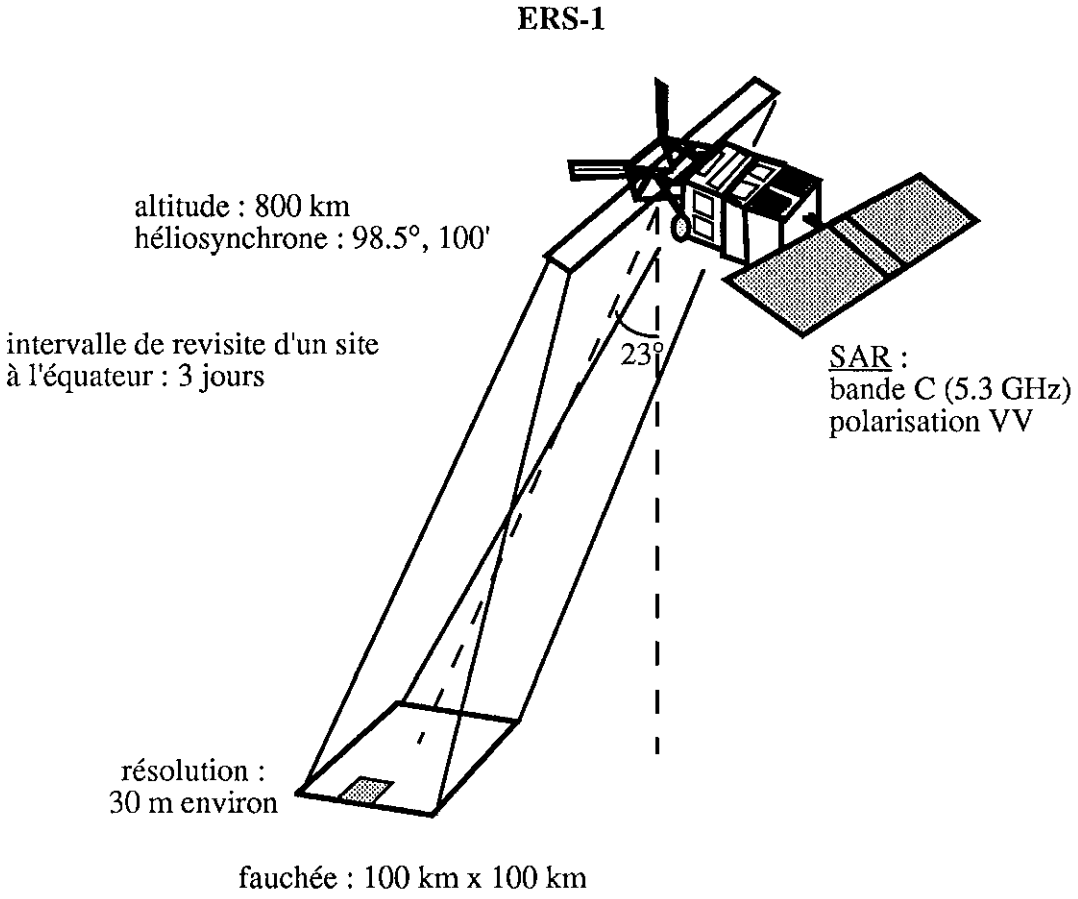


Figure 2.3 : caractéristiques synthétiques de ERS-1

Les trois capteurs précédents possèdent des géométries de prise de vue très différentes et, comme il a déjà été précisé, ils mesurent des grandeurs physiques distinctes. Préalablement à l'application de méthodes de fusion au niveau pixel, il est nécessaire d'effectuer des prétraitements destinés à rendre les images comparables au niveau du pixel.

2.2.3 Prétraitements nécessaires

Les algorithmes de fusion au niveau pixel exigent que les images sources soient rendues parfaitement superposables. Les images SPOT, Landsat, et ERS sont disponibles à différents niveaux de traitement, qui diffèrent en particulier par le système

de projection adopté. On peut commander directement au distributeur, pour chaque capteur, des images géocodées c'est-à-dire projetées dans un repère cartographique. Cependant ces images ne sont pas corrigées des effets du relief, et le traitement capteur par capteur ne garantit pas une bonne superposabilité finale de toutes les images. Deux approches sont alors possibles :

- on peut recaler géométriquement les images par prise de points d'appui qui servent à l'estimation d'un modèle de déformation, souvent polynômial. La prise de points d'appui est généralement manuelle, mais des méthodes récentes permettent sous certaines hypothèses une sélection automatique (Djamdji *et al.* 1993). La géométrie finale est celle de l'une des images de départ, et ne correspond pas nécessairement à un repère cartographique. Cette approche est la plus simple à mettre en œuvre, mais elle n'est pas performante sur les régions comportant du relief, surtout en ce qui concerne l'image radar. C'est par ce moyen qu'ont été recalées les images SPOT P et XS de Barcelone et Toulouse traitées respectivement dans le chapitre 3 et l'annexe A,
- une approche plus élaborée est la création d'ortho-images à l'aide de données relatives à la position du capteur, et d'un modèle numérique de terrain (MNT) de la scène. Cette approche, parfois superflue pour des images optiques d'une zone à faible relief, est pratiquement indispensable pour le recalage optique-radar. Le principe en est décrit par exemple dans Toutin *et al.* (1992). Une telle méthode a été adoptée pour la création des ortho-images SPOT, Landsat, et ERS sur Fos et Marignane (sud-est de la France), utilisées dans les chapitres 3, 4, et 5.

Les images recalées sont généralement rééchantillonnées à un pas proche de leur résolution d'origine. Au delà de l'adoption d'une projection géométrique commune, il faut uniformiser le pas des différentes images pour obtenir une correspondance pixel à pixel. Les solutions classiques proposent un rééchantillonnage au pas le plus fin disponible dans le jeu d'images, généralement par plus proche voisin, ou par interpolation bilinéaire ou bicubique. Nous proposons alternativement une méthode originale qui, au lieu de rééchantillonner chaque image indépendamment, fusionne les images pour simuler un jeu d'images à la meilleure résolution disponible (cf. chap. 3).

Un prétraitement souvent requis en fusion optique-radar est le filtrage du bruit de speckle de l'image radar. Le speckle peut être considéré comme un bruit

multiplicatif, décrit par exemple dans Elachi (1988). Pour les applications de fusion radiométrique (cf. chap. 4) ou de classification (cf. chap. 5) particulièrement, le speckle pollue le résultat de fusion. Nous appliquons dans ce cas un filtrage adapté, tel que l'algorithme proposé par Lopes *et al.* (1990). Par contre pour les applications qui exploitent les petites structures de l'image radar, comme la méthode de fusion décentralisée optique-radar décrite au chapitre 4, le filtrage du speckle est à proscrire car il altère ces structures.

Un dernier prétraitement, ou plutôt traitement annexe, mérite d'être signalé bien qu'il ne soit pas indispensable à toutes les applications de fusion d'images, mais plutôt dédié à la classification (cf. chap. 5). Ce traitement consiste à extraire une information sur l'environnement du pixel, c'est-à-dire des paramètres de variation locale de la radiométrie. Cette information souvent appelée texture augmente notablement les potentialités de la fusion de pixel, tout en restant un attribut de bas niveau sémantique. La texture est exploitée tant sur les images optiques que radar. Le travail de base de Haralick (1979) distingue les approches statistique et structurelle. Barber *et al.* (1993) comparent différentes textures appliquées à la classification sur des images SAR. Des approches récentes font appel au concept simplificateur d'"unité de texture" (He, Wang 1990), ou à l'outil de transformée en ondelettes (Du *et al.* 1992).

Les données utilisées ayant été présentées, nous allons maintenant décrire les différents niveaux de fusion possibles, et les grandes classes de méthodes qui s'y rattachent.

2.3 DIFFERENTS NIVEAUX DE FUSION ET METHODES ASSOCIEES

Nous avons expliqué dans le premier chapitre que la fusion peut être effectuée à différents niveaux : on parle de fusion "bas niveau" lorsque les données combinées sont proches des paramètres physiques mesurés, et de fusion "haut niveau" si l'on traite des attributs sémantiques extraits de ces mesures. Plus exactement, pour des données de type image, on distinguera la fusion de pixels, la fusion d'attributs, et la fusion de décisions. La figure 2.4 illustre ces différents types de fusion pour trois images sources notées S_1, S_2, et S_3. L'extraction de l'information, de type décisionnelle, est effectuée de bas en haut dans les trois schémas de fusion, qui se différencient par le niveau auquel intervient la fusion. On désigne ici par "attributs" des caractéristiques extraites des

images sources, qui ne sont pas localisées en chaque pixel, mais mettent en jeu des critères complexes tels que les contours définissant des objets ou des régions, la forme de ces objets ou régions, ou encore leur positionnement les uns par rapport aux autres.

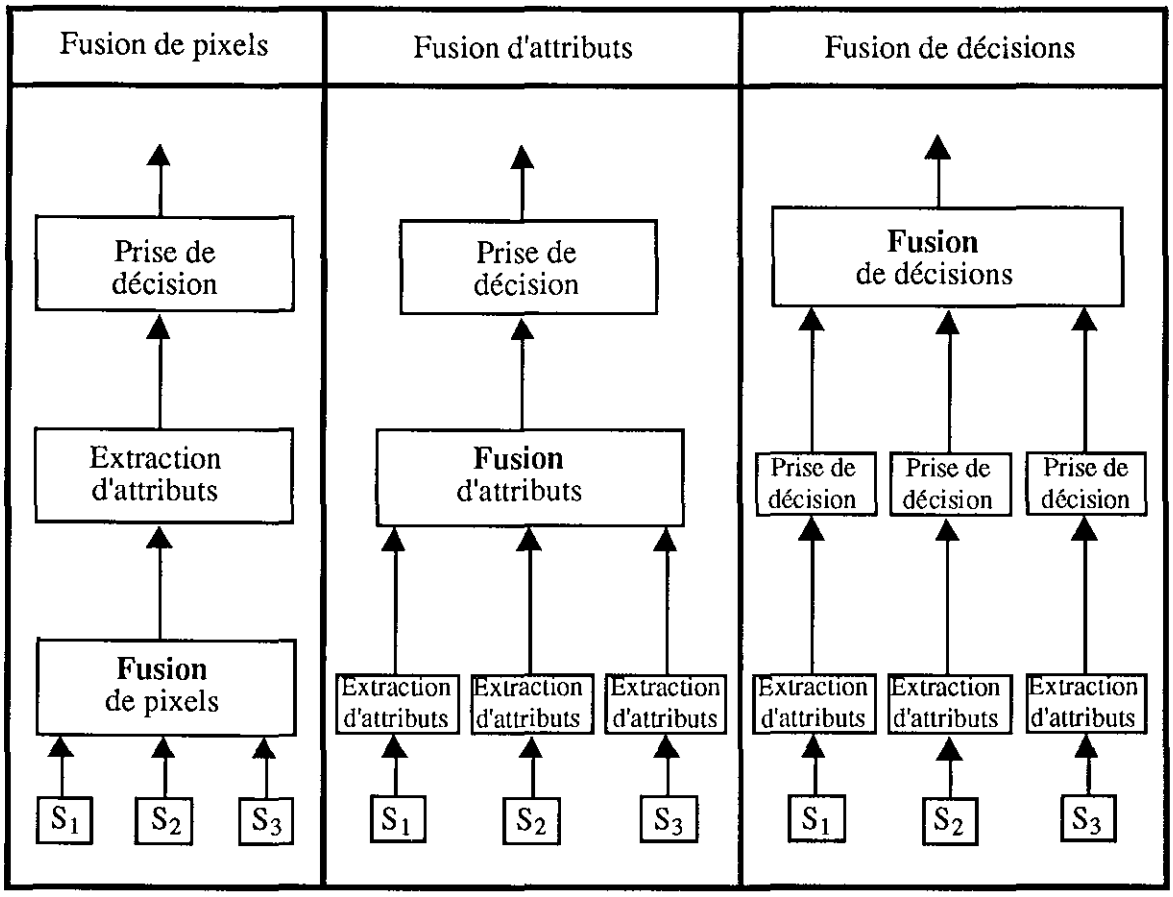


Figure 2.4 : trois types complémentaires de fusion d'images

L'avantage de la fusion de pixels sur les deux autres types de fusion est d'éviter toute perte d'information inhérente à l'extraction d'attributs. Ses contraintes tiennent à la nécessité d'un recalage géométrique précis entre les images, à la difficulté de comparaison de pixels hétérogènes (par exemple optique et radar), et à l'imposant flot de données à traiter qui nécessite de gros moyens de calcul. Le recalage géométrique est moins critique pour la fusion d'attributs, qui peut faire appel à des critères tels que la forme pour mettre en correspondance les structures extraites des différentes images. Le recalage n'est pas nécessaire à la fusion de décision, qui permet de traiter des images délocalisées spatialement. Ce facteur de délocalisation est déterminant pour le choix du type de fusion : dans le cas où les sources sont délocalisées, le volume important que représentent les images en regard des débits de transmission limités impose souvent une

fusion de décisions. Cependant, la fusion de décisions n'est pas optimale dans le sens où elle optimise chaque traitement mono-image individuellement, mais ne recherche pas la solution globalement optimale. Nous nous limitons dans le cadre de la thèse au cas de données colocalisées, traitées par des méthodes de fusion de pixels, et nous n'imposons pas de contraintes de traitements temps-réel.

Les méthodes de fusion d'images sont multiples et variées. Chaque méthode n'est pas nécessairement rattachée de manière exclusive à un type de fusion particulier, mais on peut dégager de grandes tendances. Ainsi, la fusion de pixels fait principalement appel à des techniques de traitement du signal (filtrage, transformée de Fourier, transformée en ondelettes, etc.) et d'analyse de données (analyse en composantes principales, analyse discriminante, etc.). Les approches statistiques concernent tant la fusion de pixels que la fusion d'attributs ; en classification multicapteurs par exemple, l'estimation bayésienne peut être appliquée à chaque vecteur pixel, composé des radiométries dans les différentes images, ou encore à des régions préalablement extraites de chaque image. Une généralisation de l'inférence bayésienne appliquée à des données au caractère probabiliste est proposée par la théorie de l'évidence (Shafer 1976), et peut être avantageusement utilisée en fusion de données (Appriou 1991). Les approches connexionistes faisant appel aux réseaux de neurones sont de plus en plus courantes, aussi bien en fusion de pixels (cf. annexe C) qu'en fusion d'attributs. Les principaux avantages des réseaux de neurones sur les méthodes précédentes sont leur capacité de généralisation à partir d'un nombre restreint d'exemples, et la possibilité d'une implantation sur des architectures informatiques parallèles, qui diminue les temps de calcul. Enfin les systèmes à base de connaissance, développés en intelligence artificielle, sont utilisés pour la fusion d'attributs et la fusion de décision. Ils mettent généralement en œuvre des architectures de type "tableau noir", ce terme désignant une zone à travers laquelle les différents modules échangent leurs données de manière interactive (Hayes-Roth 1983, Pau 1989).

La plupart des méthodes de fusion de pixels développées dans la thèse sont décomposables en un enchaînement de tâches élémentaires. La clarté de la description de ces méthodes et leur comparaison seront améliorées par une formalisation des entrées-sorties d'une tâche et des différentes façons d'enchaîner les tâches. Nous décrivons dans le paragraphe suivant le formalisme adopté dans le cadre de la thèse.

2.4 FORMALISME DE FUSION ADOPTE

Le formalisme retenu a été proposé par Houzelle, Giraudon (1994) pour décrire des méthodes de fusion d'attributs, mais il est également adapté aux méthodes de fusion de pixels. Chaque méthode de fusion peut être décomposée en un enchaînement de tâches élémentaires, qui se réduit dans le cas limite à une tâche unique. Chaque tâche élémentaire est modélisée par une "cellule de fusion", représentée sur la figure 2.5.

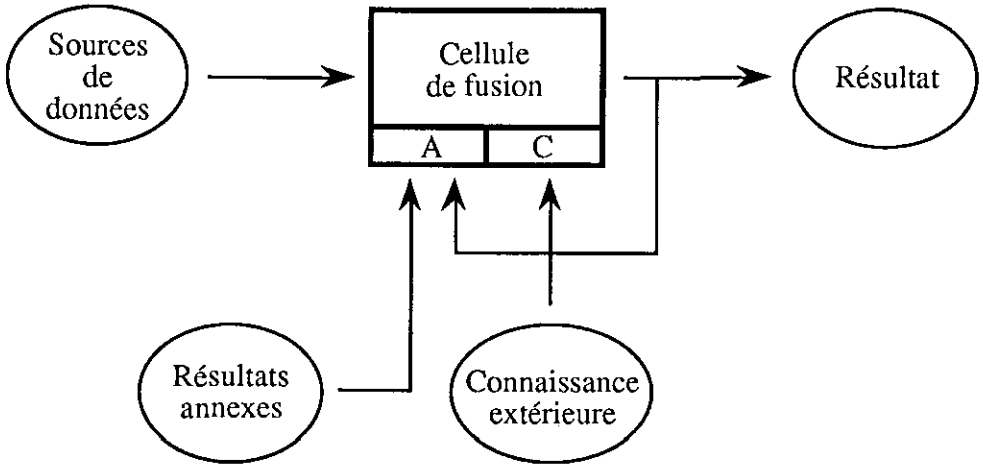


Figure 2.5 : cellule de fusion

La cellule de fusion reçoit trois sources d'informations différentes : les sources de données, qui dans notre cas sont les images sources, des résultats annexes provenant d'autres cellules de fusion, et des éléments de connaissance extérieure, tels que l'attitude du satellite, la couverture nuageuse, etc. Le rebouclage de la sortie de la cellule sur sa propre entrée permet de modéliser les processus itératifs, mais cette fonctionnalité ne sera pas utilisée dans la thèse. Le traitement réalisé par la cellule est un algorithme quelconque de fusion : filtrage, calcul statistique, réseau de neurones, etc.

La combinaison des algorithmes élémentaires qui constituent une méthode de fusion définit une architecture de fusion. Cette architecture peut être représentée par un schéma associant les cellules de fusion. Dans les chapitres suivants 4 et 5, les architectures utilisées seront décrites de cette manière. La figure 2.6 illustre l'utilisation de ce formalisme. Les images sources sont notées $\{S_i\}$, et les algorithmes de fusion $\{F_j\}$ sont indicés pour souligner qu'il peuvent être différents les uns des autres. Sur cet exemple, les sources S_1 à S_n font l'objet de traitements indépendants, qui fournissent des résultats annexes utilisés par la cellule F_0. On montre l'utilisation possible d'un résultat

annexe comme la texture, ou de connaissance extérieure telle que les coefficients d'étalonnage. La cellule F_0 utilise simultanément les sources S_{n+1} à S_p, ainsi que les résultats annexes précédemment cités, pour fournir le résultat final.

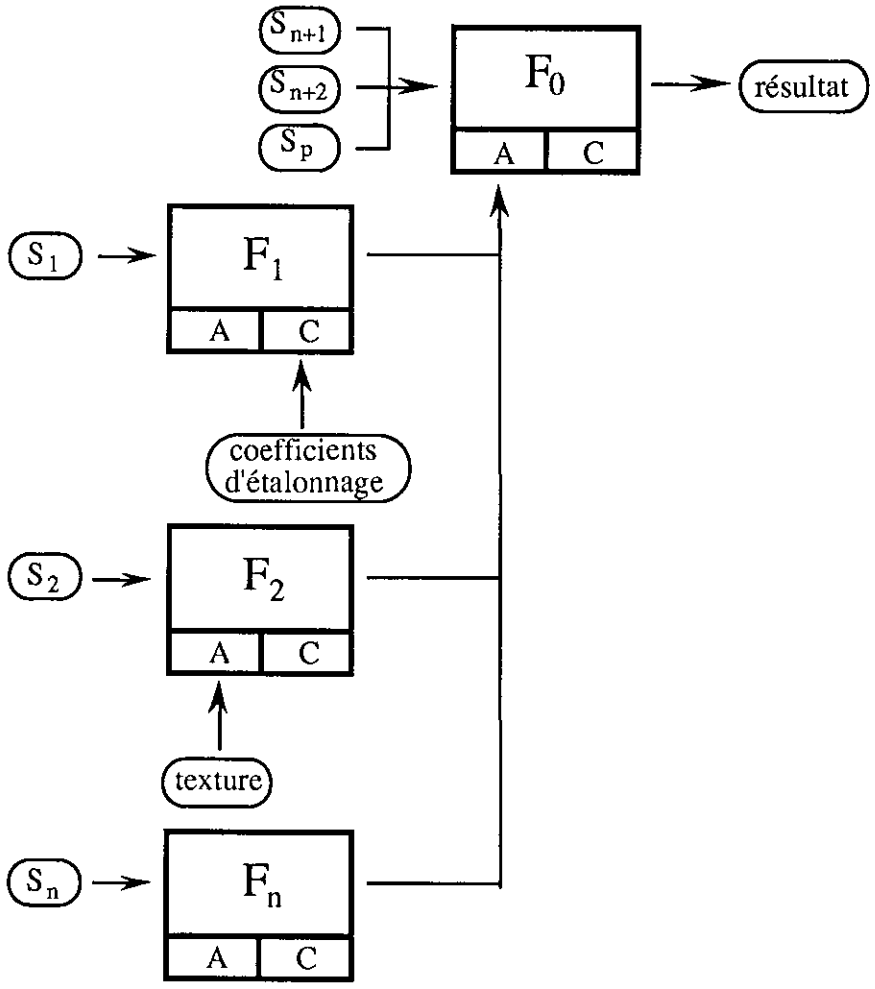


Figure 2.6 : exemple d'architecture de fusion

Chapitre 3

UNE SOLUTION A LA FUSION MULTI-RESOLUTION : LA METHODE ARSIS

3.1 INTRODUCTION

Comme il a été exposé au chapitre précédent, les images en provenance de différents capteurs satellitaires possèdent souvent des résolutions spatiales différentes. Ces images possèdent également des contenus thématiques variés, dépendant des bandes spectrales utilisées. L'utilisateur se trouve donc en possession d'un jeu d'images multispectrales de résolutions différentes, dont certaines lui apporteront une bonne résolution spatiale et d'autres une bonne résolution spectrale. Ces deux informations sont conjointement nécessaires à de nombreuses applications de télédétection, et en particulier à la cartographie : l'information spectrale permet par exemple de discriminer différentes végétations ou de caractériser la turbidité d'un fleuve, tandis que la haute résolution spatiale apporte des détails sur les zones hétérogènes, composées d'objets de taille comparable à cette résolution spatiale (par exemple bâtiments ou voies de communication sur une image SPOT panchromatique).

L'exploitation d'un tel jeu d'images numériques est peu commode, d'une part parce qu'elles n'ont pas la même taille, d'autre part parce qu'il faut extraire l'information spectrale et l'information spatiale d'images différentes. La première difficulté peut être résolue par suréchantillonnage des images de moindre résolution, de telle sorte que toutes les images possèdent le même pas, égal à celui de l'image de meilleure résolution. Il faut bien noter que le suréchantillonnage ne modifie pas la résolution des images, qui est liée au capteur lui-même, mais en change seulement le pas. On s'affranchira de la seconde difficulté d'exploitation, si l'on parvient à synthétiser des images multispectrales possédant la meilleure résolution spatiale disponible au sein du jeu de données. Les travaux de Raffy (1992, 1993), Woodcock et Strahler (1987) ou Kong et Vidal-Madjar (1988) notamment, ont montré qu'on ne pouvait pas extrapoler simplement des résultats statistiques ou des classifications acquis sur des images basse

résolution vers de plus hautes résolutions, sans erreur grossière. La méthode ARSIS, en fusionnant les images multirésolutions préalablement à leur exploitation thématique (par exemple en cartographie), permet de contourner ces difficultés.

De nombreuses méthodes ont déjà été proposées pour réaliser la fusion d'images multispectrales de résolutions spatiales différentes. Elles peuvent être classées en deux catégories : les premières combinent les radiométries (ou les luminances) des différentes bandes, alors que les secondes dissocient information spectrale et information spatiale. Nous montrons en annexe A que pour la première catégorie, l'information spectrale des images initiales est fortement altérée. Les méthodes existantes de la seconde catégorie présentent souvent le même défaut d'altération spectrale ; dans le cas contraire, elles ont l'inconvénient de requérir une très forte corrélation entre les images à fusionner, ce qui en limite le champ d'application (cf. annexe A). Nous avons donc développé une nouvelle méthode, désignée sous le nom d'ARSIS pour "accroissement de la résolution spatiale par injection de structures", élaborée dans le but de préserver l'information spectrale des images dont on améliore la résolution spatiale. Cette méthode appartient à la seconde catégorie définie ci-dessus, puisqu'elle extrait l'information spatiale de l'image de meilleure résolution pour la réinjecter dans les images multispectrales.

Après avoir décrit la méthode ARSIS, nous évaluerons ses performances dans le cas de la fusion d'images SPOT panchromatique (P) et multispectrales (XS). La méthode "P+XS" développée par le CNES (Anonyme 1986) servira alors de référence, car son exploitation commerciale par la société SPOT Image a largement contribué à sa vulgarisation. Les deux méthodes seront comparées principalement à l'aide de critères numériques, et l'aspect visuel, difficile à quantifier, ne sera que brièvement évoqué. L'utilisation d'ARSIS dans le but d'améliorer la résolution spatiale de l'indice de végétation NDVI sera alors examinée. On comparera pour cela deux mises en œuvre différentes de la méthode.

La méthode ARSIS a également été appliquée avec succès à des images issues d'autres capteurs optiques, présentant des bandes spectrales variées, et des rapports de résolution différents du cas SPOT P+XS. La fusion d'images Landsat TM et SPOT panchromatique en est un exemple. Par contre, l'application de la méthode ARSIS à la fusion hétérogène d'images optique et radar semble difficile, car les structures sont souvent trop dissimilaires pour être efficacement combinées.

Cette méthode est le résultat d'un travail commun entre Thierry Ranchin, Lucien Wald, et moi-même. Elle a déjà été publiée dans la thèse de Thierry Ranchin (1993), ainsi que dans des actes de colloques (Mangolini *et al.* 1993-a, Ranchin *et al.* 1993-a, Mangolini *et al.* 1993-b, Ranchin *et al.* 1993-b). Elle est également couverte par un brevet (Mangolini *et al.* 1992). La présentation que nous en faisons ici diffère de celle de Ranchin (1993), qui insiste davantage sur l'aspect traitement du signal et transformée en ondelettes. Le lecteur intéressé trouvera avantage à consulter les textes précités, qui donnent une description très détaillée de la méthode et de nombreux exemples.

3.2 DESCRIPTION DE LA METHODE ARSIS

3.2.1 Prétraitements et hypothèses

Le jeu d'images multispectrales est constitué d'une image au moins dont la résolution spatiale est la plus fine, et d'images de moins bonne résolution spatiale que l'on cherche à améliorer à l'aide de cette première image. Pour l'instant, la méthode ARSIS traite séparément chacune des images de moins bonne résolution spatiale. Elle extrait des informations spatiales de l'image de meilleure résolution, pour les injecter dans chacune de ces images. La méthode s'applique donc à un couple d'images de résolutions différentes. Les images doivent être recalées géométriquement, pour assurer une parfaite localisation de leurs structures au niveau pixel. Dans certains cas, aucun traitement informatique de recalage n'est nécessaire : c'est le cas lorsque les images sont issues d'un même capteur, conçu pour assurer la superposabilité géométrique dans les différentes bandes, et où la prise de vue est simultanée (par exemple pour les canaux de Landsat TM). Par contre, dans le cas de SPOT, les capteurs HRV utilisés en mode d'acquisition simultanée P et XS fournissent des images P et XS non recalées (Anonyme 1986).

De plus, la méthode ARSIS s'appuie sur les hypothèses suivantes :

- il doit exister une similitude entre les petites structures observées dans les différentes bandes spectrales. Cela n'implique pas nécessairement qu'il y ait recouvrement des bandes spectrales (comme c'est le cas pour SPOT P, XS1, et XS2), ni que le coefficient de corrélation des images soit très élevé (condition requise par d'autres méthodes décrites en annexe A),

- les images à fusionner peuvent avoir été prises à des dates différentes, à condition de conserver une bonne similitude de leurs structures. Sur certaines scènes dont le contenu thématique est lié aux saisons, la similitude des structures est parfois meilleure lorsqu'on augmente l'écart temporel afin qu'il coïncide avec la période d'évolution.

3.2.2 Description de la méthode

Comme nous avons déjà mentionné, la méthode ARSIS extrait des informations sur les structures de l'image de meilleure résolution spatiale, pour les injecter dans chacune des images multispectrales de moins bonne résolution spatiale. Elle utilise pour cela l'analyse multirésolution et la transformée en ondelettes. L'analyse multirésolution (Mallat 1989) est un concept qui permet de décrire une image par ses approximations successives à des résolutions de plus en plus grossières. Elle consiste donc à extraire de l'image les structures dont les fréquences spatiales sont contenues dans un intervalle donné. Son avantage sur les outils descriptifs que sont la transformée de Fourier ou les variogrammes (fonctions de structures) est de conserver la localisation spatiale des structures décrites. La transformée en ondelettes permet de modéliser la différence d'information entre deux approximations successives de cette analyse multirésolution. On peut ensuite effectuer une reconstruction parfaite de l'image initiale, à partir de ses approximations successives et des coefficients d'ondelettes.

La méthode ARSIS s'applique à un couple d'images de résolutions spatiales différentes ; nous noterons respectivement HR l'image de meilleure résolution et BR l'image de moins bonne résolution, et R_{HR} et R_{BR} leurs résolutions respectives. Le concept d'analyse multirésolution, qui permet de décrire chacune de ces images, peut être illustré par une pyramide. A la base de la pyramide se trouve l'image initiale, et l'on dispose de ses approximations à des résolutions de plus en plus grossières au fur et à mesure que l'on gravit la pyramide. Au sommet, l'approximation se réduit à un unique pixel. Les pyramides représentant respectivement les images HR et BR sont représentées sur la figure 3.1. La longueur du côté de leurs bases respectives diffère dans un rapport égal au rapport de leurs résolutions spatiales. Nous considérerons dans un premier temps le cas où ce rapport de résolutions est égal à 2, ce qui permet l'utilisation d'une ondelette dyadique. Dans ce cas, les résolutions successives d'une même pyramide (correspondants aux "étages") sont dans un rapport 2. Le but de la méthode est de prolonger la pyramide BR en utilisant l'information de la pyramide HR. On va pour ce faire utiliser les coefficients d'ondelettes contenant l'information sur les

structures d'échelles comprises entre les résolutions R_{HR} et R_{BR}. La méthode ARSIS va donc injecter dans l'image BR uniquement des structures manquantes aux petites échelles, sans modifier les autres structures déjà présentes.

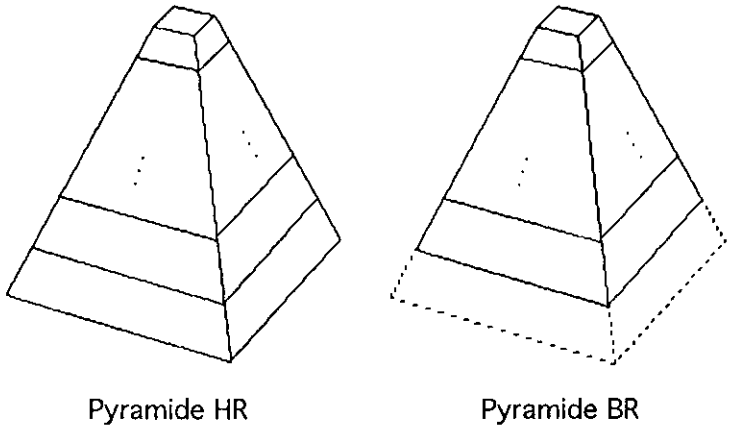


Figure 3.1 : pyramides multirésolutions

La solution la plus intuitive consiste à associer à l'image BR les coefficients d'ondelettes issus de la première décomposition de l'image HR, pour synthétiser une image BR+ de même résolution que HR, par transformée en ondelettes inverse. Cette méthode est illustrée par la figure 3.2, dans le cas où le rapport des résolutions R_{HR} et R_{BR} est de 2. L'algorithme d'analyse multirésolution utilisé fait appel à des ondelettes directionnelles (Mallat 1989). Il fournit, pour chaque décomposition, trois images de coefficients dans les directions respectivement horizontale (H), verticale (V), et diagonale (D). Les images de coefficients d'ondelettes issues de la première décomposition de l'image HR sont notées H(1), V(1), et D(1), et l'image de contexte correspondant au sous-échantillonnage de l'image HR à la résolution R_{BR} est notée HR(1). La transformée en ondelettes directe est notée T.O. et la transformée inverse $T.O.^{-1}$.

Cependant, les coefficients d'ondelettes sont fortement dépendants de l'image dont ils proviennent. Ceci explique que la méthode précédente d'identification des coefficients des images HR et BR ne donne pas des résultats très satisfaisants. Nous allons donc utiliser, en plus des coefficients de HR à la résolution R_{BR}, les décompositions des images HR et BR aux résolutions supérieures à R_{BR}. C'est le principe de la méthode ARSIS, illustré par la figure 3.3.

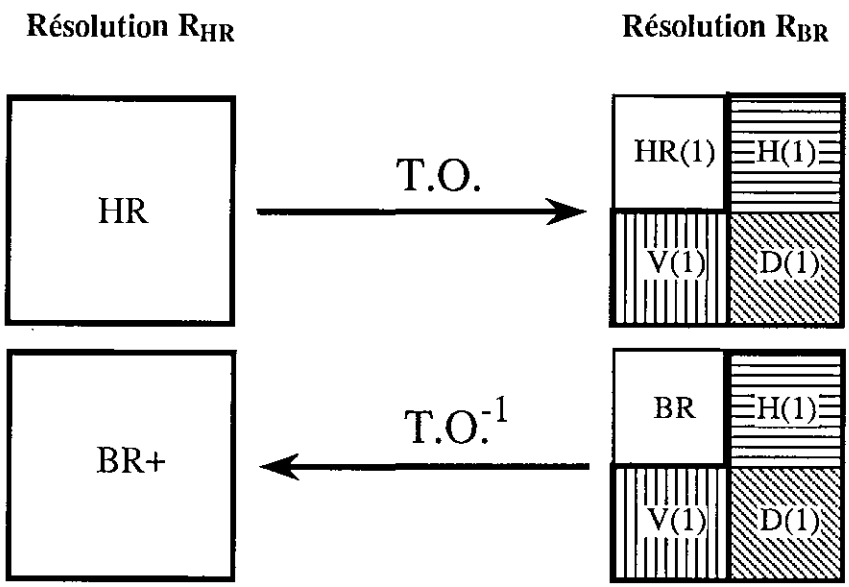


Figure 3.2 : méthode d'identification des coefficients des images HR et BR

Les coefficients d'ondelettes aux différentes résolutions des images HR et BR constituent les entrées d'un modèle, destiné à synthétiser les coefficients d'ondelettes représentant l'information aux échelles comprises entre R_{BR} et R_{HR}, que l'on souhaite ajouter à l'image BR. La résolution n° 3 correspond à une décomposition en ondelettes sur une image de résolution R_{BR} ; pour une ondelette dyadique, elle est égale à deux fois R_{BR}. Dans le cas de la fusion SPOT P+XS, les résolutions R_{HR}, R_{BR}, et la résolution n° 3 valent respectivement 10 m, 20 m, et 40 m. L'image BR+, de résolution spatiale R_{HR}, est obtenue par une transformée en ondelettes inverse sur les coefficients estimés et sur l'image BR. Il est important de noter que, par construction, l'approximation à la résolution R_{BR} de l'image BR+ est identique à l'image BR de départ.

Différents modèles ont été envisagés, et nous les regrouperons en deux catégories. Les premiers s'appliquent exclusivement aux coefficients d'ondelettes de HR et BR à une même résolution. Ils peuvent être qualifiés de modèles verticaux, relativement à la figure 3.3. Les seconds, qualifiés d'horizontaux, utilisent les coefficients d'ondelettes aux différentes résolutions des images HR et BR pour prédire les coefficients recherchés à la résolution R_{BR}. Ces derniers modèles ne sont pas décrits dans la thèse.

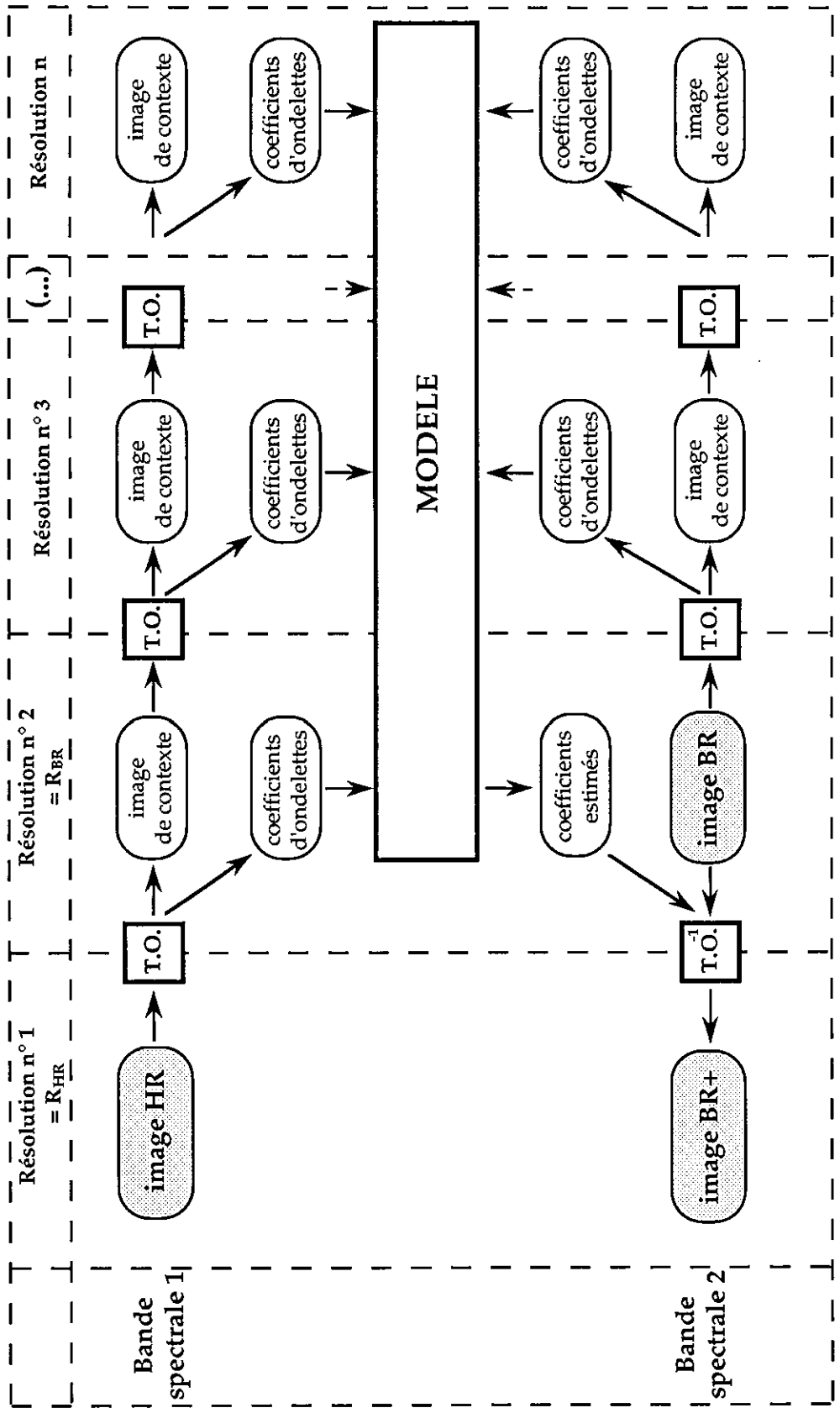


Figure 3.3 : schéma général de la méthode ARSIS

Pour un modèle de type vertical, la méthode ARSIS fait l'hypothèse que ce modèle, établi à des résolutions supérieures à R_{BR}, peut être appliqué à la résolution R_{BR}. Les paramètres du modèle peuvent être estimés à différentes résolutions supérieures à R_{BR}. Leur évolution avec la résolution est ensuite extrapolée pour en déduire leurs valeurs à la résolution R_{BR}. Dans le cas le plus simple de l'utilisation unique des coefficients d'ondelettes à la résolution n° 3, le schéma général de la méthode ARSIS se ramène à celui de la figure 3.4, lorsqu'on utilise les ondelettes directionnelles de Mallat (1989). Les coefficients estimés y sont notés $\tilde{H}_{BR}$, $\tilde{V}_{BR}$, et $\tilde{D}_{BR}$, et les coefficients calculés sont repérés par le nom de l'image dont ils sont issus (en indice) et le nombre de transformées en ondelettes effectuées (entre parenthèses).

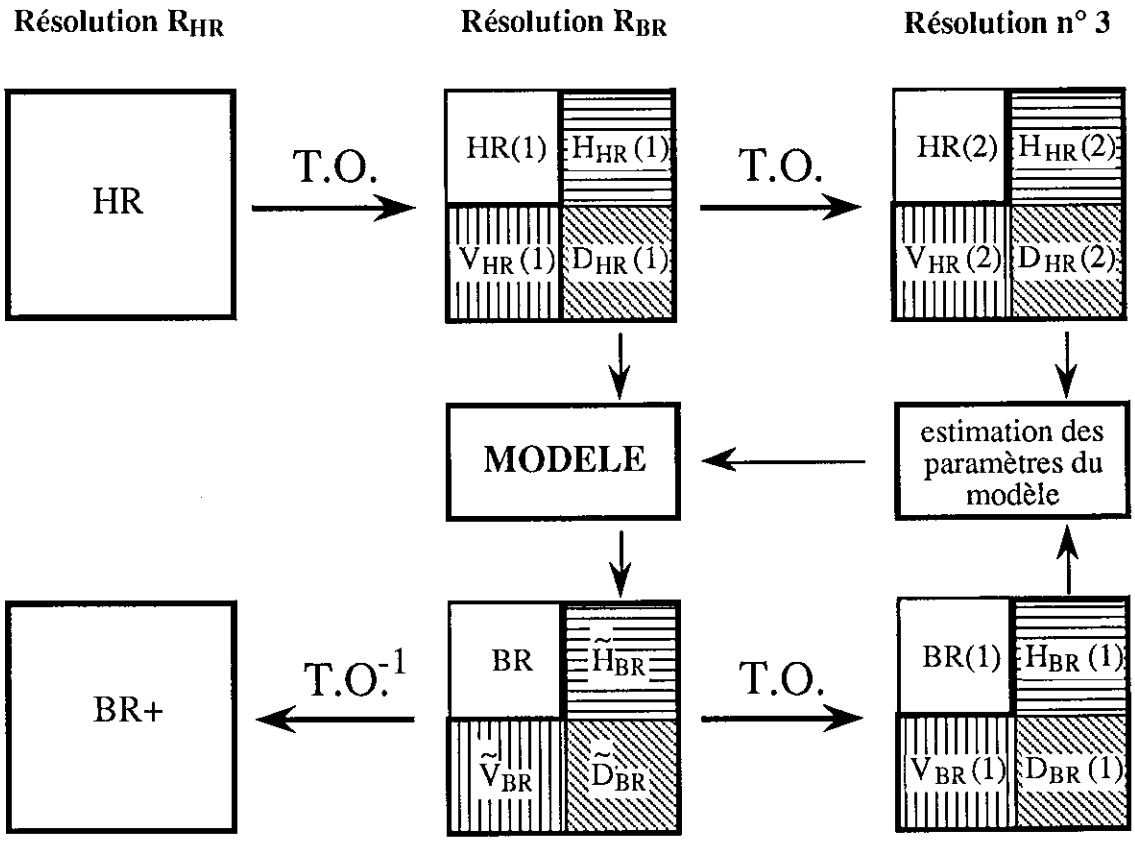


Figure 3.4 : schéma de la méthode ARSIS limitée à trois résolutions

Les modèles verticaux définis précédemment réalisent une régression sur les coefficients d'ondelettes des images HR et BR. Or on constate que les coefficients d'ondelettes des images naturelles suivent une distribution proche d'une gaussienne généralisée (Antonini *et al.* 1992). Nous avons donc retenu des modèles linéaires, qui

permettent de décrire correctement le passage d'une distribution gaussienne généralisée à une autre, tout en privilégiant la simplicité des calculs. Un tel modèle, estimé à la résolution n° 3, s'exprime par l'équation (3-1). Nous désignons par C_{HR} et C_{BR} les coefficients d'ondelettes des images HR et BR. Par exemple dans l'algorithme de Mallat, C_{HR} désigne de manière générique les coefficients H_{HR}, V_{HR}, et D_{HR}. L'indice entre parenthèses indique le nombre d'itérations de transformées en ondelettes effectuées sur l'image d'origine (cf. fig. 3.4).

$$C_{BR}(1) = a\, C_{HR}(2) + b \qquad (3-1)$$

Le rôle du modèle étant de mettre en correspondance les structures locales entre les deux images à fusionner, il est fondamental d'estimer les paramètres de ce modèle localement, et non sur la scène entière. On a vérifié que ces paramètres évoluent considérablement lorsqu'on se déplace dans la scène ; l'adoption d'une valeur moyenne décrit mal les phénomènes locaux. Le modèle est appliqué sur une fenêtre glissante, centrée sur le pixel traité, et dont la taille résulte d'un compromis entre le souci de réaliser un traitement local et la nécessité d'utiliser un nombre suffisant de pixels au sens statistique du terme.

Le premier modèle adopté, suggéré par la distribution gaussienne généralisée des coefficients d'ondelettes, consiste à ajuster la moyenne et la variance des coefficients des images HR et BR. Il est traduit par les équations (3-2) et (3-3), où *moy* et *var* désignent respectivement les opérateurs locaux de moyenne et variance. Les paramètres du modèle sont estimés à la résolution n° 3, et l'on suppose leur conservation à la résolution R_{BR} à laquelle le modèle est ensuite appliqué pour synthétiser les coefficients recherchés.

$$a = \sqrt{\frac{var(C_{BR}(1))}{var(C_{HR}(2))}} \qquad (3-2)$$

$$b = moy(C_{BR}(1)) - a\, moy(C_{HR}(2)) \qquad (3-3)$$

Un autre modèle linéaire consiste à calculer le premier axe factoriel du nuage de points composé des coefficients d'ondelettes des images HR et BR, restreints spatialement à la fenêtre de calcul. Ce modèle est représenté par les équations (3-4) et (3-5), où *cov* désigne l'opérateur local de covariance. L'avantage de ce dernier modèle sur le précédent est de faire intervenir la covariance entre les coefficients des deux

images, qui traduit le regroupement du nuage de points autour de l'axe factoriel, c'est-à-dire la validité de l'hypothèse de linéarité.

$$a = \frac{\text{var}(C_{BR}(1)) - \text{var}(C_{HR}(2))}{2\,\text{cov}\,(C_{BR}(1), C_{HR}(2))} + \frac{\sqrt{(\text{var}(C_{BR}(1)) - \text{var}(C_{HR}(2)))^2 + 4\,\text{cov}^2(C_{BR}(1), C_{HR}(2))}}{2\,\text{cov}\,(C_{BR}(1), C_{HR}(2))} \quad (3-4)$$

$$b = \text{moy}(C_{BR}(1)) - a\,\text{moy}(C_{HR}(2)) \quad (3-5)$$

Un troisième modèle de régression linéaire peut être obtenu par la droite des moindres carrés, dont l'expression des coefficients est donnée par les équations (3-6) et (3-7).

$$a = \frac{\text{cov}(C_{HR}(2), C_{BR}(1))}{\text{var}(C_{HR}(2))} \quad (3-6)$$

$$b = \text{moy}(C_{BR}(1)) - a\,\text{moy}(C_{HR}(2)) \quad (3-7)$$

Ces trois modèles linéaires diffèrent par le critère d'obtention de la droite de régression : ajustement des deux premiers moments, maximisation de la variance, ou moindres carrés des distances à la droite. Sur les exemples traités, ces modèles ont donné des résultats proches, avec un léger avantage pour le premier axe factoriel, relativement aux critères de qualité spectrale définis au paragraphe 3.3.2.

L'hypothèse effectuée par la méthode ARSIS est une similitude de la variabilité locale des deux images à fusionner, représentée par les coefficients d'ondelettes. La validité physique d'une image synthétisée par ARSIS est donc d'autant plus forte que les bandes spectrales des deux images présentent des petites structures similaires. Cependant, par définition même, aucune méthode de fusion multirésolution d'images multispectrales n'échappe à cette limitation.

3.2.3 Influence de la fonction de transfert de l'instrument et des effets atmosphériques

Nous avons identifié deux problèmes affectant la synthèse des images multispectrales à haute résolution spatiale : la fonction de transfert optique de l'instrument d'une part, et les effets atmosphériques d'autre part.

La fonction de transfert optique caractérise la réponse "géométrique" (donc la résolution spatiale) d'un instrument optique pour chaque bande spectrale. Les structures déduites de l'image HR par une transformée en ondelettes sont donc affectées des effets de la fonction de transfert de HR, qu'elles transmettent *via* le modèle à l'image synthétisée BR+. Nous n'avons pas pu estimer l'influence de cette fonction de transfert sur l'image BR+. En effet, les calculs mathématiques sont complexes et ne peuvent être simplifiés, car les fréquences spatiales de la fonction de transfert et de l'ondelette analysante sont similaires. L'application d'ARSIS à différents capteurs comme SPOT P et XS, ou SPOT P et Landsat TM, ou Landsat TM (30 m) et TM6 (120 m), donne des résultats relativement proches, ce qui semble indiquer que la fonction de transfert ne joue pas un rôle prédominant dans la synthèse de l'image BR+.

Les images HR et BR fusionnées par la méthode ARSIS sont toutes deux sujettes aux effets atmosphériques, lesquels sont spécifiques à chaque bande spectrale. Il est donc légitime de se demander si les structures extraites de l'image HR, au niveau satellite, ne vont pas entacher l'image synthétisée BR+ des effets atmosphériques propres à la bande HR. Nous examinons d'une part l'influence de l'atmosphère sur les structures de l'image HR, et d'autre part le comportement du modèle d'ARSIS qui en déduit les structures à réinjecter dans l'image BR.

En première approximation, la luminance satellite $L_{SAT}(HR)$ dans la bande spectrale de l'image HR se déduit de la luminance sol $L_{SOL}(HR)$ dans cette même bande par (Tanré *et al.* 1990) :

$$L_{SAT}(HR) = T(HR) \cdot L_{SOL}(HR) + L_a(HR)$$

où : $T(HR)$ est un facteur de transmission incluant les effets gazeux,
$L_a(HR)$ est une composante propre à l'atmosphère, incluant également les effets gazeux.

Les coefficients d'ondelettes $C_{HR\text{-}SAT}$ calculés sur l'image de luminance satellite s'écrivent, en notant TO la transformée en ondelettes et en utilisant la linéarité de cette transformée :

$$C_{HR\text{-}SAT} = TO\,[L_{SAT}(HR)] = TO\,[T(HR)\,.\,L_{SOL}(HR)] + TO\,[L_a(HR)]$$

Or $T(HR)$ et $L_a(HR)$ varient très lentement spatialement (échelles de variations de l'ordre du km), comparativement à la TO qui traite les fréquences spatiales de l'ordre du pixel, c'est-à-dire les échelles décamétriques. On en déduit :

$$C_{HR\text{-}SAT} = T(HR)\,.\,C_{HR\text{-}SOL}$$

où $C_{HR\text{-}SOL}$ désigne les coefficients d'ondelettes issus de la luminance sol.

En appliquant un raisonnement analogue à la luminance satellite $L_{SAT}(BR)$ dans la bande spectrale de l'image BR et à la luminance sol $L_{SOL}(BR)$, il vient :

$$C_{BR\text{-}SAT} = T(BR)\,.\,C_{BR\text{-}SOL}$$

où $C_{BR\text{-}SAT}$ et $C_{BR\text{-}SOL}$ désignent respectivement les coefficients d'ondelettes issus des luminances satellite et sol, et $T(BR)$ est un facteur de transmission dans la bande considérée.

Pour chaque pixel et dans chaque bande spectrale, l'effet de l'atmosphère est de multiplier par une constante la valeur des coefficients d'ondelettes calculés sur la luminance sol. On a évalué analytiquement l'influence sur BR+ des effets atmosphériques affectant l'image HR. Cette influence est fonction du modèle utilisé dans ARSIS. Les modèles 1 et 3 (cf. § 3.2.2) entraînent une indépendance vis-à-vis des effets atmosphériques de l'image HR, c'est-à-dire que l'utilisation des coefficients $C_{HR\text{-}SAT}$ se ramène à celle des coefficients $C_{HR\text{-}SOL}$, pondérés par le facteur de transmission $T(BR)$ pour la synthèse de BR+. Dans le cas général, incluant le modèle n°2 d'ARSIS, l'influence des effets atmosphériques est réduite au facteur de transmission $T(HR)$, mais ne disparaît pas lors de la synthèse de BR+. De manière empirique, en comparant les résultats des trois modèles sur les mêmes scènes selon les critères exposés au § 3.3.2, nous avons constaté des écarts relativement faibles d'un modèle à l'autre, tout en notant que nous ne savons pas déterminer la part de l'influence atmosphérique.

En conclusion, les effets atmosphériques semblent influer peu sur les structures haute résolution injectées dans l'image BR, qui sont représentées par les coefficients d'ondelettes estimés par le modèle d'ARSIS. Ce résultat découle de la sélectivité de la méthode ARSIS, qui ne traite que les petites échelles, ce qui permet de s'affranchir des phénomènes affectant les grandes échelles.

3.3 APPLICATION A LA FUSION SPOT P+XS

3.3.1 Introduction

Les images SPOT panchromatique (P) et multispectrales (XS) se prêtent bien aux méthodes de fusion d'images multispectrales multirésolutions : elles sont issues d'une même plate-forme, leurs prises de vue peuvent être simultanées (mode d'acquisition spécifique), et elles présentent une grande similitude de contenu thématique. Malheureusement, leurs caractéristiques géométriques ne sont pas identiques en raison de l'écart de visée entre les images P et XS, dans le mode d'acquisition simultanée P et XS (Anonyme 1986). Un recalage géométrique est de ce fait nécessaire préalablement à la fusion ; on rééchantillonne de préférence les images XS plutôt que l'image P, afin de préserver les petites structures qui seront ensuite extraites de l'image P. La mise en œuvre de la méthode ARSIS sur ces images est facilitée par le rapport de résolutions spatiales, égal à 2 entre les images P (résolution 10 m) et XS (résolution 20 m), qui permet l'utilisation d'ondelettes dyadiques. Les ondelettes utilisées sont des ondelettes orthogonales, à support compact, de régularité 2 (Daubechies 1988). Elles sont exploitées dans le cadre de l'analyse multirésolution définie par Mallat (1989). Enfin, la proximité des bandes spectrales (cf. fig. 3.5) garantit une assez bonne similitude entre les structures observées dans ces différentes bandes. Cette similitude peut être estimée par le coefficient de corrélation, calculé non pas sur les images elles-mêmes, mais sur leurs coefficients d'ondelettes aux différentes échelles.

Les résultats présentés dans ce chapitre sont issus des traitements appliqués à des images de la ville de Barcelone (Espagne). La figure 3.6 montre une vignette de 18 km de côté, extraite de l'image panchromatique sur la scène entière de Barcelone. Le contenu thématique de cette vignette est très varié, puisqu'on y trouve un riche paysage urbain, composé de plusieurs agglomérations, d'un port et d'un aéroport, ainsi que des parcelles agricoles et des zones montagneuses. Cette vignette recouvre elle-même trois

vignettes plus petites, marquées par des cadres et indicées, qui seront utilisées ultérieurement afin d'estimer la qualité des produits fusionnés. D'autres jeux d'images, localisés sur des sites thématiquement très différents, ont abouti à des résultats comparatifs identiques.

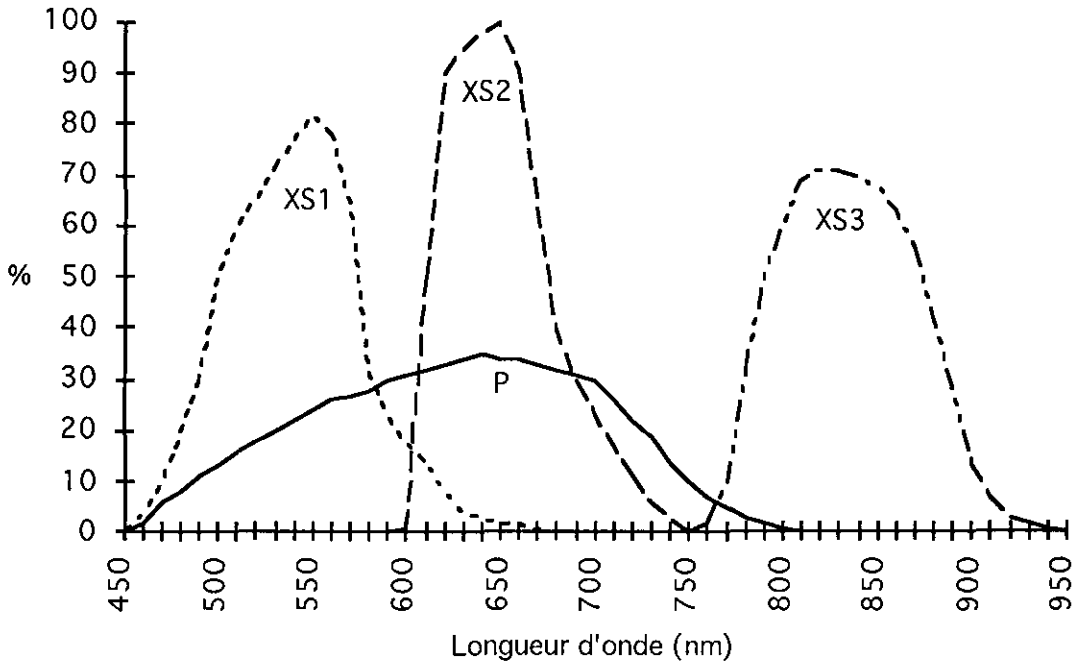


Figure 3.5 : bandes spectrales de SPOT

3.3.2 Résultats numériques

Les performances de la méthode ARSIS sont évaluées comparativement à celles de la méthode "P+XS" développée par le CNES (Anonyme 1986), qui fait référence en la matière (description de la méthode en annexe A). Les images XS simulées à la résolution de 10 m par la méthode P+XS sont notées XP, et celles produites par ARSIS sont notées XS-HR.

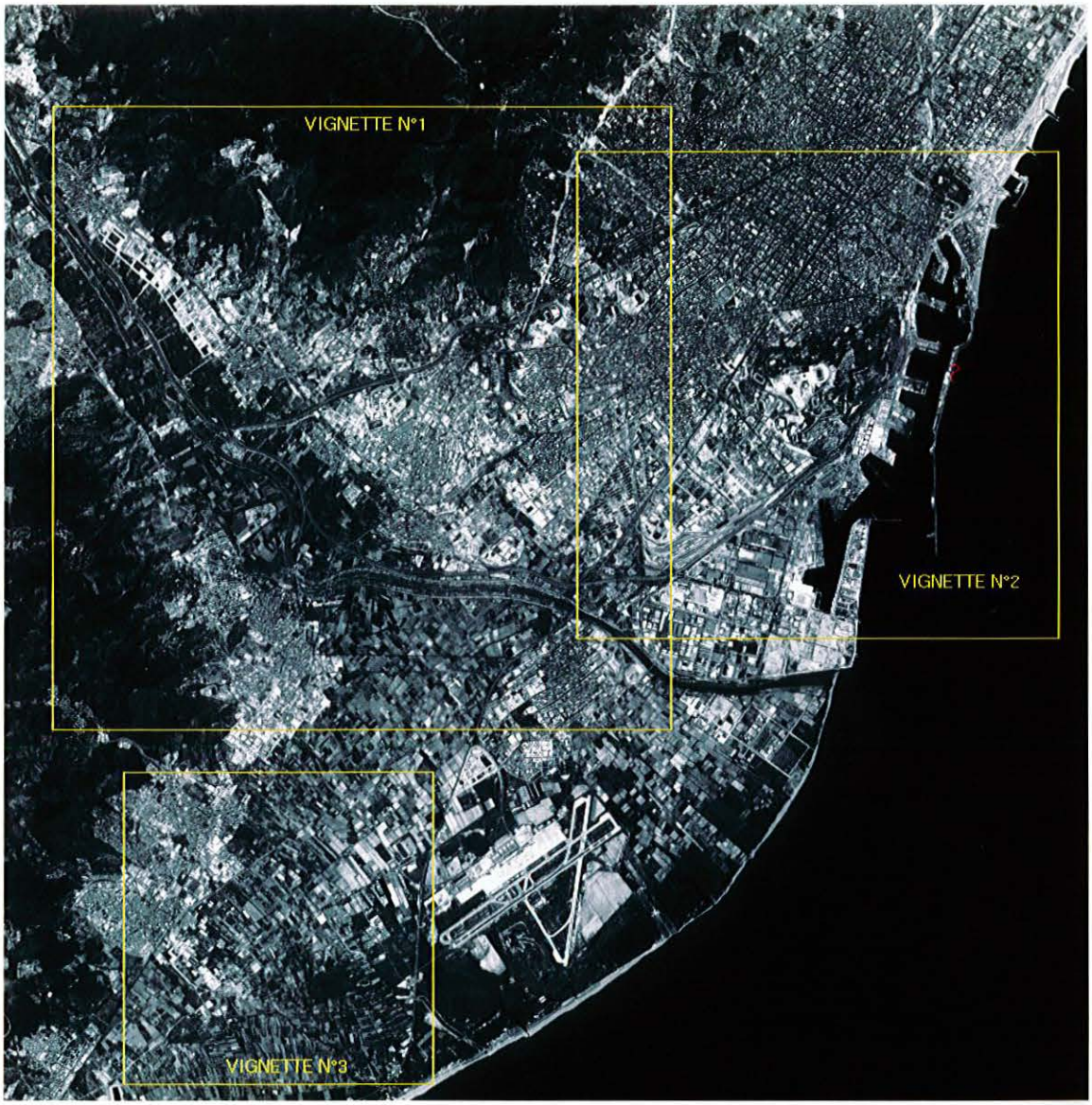


Figure 3.6 : extrait de la scène panchromatique sur la ville de Barcelone, et emplacements des 3 vignettes utilisées ultérieurement

La définition de critères de qualité pour ces images multispectrales simulées à 10 m de résolution pose un problème majeur, dû à l'absence de vérité terrain à 10 m. Cette difficulté peut être contournée en translatant les traitements à une résolution deux fois plus grande (c'est-à-dire plus grossière). Nous avons donc synthétisé des pseudo-images XS* de résolution 20 m à partir de l'image P dégradée à 20 m et des images XS dégradées à 40 m, en utilisant successivement les deux méthodes. Il est alors possible de comparer les images construites XS* aux images réelles XS.

Les tableaux de ce paragraphe présentent des résultats numériques relatifs à la vignette n°1 sur Barcelone (cf. fig. 3.6). Cette vignette a été choisie en raison de sa grande variété thématique, souhaitable pour effectuer la comparaison des deux méthodes du point de vue spectral. Elle comprend des zones urbaines, des paysages montagneux et des surfaces cultivées. Le tableau 3.1 donne les luminances moyennes, leurs écarts-type, et l'entropie des images XS et P originales, ainsi que les coefficients d'étalonnage utilisés pour le calcul de l'indice de végétation.

	XS1	XS2	XS3	P
Moyenne	58	48	55	53
Ecart-type	12	15	9	15
Entropie	4,02	4,26	3,87	4,40
Coefficient d'étalonnage	1,22181	1,22545	1,29753	1,39198

Tableau 3.1 : statistiques sur les images originales exprimées en luminance (en W m^{-2} sr^{-1} μm^{-1}), pour la vignette n°1 sur Barcelone

Le tableau 3.2 rassemble des résultats statistiques relatifs à la comparaison des images XS* et XS. Le biais indique une erreur systématique de la méthode. La différence des variances ou des entropies permet d'apprécier la quantité d'information introduite ou perdue par une méthode par rapport à l'image originale. Si une méthode introduit trop d'information, en provenance de l'image P, dans l'image XS*, les différences de variances et d'entropie seront positives. Le coefficient de corrélation entre l'image originale et l'image estimée doit être le plus proche possible de 1. L'écart-type de l'image différence [XS* - XS] traduit la précision globale de l'estimée XS*.

	XS1		XS2		XS3	
	P+XS	ARSIS	P+XS	ARSIS	P+XS	ARSIS
Biais (idéal : 0) relativement à la moyenne de XS	- 0,35 - 0,6%	0,00 0,0 %	- 0,26 - 0,5 %	0,00 0,0 %	0,00 0,0 %	0,00 0,0 %
Variance XS* - variance vraie (valeur idéale : 0) relativement à la variance vraie	50 35 %	- 6 - 5 %	42 19 %	- 7 - 3 %	- 9 - 11 %	- 7 - 9 %
Entropie XS* - entropie vraie (valeur idéale : 0) relativement à l'entropie vraie	0,15 3,6 %	- 0,01 - 0,3%	0,08 1,8 %	- 0,01 - 0,3 %	- 0,06 - 1,5 %	- 0,05 - 1,2 %
Coefficient de corrélation entre XS* et XS (valeur idéale : 1)	0,97	0,99	0,98	0,99	0,91	0,95
Ecart-type des différences (valeur idéale : 0) relativement à la moyenne de XS	3,8 7 %	1,9 3 %	3,1 6 %	1,9 4 %	3,8 7 %	2,7 5 %

Tableau 3.2 : statistiques sur les différences entre XS* et XS (en luminance ou valeur relative) pour la vignette n°1 sur Barcelone

La méthode ARSIS donne un biais nul à la troisième décimale près pour toutes les bandes, alors que la méthode P+XS introduit un biais de l'ordre d'un pour cent. Les différences de variance et d'entropie fortement positives pour les deux premières bandes, avec la méthode P+XS, traduisent un fort accroissement de la quantité d'information, dû au renforcement des contours apportés par l'image P. A l'inverse, la méthode ARSIS diminue l'information des deux premières bandes, mais elle s'approche beaucoup plus des images originales. Ce résultat découle de la définition même de la méthode, qui ne renforce que les petites structures, dans des proportions adaptées à chaque bande spectrale. Pour la bande XS3, les deux méthodes diminuent la quantité d'information de l'image initiale ; ceci s'explique différemment pour chacune d'elles. L'image XS3* réalisée par la méthode P+XS résulte d'une simple duplication des pixels de l'image XS3 dégradée à 40 m (cf. annexe A), donc son contenu informationnel est plus pauvre que celui de l'image originale XS3. La méthode ARSIS, sensible à la présence de chutes locales de la similitude des structures P et XS3, ne parvient pas à restituer la totalité des structures 20-40 m de l'image XS3. Elle est cependant la plus proche du résultat idéal, comme pour les deux premières bandes.

Le coefficient de corrélation entre l'image originale et l'image estimée est toujours plus élevé avec la méthode ARSIS qu'avec la méthode P+XS. Il avoisine, avec cette première méthode, la valeur idéale de 1 pour les bandes XS1 et XS2, et chute quelque peu pour la bande XS3, en raison de la moins grande similitude des structures dans cette bande et dans le panchromatique. L'écart-type des différences est du même ordre pour les deux méthodes, avec un avantage à ARSIS pour chacune des bandes.

L'erreur au niveau du pixel peut être davantage détaillée. Ainsi le tableau 3.3 présente la probabilité d'avoir, en un pixel, une erreur relative inférieure en valeur absolue à un seuil donné. Cette erreur relative, calculée pour chaque pixel, est égale à la différence entre la valeur originale et la valeur estimée, divisée par la valeur originale. L'histogramme cumulé de ces erreurs relatives permet de déduire la probabilité qu'un pixel présente une erreur relative inférieure à un seuil donné. Une méthode est d'autant meilleure que cette probabilité croît rapidement avec l'erreur et atteint 100 % pour une erreur faible.

Avec la méthode ARSIS, la quasi totalité des pixels de l'image (au moins 99,5 %) présentent une erreur relative inférieure à 20 % en valeur absolue. Au moins 95 % des pixels présentent une erreur relative inférieure à 10 % en valeur absolue. En ce qui concerne la méthode P+XS, cette probabilité croît moins vite avec le seuil d'erreur, et la proportion de pixels d'erreur nulle est plus faible qu'avec la méthode ARSIS.

	XS1		XS2		XS3	
Seuils (%)	P+XS	ARSIS	P+XS	ARSIS	P+XS	ARSIS
0	9	28	12	26	13	16
1	9	28	12	26	13	16
2	27	63	26	47	35	43
5	58	92	60	86	66	76
10	90	99	91	99	88	95
20	100	100	100	100	98	100
50	100	100	100	100	100	100
100	100	100	100	100	100	100

Tableau 3.3 : probabilité (en pour cent) d'avoir en un pixel une erreur relative sur XS* (en valeur absolue et en pour cent) inférieure au seuil indiqué en première colonne.

Tous ces critères statistiques montrent que la méthode ARSIS est radiométriquement meilleure que la méthode P+XS. Les images calculées par la méthode P+XS, en dépit d'un aspect visuel satisfaisant, sont mal adaptées à des traitements ultérieurs nécessitant une bonne qualité spectrale.

3.3.3 Résultats visuels

Afin de montrer l'apport visuel de la méthode ARSIS, nous présentons des images localisées sur la vignette n°2 de Barcelone (cf. fig. 3.6). Cette zone est particulièrement riche en petites structures, tant sur la partie urbaine que sur les infrastructures portuaires. C'est pourquoi elle a été préférée à la vignette n°1 pour mettre en valeur l'accroissement de résolution spatiale apporté par la méthode ARSIS. La figure 3.7 représente la vignette n°2 extraite de l'image panchromatique, et les figures 3.8 et 3.9 les compositions colorées effectuées respectivement sur les images XS et XS-HR correspondantes. Un grossissement d'une petite partie du port est présenté sous chaque image dont il est tiré, pour mieux mettre en évidence les différences de qualité géométrique. L'image XS a été ramenée à la même taille que les deux autres, pour faciliter leur comparaison, par une duplication de pixels qui n'en modifie pas les couleurs. Les compositions colorées affectent classiquement XS1 au canal bleu, XS2 au vert, et XS3 au rouge. Les tables de transcodage, servant à adapter la dynamique de chaque bande au restituteur, sont identiques pour les images XS et XS-HR, afin de permettre leur comparaison.

On constate que les couleurs sont semblables sur les compositions colorées XS et XS-HR, ce qui traduit que les images XS-HR sont radiométriquement très proches des images XS d'origine. L'apport d'ARSIS en petites structures est notable sur l'ensemble de la vignette ; il est en particulier très net sur les infrastructures du port dont certaines sont discernables sur XS-HR et non sur XS. Cette remarque s'applique également à la zone urbaine, qui comporte beaucoup de structures d'échelles 10-20 m.

3.3.4 Conclusion

Nous avons montré, sur l'exemple de SPOT, que la méthode ARSIS préserve l'information spectrale des images originales, conformément au but que l'on s'était fixé. Pour mettre en évidence l'intérêt de cette propriété, nous avons effectué une classification non-dirigée, successivement sur les images XS et sur les images XS-HR déduites par la méthode ARSIS. L'algorithme utilisé est la classification ascendante hiérarchique (Lerman 1981). On constate que les classes obtenues, dont le nombre est prédéterminé, sont très voisines sur les deux classifications, ce qui témoigne de la proximité spectrale des images XS et XS-HR. Grâce au gain en résolution apporté par ARSIS, la classification effectuée sur les images XS-HR possède des frontières spatiales beaucoup plus précises que celle utilisant les images XS.



Figure 3.7 : image panchromatique du port de Barcelone (vignette n°2).

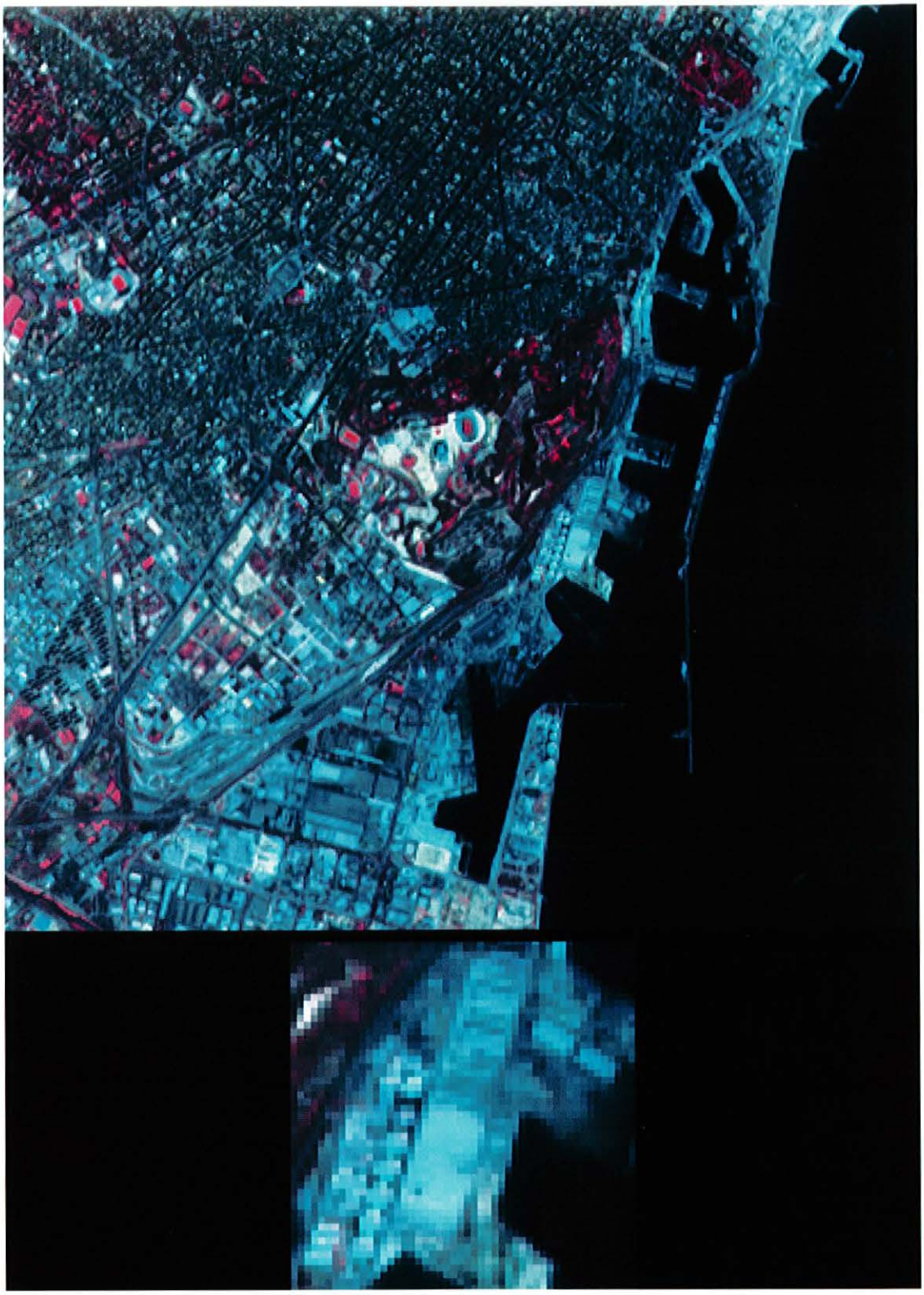


Figure 3.8 : composition colorée XS sur le port de Barcelone (vignette n°2).

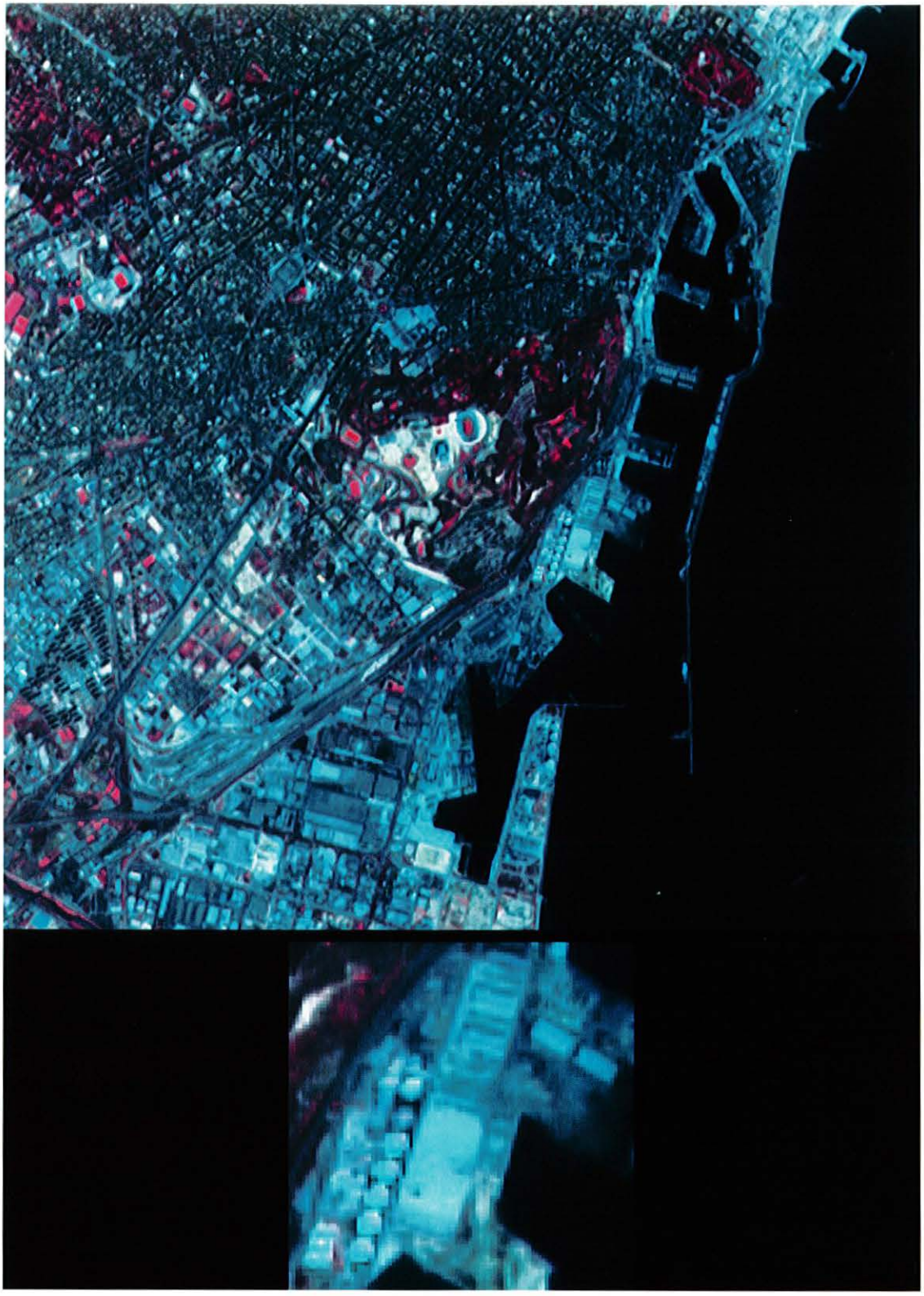


Figure 3.9 : composition colorée XS-HR sur le port de Barcelone (vignette n°2).

3.4 APPLICATION A L'ESTIMATION DE L'INDICE DE VEGETATION

3.4.1 Définition du NDVI

L'exploitation des images satellitaires pour des applications relatives à la végétation utilise couramment des indices de végétation, qui permettent de discriminer différents types de végétations. L'un des plus usités est le NDVI (normalized difference vegetation index), défini par :

$$NDVI = \frac{\rho_{PIR} - \rho_R}{\rho_{PIR} + \rho_R}$$

où ρ_{PIR} et ρ_R désignent les réflectances mesurées respectivement dans le proche infrarouge et dans le rouge.

Dans le cas de SPOT, cet indice s'écrit :

$$NDVI = \frac{\rho_{XS3} - \rho_{XS2}}{\rho_{XS3} + \rho_{XS2}}$$

où ρ_{XS2} et ρ_{XS3} désignent les réflectances dans les bandes XS2 et XS3.

La réflectance ρ_{XSi} dans la bande XSi (i=1 à 3) est obtenue à partir de la luminance équivalente SPOT L_i définie dans Anonyme (1986), de l'éclairement solaire E_i dans cette bande, et de l'angle zénithal solaire θ_S par :

$$\rho_{XSi} = \frac{\Pi L_i}{E_i \cos \theta_s}$$

ρ_{XSi} est la réflectance au niveau du satellite et non la réflectance sol, c'est-à-dire que les effets atmosphériques n'ont pas été corrigés. Les luminances équivalentes SPOT sont restituées à partir des comptes numériques et des coefficients d'étalonnage, dont les valeurs pour l'image de Barcelone sont données dans le tableau 3.1. Les éclairements solaires E_2 et E_3 dans les bandes 2 et 3 de SPOT ont été calculés par convolution de la densité spectrale d'éclairement solaire par les filtres de SPOT puis normalisation ; on trouve :

$L_2 = 1594$ W m^{-2} µm^{-1}
$L_3 = 1057$ W m^{-2} µm^{-1}

3.4.2 Mise en œuvre d'ARSIS

L'exploitation du NDVI, calculé sur XS, est parfois limitée par la résolution de 20 m, notamment pour le petit parcellaire. La méthode ARSIS permet de synthétiser un NDVI à 10 m de résolution, à l'aide de l'image panchromatique. Nous avons identifié deux mises en œuvre possibles de la méthode :

- synthèse indépendante des deux images XS2-HR et XS3-HR par ARSIS, puis calcul du NDVI à 10 m, noté NDVI-1,
- calcul préalable du NDVI à 20 m sur les images XS, puis application d'ARSIS à ce NDVI basse résolution et à l'image panchromatique, pour obtenir le NDVI-2.

Dans le premier cas, ARSIS exploite les similitudes sur les petites structures entre chaque bande XS et la bande P. On a déjà estimé la qualité du résultat sur Barcelone au § 3.3. Dans le second cas, on cherche à modéliser les similitudes entre le NDVI et l'image panchromatique. Ces deux NDVI ne sont pas identiques, car si la transformée en ondelettes est bien linéaire, ce n'est pas le cas du calcul de NDVI. La synthèse de NDVI-2 présente la particularité de modéliser les similitudes sur des structures extraites de deux grandeurs physiquement différentes : l'image P représente une luminance, tandis que le NDVI est lié à la biomasse (masse par unité de surface), en particulier pour les surfaces agricoles. Toutefois, le NDVI est par définition sans dimension. Une autre différence entre NDVI-1 et NDVI-2 concerne l'influence de l'atmosphère. Dans le premier cas, l'influence atmosphérique sur NDVI-1 est la même que sur le NDVI basse résolution, puisqu'elle est identique entre chaque bande XSi-HR et XSi (cf. § 3.2.3). Dans le second cas, la non linéarité de cet indice entraine que le modèle ne compense pas les effets atmosphériques qui affectent l'image panchromatique et NDVI-2, donc NDVI-2 risque de s'écarter du NDVI basse résolution.

En raison de l'absence de référence à 10 m, nous allons estimer numériquement la qualité du NDVI synthétisé à 20 m de résolution, et non celle du NDVI à 10 m, en suivant la même démarche qu'au § 3.3.2. Les NDVI synthétisés à 20 m suivant chaque méthode sont notés respectivement NDVI*-1 et NDVI*-2 : NDVI*-1 est calculé d'après XS2* et XS3*, et NDVI*-2 à partir du NDVI dégradé à 40 m et de l'image panchromatique dégradée à 20 m. Les paramètres statistiques reflétant la qualité du NDVI* sont les mêmes que ceux utilisés pour XS* (cf. § 3.3.2).

L'exploitation du NDVI étant spécifique à la végétation, les traitements ont été effectués sur une zone comportant presque exclusivement de la végétation : il s'agit de la vignette n°3 sur Barcelone (cf. fig. 3.6). Sur cette vignette, le NDVI à 20 m est négatif sur 0,08 % des pixels, qui s'avèrent correspondre à des bâtiments, des infrastructures, et des étendues d'eau. En excluant ces points des calculs de paramètres de qualité, on se restreint à des données adaptées à la comparaison des deux NDVI*. Les résultats statistiques pour les deux méthodes sont rassemblés dans les tableaux 3.4 et 3.5.

	NDVI*-1	NDVI*-2
Biais (idéal : 0) relativement à la moyenne de NDVI	- 0,03 - 0,1%	0,00 0,0 %
Variance NDVI* - variance vraie (valeur idéale : 0) relativement à la variance vraie	7 5 %	7 5 %
Entropie NDVI* - entropie vraie (valeur idéale : 0) relativement à l'entropie vraie	0,01 0,4 %	0,00 0,2%
Coefficient de corrélation entre NDVI* et NDVI (valeur idéale : 1)	0,98	0,98
Ecart-type des différences (valeur idéale : 0) relativement à la moyenne de NDVI	2,6 10 %	2,6 10 %

Tableau 3.4 : statistiques sur les différences entre NDVI* et NDVI (en comptes numériques ou en valeur relative), pour la vignette n°3 sur Barcelone

Seuils (%)	NDVI*-1	NDVI*-2
0	20	19
1	20	19
2	21	20
5	44	44
10	71	71
20	91	91
50	99	99
100	100	100

Tableau 3.5 : probabilité (en pour cent) d'avoir en un pixel une erreur relative sur NDVI* (en valeur absolue et en pour cent) inférieure au seuil indiqué en première colonne, pour la vignette n°3 sur Barcelone

D'après ces critères, les deux mises en œuvre d'ARSIS pour l'amélioration de la résolution du NDVI donnent des résultats quasiment identiques. Visuellement, les deux images de NDVI à 10 m de résolution sont également similaires. Ces conclusions illustrent la flexibilité d'ARSIS, qui s'applique aussi bien au couple hétérogène (NDVI, P) qu'au cas (XS, P), grâce à la modélisation des similitudes entre les petites structures contenues dans les deux images.

3.5 APPLICATION A D'AUTRES RAPPORTS DE RESOLUTION

Les applications de la méthode ARSIS présentées dans ce chapitre concernaient jusqu'ici des images dont le rapport des résolutions spatiales était égal à 2. La méthode est cependant utilisable pour d'autres rapports de résolutions. Comme il a déjà été précisé, l'utilisation d'une ondelette dyadique impose que ce rapport soit égal à une puissance de 2. Avec une telle ondelette, le cas d'un rapport de résolutions égal à 4 peut être illustré sur des images SPOT, en fusionnant une image XS dégradée à 80 m avec une image P dégradée à 20 m, pour obtenir une image XS** à 20 m. On réalise deux itérations avec ARSIS, pour obtenir successivement une image à 40 m, puis une image à 20 m. Les résultats statistiques comparant l'image XS1** à la référence XS1, sur la vignette n°1 de Barcelone, sont donnés par les tableaux 3.6 et 3.7. On y rappelle également les valeurs obtenues pour XS1*, provenant de XS1 à 40 m (cf. § 3.3.2).

	XS1*	XS1**
Biais (idéal : 0) relativement à la moyenne de XS	0,00 0,0 %	0,00 0,0 %
Variance estimée - variance vraie (valeur idéale : 0) relativement à la variance vraie	- 6 - 5 %	- 21 -15 %
Entropie estimée - entropie vraie (valeur idéale : 0) relativement à l'entropie vraie	- 0,01 - 0,3%	- 0,05 - 1,3 %
Coefficient de corrélation entre XS estimé et XS (valeur idéale : 1)	0,99	0,96
Ecart-type des différences (valeur idéale : 0) relativement à la moyenne de XS	1,9 3 %	3,3 6 %

Tableau 3.6 : comparaison de l'erreur commise par ARSIS pour deux rapports de résolutions différents, pour la vignette n°1 sur Barcelone

Seuils (%)	XS1*	XS1**
0	28	16
1	28	16
2	63	42
5	92	75
10	99	95
20	100	100
50	100	100
100	100	100

Tableau 3.7 : probabilité (en pour cent) d'avoir en un pixel une erreur relative sur XS1* et XS1** (en valeur absolue et en pour cent) inférieure au seuil indiqué en première colonne, pour la vignette n°1 sur Barcelone

On constate que l'image XS1** est moins proche de l'image originale que ne l'est XS1*. Ce résultat était attendu, puisque XS1** est réalisée à partir d'une information plus pauvre (XS1 à 80 m) que XS1* (issue de XS1 à 40 m). Néanmoins, la qualité de l'image XS1** reste correcte, ce qui montre l'applicabilité d'ARSIS avec un rapport de résolutions égal à 4.

Il arrive que le rapport des résolutions des images à fusionner ne soit pas une puissance de 2 : c'est le cas par exemple pour des images SPOT panchromatique (résolution de 10 m) et Landsat TM (résolution de 30 m). Il faudrait en principe utiliser une ondelette non dyadique, qui permette de passer de la résolution de la première image à celle de la seconde. Un exemple d'ondelette non dyadique est celle de Feauveau (1990), qui permet des rapports de résolutions en puissances de $\sqrt{2}$; cependant les images de coefficients d'ondelettes qui en résultent ne sont pas superposables d'une échelle à la suivante, ce qui complique la mise en œuvre du modèle d'ARSIS. Lorsqu'on ne dispose pas d'une ondelette fournissant le bon rapport de résolutions, on peut se ramener à l'utilisation d'une ondelette dyadique par rééchantillonnage préalable d'une des images, de telle sorte que le rapport de résolutions atteigne une puissance de 2. Il est préférable de ne pas rééchantillonner l'image haute résolution, car cela altérerait les petites structures exploitées ensuite par ARSIS. On rééchantillonne donc l'image basse résolution, à un pas qui reste à déterminer. Dans le cas de la fusion SPOT P et Landsat TM par exemple, on peut rééchantillonner l'image TM à 20 m ou à 40 m, pour se ramener à un rapport de résolution de 2 ou 4 respectivement. Si l'on utilise l'image TM à 20 m, le modèle d'ARSIS exploite les similitudes entre les structures d'échelles 20-40 m ; si l'on choisit l'image TM à 40 m, le modèle se base sur les échelles 40-80 m. Nous avons observé sur

plusieurs sites, par calcul de la corrélation des coefficients d'ondelettes, que les structures d'échelles 40-80 m sont plus corrélées, entre les images P et TM1, que les structures d'échelles 20-40 m. Cela incite à utiliser plutôt l'image TM1 à 40 m, dans laquelle ARSIS va injecter successivement les structures 20-40 m puis 10-20 m extraites de l'image P. Nous avons testé les deux possibilités de rééchantillonnage sur des images de Marignane (cf. chap. 5). En partant de l'image TM1 à 40 m, on synthétise une image à 10 m beaucoup plus riche en petites structures qu'avec l'image TM1 à 20 m, mais également plus bruitée. Cette augmentation du bruit avec le rapport de résolutions avait déjà été constatée sur les images SPOT (cf. tableau 3.6).

3.6 VALIDITE DE L'ESTIMATION DE LA QUALITE

Afin d'estimer quantitativement la qualité radiométrique des images synthétisées par ARSIS, il est nécessaire de posséder une image de référence à laquelle comparer l'image estimée. Dans le cas de SPOT par exemple, on ne dispose bien évidemment pas d'images multispectrales de résolution 10 m pour évaluer la qualité des images XS-HR. Nous avons donc décidé de translater les traitements d'un facteur 2 en résolution, et de comparer les images estimées XS*, de résolution 20 m, aux images de référence XS. On ne peut cependant pas en déduire que les images estimées XS-HR, de résolution 10 m, possèdent la même qualité que les images XS*. Intuitivement, l'erreur de modélisation d'un objet croît lorsqu'on affine la résolution, sauf si l'objet possède une structure caractéristique d'échelles supérieures à la résolution. Pourtant, des travaux récents, théoriques comme ceux de Raffy (1993) ou d'analyse de mesures satellitaires, comme ceux de Welch *et al.* (1989) Kong, Vidal-Madjar (1988), ou Woodcock, Strahler (1987) ont montré que l'estimation d'un paramètre est une fonction non prédictible de l'échelle d'observation et également des types d'objets observés. Il s'ensuit qu'il est difficile de déduire la qualité des images XS-HR (10 m) de la qualité mesurée sur les images XS* (20 m). La première peut être meilleure ou moins bonne que la dernière, car la qualité de l'estimée n'est pas nécessairement une fonction monotone de la résolution. Pour illustrer ce propos, voici l'erreur commise sur une image XS1*, estimée à 20 m à partir d'une image XS1 à 40 m, comparée à celle d'une image notée XS140*, estimée à 40 m à partir d'une image XS1 à 80 m. Les images P et XS1 utilisées correspondent à la vignette n°1 de Barcelone (cf. fig. 3.6), assez riche en petites structures. Les résultats sont présentés par les tableaux 3.8 et 3.9.

	XS1*	XS140*
Biais (idéal : 0) relativement à la moyenne de XS	0,00 0,0 %	0,00 0,0 %
Variance estimée - variance vraie (valeur idéale : 0) relativement à la variance vraie	- 6 - 5 %	- 12 - 9 %
Entropie estimée - entropie vraie (valeur idéale : 0) relativement à l'entropie vraie	- 0,01 - 0,3%	- 0,03 - 0,8 %
Coefficient de corrélation entre XS estimé et XS (valeur idéale : 1)	0,99	0,98
Ecart-type des différences (valeur idéale : 0) relativement à la moyenne de XS	1,9 3 %	2,2 4 %

Tableau 3.8 : comparaison de l'erreur commise par ARSIS à deux résolutions différentes, pour la vignette n°1 sur Barcelone

Seuils (%)	XS1*	XS140*
0	28	23
1	28	24
2	63	56
5	92	88
10	99	99
20	100	100
50	100	100
100	100	100

Tableau 3.9 : probabilité (en pour cent) d'avoir en un pixel une erreur relative sur XS1* et XS140* (en valeur absolue et en pour cent) inférieure au seuil indiqué en première colonne, pour la vignette n°1 sur Barcelone

On constate que la qualité de l'estimée à 20 m XS1* est meilleure que celle de l'estimée à 40 m XS140*. Il en est de même pour les bandes XS2 et XS3. Ce résultat a également été constaté sur les autres sites traités par ARSIS, pour les trois bandes XS. On ne peut cependant le généraliser à tous les sites, et encore moins en déduire que l'estimée XS-HR à 10 m est meilleure que l'estimée XS* à 20 m.

3.7 CONCLUSION

La méthode ARSIS offre l'avantage d'être applicable à de nombreux capteurs, fournissant des images dans d'autres bandes spectrales et à d'autres résolutions spatiales

que celles de SPOT. La diversité des bandes utilisables peut être illustrée par exemple par la fusion d'images Landsat TM1 et TM7 : ces images possédant initialement la même résolution (30 m), on a artificiellement dégradé celle de l'image TM7 à 60 m. La méthode ARSIS permet de simuler une image TM7* à 30 m de résolution, à partir de l'image TM7 à 60 m et de l'image TM1 à 30 m. Les critères statistiques de qualité de l'image TM7*, comparée à l'image de référence TM7, sont donnés par les tableaux 3.10 et 3.11, pour le site de Marignane (cf. chap. 5).

	TM7*
Biais (idéal : 0) relativement à la moyenne de TM7	0,03 0,1 %
Variance estimée - variance vraie (valeur idéale : 0) relativement à la variance vraie	 - 13 - 2 %
Entropie estimée - entropie vraie (valeur idéale : 0) relativement à l'entropie vraie	 - 0,02 - 0,6%
Coefficient de corrélation entre TM7 estimé et TM7 (valeur idéale : 1)	0,99
Ecart-type des différences (valeur idéale : 0) relativement à la moyenne de TM7	 5 14 %

Tableau 3.10 : statistiques sur les différences entre TM7* et TM7 (en luminance ou valeur relative), pour l'image de Marignane

Seuils (%)	TM7*
0	22
1	22
2	28
5	44
10	61
20	85
50	97
100	100

Tableau 3.11 : probabilité (en pour cent) d'avoir en un pixel une erreur relative sur TM7* (en valeur absolue et en pour cent) inférieure au seuil indiqué en première colonne, pour l'image de Marignane

Ces résultats sont comparables à ceux obtenus sur SPOT P et XS, à l'exception de l'écart-type de l'image différence qui est ici beaucoup plus grand, et de la probabilité

de justesse radiométrique qui croît moins vite avec le seuil d'erreur. Cet écart peut être dû à une moindre similitude des petites structures entre les bandes TM1 et TM7, qu'entre les bandes P et XSi, cette similitude étant mesurée par exemple par la corrélation des coefficients d'ondelettes. Ceci met en évidence les limites d'utilisation d'ARSIS, et pose la question du contenu physique de l'image synthétisée : l'injection de structures extraites du visible dans une image infrarouge peut aboutir à une image infrarouge haute résolution peu réaliste. *A fortiori*, la fusion de données hétérogènes telles que des images optique et radar donne un résultat encore plus difficile à interpréter. Une autre difficulté s'ajoute si les images fusionnées ont été prises à des dates différentes ; lors de l'interprétation, il est alors nécessaire d'utiliser les images sources conjointement à l'image calculée.

L'application d'ARSIS à une image SAR et une image optique est réalisable simplement, à condition qu'il existe suffisamment de similitudes entre leurs structures, ce qui se produit par exemple sur des paysages de petit parcellaire agricole. Dans le cas contraire, la fusion hétérogène impose de perfectionner le modèle d'ARSIS, qui peut être statistique ou déterministe, afin qu'il intègre les relations entre les deux physiques exploitées. On peut envisager un modèle non-paramétrique à base de réseau de neurones, qui serait vraisemblablement plus robuste qu'une régression statistique (Iverson, Lersch 1994). De manière générale, le modèle mis en œuvre par ARSIS peut être complexifié pour exploiter simultanément plus de deux bandes. Dans l'exemple de fusion de TM1 et TM7 évoqué ci-dessus, on a choisi arbitrairement la bande TM1 pour améliorer la résolution de l'image TM7. Il aurait été préférable que le modèle exploitât les bandes 1 à 5 simultanément.

L'intérêt de la méthode ARSIS tient à sa richesse physique et conceptuelle, qui offre un cadre propice à des développements ultérieurs. L'originalité par rapport aux méthodes de fusion existantes réside dans l'injection sélective des structures, caractérisées par leurs échelles, et dans le respect de la radiométrie des images d'origine. Les évolutions majeures envisagées portent sur la modélisation des similitudes de structures, qui doit être adaptée à la physique des images traitées. Dans les cas difficiles de fusion hétérogène optique-radar, nous proposons également une autre méthode utilisant l'analyse multirésolution et la transformée en ondelettes. Cette méthode ne combine pas les structures issues d'images de natures différentes, mais sert à la présentation visuelle du volume de données, et constitue une aide à la photo-interprétation (cf. chap. 4).

Chapitre 4

PRESENTATION VISUELLE DU VOLUME DE DONNEES

4.1 INTRODUCTION

Nous avons jusqu'ici développé des méthodes relatives à la constitution et au pré-traitement du volume de données : il s'agissait de faire en sorte que les images satellitaires, en provenance de différents capteurs, soient aisément comparables et exploitables par des méthodes numériques qui combinent leurs informations. Le présent chapitre aborde les traitements ultérieurs destinés à la représentation visuelle du volume de données. Cette étape finale est essentielle, car elle conditionne l'efficacité du photo-interprète à extraire l'information qu'il recherche, au sein d'un volume foisonnant de données complémentaires. Les méthodes présentées ici peuvent donc être considérées comme une aide à la photo-interprétation. Elles sont avant tout destinées à mettre en évidence l'information recherchée, en synthétisant des produits images, à partir de l'ensemble du volume de données, qui permettent une meilleure extraction du renseignement que l'utilisation indépendante de chaque image source. Cette fonction correspond à la transformation d'un ensemble d'images, traduisant des mesures physiques, en images synthétiques adaptées à la perception visuelle de l'homme. De plus, le photo-interprète n'est pas toujours en mesure de consulter le volume croissant d'images à sa disposition, en raison des contraintes opérationnelles. Un second objectif est donc de concentrer le volume de données en un minimum d'images.

Nous exposons, dans une première partie, les méthodes existantes de présentation visuelle du volume de données, qui reposent sur le choix d'un espace de représentation adapté. Ces méthodes réalisent une fusion de type centralisé, car elles exploitent simultanément l'ensemble des images sources. Nous verrons qu'une telle architecture de fusion n'est pas toujours optimale lorsqu'elle s'applique à des données hétérogènes. C'est pourquoi nous proposons, dans une seconde partie, une méthode originale de fusion d'images optique et radar, qui utilise une architecture décentralisée :

l'image radar y est traitée indépendamment de l'image optique avant l'ultime étape de fusion. Cette méthode exploite la transformée en ondelettes et l'analyse multirésolution, comme la méthode ARSIS (cf. chap. 3), mais elle s'en distingue par l'absence de modélisation des similitudes structurelles entre les images.

4.2 CHANGEMENT D'ESPACE DE REPRESENTATION

Nous cherchons à extraire des informations d'un volume de données, constitué de n images, pour les présenter sous forme d'un nombre aussi réduit que possible d'images synthétiques. Il est d'usage de représenter chaque grandeur physique mesurée, fonction d'une variable spatiale, par une image monochrome (*i.e.* en niveaux de gris) dont l'intensité traduit les valeurs prises par cette grandeur. Or la perception visuelle humaine permet de distinguer trois composantes indépendantes dans une seule image, grâce à l'utilisation de la couleur (Buchanan 1979, Foley, Van Dam 1982, Pratt 1978) :

- la luminance (L) représente l'intensité de la couleur (ou l'énergie totale),
- la teinte (T) correspond à la longueur d'onde moyenne,
- la saturation (S) définit la pureté de la couleur (ou chroma), c'est-à-dire sa teneur en blanc.

Il est donc possible d'accroître d'un facteur 3 la quantité d'information contenue dans une image, par l'utilisation de la couleur. L'espace des couleurs peut être représenté par d'autres bases que la base de perception humaine LTS (IHS en anglais) définie ci-dessus. La plus communément utilisée en traitement d'images numériques est la base rouge, vert, bleu (RVB) qui permet de composer l'image finale sur les écrans d'ordinateurs et moniteurs de télévision. On devra donc toujours, en fin de traitement, projeter l'image couleur dans la base RVB pour permettre son affichage. Le passage de la base RVB à la base LTS correspond au passage des coordonnées cartésiennes aux coordonnées sphériques, où l'axe polaire coïncide avec l'axe achromatique R=V=B (Gillespie 1980). La luminance est définie par le rayon du "vecteur couleur", la teinte par la longitude qui varie de 0° à 360°, et la saturation par la colatitude. D'autres coordonnées sont parfois adoptées pour représenter la base LTS, comme les coordonnées cylindriques (Robertson, O'Callaghan 1988), ou des représentations simplifiées (Smith 1978). Nous avons choisi le modèle sphérique de King *et al.* (1984), qui tient compte de l'absence d'invariance spatiale de la transformation RVB en LTS pour définir la transformation inverse.

Nous allons tout d'abord montrer comment exploiter les trois composantes indépendantes de la couleur, pour visualiser simultanément trois images monochromes. Puis nous examinerons des méthodes qui permettent de concentrer plus de trois images en une seule. Les images utilisées pour illustrer ces méthodes proviennent des capteurs SPOT HRV en mode panchromatique, Landsat TM, et ERS-1 ; elles permettront de souligner les difficultés du mélange de données hétérogènes, issues du radar (micro-ondes) et de l'optique visible et infrarouge. Un des problèmes rencontrés en mélangeant ces images est la forte altération des produits composites par le bruit de speckle présent dans l'image ERS-1. Nous avons donc choisi de filtrer l'image radar, préalablement à toute composition colorée, à l'aide du filtre de Lopes *et al.* (1990). Le site retenu ici pour illustrer les méthodes de changement d'espace de représentation est localisé à Fos-sur-Mer, dans le sud-est de la France, et s'étend sur 15 km de côté (cf. § 5.2.2). Les produits composites réalisés ont été soumis à des photo-interprètes, qui ont formulé des remarques et commentaires que nous reprenons pour étayer notre propos.

Un moyen simple de fusionner trois images monochromes est de les affecter respectivement aux composantes R, V, et B de l'espace des couleurs. Cette démarche est parfaitement justifiée lorsqu'elle est appliquée à trois images prises respectivement dans les bandes spectrales rouge, verte, et bleue : il s'agit d'une composition colorée en vraies couleurs. Une telle composition peut être réalisée sur les bandes 1, 2, et 3 de Landsat TM, comme illustré par la figure 4.1. Les couleurs sont naturelles à l'œil puisqu'il s'agit des vraies couleurs, à la différence près que les fonctions de contraste appliquées à chaque composante RVB permettent de doser le poids de chaque couleur primaire : il est par exemple nécessaire d'atténuer le canal bleu (TM1), dans la composition Landsat TM 1-2-3, car son niveau est artificiellement accru par la diffusion de Rayleigh atmosphérique. La facilité d'interprétation des compositions colorées en vraies couleurs en fait un produit fort apprécié des photo-interprètes. Par contre, si l'on affecte trois autres bandes spectrales de Landsat TM aux composantes RVB, l'image composite est une composition en fausses couleurs. Son interprétation est très difficile, dans la mesure où l'œil est incapable de restituer correctement la proportion de chaque couleur primaire R, V, et B formant la couleur finale. De nombreux capteurs satellitaires ne fournissent pas d'images dans les trois bandes optiques rouge, verte, et bleue conjointement ; il est alors impossible de synthétiser une image en vraies couleurs. Les images SPOT XS par exemple sont fréquemment utilisées pour des compositions en fausses couleurs où (R, V, B) = (XS3, XS2, XS1). Cette composition consiste à translater les images dans le spectre : XS3 (proche infrarouge) → rouge, XS2 (rouge) → vert, et XS1 (vert) → bleu. Elle est illustrée, dans le chapitre 3, par les

figures 3.8 et 3.9. La plupart des photo-interprètes, incluant notamment les géologues, les urbanistes, ou encore les spécialistes de la végétation sont habitués à exploiter une telle composition en fausses couleurs ; néanmoins certains photo-interprètes, davantage sensibilisés à la qualité géométrique qu'aux propriétés spectrales des images, sont parfois déroutés par les couleurs : il est par exemple peu naturel de faire apparaître la végétation en rouge. Il faut bien distinguer ces compositions en fausses couleurs des compositions en pseudo-couleurs : les compositions en fausses couleurs correspondent à la présentation d'une information spectrale avec déformation des couleurs, par exemple par translation du spectre dans le cas d'une composition RVB sur SPOT XS. Par contre, une composition en pseudo-couleurs utilise la couleur pour améliorer la présentation d'une information monospectrale : les couleurs, arbitrairement choisies, sont appliquées à une image monochrome via une "table de couleurs".

La difficulté d'interprétation croît encore lorsque les trois images affectées aux canaux RVB sont issues de capteurs hétérogènes : il est tentant d'associer respectivement les données radar, infrarouge et visible aux couleurs primaires RVB. La figure 4.2 montre la composition (R, V, B) = (ERS, TM5, P), sur laquelle les forts échos radar apparaissent en rouge, et la végétation en vert. Les pixels dont la couleur n'est pas proche d'une couleur primaire, mais résulte d'un mélange équilibré, sont difficiles à interpréter. De plus, la composition RVB d'images de résolutions spatiales différentes aboutit à une image couleur aux qualités géométriques médiocres : les images ERS (résolution 20 m) et TM5 (résolution 30 m) dégradent la résolution géométrique (10 m) de l'image panchromatique. C'est alors qu'il est intéressant d'exploiter la base LTS qui repose sur la perception visuelle humaine. On affecte désormais les images P, TM5, et ERS aux composantes luminance, teinte, et saturation. Il a été constaté que l'œil est davantage sensible à la couleur sur des objets de grande taille que sur les petits objets. Cette propriété a d'ailleurs guidé la définition du système NTSC de transmission télévisée, pour exploiter au mieux une bande passante limitée (Travis 1991) : la résolution accordée à la luminance est supérieure à celle des composantes de couleurs. De manière analogue, nous attribuons l'image de meilleure résolution spatiale (P) à la luminance pour obtenir une composition colorée de bonne qualité géométrique. L'image ERS est associée à la composante saturation, comme proposé par Welch, Ehlers (1988), ce qui permet d'injecter l'information radar sans altérer la résolution géométrique de l'image panchromatique. Enfin l'information infrarouge TM5 est attribuée à la teinte. Le résultat est présenté par la figure 4.3.



Figure 4.1 : composition colorée RVB sur Landsat TM 3-2-1

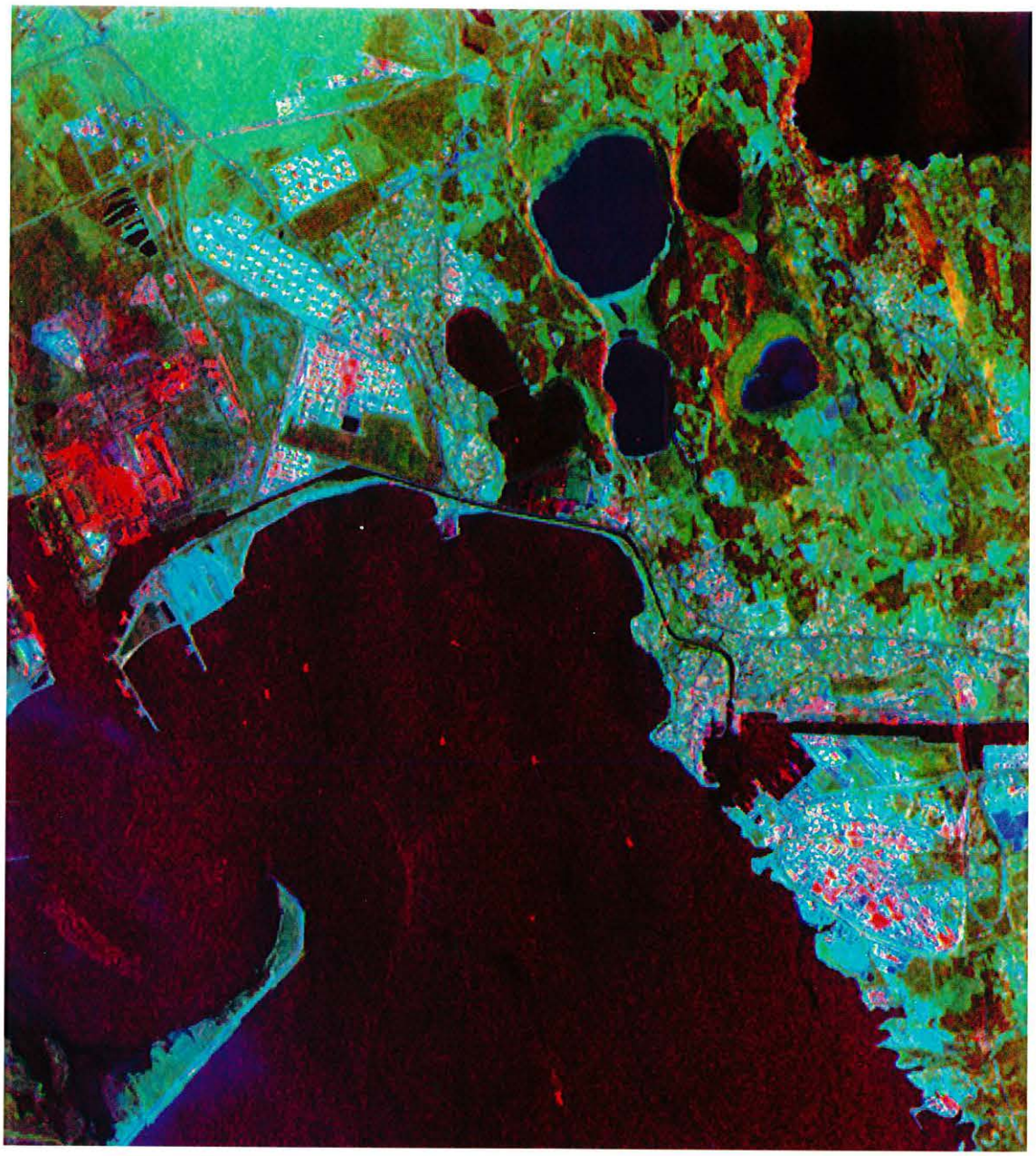


Figure 4.2 : composition colorée RVB sur ERS, TM5, et P

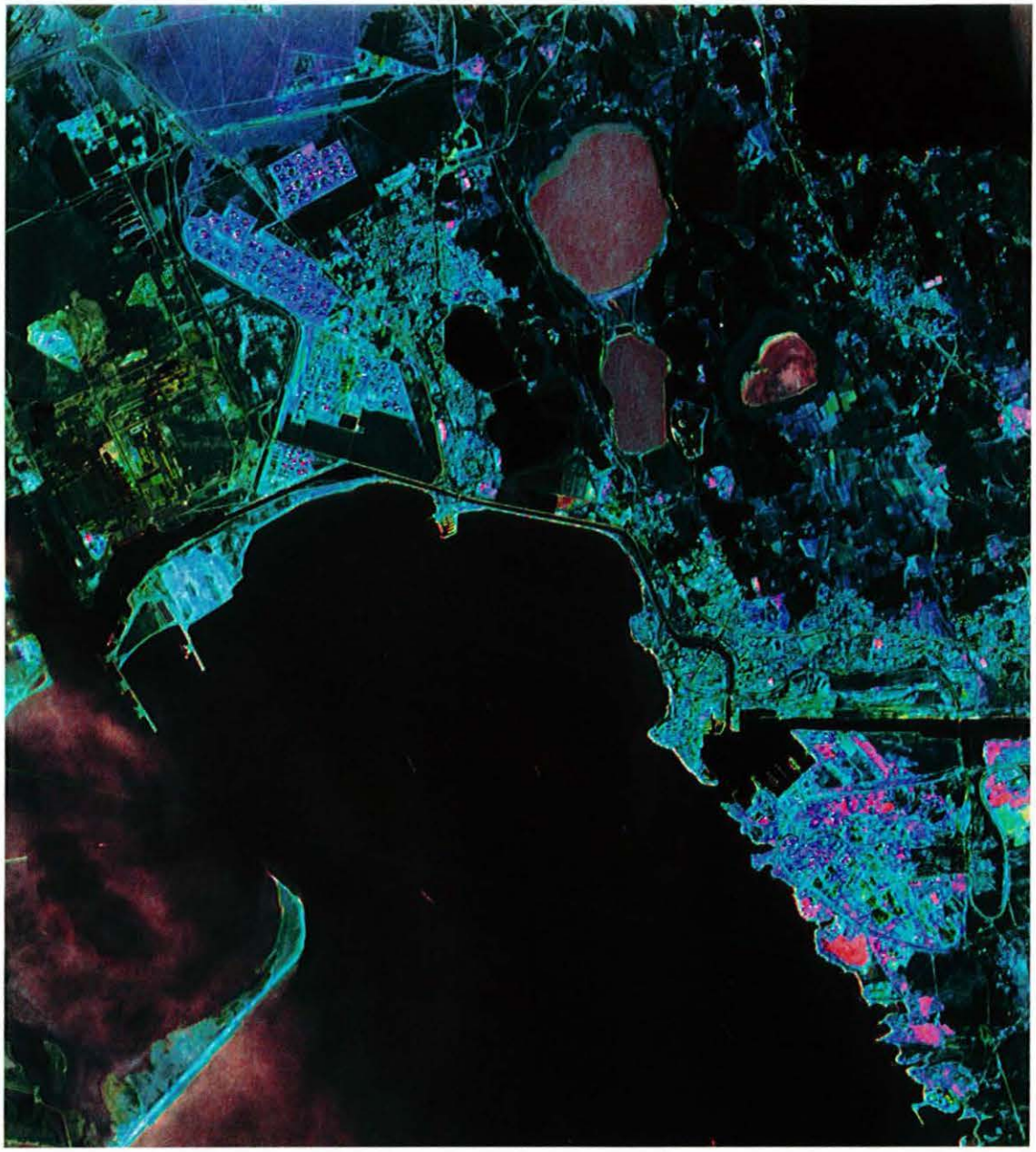


Figure 4.3 : composition colorée LTS sur P, TM5, et ERS

On peut apprécier la bonne qualité géométrique de la composition colorée de la figure 4.3, et les forts contrastes entre les couleurs qui permettent de séparer efficacement les parcelles. Cependant les couleurs très artificielles rendent difficile l'interprétation thématique des constituants de l'image. Afin de remédier à cet inconvénient, nous allons maintenant présenter une méthode qui permet de combiner les couleurs naturelles fournies par TM 1, 2, et 3, à l'image radar ERS et l'image haute résolution P.

Une telle méthode doit permettre de fusionner plus de trois images, et nous avons vu que l'espace des couleurs était limité à trois composantes indépendantes. Le principe est d'isoler la teinte déduite des images TM 1-2-3, et de substituer l'information P et/ou ERS aux autres composantes de l'image couleur (luminance et saturation). La figure 4.4 détaille les différentes étapes :

- une transformée RVB $\rightarrow$ LTS est appliquée aux images TM 1-2-3,
- l'image P ou ERS est substituée à l'image luminance ou saturation, mais la teinte représentant les couleurs naturelles est préservée,
- une transformée LTS $\rightarrow$ RVB fournit l'image couleur fusionnée.

De multiples combinaisons sont possibles : on peut injecter soit P seule, soit ERS seule, par substitution à la luminance ou à la saturation, ou encore injecter les deux simultanément. Comme nous l'avons expliqué ci-dessus, il est conseillé d'affecter l'image de meilleure résolution spatiale à l'image de luminance. La fusion ainsi réalisée sur TM 1-2-3 et SPOT P est une méthode classique d'amélioration de la résolution spatiale d'un jeu de trois images multispectrales (cf. méthodes de changement de base de l'annexe A). Dans le cas où l'on injecte l'image radar ERS seule, sa résolution spatiale (20 m) n'est plus critique vis-à-vis de celle des images Landsat TM 1-2-3 (30 m). On peut donc la substituer à l'image de luminance (Harris, Murray 1990), ce qui donne une image fusionnée représentée sur la figure 4.5. On constate que les couleurs naturelles de la composition initiale TM 1-2-3 (cf. fig. 4.1) sont bien conservées, et que les forts échos radar représentés par une luminance élevée ressortent distinctement.

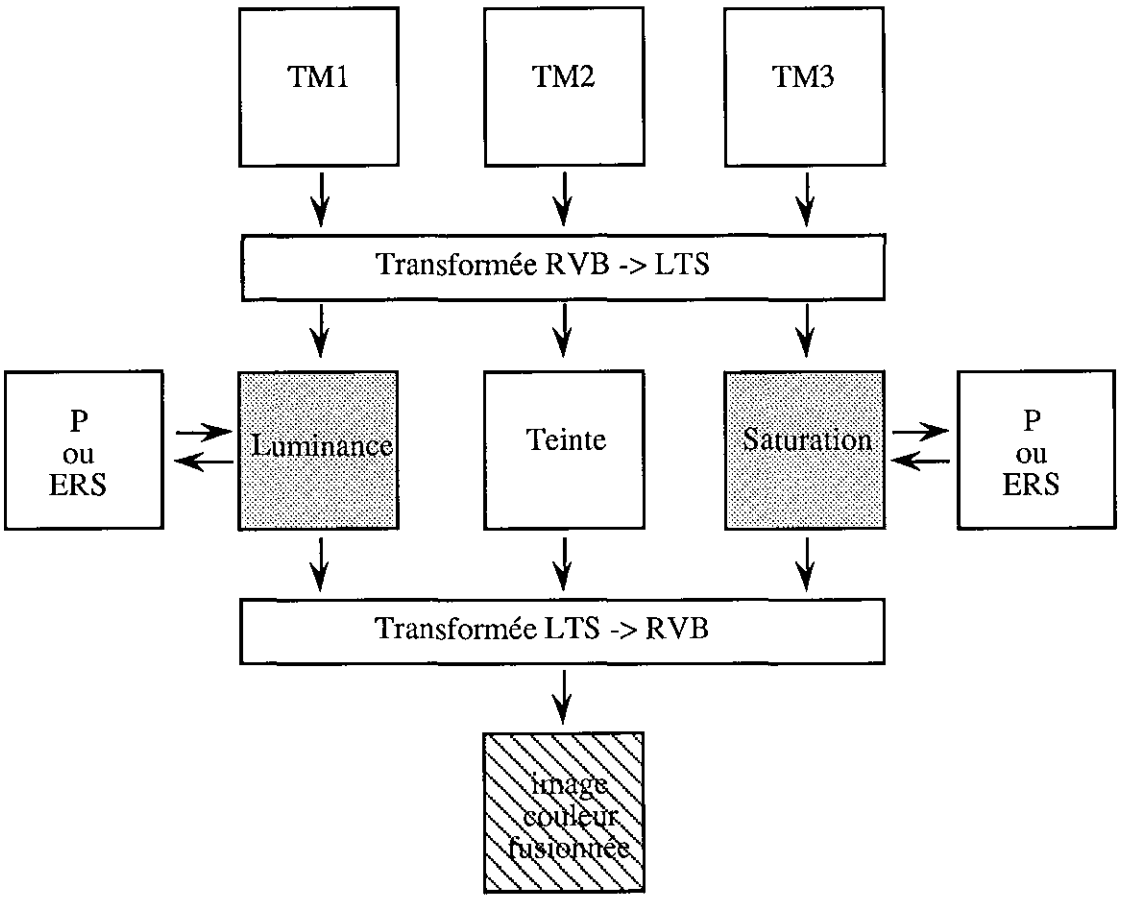


Figure 4.4 : principe de la méthode de substitution dans le repère LTS

Lorsqu'on souhaite fusionner davantage d'images, il est possible de substituer à la luminance ou à la saturation non plus une image, mais une combinaison de plusieurs images. On peut par exemple remplacer la luminance, calculée sur TM 1-2-3, par une combinaison linéaire de P, ERS, et des bandes TM dans l'infrarouge. Si le produit fusionné peut être agréable à l'œil, il en devient néanmoins pratiquement inexploitable en termes de caractérisation thématique : la complexité de la composition est un obstacle à la restitution de ses composantes.

Il est alors d'usage de faire appel à des techniques issues de l'analyse statistique, telles que l'analyse en composantes principales (ACP). Le principe est le suivant : chaque vecteur pixel, dont les composantes sont les radiométries du pixel dans les différentes images à fusionner, est représenté par un point de l'espace $\Re^n$. On cherche à extraire un sous espace de $\Re^n$ qui décrive au mieux l'information du nuage de points, c'est-à-dire à concentrer cette information en un minimum d'images. Le critère adopté pour mesurer la quantité d'information est la variance. L'ACP fournit un nouveau repère, dont les axes orthogonaux, appelés axes factoriels, sont tels que la

projection du nuage de points sur chacun d'eux possède une variance maximale (voir par exemple Escoffier, Pagès 1988). La projection du nuage sur chaque axe factoriel constitue une image appelée composante principale (CP). Les CP sont ordonnées dans l'ordre décroissant des variances qui leur sont associées : la première CP est donc celle qui représente le plus d'information. La somme des variances associées aux CP est égale à la variance initiale totale du nuage. On constate souvent que la variance diminue très vite d'une CP à la suivante, ce qui permet de décrire l'essentiel de l'information en se restreignant à un nombre de CP inférieur à n. Dans le cas où l'ACP est appliquée aux images P, ERS, et TM 1-2-3-4-5-7 de Fos, la première CP décrit 75 % de l'information, la deuxième 11 %, la troisième 8 %, etc. La concentration de l'information des huit images initiales en trois CP reflète donc 94 % de l'information. L'ACP est également considérée comme une technique de transformation d'un jeu d'images sources intercorrélées en un ensemble de CP décorrélées (Gillespie *et al.* 1986). Les trois premières CP peuvent être visualisées à l'aide d'une composition colorée, qui à elle seule concentre 94 % de l'information des huit images initiales. Une telle composition colorée est cependant difficilement exploitable à des fins de photo-interprétation, car elle s'éloigne trop des grandeurs physiques mesurées.

L'ACP appliquée aux huit images précédentes de Fos aboutit à une première CP qui est approximativement la moyenne des images ; en conséquence le capteur TM, représentant six images sur huit, va masquer l'information en provenance des capteurs SPOT et ERS. Cette remarque s'étend aux deuxième et troisième CP pour lesquelles les poids comparables des bandes visibles s'ajoutent, de même que les poids des bandes infrarouge dont le signe est opposé. Pour ne privilégier aucun capteur, nous sélectionnons une seule image en provenance de chacun d'eux : l'ACP sera calculée sur l'image SPOT P, l'image ERS, et une bande de Landsat TM. La matrice de corrélation révèle que l'image TM la moins corrélée avec l'image P est TM4. Nous choisissons par conséquent cette bande qui, selon le critère de corrélation, est la plus complémentaire de l'image P. L'ACP appliquée à ces trois images de nature différente (visible, infrarouge, et radar) donne les résultats du tableau 4.1.

Le tableau se lit verticalement : chaque CP est une combinaison linéaire des images sources, dont les coefficients constituent une colonne. Les CP représentent respectivement 65 % (CP1), 23 % (CP2), et 12 % (CP3) de l'information totale. Il est judicieux de chercher à interpréter physiquement les CP (Yésou *et al.* 1993). Sur cet exemple, la première CP est approximativement égale à la moyenne des images. La deuxième représente la différence entre les images optiques (visible et infrarouge) et

l'image radar. Enfin, CP3 est pratiquement égale à la différence entre l'image TM4 et l'image P ; cette différence entre une bande proche infrarouge et une bande visible peut être assimilée à un indice de végétation (Clevers 1988). Les images CP1 et CP3 ont été les plus appréciées des photo-interprètes ; elles sont représentées respectivement par les figures 4.6 et 4.7.

	CP1	CP2	CP3
P	0,83	0,39	-0,40
ERS	0,71	-0,70	-0,09
TM4	0,87	0,20	0,46

Tableau 4.1 : résultats d'ACP sur les images P, ERS, et TM4 de Fos

L'ACP peut également être utilisée en fusion d'images, pour améliorer la résolution spatiale d'un jeu de trois images multispectrales à l'aide d'une image de meilleure résolution, de manière analogue à la méthode utilisant la transformation LTS (cf. méthodes de changement de base de l'annexe A).

Les méthodes centralisées de projection des images sources dans un nouvel espace de représentation facilitent dans certains cas le travail du photo-interprète, mais elles conduisent parfois à certaines critiques. Tout d'abord, les images en fausses couleurs sont peu appréciées de certains photo-interprètes, qui leur préfèrent les compositions en vraies couleurs (couleurs naturelles). Une autre critique concerne les mélanges d'images de résolutions spatiales différentes, qu'il s'agisse de compositions colorées ou d'ACP : l'image résultante possède en général une qualité géométrique inférieure à celle de l'image source de meilleure résolution spatiale. Enfin, pour les compositions hétérogènes optique-radar, l'image radar "pollue" les images optiques dans le sens où elle en dégrade les détails sur une grande partie de la scène. On peut en effet considérer, pour de nombreuses applications, que l'information utile du radar est localisée sur les objets de petite taille présentant un écho particulier. Sur les surfaces étendues homogènes où l'image radar comporte essentiellement un bruit de speckle, cette image altère les détails présents dans les images optiques lorsqu'on les combine pixel à pixel. En l'absence de filtrage du speckle, cette dégradation est quasiment rédhibitoire, car les détails des images optiques sont brouillés par le speckle.

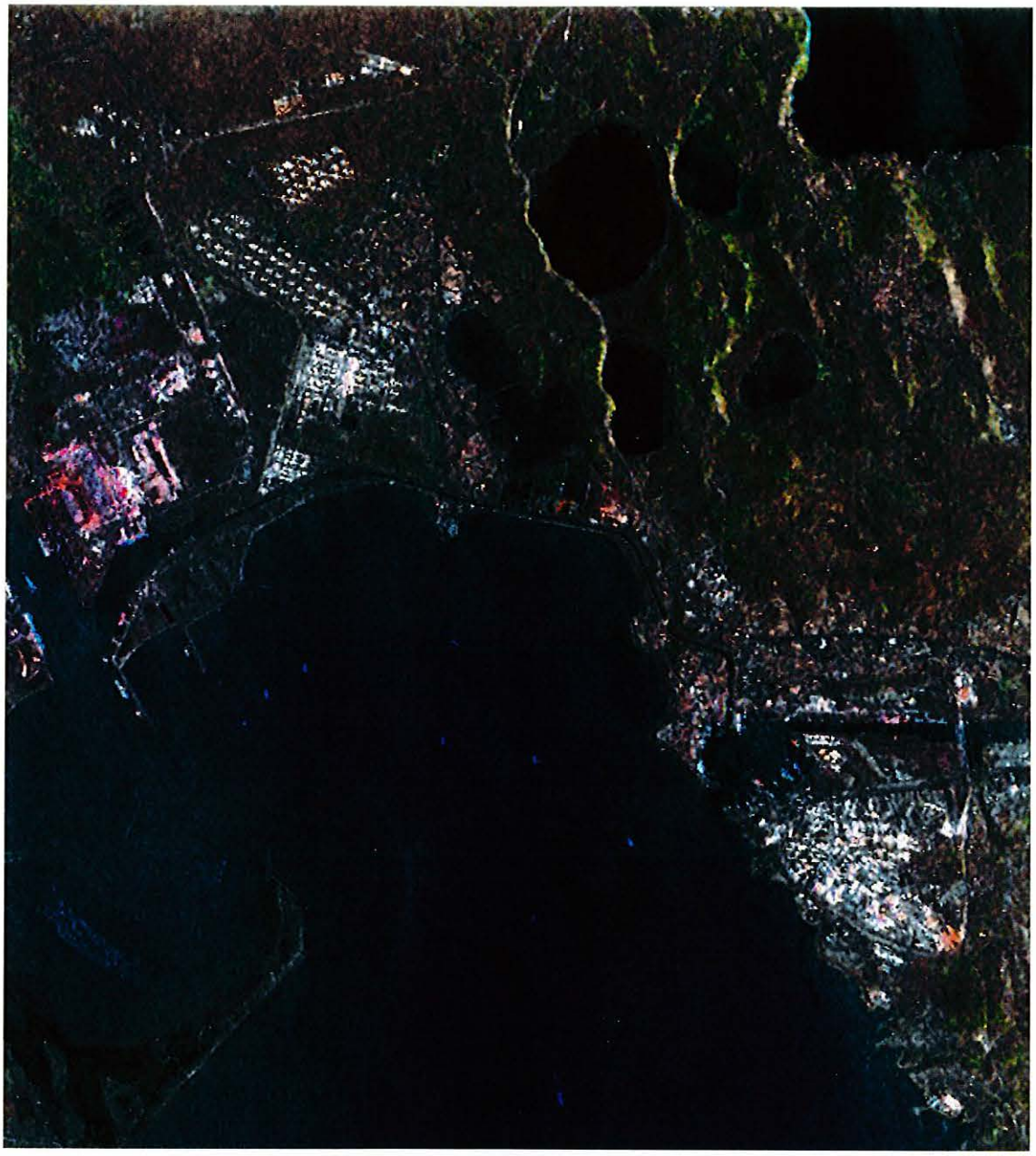


Figure 4.5 : composition colorée TM 3-2-1 avec transformation LTS et substitution de l'image ERS à la luminance



Figure 4.6 : première composante principale calculée sur P, ERS, et TM4



Figure 4.7 : troisième composante principale calculée sur P, ERS, et TM4

Si l'image radar est filtrée, comme dans les exemples précédents, sa radiométrie sur les surfaces homogènes devient constante, donc les détails des images optiques ne sont souvent qu'atténués. Afin d'éviter une telle perte de qualité géométrique, nous suggérons d'extraire des informations ponctuelles de l'image SAR et de les injecter dans les images optiques, qui constituent une référence pour les photo-interprètes.

4.3 UNE METHODE DE FUSION D'IMAGES OPTIQUE ET RADAR

La méthode de fusion d'images optique-radar développée consiste à extraire une information caractéristique des objets de petite taille qui peuvent être détectés dans l'image radar, et à la présenter visuellement par superposition au contexte optique. Le seuillage des points brillants dans l'image radar ne fournit qu'une partie de l'information ponctuelle détectable. On peut caractériser les petits objets de manière plus complète par leurs structures caractéristiques à différentes échelles. Nous utilisons à cette fin l'analyse multirésolution et la transformée en ondelettes, qui sont également exploitées par la méthode ARSIS (cf. chap. 3). La méthode de fusion optique-radar a été appliquée à des ortho-images SPOT panchromatique et ERS-1 sur le site de Marignane, dans le sud-est de la France (cf. § 5.2.2). Il aurait été préférable de travailler sur l'image ERS-1 brute (niveau PRI par exemple), pour éviter l'altération des petites structures induite par le rééchantillonnage. Mais la méthode de fusion requiert deux images parfaitement recalées, et nous ne disposons pas de l'image SPOT dans la géométrie de l'image ERS. Les illustrations sont restreintes géographiquement à l'aéroport, pour lequel de nombreux objets sont détectables sur l'image ERS, tout en présentant une densité suffisamment faible pour être séparés les uns des autres (contrairement aux concentrations urbaines).

L'image ERS est traitée indépendamment de l'image SPOT P, par opposition aux méthodes de projection des données dans un nouvel espace. Le résultat du traitement est une image d'attributs au niveau pixel, qui caractérisent les objets ponctuels détectables dans l'image ERS. Dans un second temps, l'image d'attributs est fusionnée avec l'image P. On peut donc parler de fusion décentralisée dont le schéma d'intégration, basé sur le formalisme adopté au chapitre 2, est présenté sur la figure 4.8. La cellule F_1 désigne l'algorithme d'extraction de petites structures, basé sur la

transformée en ondelettes. La fonction réalisée par F_2 est une incrustation graphique des attributs pixels de ERS dans l'image SPOT P.

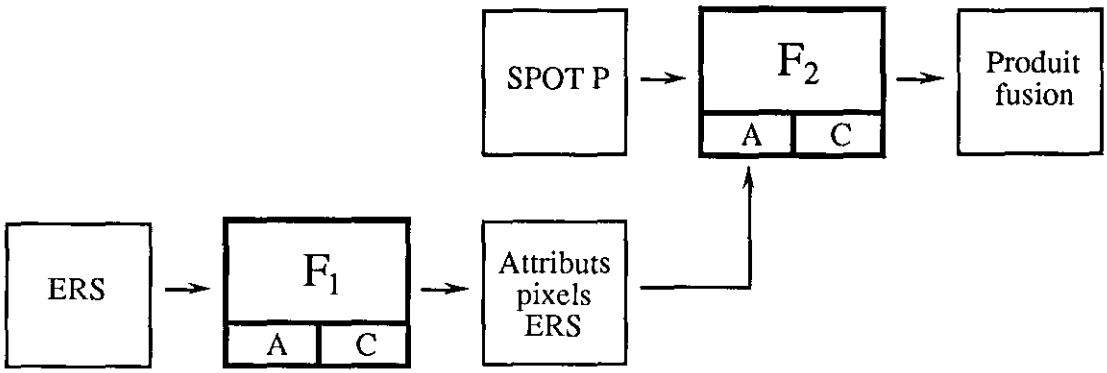


Figure 4.8 : schéma d'intégration de la méthode de fusion optique-radar

L'image d'attributs pixels ERS est le résultat du filtrage passe-haut des coefficients d'ondelettes, obtenus par l'algorithme à trous de Dutilleux (1987) qui fournit une unique image de coefficients à chaque résolution. Puisque les structures d'intérêt sont les petites structures, on se limite à la première décomposition en ondelettes. La distribution des coefficients d'ondelettes est une gaussienne généralisée (Antonini *et al.* 1992), ce qui nous a amenés à utiliser l'écart-type comme base pour le seuillage. Le seuil peut être ajusté en fonction de l'image traitée et des structures désirées ; on constate en général qu'un seuil égal à trois fois l'écart-type de la distribution isole correctement les structures d'intérêt, et élimine l'essentiel du speckle. Le résultat du filtrage est une image de coefficients d'ondelettes, dont les valeurs sont non nulles sur les petites structures de forte amplitude, et nulles sur le reste des pixels. Il existe différentes possibilités pour réinjecter cette information radar dérivée dans l'image optique.

De manière similaire à la méthode ARSIS, on peut envisager de faire appel à un modèle qui s'applique aux coefficients d'ondelettes de l'image P et aux coefficients filtrés de ERS. Ce modèle combinerait les coefficients des deux images pour synthétiser une image fusionnée de coefficients, destinée à la reconstruction du produit final par transformée en ondelettes inverse à partir du contexte optique. Malheureusement, on constate que les coefficients d'ondelettes issus des images P et ERS sont très décorrélés, même en se limitant aux structures isolées par le filtrage des coefficients de ERS. Cela traduit la nature physique différente des structures issues de l'optique et du radar. La combinaison numérique des coefficients d'ondelettes optique et radar n'est donc pas

justifiée, mais il est possible de substituer les coefficients filtrés non nuls de ERS à ceux de l'image P. Le résultat de la reconstruction est une image P sur laquelle ont été ajoutés localement des points brillants ou sombres, suivant le signe de la transition détectée dans l'image ERS. Un tel produit est difficile à exploiter car on ne peut identifier la contribution du radar à la radiométrie d'un pixel. Nous proposons par conséquent de ne pas combiner les radiométries des images optique et radar, mais de présenter les structures radar détectées sous forme d'un plan graphique superposable à l'image optique.

Il est préférable d'utiliser la couleur pour incruster dans l'image SPOT P les structures détectées sur l'image ERS, afin de bien séparer visuellement les deux types d'information. L'image d'attributs radar est exploitée de manière binaire : F_2 marque chaque pixel dont la valeur dans l'image filtrée de coefficients d'ondelettes est non nulle par un pixel coloré sur l'image P. Les autres pixels, qui correspondent à l'absence de détection en radar, conservent la radiométrie initiale de l'image P. Les images sources SPOT panchromatique et ERS-1 de l'aéroport de Marignane sont présentées par les figures 4.9 et 4.10. Le résultat de la fusion est illustré par la figure 4.11, sur laquelle les structures radar sont marquées en rouge.

On remarque tout d'abord que la plupart des structures radar détectées, représentées en rouge sur le résultat de fusion, ne sont pas visibles à l'œil sur l'image radar. L'algorithme F_1 d'extraction d'attributs radar ne peut donc pas être remplacé par l'œil du photo-interprète. Par ailleurs, les échos radar multiples sur les bâtiments se mélangent et sont difficilement exploitables sur l'image fusionnée. Les structures radar détectées présentant un intérêt sont localisées sur des zones homogènes telles que les pistes d'aéroport, les taxiways, les champs, les parkings, etc. Nous avons cherché à interpréter les structures détectées, qui résultent de traitements successifs complexes : traitement SAR, création d'une ortho-image puis transformation en ondelettes. Monsieur L. Wald s'est rendu sur le terrain où il a pu effectuer une reconnaissance, en particulier sur la portion nord de l'aéroport. Certaines structures radar détectées au bord des pistes sur l'image ERS ont été identifiées comme provenant d'antennes, de réflecteurs radar, d'un transformateur ou encore de véhicules en stationnement. D'autres n'ont pu être identifiées, en particulier les structures issues d'objets mobiles, puisque leur configuration à la date de la prise de vue ERS n'est pas connue. Lors de l'exploitation du produit de fusion, il faut bien garder à l'esprit que les images SPOT et ERS-1 n'ont pas été prises à la même date. Cette contrainte due à l'indépendance des deux plates-formes satellitaires n'est pas un handicap majeur, mais il faut dater les

objets en fonction de l'image dans laquelle ils sont détectés. Nous remarquerons enfin que l'apport de la méthode se limite à la détection d'objets ponctuels, car l'identification des structures radar mises en évidence parait très difficile.

4.4 CONCLUSION

Les méthodes de présentation visuelle du volume de données représentent une aide précieuse à la photo-interprétation. L'objectif est à la fois d'améliorer la qualité visuelle de l'information, par rapport à l'exploitation indépendante des images, et de concentrer cette information en un minimum d'images pour alléger le travail de manipulation d'images du photo-interprète. Les méthodes centralisées de projection des images dans un nouvel espace de représentation sont adaptées à des données homogènes, c'est-à-dire de même nature physique. Les compositions colorées sur des images optiques en sont un exemple, dont l'apport en photo-interprétation a été démontré. Les limites de ces méthodes sont atteintes lors de la fusion de données hétérogènes, telles que les images optiques et radar. Des méthodes décentralisées, plus spécifiques à l'objectif recherché, sont alors préférables. Nous avons développé une telle méthode, qui extrait des attributs pixels d'une image radar et les réinjecte dans une image optique, laquelle fournit un contexte favorisant l'interprétation des attributs détectés.

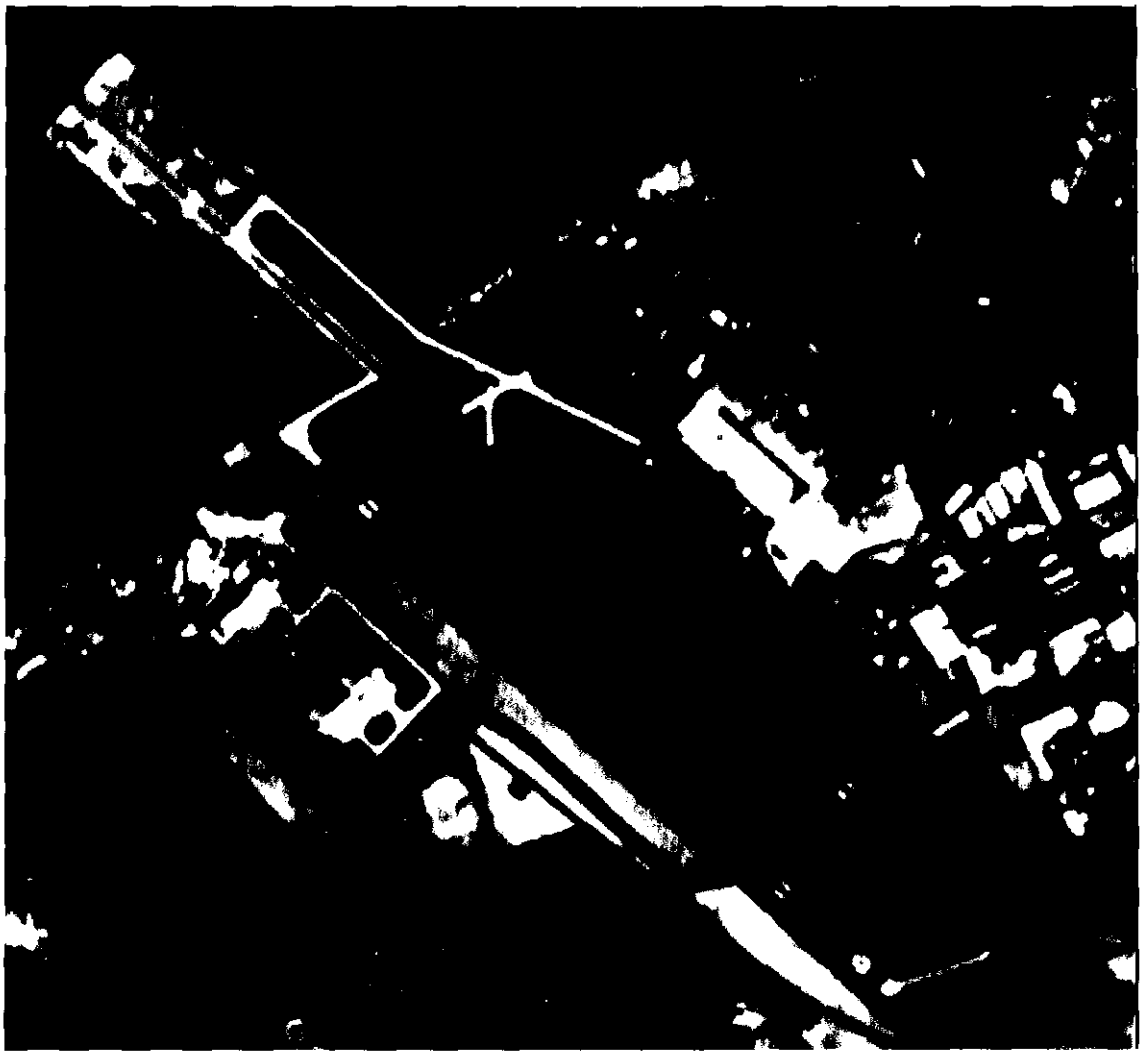


Figure 4.9 : image SPOT P de l'aéroport de Marignane

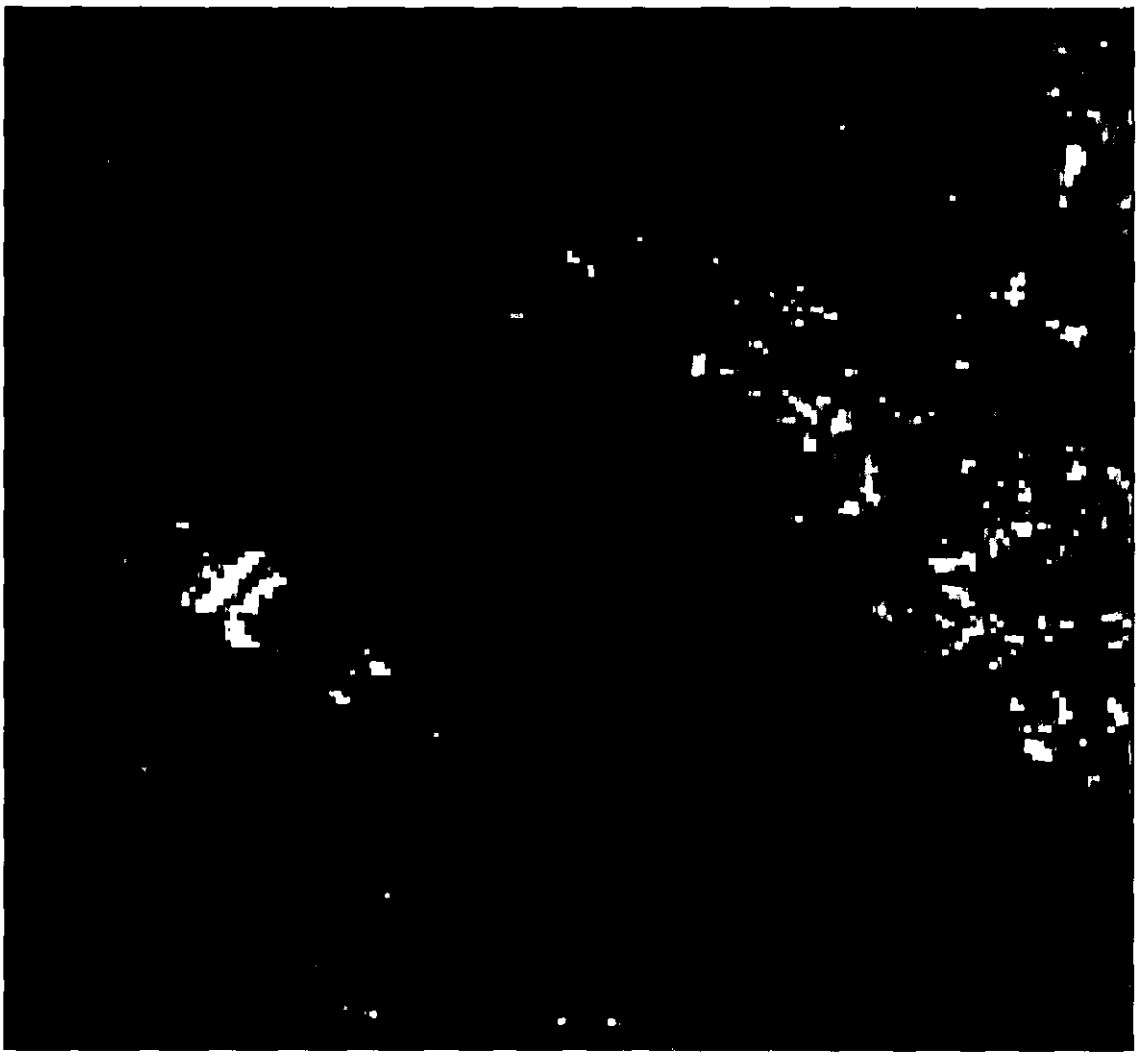


Figure 4.10 : image ERS-1 de l'aéroport de Marignane

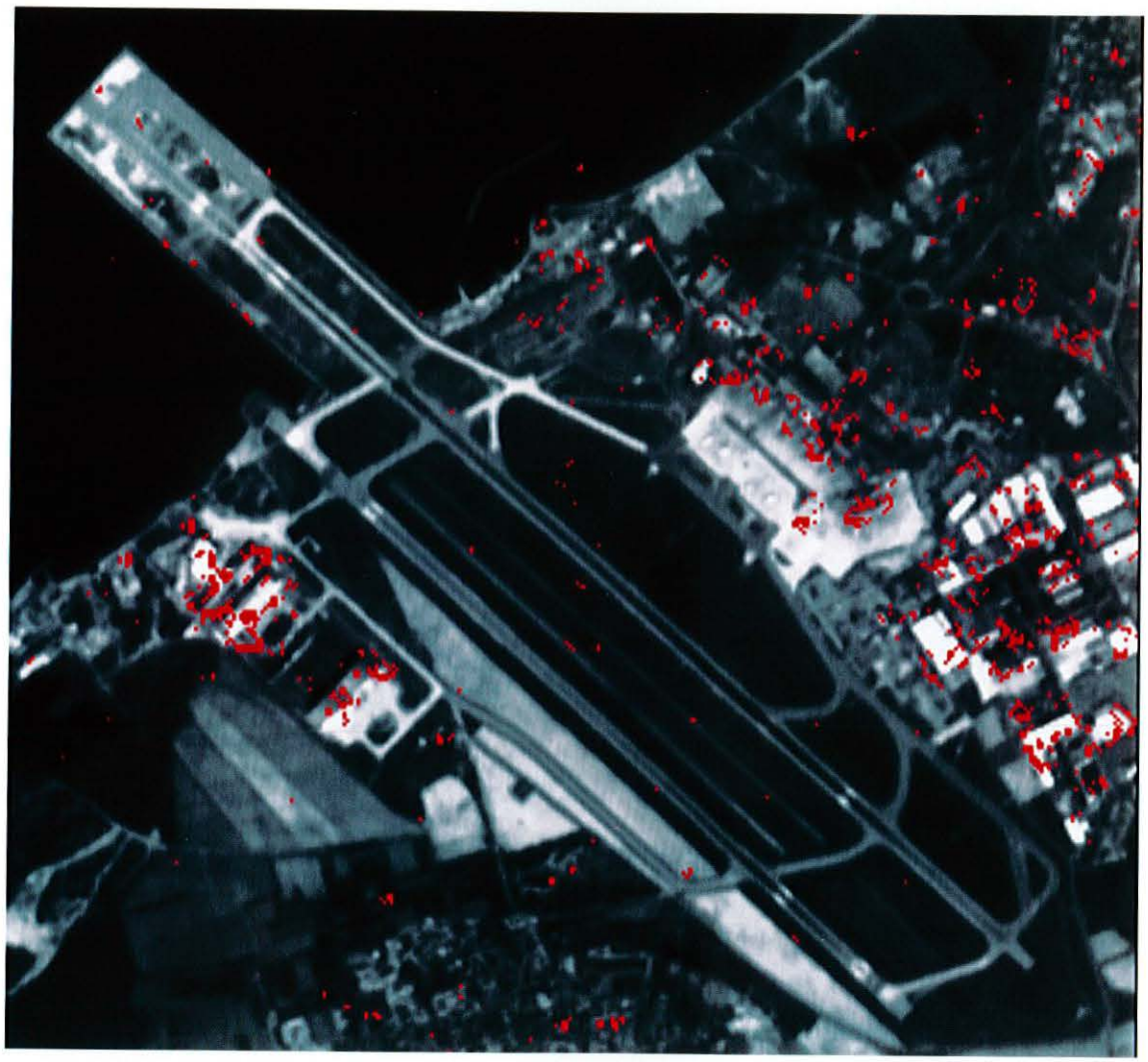


Figure 4.11 : image fusionnée SPOT P + ERS-1 de l'aéroport de Marignane

Chapitre 5

CLASSIFICATION - ANALYSE DE DIFFERENTES ARCHITECTURES DE FUSION

5.1 INTRODUCTION

La classification d'images fait partie des méthodes couramment utilisées pour exploiter des images satellitaires d'observation de la Terre. Elle consiste à affecter à chaque entité de la scène traitée une étiquette indiquant son appartenance à une classe particulière. L'entité utilisée peut être par exemple le pixel ou une région issue d'une segmentation préalable. Par souci d'homogénéité avec l'ensemble des applications traitées dans la thèse, nous avons privilégié l'approche au niveau pixel. La classification d'images revient alors à réaliser, à partir d'une ou plusieurs images numériques, une carte dont chaque pixel a pour valeur une étiquette caractéristique de la classe à laquelle il appartient. Ainsi, pour des applications d'analyse du paysage par exemple, la classification permet de quantifier l'occupation des sols.

Il existe de nombreux algorithmes de classification (voir par exemple Schowengerdt 1983), mais on ne cherchera pas ici à comparer leurs performances. Au contraire, on s'attachera à en mettre en œuvre quelques uns, de différentes façons, sur les données à classifier, et à tirer des conclusions indépendantes de l'algorithme utilisé. Les algorithmes de classification ont en commun une mise en œuvre d'autant plus lourde que le volume de données augmente : pour n images d'entrée, la classification revient à chercher des surfaces de décision dans un espace de dimension n (Gibson, Cowan 1990). L'espace mémoire nécessaire et le temps de calcul augmentent donc beaucoup avec le nombre d'images n. Un moyen de contourner ces contraintes consiste à réaliser des classifications indépendantes sur des sous-ensembles d'images d'entrée, nécessitant chacune une charge de calcul limitée, puis à fusionner ces classifications. Ces deux approches illustrent la mise en œuvre d'architectures de fusion, dites respectivement centralisée et décentralisée (Hall 1992), à des fins de classification. Nous proposerons

ensuite une architecture dite "mixte", dans le but de tirer parti des avantages des deux architectures précédentes.

Au delà du choix des algorithmes de classification, il est donc nécessaire de définir une architecture de fusion qui intègre ceux-ci, et dont l'incidence sera traduite tant en termes de performances qu'en termes de flexibilité de mise en œuvre. Les images dont nous disposons pour mettre en œuvre ces architectures sont des images SPOT panchromatiques, ERS-1, et Landsat-TM. Notre propos étant la comparaison des architectures, nous chercherons à nous affranchir autant que possible du classificateur utilisé et de la scène traitée ; deux classificateurs différents et deux sites seront donc examinés. Les architectures sont comparées tant d'un point de vue qualitatif que quantitatif : nous envisageons donc la flexibilité de mise en œuvre ainsi que les performances traduites par les taux de classification. Ce dernier point requiert l'utilisation d'une vérité terrain, que nous avons réalisée sur des zones limitées extraites des images.

Nous examinerons d'une part l'apport de la fusion, utilisant chacune des architectures précédentes, vis-à-vis des résultats fournis par chaque capteur. D'autre part, nous tenterons de déterminer quelle est l'architecture la mieux adaptée à chaque cas, compte-tenu des données disponibles et des contraintes de l'utilisateur.

5.2 CLASSIFICATION D'IMAGES

5.2.1 Choix des classificateurs

Les algorithmes de classification peuvent être regroupés en deux grandes familles (Duda, Hart 1973). La première, appelée classification non-dirigée, ne fait pas appel à une connaissance extérieure du problème : les pixels sont classés seulement à partir d'un critère de distance défini dans l'espace spectral. Elle est donc guidée par les données et non par les buts à atteindre. La classification non-dirigée fournit une classification contenant un nombre prédéterminé ou non de classes, qui doivent être identifiées a posteriori. Une telle classification ne fournit pas à l'utilisateur une "carte" dont les étiquettes sont thématiquement connues à l'avance. C'est pourquoi nous avons préféré, pour notre application, une autre famille de classificateurs réalisant une classification dirigée (ou supervisée).

La classification dirigée utilise une connaissance partielle du problème, fournie par l'utilisateur. L'intégration de cette connaissance est appelée apprentissage, et consiste à sélectionner dans les images quelques échantillons représentatifs de chaque classe recherchée. L'algorithme comparera alors les pixels à classer aux échantillons connus, pour fournir une classification contenant les classes présélectionnées lors de l'apprentissage. Une classe supplémentaire de rejet est généralement utilisée pour regrouper les pixels dont la ressemblance avec les échantillons d'apprentissage est trop faible.

Dans un but de commodité opérationnelle, on utilisera donc des algorithmes de classification dirigée. On considérera de plus une autre subdivision majeure des classificateurs, aboutissant à des méthodes dites paramétriques ou non-paramétriques (Duda, Hart 1973). Les méthodes paramétriques, souvent de type statistique, cherchent à modéliser les classes d'intérêt pour chaque capteur, puis à combiner ces modèles pour effectuer une classification multicapteurs. Leur inconvénient majeur est de nécessiter des hypothèses statistiques très fortes sur les données. Ainsi, la classification bayésienne, pour être applicable, modélise généralement les classes par des distributions gaussiennes, alors qu'on sait que cette hypothèse n'est pas vérifiée en général sur les données de télédétection. Cette remarque s'applique en particulier aux classes et capteurs choisis pour cette étude (cf. § 5.2.2). De plus, les méthodes paramétriques requièrent des données multiples souvent inconnues de l'utilisateur, comme les probabilités *a priori* en classification bayésienne, qui sont pourtant déterminantes quant au résultat.

Pour affranchir le photo-interprète de ces contraintes, on utilisera ici des méthodes non-paramétriques, qui ne modélisent pas les classes pour chaque capteur. Nous avons développé et testé deux classificateurs non-paramétriques. Le premier est connu sous le nom de K-plus proches voisins (Milgram 1993), et sera noté K-PPV. L'autre, appartenant aux méthodes connexionnistes, est un réseau de neurones de type perceptron multicouches (Widrow, Lehr 1990). Ces deux classificateurs sont décrits en annexes B et C respectivement.

5.2.2 Capteurs disponibles et classes choisies

Les images dont nous disposons sont des images SPOT panchromatiques (P), des images Landsat-TM prises dans les 7 bandes spectrales, ainsi que des images SAR prises par ERS-1. Ces images ont été recalées précisément à l'aide d'un modèle numérique de terrain (MNT), très utile sur le relief pour obtenir des ortho-images de

bonne qualité ; il s'agit de produits commerciaux. Les résolutions spatiales de ces images sont de 10 m pour SPOT P, 25 m environ pour ERS-1, et 30 m pour les bandes 1-2-3-4-5-7 de Landsat. La bande TM6 (infrarouge thermique) n'est pas utilisée ici en raison de sa résolution spatiale de 120 m. Afin d'effectuer une classification au niveau pixel, il est nécessaire d'avoir une correspondance pixel à pixel entre les différentes images. Les ortho-images possèdent approximativement le même pas que celui des images brutes, à savoir 10 m pour SPOT P, 20 m pour ERS-1, et 30 m pour Landsat. Un rééchantillonnage à un pas commun s'impose donc. Le pas retenu est le plus fin (10 m) pour éviter une perte d'information ; plusieurs techniques permettent d'amener toutes les images à ce pas. On souhaite dans ce chapitre mettre l'accent sur les architectures. On a donc préféré utiliser des techniques standards pour les traitements annexes, dont ce rééchantillonnage, alors qu'une application de la méthode dite ARSIS (cf. chap. 3) aurait été possible. Un rééchantillonnage par simple duplication de pixels a été effectué. Nous disposons finalement, pour chaque pixel à classifier, d'un vecteur composé de ses radiométries dans les différentes images. A ces composantes spectrales, nous avons ajouté une composante spatiale qui traduit la variation locale de la radiométrie autour du pixel considéré. Ce paramètre, souvent appelé texture, permet de discriminer des classes spectralement identiques. En raison de sa très bonne résolution spatiale, l'image SPOT P fournit une information texturale riche. C'est donc sur cette image que la texture sera calculée. De très nombreux paramètres de texture, plus ou moins complexes, sont présentés dans la littérature. Notre objectif n'étant pas de comparer ces paramètres, mais de les utiliser pour lever une ambiguïté sur des classes, nous avons retenu la variance locale qui remplit correctement cet office sur les cas traités. Un essai a été également effectué à l'aide des coefficients d'ondelettes.

En fait, la mise en œuvre d'une architecture décentralisée nécessite d'éclater les composantes du vecteur défini précédemment. On pourrait envisager toutes les combinaisons possibles des différentes composantes, mais cette multiplicité de cas risquerait de nuire à la clarté de la comparaison des architectures. Nous envisagerons donc le point de vue opérationnel de la disponibilité de chaque capteur, et non la disponibilité d'un ensemble quelconque de bandes spectrales. Ainsi, le capteur SPOT sera représenté par sa bande panchromatique P, Landsat-TM par l'ensemble de ses six canaux à 30 m, et ERS-1 par l'image SAR exprimée en module. Nous montrerons plus loin l'apport du filtrage du speckle de cette dernière, en termes de taux de classification.

Les deux sites choisis sont localisés dans le sud-est de la France, à Fos-sur-Mer et Marignane (cf. fig. 5.1), et leurs étendues respectives sont de 15 km x 15 km et

11,5 km x 11,5 km. Le site de Fos comporte de la mer, des étangs, des ports, des complexes sidérurgiques et chimiques, des agglomérations urbaines et un peu de végétation (cf. fig. 5.2). Celui de Marignane comprend un aéroport, une portion de l'étang de Berre, des salins, ainsi que divers types d'agglomérations et des zones boisées (cf. fig. 5.3).



Figure 5.1 : localisation des sites traités (France)

Les classes retenues sont très générales et volontairement limitées en nombre, d'une part parce que nous ne nous sommes pas attachés à une application spécifique, d'autre part afin de retrouver ces mêmes classes sur les deux sites. Il s'agit des classes eau, sols boisés, sols non boisés, et ville. Par sols boisés, nous désignons les forêts, les broussailles, et la végétation chlorophyllienne en général dont la signature est caractéristique en proche infrarouge. La classe "sols non boisés" comprend les autres types de sols, nus ou recouverts de végétation non chlorophyllienne. La difficulté inhérente à l'utilisation de classes aussi générales réside dans leur grande hétérogénéité, qui se manifeste à différentes échelles, définies relativement à la résolution des capteurs (voir par exemple Fung, Chan 1994).

La radiométrie du pixel résulte de l'intégration spatiale des radiométries des éléments qui le composent, et n'est donc pas nécessairement caractéristique d'un objet précis. A l'échelle de quelques pixels, une classe peut présenter une grande variabilité spectrale, faisant que deux pixels voisins traduisent les luminances de deux terrains différents. C'est le cas de la classe urbaine, composée notamment d'immeubles, de maisons, de routes, d'arbres, etc. Enfin, une classe peut être localement homogène, comme l'eau, mais la radiométrie de cette classe peut varier dans l'image. Ainsi, pour le site de Fos-sur-Mer, la radiométrie de l'eau est très différente sur la mer et sur chacun des nombreux étangs (cf. fig. 5.2).

5.2.3 Sélection des données d'apprentissage et des données test

L'estimation des performances des classificateurs dans les différentes architectures sera fondée sur les taux de classification (cf. § 5.2.4). Pour calculer ces taux, il est nécessaire de posséder une vérité terrain sur les deux sites à traiter. Cette vérité terrain a été réalisée sur des zones limitées, à l'aide de cartes IGN et des images satellitaires elles-mêmes. Ces zones se présentent sous la forme de polygones dont les pixels intérieurs sont ensuite extraits. La sélection des zones est faite de telle sorte que chaque classe soit représentée par plusieurs polygones qui traduisent au mieux la variabilité de la classe. On a veillé de plus à ce que chaque classe comporte un grand nombre de pixels au sens statistique du terme. Ces données de vérité terrain sont ensuite partitionnées en deux pour constituer :

- un ensemble de données d'apprentissage pour le classificateur,
- un ensemble de données test, sur lequel seront comparées les architectures de fusion.

Notre première démarche a été de partitionner chaque polygone, en y tirant un pixel sur n pour l'apprentissage, les autres pixels étant affectés aux données test. La variabilité des classes était alors très bien représentée dans les données d'apprentissage. Sur une telle partition, un classificateur de type réseau de neurones donne des taux de classifications très élevés (supérieurs à 95 %), et la comparaison des architectures de fusion à partir de ces taux très proches les uns des autres devient délicate.



Figure 5.2 : image panchromatique SPOT de Fos-sur-Mer (France)



Figure 5.3 : image panchromatique SPOT de Marignane (France)

Nous avons donc préféré partitionner les données à l'échelle du polygone et non plus à l'échelle du pixel : pour chaque classe, certains polygones dans leur intégralité sont affectés aux données d'apprentissage. C'est la démarche qui est effectivement adoptée par l'utilisateur lorsqu'il doit classer un ensemble d'images. Les données d'apprentissage ne sont plus représentatives de la globalité de l'information de chaque classe, et par conséquent les taux de classification chutent sensiblement. Les écarts entre les taux associés aux différentes architectures sont amplifiés, et la comparaison des architectures prend davantage de sens.

La composition des données de vérité terrain sur les sites de Fos et Marignane est donnée par le tableau 5.1. Le nombre de pixels d'apprentissage pour chaque classe est très variable : il dépend en pratique de la capacité de l'utilisateur à sélectionner ces pixels dans les images. On veillera toutefois à limiter le nombre de pixels d'apprentissage pour conserver des temps de calcul raisonnables. Par ailleurs, nous souhaitons que les résultats de classification sur les données test soient aussi représentatifs que possible des résultats que l'on obtiendrait sur un cas opérationnel, où les images entières seraient traitées. C'est pourquoi les données test sont choisies de telle sorte que la représentativité relative des classes y soit proche de celle des scènes entières. Ainsi, sur le site de Fos, la représentativité relative de la classe eau dans les données test est très élevée (45 %) conformément à la composition de l'image.

	Fos		Marignane	
référence	apprentissage	test	apprentissage	test
eau	977	12565	441	3167
sols boisés	1560	3536	151	3173
sols non boisés	908	6508	164	2678
urbain	996	5293	549	4865

Tableau 5.1 : composition des données de vérité terrain (en nombre de pixels)

Le fait de travailler sur des zones présélectionnées et non sur les images entières présente un autre avantage. Puisque les données ont été extraites sur le critère d'appartenance à l'une des classes d'intérêt, on ne sera pas embarrassé, lors de la classification, par des données représentées dans l'image et n'appartenant pas à ces classes d'intérêt. Cette remarque fait apparaître une difficulté liée à la définition du

complément d'une classe, dont nous aurons besoin pour l'utilisation de l'architecture mixte. Le complément d'une classe d'intérêt ne peut être défini, à partir des données des autres classes d'intérêt, qu'à condition que les classes d'intérêt forment une partition de l'ensemble des classes représentées. Cette condition est vérifiée sur les échantillons de vérité terrain puisque les seules classes représentées sont les classes d'intérêt. Par contre, sur l'image entière, d'autres classes que les classes d'intérêt (eau, sols boisés, sols non boisés, et ville) peuvent être représentées, comme les routes, les ponts, etc.

5.2.4 Critères d'évaluation de la qualité des classifications

Un excellent critère d'évaluation de la qualité d'une classification est la matrice de confusion. Cette matrice permet d'estimer non seulement les taux de bonne classification, mais également les erreurs d'inclusion (ou commission) et d'exclusion (ou omission) (voir par exemple Congalton 1991). Il s'agit d'un tableau indiquant le nombre de pixels affectés à chaque classe, relativement à leur appartenance aux échantillons de vérité terrain. Nous choisissons arbitrairement de disposer en ligne la référence (polygones test) et en colonnes le résultat de classification. Un exemple de matrice de confusion est donné par le tableau 5.2.

classé / référence	eau	boisé	non boisé	urbain
eau	5313	3962	1674	1616
boisé	36	2528	313	659
non boisé	0	146	2001	4361
urbain	0	0	0	5293

Tableau 5.2 : exemple de matrice de confusion

On constate qu'il est malaisé de lire cette matrice pour effectuer une comparaison inter-classe, parce qu'il faudrait s'affranchir de la représentativité relative des classes. La démarche traditionnelle propose une normalisation en lignes, c'est-à-dire relativement aux nombres totaux de pixels de chaque catégorie de référence. La matrice ainsi obtenue (tableau 5.3) indique la probabilité qu'un pixel de référence soit classé dans telle ou telle classe. Ces valeurs sont souvent appelées "taux producteur", parce qu'elles indiquent au producteur de la classification de quelle façon chaque catégorie est classifiée.

classé / référence	eau	boisé	non boisé	urbain
eau	42%	32 %	13 %	13 %
boisé	1 %	71 %	9 %	19 %
non boisé	0 %	2 %	31 %	67 %
urbain	0 %	0 %	0 %	100 %

Tableau 5.3 : exemple de matrice de confusion normalisée en lignes

A l'inverse, une normalisation en colonnes, c'est-à-dire relativement aux nombres totaux de pixels classés dans chaque catégorie, fournit des "taux utilisateur" (cf. tableau 5.4). Ces taux donnent à l'utilisateur de la classification la probabilité qu'un pixel classé dans une catégorie y appartienne réellement.

classé / référence	eau	boisé	non boisé	urbain
eau	99 %	60 %	42 %	14 %
boisé	1 %	38 %	8 %	6 %
non boisé	0 %	2 %	50 %	37 %
urbain	0 %	0 %	0 %	44 %

Tableau 5.4 : exemple de matrice de confusion normalisée en colonnes

Les termes diagonaux sont appelés taux de bonne classification ou précision de la classe considérée. On serait tenté de conclure, au seul vu de la "matrice producteur" (tableau 5.3), que la précision de la classification est optimale pour la ville (100 %). Ce serait contraire à l'opinion de l'utilisateur, puisque la "matrice utilisateur" (tableau 5.4) donne une précision de 44 % pour la ville, ce qui signifie que seulement 44 % des pixels classés urbain sont justes.

Une analogie peut être faite entre ce formalisme et l'approche bayésienne. Notons :

H_i (i=1 à 4) les 4 hypothèses : eau, boisé, non boisé, urbain

D_i (i=1 à 4) les 4 décisions possibles du capteur

X_{ij} le terme général de la matrice de confusion non normalisée
X_{i+} les totaux marginaux en lignes
X_{+j} les totaux marginaux en colonnes
N le nombre total de pixels de la matrice

Par définition :

$$N = \sum_{i,j} X_{ij} , \qquad X_{i+} = \frac{\sum_{j} X_{ij}}{N} , \text{ et } \quad X_{+j} = \frac{\sum_{i} X_{ij}}{N}$$

La probabilité conditionnelle $p(D_j/H_i)$ caractérise la réalisation de D_j quand H_i est vérifiée. Elle correspond au taux producteur, défini par la normalisation de la matrice en lignes :

$$p(D_j \mid H_i) = \frac{X_{ij}}{X_{i+}}$$

La probabilité *a posteriori* $p(H_i/D_j)$ est la probabilité de H_i connaissant la réalisation de D_j. Elle correspond au taux utilisateur, défini par la normalisation de la matrice en colonnes :

$$p(H_i \mid D_j) = \frac{X_{ij}}{X_{+j}}$$

Les probabilités *a priori* $p(H_i)$ des différentes hypothèses sont définies à partir des fréquences relatives des classes dans les données test. Elles s'écrivent comme le rapport des totaux marginaux en lignes au nombre total de pixels :

$$p(H_i) = \frac{X_{i+}}{N}$$

On peut vérifier le théorème de Bayes qui s'écrit :

$$p(H_i \mid D_j) = \frac{p(D_j \mid H_i).p(H_i)}{\sum_{k} p(D_j \mid H_k).p(H_k)}$$

En effet :

$$\frac{p(D_j \mid H_i).p(H_i)}{\sum_k p(D_j \mid H_k).p(H_k)} = \frac{\frac{X_{ij}}{X_{i+}} \cdot \frac{X_{i+}}{N}}{\sum_k \frac{X_{kj}}{X_{k+}} \cdot \frac{X_{k+}}{N}} = \frac{X_{ij}}{X_{+j}} = p(H_i \mid D_j)$$

Nous présenterons par la suite la matrice de confusion sous la forme producteur, qui est la forme usuelle dans la littérature. Cependant le critère de "fiabilité statistique" mis en œuvre dans les architectures décentralisée et mixte utilisera la "précision utilisateur" ou probabilité *a posteriori* de bonne classification. La vérité terrain sur les données test n'étant pas utilisable par le processus de fusion, les précisions utilisateur ne seront pas calculées sur ces données. La solution retenue consiste à estimer celles-ci sur les données d'apprentissage.

Enfin, le critère synthétique retenu pour comparer différentes classifications est la précision moyenne, qui n'est liée ni à l'approche producteur, ni à l'approche utilisateur. La précision moyenne, ou taux moyen de bonne classification, est définie comme le rapport du nombre de pixels correctement classés au nombre total de pixels. C'est donc la somme des termes diagonaux de la matrice non normalisée divisée par le nombre total de pixels de cette matrice :

$$T = \frac{\sum_i X_{ii}}{N}$$

Il est important de remarquer que ce taux moyen de bonne classification ne s'affranchit pas des différences de représentativité relative des classes dans les données test. Une classe influencera donc d'autant plus ce taux que sa représentativité relative dans les données test est importante. Sur Fos par exemple, la classe eau pèsera plus de 3 fois plus que la classe sols boisés (cf. tableau 5.1). Le taux moyen de bonne classification sera donc différent de la moyenne des précisions producteur qui constituent les termes diagonaux de la matrice de confusion producteur.

Par souci d'allégement de la présentation des résultats de classification, la matrice de confusion ne sera fournie que lorsque le détail des différentes classes sera envisagé. Dans le cas contraire, on se limitera au paramètre synthétique qu'est le taux moyen de bonne classification.

5.2.5 Prétraitements des images

5.2.5.1 Filtrage du speckle

Les images ERS sont affectées du bruit multiplicatif de speckle caractéristique des images radar. Il en découle que deux pixels voisins représentant un terrain homogène donné posséderont des radiométries différentes. On comprend donc que ce bruit altère fortement la qualité d'une classification au niveau pixel. Nous allons mettre en évidence sur un exemple l'apport du filtrage du speckle. Différents filtres de speckle ont été testés : le filtre sigma (Lee 1985), le filtre de Lee (Lee 1985), et le filtre de Lopes *et al.* (1990). Ce dernier a été finalement retenu notamment parce qu'il ne lisse pas les contours, même pour des objets de petite taille.

Afin de ne pas faire intervenir d'autres données que le radar, ce qui pourrait fausser la comparaison, nous illustrons l'apport du filtrage sur la classification monocapteur de l'image ERS. Pour le classificateur réseau de neurones appliqué au site de Marignane, le taux moyen de bonne classification est de 44 % pour l'image ERS non filtrée, et de 60 % pour l'image filtrée. Le filtrage apporte donc sur cet exemple un gain significatif de 16 % sur le taux moyen de bonne classification. L'autre site et l'autre classificateur donnent des conclusions comparables.

5.2.5.2 Texture

Rappelons que l'objectif de cette étude n'est pas de comparer les paramètres de texture, mais d'utiliser celle-ci pour lever une ambiguïté sur des classes. Le tableau 5.5 représente la matrice de confusion de la classification sur l'image panchromatique seule, pour le site de Fos et le classificateur réseau de neurones. On constate deux principales confusions portant respectivement sur les classes eau et boisé, et les classes non boisé et urbain.

Le premier paramètre de texture utilisé est la variance locale de l'image, calculée sur une fenêtre glissante dont la taille a été fixée de manière empirique à 9x9 pixels. La variance locale de l'image est un paramètre très usité pour refléter la richesse structurale de celle-ci. Le résultat de classification est donné par le tableau 5.6. Cette texture apporte un gain de 30 % sur le taux moyen de classification, et permet de résoudre correctement les deux principales confusions sur les classes, mentionnées ci-dessus. La classe "non boisé" est maintenant bien rendue. On peut cependant noter que ce gain global ne

correspond pas à un gain pour chaque classe, comme le montre la diminution du taux moyen de classification sur la classe "boisé".

classé / référence	eau	boisé	non boisé	urbain
eau	18 %	46 %	0 %	36 %
boisé	2 %	92 %	0 %	6 %
non boisé	0 %	5 %	0 %	95 %
urbain	0 %	1 %	0 %	99 %

Taux moyen de bonne classification : 39 %

Tableau 5.5 : matrice de confusion de la classification sur l'image P seule

classé / référence	eau	boisé	non boisé	urbain
eau	50 %	27 %	10 %	13 %
boisé	3 %	84 %	0 %	12 %
non boisé	0 %	6 %	72 %	22 %
urbain	0 %	0 %	0 %	100 %

Taux moyen de bonne classification : 69 %

Tableau 5.6 : matrice de confusion de la classification utilisant l'image P et la texture "variance locale"

Il est étonnant qu'un paramètre aussi simple que la variance locale améliore autant la classification. Il est en effet bien connu (voir par exemple Woodcock *et al.* 1988) que chaque classe est caractérisée à des échelles qui lui sont spécifiques, alors que la variance locale représente les mêmes échelles pour toutes les classes. Nous nous sommes donc demandés si les petites structures de l'image panchromatique ne suffisaient pas à lever les ambiguïtés sur les quatre classes adoptées. Un outil adapté à l'extraction de telles structures est la transformée en ondelettes, à l'aide de laquelle nous sélectionnons les structures d'échelles 10-20 m. Les coefficients d'ondelettes expriment l'amplitude (positive ou négative) des transitions sur les structures de l'image, au niveau de chaque pixel. Or les classes retenues sont étendues, relativement à la résolution de l'image

panchromatique : l'urbain par exemple n'est pas caractérisé à l'échelle du pixel, qui peut représenter un bâtiment ou une rue, mais par l'organisation locale des structures. C'est pourquoi nous choisissons comme paramètre de texture la variance locale des coefficients d'ondelettes 10-20 m, qui reflète la densité des structures d'échelles 10-20 m. De même que pour la variance locale de l'image (première texture utilisée), la taille de la fenêtre de calcul retenue pour la variance locale des coefficients d'ondelettes est de 9x9 pixels. L'algorithme de transformation en ondelettes utilisé est l'algorithme à trous de Dutilleux (1987) qui présente l'avantage, par rapport à l'algorithme de Mallat (1989) employé pour la méthode ARSIS (cf. chap. 3), de fournir une unique image de coefficients à une résolution donnée, superposable à l'image panchromatique du fait que cet algorithme n'effectue pas de sous-échantillonnage.

Ce second paramètre de texture donne un taux moyen de classification meilleur que celui fourni par la variance locale (cf.tableau 5.7). Les structures d'échelles 10-20 m permettent donc d'augmenter la discrimination entre l'eau et les sols boisés, et entre les sols non boisés et la ville. Cette texture a été également testée sur le site de Marignane, où son apport est comparable à celui mesuré sur Fos.

classé / référence	eau	boisé	non boisé	urbain
eau	55 %	28 %	15 %	2 %
boisé	7 %	79 %	1 %	13 %
non boisé	0 %	8 %	84 %	8 %
urbain	0 %	0 %	1 %	99 %

Taux moyen de bonne classification : 73 %

Tableau 5.7 : matrice de confusion de la classification utilisant l'image P et la texture basée sur les coefficients d'ondelettes

Ces résultats illustrent le fait que la transformée en ondelettes est un outil adapté à la détermination de texture pour la classification, comme mentionné déjà par Ranchin, Wald (1993). Nous utiliserons cependant, dans la suite de cette étude, la texture définie par la variance locale de l'image afin de respecter notre souhait, déjà exprimé, d'utiliser autant que faire se peut des approches "classiques".

5.3 DIFFERENTES ARCHITECTURES DE FUSION POUR LA CLASSIFICATION

5.3.1 Généralités

L'architecture de fusion qui utilise l'ensemble des images simultanément pour fournir un résultat en une seule étape est désignée sous le nom d'architecture centralisée. Dans le cas de la classification de pixels, cela revient à prendre une décision pour chaque pixel, à partir du vecteur composé de ses radiométries dans les différentes images et éventuellement de la texture. On peut procéder différemment, en effectuant une classification sur chaque image, puis en fusionnant ces classifications : il s'agit là d'une architecture décentralisée. Comme il a été exposé précédemment, l'architecture décentralisée que nous avons choisie n'utilise pas chaque image séparément, mais chaque capteur (un capteur pouvant fournir plusieurs images). Enfin, nous proposerons une architecture mixte, qui exploite la spécificité des classes pour sélectionner les capteurs utiles à la détermination de chaque classe.

5.3.2 Architecture centralisée

Désignons par S_{SPOT}, S_{TM}, S_{ERS} les ensembles d'images sources fournies par chaque capteur, et par F_1 l'algorithme de classification, par exemple un réseau de neurones. S_{SPOT} comporte l'image panchromatique P, S_{TM} les six canaux TM, et S_{ERS} l'image SAR filtrée. Le schéma de l'architecture centralisée, utilisant le formalisme de fusion décrit au chapitre 2, est illustré sur la figure 5.4. L'image "texture P" est le résultat d'un traitement appliqué à l'image P ; elle intervient donc comme résultat annexe, en entrée de la cellule de fusion, et non comme image source. Lorsqu'il s'agira de calculer des taux de classification pour comparer les performances des diverses architectures, les entités manipulées ne seront plus les images entières, mais les pixels appartenant aux données test (cf. § 5.2.3).

Parallèlement à la classification elle-même, le classificateur F_1 réalise une image de "confiance" qui donne, pour chaque pixel, une information sur la fiabilité de sa classification. Ce paramètre de confiance peut être défini de différentes façons dépendant du classificateur, mais il est fourni de manière autonome par l'algorithme de classification sans apport de connaissance extérieure. Dans le cas de l'architecture centralisée, l'image de confiance n'est qu'une aide à l'interprétation de la classification. Elle joue par contre un rôle essentiel dans le fonctionnement des architectures décentralisée et mixte. Pour le

classificateur réseau de neurones, les deux paramètres de confiance définis à partir de l'activation des neurones de sortie (cf. annexe C) donnent des résultats qui aboutissent aux mêmes conclusions quant à la comparaison des architectures. Nous ne présenterons donc que les résultats obtenus avec l'un des deux paramètres de confiance.

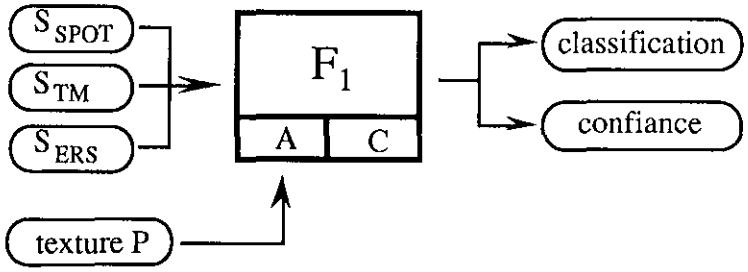


Figure 5.4 : architecture centralisée

L'avantage de l'architecture centralisée est de fournir théoriquement un résultat optimal, puisque la décision est prise à partir de l'ensemble de l'information contenue dans les images sources (Hall 1992). Par contre, cette architecture impose une charge de calcul élevée, renouvelée à chaque modification d'une donnée source. C'est pourquoi une architecture décentralisée, qui réduit la charge de calcul et ajoute les avantages de la modularité, est souvent adoptée.

5.3.3 Architecture décentralisée

L'architecture décentralisée réalise des classifications monocapteurs en parallèle, puis fusionne celles-ci en une seconde étape. Les classifications monocapteurs visent à extraire de chaque capteur l'information recherchée, pour réduire le volume de données à fusionner (Thomopoulos *et al.* 1987). On peut donc considérer la seconde étape de cette architecture comme une fusion d'attributs au niveau pixel, les attributs étant des étiquettes de classification. Différentes règles de fusion peuvent être adoptées pour réaliser la classification finale à partir des classifications monocapteurs. Les deux règles suivantes ont été envisagées :

- la première utilise les valeurs de confiance, fournies pour chaque pixel par les classifications monocapteurs. Sa mise en œuvre dans l'architecture décentralisée est illustrée par la figure 5.5. Les classifications monocapteurs sont effectuées par l'algorithme F_1, puis la cellule F_2 réalise la fusion des

classifications. Pour chaque pixel, F_2 dispose des décisions des différents capteurs assorties de leurs valeurs de confiance. La règle de fusion consiste à sélectionner la décision monocapteur dont la confiance associée est maximale. Dans les termes du formalisme adopté, les classifications monocapteurs sont les sources de données de la cellule de fusion F_2, et les images de confiance constituent des résultats annexes. Ces valeurs de confiance sont appelées "confiances pixels".

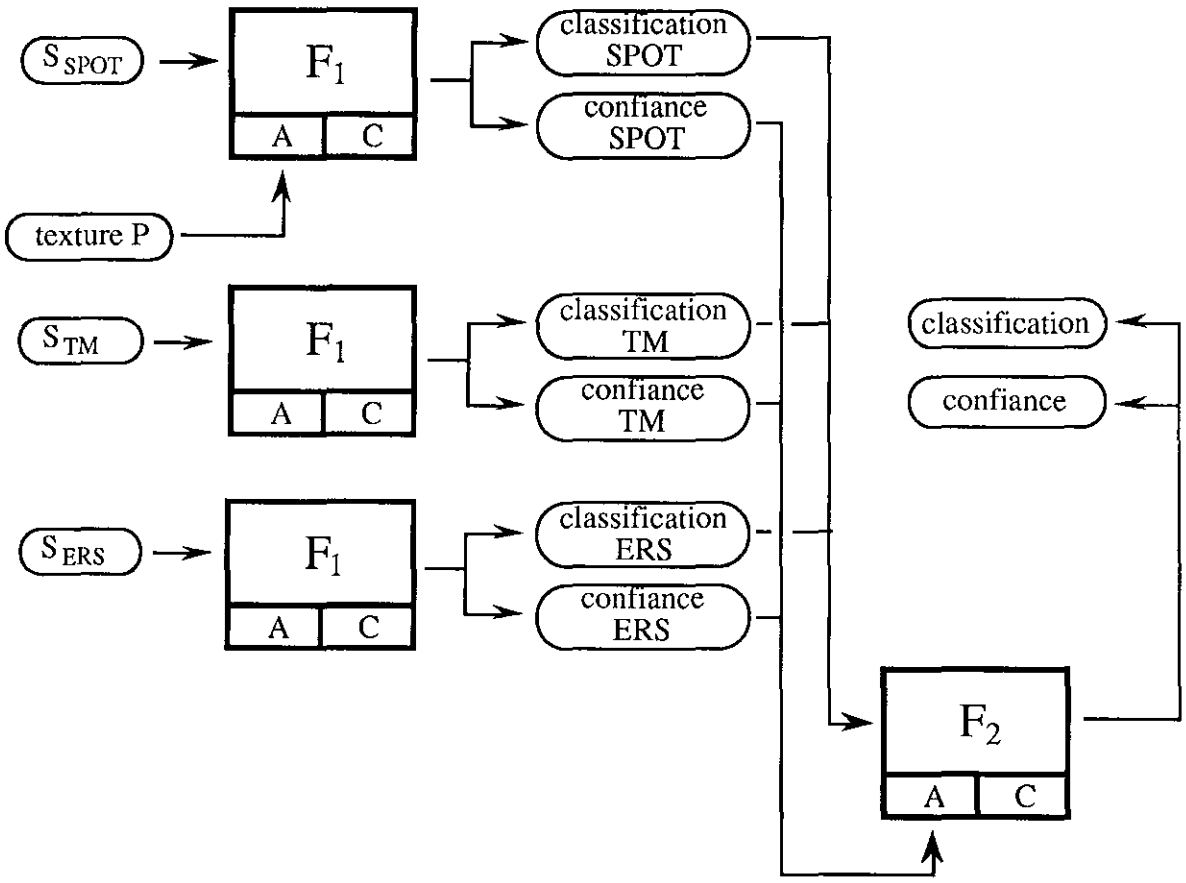


Figure 5.5 : architecture décentralisée utilisant les confiances pixel

- la seconde règle utilise non plus un critère de confiance au niveau pixel fourni par le classificateur, mais un critère statistique global, basé sur la probabilité que chaque capteur fournisse une décision juste, compte tenu de la classe qu'il sélectionne. Ce critère, désigné par la suite sous le nom de "fiabilité statistique", compare les probabilités *a posteriori* de bonne classification des différents capteurs (cf. § 5.2.4). La vérité terrain sur les données test n'étant pas utilisable dans le processus de fusion, puisqu'inconnue dans le cas

opérationnel, on estime ces probabilités *a posteriori* de bonne classification sur les données d'apprentissage. La règle de fusion, utilisée par la cellule F_3 (cf. fig. 5.6), consiste à sélectionner la décision monocapteur correspondant à la plus forte probabilité de bonne classification.

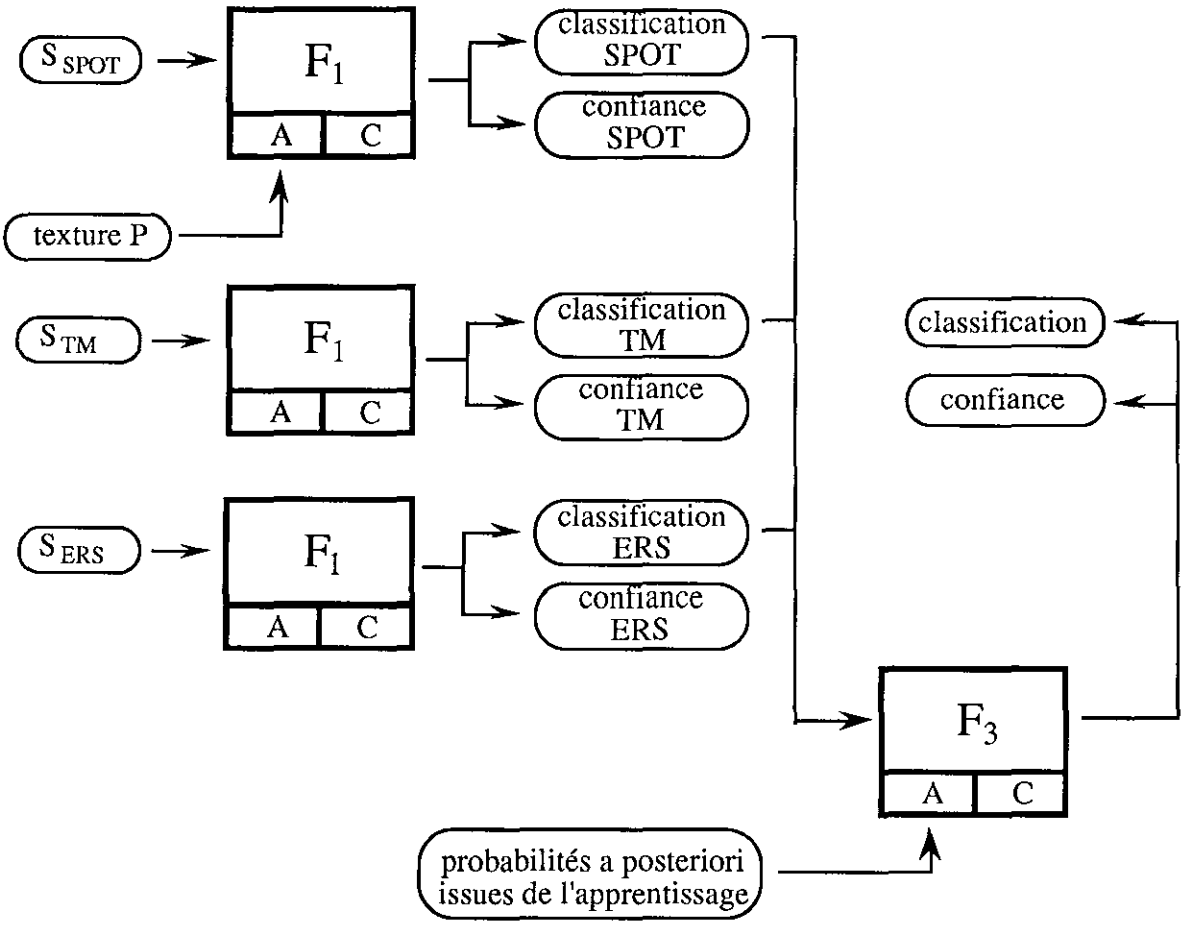


Figure 5.6 : architecture décentralisée utilisant les fiabilités statistiques

Cette architecture n'est théoriquement pas optimale, car elle n'utilise, au niveau de la cellule de fusion (F_2 ou F_3), qu'une partie de l'information totale contenue dans les images sources (Hall 1992). Chaque classification monocapteur est obtenue en optimisant un critère spécifique à ce capteur, alors que la classification centralisée optimise un critère global (Tsitsiklis, Athans 1985). Ainsi il arrive que chaque capteur exploité séparément ne puisse pas distinguer deux classes, mais que l'utilisation conjointe des capteurs lève cette ambiguïté. C'est pourquoi l'architecture centralisée donne théoriquement de meilleurs résultats que l'architecture décentralisée (Kam *et al.* 1991).

Le premier avantage de l'architecture décentralisée est de réduire la charge de calcul. Chaque classification monocapteur est en effet nettement moins complexe que la classification centralisée, puisqu'elle consiste à chercher des surfaces de décision dans un espace de dimension inférieure. De plus la fusion des classifications, à l'aide de l'une des deux règles décrites précédemment, est peu gourmande en calcul. L'autre avantage de cette architecture est sa modularité, qui permet par exemple d'ajouter un capteur tout en conservant les classifications monocapteurs existantes, alors que le modèle centralisé impose de tout recalculer. Sur le plan matériel, cette modularité permet une implantation sur des architectures parallèles, qui accélérera l'obtention du résultat final. L'implantation de chaque classification monocapteur peut aussi être délocalisée, d'où une grande souplesse d'utilisation.

5.3.4 Architecture mixte

Une architecture mixte a également été développée, dans le but de bénéficier en partie de la modularité de l'architecture décentralisée, tout en conservant des taux de bonne classification comparables à ceux de l'architecture centralisée. Le principe est de sélectionner pour chaque classe les sources riches en information, et d'éliminer les sources inutiles, qui augmentent la charge de calcul et dégradent éventuellement les performances. On réalise sur ces sources une classification binaire par classe, décidant de l'appartenance ou non de chaque pixel à la classe considérée. Ces classifications binaires sont ensuite fusionnées à l'aide de l'une des deux règles de fusion définies au § 5.3.3.

L'architecture mixte utilisant les confiances pixel est illustrée par la figure 5.7, et celle basée sur les fiabilités statistiques par la figure 5.8. On note S_1, S_2, S_3, et S_4 les ensembles constitués des capteurs sélectionnés respectivement pour les 4 classes. Chaque ensemble S_i (i=1 à 4) est donc une combinaison particulière des données capteurs S_{SPOT}, S_{TM}, S_{ERS} précédemment définies. La classification d'un pixel effectuée sur S_i fournit une information binaire d'appartenance ou de non-appartenance à la classe i, assortie d'une valeur de confiance.

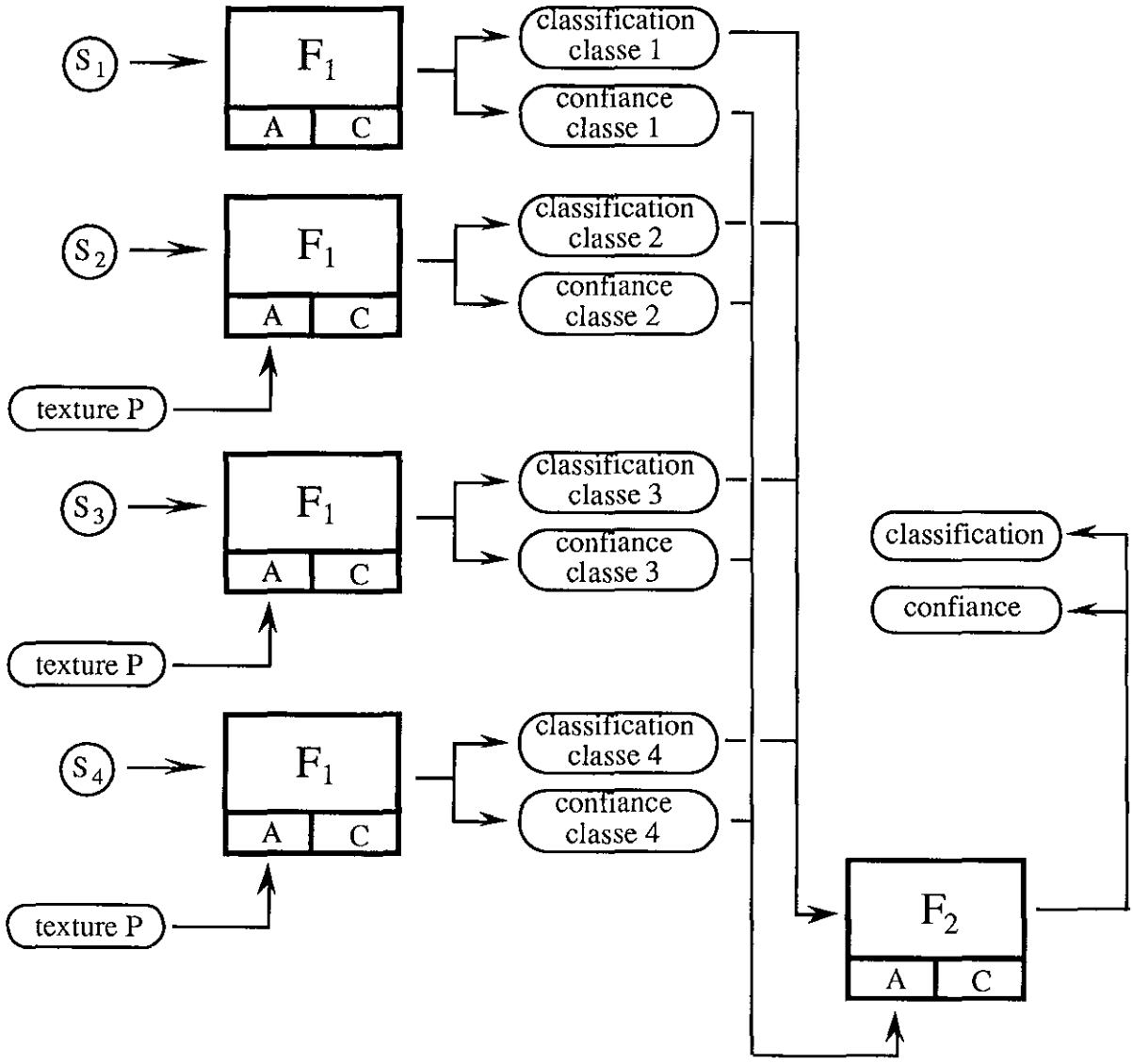


Figure 5.7 : architecture mixte utilisant les confiances pixel

La sélection des capteurs qui détermine les ensembles S_i repose sur la connaissance thématique des classes. Ainsi, il est communément montré que le proche infrarouge permet une excellente discrimination de l'eau ; nous utiliserons donc uniquement les images S_{TM} pour classifier l'eau (ensemble S_1). Les sols boisés et non boisés sont caractérisés tant par leur réponse spectrale que par leur texture. Nous utiliserons donc pour les classifier l'ensemble $S_{TM} + S_{SPOT}$ (ensembles S_2 et S_3). La classe urbaine gagne à exploiter les informations spectrale, texturale, et radar, donc l'ensemble des capteurs sera utilisé (ensemble S_4). La texture étant calculée sur l'image SPOT, elle ne sera utilisée que conjointement à S_{SPOT}, c'est-à-dire avec les ensembles S_2, S_3, et S_4 (cf. figs. 5.7 et 5.8). Ces considérations thématiques peuvent être appuyées par

les résultats des classifications monocapteurs : on peut estimer, à l'aide des taux de classifications monocapteurs, si un capteur fournit une information fiable pour une classe donnée (cf. § 5.4.2.1).

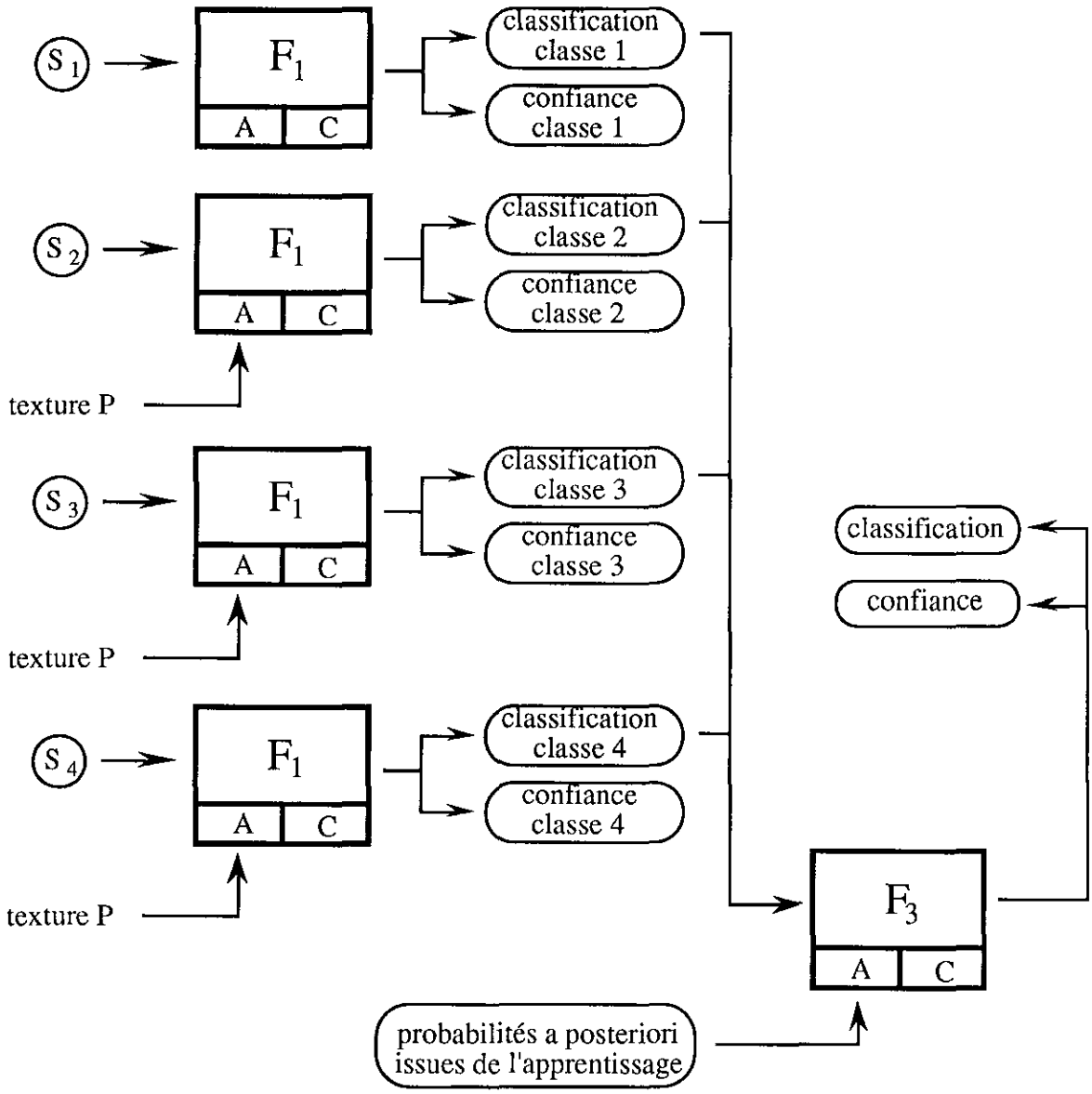


Figure 5.8 : architecture mixte utilisant les fiabilités statistiques

Le cas où tous les capteurs sont utilisés par chaque classification binaire est également envisagé. L'architecture mixte utilise alors la même information que le modèle centralisé, mais les conflits entre les classes sont traités en une seconde étape (F_2 ou F_3) et non plus par l'algorithme F_1. Les résultats de ce traitement serviront à évaluer l'influence de la sélection de capteurs sur la qualité de la classification finale.

Cette architecture peut être qualifiée de "mixte" : elle s'apparente au modèle décentralisé parce qu'elle réalise la classification en deux étapes, mais les classifications binaires multicapteurs F_1 s'apparentent à l'architecture centralisée.

5.4 ANALYSE DES RESULTATS

5.4.1 Introduction

Les résultats comparatifs des architectures centralisée, décentralisée, et mixte sont maintenant présentés et analysés. Afin de ne privilégier ni le site, ni le classificateur utilisé, et de tirer des conclusions aussi générales que possible, les architectures ont été mises en œuvre sur deux sites avec deux classificateurs différents. Pour alléger la présentation, les résultats complets ne sont présentés que sur un site (Fos) avec un unique classificateur (réseau de neurones), ces choix étant arbitraires. Dans les autres cas, seuls les taux moyens de bonne classification utiles à la comparaison des architectures sont analysés. Des résultats plus détaillés sont rassemblés en annexe D.

Deux règles de fusion ont été définies pour les architectures décentralisée et mixte (cf. § 5.3.3). On constate, avec l'architecture mixte, que les taux moyens de bonne classification sont quasiment toujours égaux pour les deux règles, et que les matrices de confusion sont très proches. Ceci s'explique par la très faible proportion de pixels conflictuels entre les différentes classifications binaires, qui réduit les écarts de performances des deux règles de fusion. Par conséquent, seuls les résultats fournis par la règle de confiances pixel seront donnés pour l'architecture mixte. Appliquées à l'architecture décentralisée, les deux règles de fusion donnent des résultats sensiblement différents. Lorsqu'un des capteurs est très peu performant sur au moins une classe, l'avantage du critère global de fiabilités statistiques est d'inhiber la décision de ce capteur pour cette classe (ou ces classes). Malheureusement ce critère statistique utilise en pratique les probabilités *a posteriori* estimées sur les données d'apprentissage, qui sont différentes des probabilités optimales correspondant aux données test. Ceci est une illustration de la difficulté générale, en fusion multicapteurs, à définir des valeurs *a priori* de fiabilité des capteurs.

Enfin, la comparaison des architectures de fusion sera effectuée sur différentes combinaisons de capteurs, pour évaluer si la complémentarité des capteurs exploités,

relativement aux classes souhaitées, privilégie une architecture particulière. On définit ici la complémentarité de deux capteurs comme étant la capacité de l'un à discriminer les classes non séparées par l'autre. Le premier cas traité, exploitant les trois capteurs SPOT, ERS-1, et Landsat, révèle que Landsat joue un rôle primordial sur les classes eau, sols boisés, et sols non boisés. On examine donc, dans un second temps, ce qu'il advient lorsque seuls SPOT et ERS-1 sont disponibles. Cette combinaison montre, conformément à ce que l'examen visuel des images pouvait suggérer, que la décision est prise sur la plupart des pixels par le capteur SPOT. Afin d'accroître artificiellement la complémentarité des images SPOT et ERS-1, on réalisera une troisième configuration dans laquelle l'image SPOT sera artificiellement bruitée.

5.4.2 Classification sur SPOT, ERS-1, et Landsat

5.4.2.1 Classifications monocapteurs

Les classifications monocapteurs sur SPOT, ERS-1, et Landsat-TM utilisant un réseau de neurones donnent, sur Fos, les résultats de classification rassemblés dans les tableaux 5.8 à 5.10.

On constate que Landsat fournit un taux moyen de bonne classification nettement supérieur à celui des autres capteurs. ERS donne des résultats très médiocres, comme on pouvait s'y attendre en examinant visuellement l'image : les taux de classification sur l'eau et l'urbain sont d'environ 50%, et les classes boisé et non boisé ne sont pas distinguées. On attend malgré tout de ce capteur qu'il apporte une information complémentaire de celle de SPOT et Landsat, c'est-à-dire qu'il discrimine les pixels non déterminés par ces derniers. Les taux de bonne classification de "type producteur" présentés ici révèlent que Landsat est le plus performant sur les classes eau, boisé et non boisé, et que SPOT est meilleur sur la ville. Il ne faut cependant pas occulter le fait que les résultats médiocres de SPOT sur les autres classes vont dégrader sa précision utilisateur sur la ville, réduisant le taux utilisateur à 60 % sur cette classe (cf. § 5.2.4).

classé / référence	eau	boisé	non boisé	urbain
eau	44 %	18 %	37 %	0 %
boisé	0 %	89 %	7 %	4 %
non boisé	5 %	83 %	9 %	3 %
urbain	2 %	41 %	3 %	54 %

Taux moyen de bonne classification : 43 %

Tableau 5.8 : matrice de confusion sur les données ERS

classé / référence	eau	boisé	non boisé	urbain
eau	50 %	27 %	10 %	13 %
boisé	3 %	84 %	0 %	12 %
non boisé	0 %	6 %	72 %	22 %
urbain	0 %	0 %	0 %	100 %

Taux moyen de bonne classification : 69 %

Tableau 5.9 : matrice de confusion sur les données SPOT (incluant la texture)

classé / référence	eau	boisé	non boisé	urbain
eau	100 %	0 %	0 %	0 %
boisé	0 %	94 %	6 %	0 %
non boisé	0 %	11 %	83 %	6 %
urbain	0 %	1 %	46 %	53 %

Taux moyen de bonne classification : 86 %

Tableau 5.10 : matrice de confusion sur les données Landsat

5.4.2.2 Architecture centralisée

L'architecture centralisée utilisant un réseau de neurones donne, sur le site de Fos, les résultats de classification présentés dans le tableau 5.11. On constate que, pour trois classes sur quatre (eau, boisé, non boisé), les taux de classification dépassent ou égalent le meilleur taux monocapteur. De plus le taux moyen dépasse tous les taux moyens monocapteurs. Ceci montre que l'architecture centralisée exploite bien la complémentarité des capteurs, pour surclasser les traitements monocapteurs. Par ailleurs, la précision de 100 % obtenue sur l'eau montre que le capteur ERS, bien qu'inadapté à la détermination de cette classe, ne pollue pas la décision du capteur le plus performant sur l'eau (Landsat). Ces conclusions sont généralisables à l'ensemble des cas traités.

classé / référence	eau	boisé	non boisé	urbain
eau	100 %	0 %	0 %	0 %
boisé	0 %	95 %	5 %	0 %
non boisé	0 %	11 %	87 %	2 %
urbain	0 %	1 %	19 %	80 %

Taux moyen de bonne classification : 93 %

Tableau 5.11 : matrice de confusion de l'architecture centralisée

5.4.2.3 Architecture décentralisée

L'architecture décentralisée effectue la fusion des classifications monocapteurs, à l'aide d'une des règles de fusion définies au §.5.3.3. Le tableau 5.12 donne les résultats obtenus par la règle basée sur les confiances au niveau pixel.

En comparant ce tableau aux tableaux 5.8 à 5.10, on constate que, pour chaque classe, il existe un capteur fournissant une précision au moins égale à celle de l'architecture décentralisée. Globalement, le taux moyen obtenu est identique à celui du capteur Landsat seul. La bonne compréhension de ces résultats impose de connaître, en complément de la matrice de confusion, la contribution de chaque capteur à la prise de décision finale. On met alors en évidence que Landsat est majoritaire dans la prise de

décision sur les quatre classes, où il intervient pour 64 % des pixels eau, 92 % du boisé, 98 % du non boisé, et 78 % de l'urbain. Or les classifications monocapteurs montrent qu'il aurait été optimal que Landsat prenne la décision sur tous les pixels eau, et SPOT sur les pixels ville. On voit donc que le paramètre de confiance pixel, évalué pour chaque capteur séparément, reflète mal quel est le meilleur capteur pour un pixel donné. C'est un bon critère intra-classification, c'est-à-dire qu'il est utile pour comparer les pixels d'une même classification, mais il est mal adapté à la comparaison inter-classifications effectuée lors de la fusion.

classé / référence	eau	boisé	non boisé	urbain
eau	94 %	6 %	0 %	0 %
boisé	0 %	94 %	5 %	1 %
non boisé	0 %	10 %	83 %	7 %
urbain	0 %	1 %	33 %	66 %

Taux moyen de bonne classification : 86 %

Tableau 5.12 : matrice de confusion de l'architecture décentralisée utilisant les confiances pixel

Les résultats obtenus par la règle utilisant les fiabilités statistiques sont présentés par le tableau 5.13.

classé / référence	eau	boisé	non boisé	urbain
eau	100 %	0 %	0 %	0 %
boisé	0 %	94 %	7 %	0 %
non boisé	0 %	11 %	83 %	6 %
urbain	0 %	1 %	46 %	53 %

Taux moyen de bonne classification : 86 %

Tableau 5.13 : matrice de confusion de l'architecture décentralisée utilisant les fiabilités statistiques

Ces résultats sont rigoureusement identiques à ceux fournis par le capteur Landsat seul, bien que ce dernier ne prenne pas toujours la décision. Ceci montre que l'architecture décentralisée n'utilise pas correctement la complémentarité des capteurs. On remarquera que le critère de fiabilités statistiques aboutit à une prise de décision exclusive de Landsat sur l'eau, qui permet de maintenir un taux de 100 % sur cette classe. De même, les autres cas révèlent que dans l'architecture décentralisée, Landsat joue un rôle dominant dans la prise de décision sur la plupart des classes, et ce pour les deux règles de fusion.

5.4.2.4 Architecture mixte

L'architecture mixte consiste à réaliser autant de classification binaires qu'il y a de classes, puis à fusionner celles-ci à l'aide d'une règle de fusion. Chaque classification binaire utilise une combinaison de capteurs spécifique, dont la sélection repose sur la connaissance thématique des classes, comme expliqué au § 5.3.4. Ces considérations thématiques peuvent être appuyées par les résultats des classifications monocapteurs. Les taux utiles à cette sélection sont les "taux producteur", utilisés dans les matrices de confusion, puisqu'on s'intéresse à la faculté d'un capteur à classifier un type de terrain donné. Les classifications binaires, réalisées à l'aide de la sélection de capteurs précédente, donnent les résultats présentés dans les tableaux 5.14 à 5.17. Les taux moyens de bonne classification ne sont pas définis pour les classifications binaires, puisque leur matrice de confusion n'est pas carrée.

classé / référence	≠ eau	eau
eau	1 %	99 %
boisé	100 %	0 %
non boisé	100 %	0 %
urbain	100 %	0 %

Tableau 5.14 : matrice de confusion pour la classification binaire "eau"

classé / référence	≠ boisé	boisé
eau	100 %	0 %
boisé	8 %	92 %
non boisé	80 %	20 %
urbain	98 %	2 %

Tableau 5.15 : matrice de confusion pour la classification binaire "boisé"

classé / référence	≠ non boisé	non boisé
eau	100 %	0 %
boisé	98 %	2 %
non boisé	21 %	79 %
urbain	97 %	3 %

Tableau 5.16 : matrice de confusion pour la classification binaire "non boisé"

classé / référence	≠ urbain	urbain
eau	97 %	3 %
boisé	100 %	0 %
non boisé	89 %	11 %
urbain	3 %	97 %

Tableau 5.17 : matrice de confusion pour la classification binaire "urbain"

Les taux de bonne classification des classifications binaires sont assez élevés, mais restent toutefois inférieurs à ceux de l'architecture centralisée, excepté pour la classe urbaine. On peut attribuer cet écart à la sélection de capteurs qui, si elle allège les calculs, prive malgré tout la classification d'une partie de l'information. Pour comparaison, les classifications binaires ont également été faites sur l'ensemble des capteurs, et leurs taux sont alors meilleurs que ceux de l'architecture centralisée. Cependant, les conflits entre ces classifications binaires restent à résoudre lors de l'étape de fusion, alors qu'ils l'ont

déjà été par l'architecture centralisée. Or, la précision de chaque classe ne peut que se dégrader avec la fusion, parce qu'aucune classification binaire ne peut corriger l'erreur d'une autre.

Comme il a été précisé au § 5.4.1, les résultats de classification de l'architecture mixte sont quasiment identiques avec les deux règles de fusion envisagées. En conséquence, seuls les résultats obtenus avec la règle de confiances pixel sont donnés ici (cf. tableau 5.18). On remarquera qu'il est nécessaire d'ajouter à la matrice de confusion de l'architecture mixte une colonne supplémentaire intitulée "indécision". Cette classe est utilisée lorsqu'aucune classification binaire ne sélectionne l'une des quatre classes de référence. Ce cas ne se produit pas sur le modèle décentralisé, parce que les classifications monocapteurs sélectionnent toujours une des quatre classes.

La précision moyenne de l'architecture mixte est meilleure que celle des classifications monocapteurs et que celle de l'architecture décentralisée, et elle est très proche de la précision de l'architecture centralisée. Lorsqu'elle utilise l'ensemble des capteurs pour effectuer les classifications binaires, l'architecture mixte permet même d'égaler l'architecture centralisée. L'architecture mixte exploite donc correctement la complémentarité des capteurs. Une analyse plus fine montre que les taux de classification de chaque classe diffèrent notablement de ceux de l'architecture centralisée. La règle de fusion des classifications binaires, qui gère les conflits entre les classes, fonctionne en effet différemment de l'unique cellule de classification du modèle centralisé.

classé / référence	eau	boisé	non boisé	urbain	indécision
eau	99 %	0 %	0 %	0 %	1 %
boisé	0 %	91 %	2 %	0 %	7 %
non boisé	0 %	15 %	73 %	8 %	4 %
urbain	0 %	1 %	0 %	97 %	2 %

Taux moyen de bonne classification : 92 %

Tableau 5.18 : matrice de confusion de l'architecture mixte utilisant les confiances pixel

5.4.3 Classification sur SPOT et ERS-1

Examinons désormais le cas où seules les images provenant des capteurs SPOT et ERS-1 sont disponibles. Les taux moyens de bonne classification obtenus sur Fos avec les différentes architectures, à l'aide d'un réseau de neurones, sont rassemblés dans le tableau 5.19.

architecture	centralisée	décentralisée confiances pixel	décentralisée fiabilités stat.	mixte
taux moyen	72 %	70 %	69 %	70%

Tableau 5.19 : taux moyens de bonne classification sur SPOT et ERS

Les taux de bonne classification chutent en comparaison de la classification sur les trois capteurs ; Landsat contribuait donc grandement à la qualité de cette classification. L'architecture décentralisée adopte tantôt la décision de SPOT, tantôt celle de ERS sur la classe urbaine, mais impose presque exclusivement celle de SPOT sur les trois autres classes. Cette constatation confirme la plus grande richesse informationnelle de SPOT, relativement aux classes sélectionnées. Dans ce cas où l'un des capteurs surclasse l'autre, il s'avère que l'architecture décentralisée a des performances plus proche de l'architecture centralisée que dans le cas de l'utilisation des trois capteurs. L'architecture mixte utilise ici les deux capteurs pour réaliser toutes les classifications binaires. Chaque classification binaire représente alors un modèle centralisé à 2 capteurs, dont on constate qu'il est plus performant (taux moyen centralisé de 72 %) que les classifications monocapteurs (taux moyens respectifs de 43 % et 69 %). L'architecture mixte possède ici encore des performances inférieures mais très proches de celles de l'architecture centralisée.

L'utilisation des seuls capteurs SPOT et ERS ne change donc pas la hiérarchie des performances des différentes architectures, établie sur les trois capteurs. Cependant, dans ces deux cas, l'un des capteurs (Landsat, puis SPOT) possédait un rôle prédominant sur la décision prise par la cellule de fusion. On va donc chercher à simuler une troisième configuration, dans laquelle deux capteurs d'importance égale seront utilisés. Ces deux capteurs sont ERS d'une part, et SPOT auquel on a volontairement ajouté un fort bruit blanc gaussien.

5.4.4 Classification sur SPOT bruité et ERS-1

Le bruit ajouté à l'image SPOT panchromatique est un bruit blanc gaussien d'écart-type 5 comptes numériques. Cet écart-type est à comparer à la dynamique totale de l'image qui est de 200 comptes numériques environ (il s'agit d'ortho-images). Le taux moyen de bonne classification de SPOT passe de 69 % à 60 %, suite à l'ajout de bruit, à comparer au taux de 43 % de ERS. Le tableau 5.20 présente les résultats sur Fos des différentes architectures, obtenus avec un classificateur réseau de neurones.

architecture	centralisée	décentralisée confiances pixel	décentralisée fiabilités stat.	mixte
taux moyen	71 %	64 %	65 %	67 %

Tableau 5.20 : taux moyens de bonne classification sur SPOT bruité et ERS

On constate, par comparaison au tableau 5.19, que les architectures centralisée et mixte sont moins sensibles au bruitage des données SPOT que l'architecture décentralisée. Cette constatation est encore plus évidente pour le classificateur K-PPV. L'architecture décentralisée semble donc moins adaptée à des sources très complémentaires que les architectures centralisée et mixte. L'architecture centralisée reste la plus performante, devant les architectures mixte et décentralisée. Les trois architectures donnent encore des taux moyens de classification plus élevés que celui de la meilleure classification monocapteur (sur SPOT).

5.4.5 Synthèse des résultats

Tous les résultats précédents proviennent des classifications réalisées sur Fos avec un classificateur réseau de neurones. Afin d'évaluer les variations liées au site traité et au classificateur utilisé, ces traitements ont été effectués sur Fos et Marignane, avec un réseau de neurones ou un classificateur K-PPV. Ces deux classificateurs se sont avérés simples d'emploi, et ils ont requis des temps de calcul comparables, d'au plus quelques minutes sur les données issues des polygones de vérité terrain. Les tableaux 5.21 à 5.24 rassemblent les taux moyens de classification pour les quatre combinaisons ainsi engendrées. Les résultats de l'architecture centralisée, et des architectures décentralisée et mixte utilisant la règle de fusion basée sur les confiances pixel, sont donnés successivement pour les trois capteurs, pour SPOT et ERS, et pour SPOT bruité et ERS.

architecture / capteurs	centralisée	décentralisée	mixte
SPOT, ERS, Landsat	93 %	86 %	92 %
SPOT, ERS	72 %	70 %	70 %
SPOT bruité, ERS	71 %	64 %	66 %

Tableau 5.21 : taux moyens de bonne classification sur Fos avec le classificateur réseau de neurones

architecture / capteurs	centralisée	décentralisée	mixte
SPOT, ERS, Landsat	93 %	72 %	93 %
SPOT, ERS	72 %	67 %	72 %
SPOT bruité, ERS	70 %	58 %	70 %

Tableau 5.22 : taux moyens de bonne classification sur Fos avec le classificateur K-PPV

architecture / capteurs	centralisée	décentralisée	mixte
SPOT, ERS, Landsat	97 %	93 %	95 %
SPOT, ERS	88 %	82 %	86 %
SPOT bruité, ERS	83 %	75 %	81 %

Tableau 5.23 : taux moyens de bonne classification sur Marignane avec le classificateur réseau de neurones

architecture / capteurs	centralisée	décentralisée	mixte
SPOT, ERS, Landsat	97 %	86 %	97 %
SPOT, ERS	88 %	85 %	88 %
SPOT bruité, ERS	85 %	77 %	85 %

Tableau 5.24 : taux moyens de bonne classification sur Marignane avec le classificateur K-PPV

5.5 CONCLUSION

De l'examen des résultats précédents se dégagent un certain nombre de conclusions générales. L'architecture centralisée est, dans tous les cas traités, celle qui aboutit au taux moyen de bonne classification le plus élevé. Le principe selon lequel l'augmentation du nombre de composantes du vecteur de données lève des ambiguïtés que ne peut résoudre un modèle décentralisé, s'applique donc aux données traitées. On constate qu'avec les deux classificateurs utilisés, la qualité de la classification centralisée n'est pas dégradée par un capteur bruité ou inapte à la détermination d'une classe. L'architecture décentralisée est aussi surclassée par l'architecture mixte, dont les performances avoisinent celles du modèle centralisé. Il semble que l'écart de performances entre les architectures centralisée ou mixte, et l'architecture décentralisée s'accroisse avec la complémentarité des données sources. Autrement dit, plus l'importance relative des sources dans la classification est inégale, plus les architectures centralisée et mixte surclassent l'architecture décentralisée. On notera également que les deux règles de fusion donnent des résultats de classification comparables.

La mise en œuvre de l'architecture décentralisée est souvent délicate car elle exige du classificateur une décision, alors que les données monocapteurs sont parfois insuffisantes pour séparer deux classes. Une des conséquences pratiques, lors de l'utilisation d'un réseau de neurones, est une convergence difficile qui augmente considérablement le temps de calcul. A l'inverse, la mise en œuvre de l'architecture mixte est plus facile, puisque les classes sont traitées séparément, et que chaque classification binaire dispose de l'ensemble des capteurs utiles. L'avantage des architectures décentralisée et mixte par rapport à la centralisée tient à leur modularité qui permet, à faible coût, d'ajouter respectivement un capteur ou une classe. Cette modularité est aussi exploitable lors d'une approche hiérarchique, puisque la maîtrise de la règle de fusion

permet d'ordonner (hiérarchiser) les capteurs (architecture décentralisée), ou les classes (architecture mixte). L'architecture décentralisée a l'indéniable qualité d'une grande souplesse d'utilisation, en particulier dans les cas où les données issues des capteurs ne sont pas traitées aux mêmes endroits, et où seules les décisions sont rassemblées "physiquement" afin de procéder à la fusion.

Les conclusions de ce chapitre sont très liées à l'application envisagée (classification) ainsi qu'aux classes, sites, images, et capteurs choisis. La fluctuation des écarts de performances entre les architectures, constatée sur les deux sites traités, avec les deux classificateurs, et sur diverses combinaisons de capteurs, abonde en ce sens. On a donc volontairement limité le niveau des conclusions tirées, en se bornant aux plus générales. Celles-ci étaient pressenties dès l'origine du travail et ne constituent pas une surprise. Les résultats donnés ici constituent donc davantage une illustration des qualités et des défauts des architectures, plutôt qu'une démonstration.

Chapitre 6

CONCLUSION

Cette étude a permis de développer certains aspects particuliers de la fusion de données, en se limitant à la fusion de pixels pour des applications de télédétection et de photo-interprétation. Nous avons mis en évidence la grande variété des techniques exploitées, qui est encore plus vaste lorsqu'on élargit le champ d'application de la fusion à d'autres domaines. Pourtant, il semble qu'on ne puisse dégager une théorie de la fusion, mais que la fusion de données soit plutôt un concept faisant appel à des théories diverses comme le traitement du signal, l'automatique, la théorie de l'information, ou encore l'intelligence artificielle. L'idée d'utiliser différents capteurs simultanément pour améliorer les performances d'un système ne date pas d'aujourd'hui. L'innovation majeure apportée par le concept de fusion est de proposer un formalisme qui décrive de manière générique des problèmes communs à différents domaines, qui utilisaient jusqu'alors des formulations différentes. Nous avons montré l'apport de la fusion à plusieurs niveaux, comme nous le rappelons ci-après.

La comparaison de plusieurs architectures de fusion a confirmé que l'architecture centralisée est la plus performante en termes de taux de classification, conformément aux prévisions théoriques. Dans certains cas où les différentes architectures testées fournissent des résultats proches, le choix d'une architecture sera guidé par d'autres contraintes que celle des performances, comme par exemple la délocalisation des données ou le temps de calcul limité. Un résultat important de cette comparaison est l'illustration qu'une architecture appropriée permet de surclasser tous les traitements monocapteurs.

Dans le domaine de la photo-interprétation, la fusion permet d'élaborer des produits de présentation visuelle de l'information, qui prennent en compte les caractéristiques de la perception visuelle humaine. Différentes méthodes basées sur des critères d'analyse de données ou sur la perception des couleurs ont été testées. En soumettant les produits à de nombreux photo-interprètes, nous avons confirmé l'importance de la présentation visuelle de l'information. Les difficultés identifiées sont

d'une part le nombre élevé et croissant d'images disponibles à exploiter en un temps limité, et d'autre part l'utilisation de bandes spectrales (infrarouge ou microondes) hors du spectre visible, dont l'interprétation pose souvent problème. Au delà de la bande spectrale, les données hétérogènes optiques et radar diffèrent par le principe de fonctionnement du capteur (passif ou actif), que le photo-interprète se doit de connaître pour reconstituer une réalité terrain à partir des images. Cette constatation suggère une nouvelle voie d'investigation pour la photo-interprétation assistée par ordinateur (PIAO) : à l'utilisation actuelle de méthodes de traitement d'images et de systèmes d'information géographique (SIG), il serait bénéfique d'associer des critères reposant sur la physique des capteurs. Nous avons proposé une telle approche pour la fusion d'images optique et radar, par extraction de structures caractéristiques de l'image radar qui sont ensuite réinjectées dans l'image optique. Cette méthode, qui constitue une aide à la détection pour le photo-interprète, ouvre la voie à d'autres développements en fusion hétérogène optique-radar.

Enfin, nous avons mis au point une méthode originale de fusion d'images multispectrales multirésolutions nommée ARSIS. Cette méthode vise en priorité à préserver l'information radiométrique des images multispectrales d'origine, tout en y injectant des détails extraits d'une image de meilleure résolution spatiale. Les essais effectués sur plusieurs capteurs et différentes scènes sont probants, et la méthode doit entrer prochainement en phase de production. L'apport de la fusion se traduit ici par la création d'un nouveau produit, exploitable tel quel par les photo-interprètes, ou utilisable en entrée d'algorithmes automatiques basés sur la signature multispectrrale des objets.

On peut penser que l'apport de la fusion en télédétection peut être encore accru par une spécialisation de la tâche à accomplir, qui permettrait une adéquation plus fine des méthodes utilisées. Les voies de recherche suggérées concernent la fusion de plus haut niveau sémantique, faisant appel à l'intelligence artificielle, mais dont l'application à la télédétection est ardue. N'oublions pas que l'ambition de telles méthodes, dans l'état actuel des connaissances, n'est pas de reproduire le raisonnement du photo-interprète qui fait appel à une analyse très complexe, mais de fournir à celui-ci des outils performants d'aide à la photo-interprétation.

REFERENCES BIBLIOGRAPHIQUES

Anonyme, 1986, *Guide des utilisateurs de données SPOT*, Editeurs CNES et SPOT Image, Toulouse, France, 3 volumes.

Antonini M., M. Barlaud, P. Mathieu, I. Daubechies, 1992, Image coding using wavelet transform, *IEEE Transactions on Image Processing*, **1**, 2, 205-220.

Appriou A., 1991, Probabilités et incertitude en fusion de données multi-senseurs, *Revue Scientifique et Technique de la Défense*, 11, 27-40.

Barber D. G., M. E. Shokr, R. A. Fernandes, E. D. Soulis, D. G. Flett, E. F. LeDrew, 1993, A comparison of second-order classifiers for SAR sea ice discrimination, *Photogrammetric Engineering & Remote Sensing*, **59**, 9, 1397-1408.

Bischof H., W. Schneider, A. J. Pinz, 1992, Multispectral classification of Landsat images using neural networks, *IEEE Transactions on Geoscience and Remote Sensing*, **30**, 3, 482-490.

Buchanan M. D., 1979, Effective utilization of color in multidimensional data representation, In *Advances in Display Technology*, SPIE, **199**, 9-18.

Carper W. J., T. M. Lillesand, R.W. Kiefer, 1990, The use of Intensity-Hue-Saturation transformations for merging SPOT Panchromatic and multispectral image data, *Photogrammetric Engineering & Remote Sensing*, **56**, 4, 459-467.

Chavez P. S. Jr., S. C. Sides, J. A. Anderson, 1991, Comparison of three different methods to merge multiresolution and multispectral data : Landsat TM and SPOT Panchromatic, *Photogrammetric Engineering & Remote Sensing*, **57**, 3, 265-303.

Clevers J. G. P. W., 1988, The derivation of a simplified reflectance model for the estimation of leaf area index, *Remote Sensing of Environment*, **25**, 1, 53-70.

Congalton R. G., 1991, A review of assessing the accuracy of classifications of remotely sensed data, *Remote Sensing of Environment*, **37**, 1, 35-46.

Daubechies I., 1988, Orthonormal bases of compactly supported wavelets, *Communications on Pure and Applied Mathematics*, **41**, 909-996.

Davalo E., P. Naim, 1990, *Des réseaux de neurones*, Editions Eyrolles, Paris, 232 p.

Djamdji J. P., A. Bijaoui, R. Manière, 1993, A new method of automatic registration, based on a multiresolution decomposition of the images using the wavelet transform, is presented, *Photogrammetric Engineering & Remote Sensing*, **59**, 5, 645-653.

Du L. J., J. S. Lee, S. A. Mango, 1992, Texture segmentation of SAR images using the wavelet transform, In Proceedings of IGARSS'92, **2**, 1108-1110, IEEE catalog number 92CH3041-1.

Duda R. O., P. E. Hart, 1973, *Pattern classification and scene analysis*, Wiley-Interscience Publication, 482 p.

Dutilleux P., 1987, An implementation of the "algorithme à trous" to compute the wavelet transform, Comptes-rendus du *Congrès Ondelettes et méthodes temps-fréquence et espace des phases*, Marseille, France, 14-18 sept. 1987, Springer Verlag Editeurs, 298-304.

Elachi C., 1988, *Spaceborne radar remote sensing : applications and techniques*, IEEE Press, New York, 255 p.

Escoffier B., J. Pagès, 1988, *Analyses factorielles simples et multiples*, Dunod, Paris, ISBN 2-04-018782-0, 241 p.

Feauveau J. C., 1990, Analyse multirésolution pour les images avec un facteur de résolution $\sqrt{2}$, *Traitement du Signal*, **7**, 2, 117-128.

Foley J. D., A. Van Dam, 1982, *Fundamentals of interactive computer graphics*, Addison-Wesley, Reading, MA, 664 p.

Fung T., K. C. Chan, 1994, Spatial composition of spectral classes : a structural approach for image analysis of heterogeneous land-use and land-cover types, *Photogrammetric Engineering & Remote Sensing*, **60**, 2, 173-180.

Gibson G. J., C. F. N. Cowan, 1990, On the decision regions of multilayer perceptrons, In Proceedings of the IEEE, **78**, 10, 1590-1594.

Gillespie A. R., 1980, *Digital techniques of image enhancement,* in *Remote Sensing in Geology*, (B. S. Siegal and A. R. Gillespie Eds.), Wiley, New York, 702 p.

Gillespie A. R., A. B. Kahle, R. E. Walker, 1986, Color enhancement of highly correlated images, I. decorrelation and HSI contrast stretches, *Remote Sensing of Environment*, **20**, 209-235.

Hall D. L., 1992, *Mathematical techniques in multisensor data fusion*, Editeur Artech House, Boston-London, 301 p.

Haralick R. M., 1979, Statistical and structural approaches to texture, In Proceedings of the IEEE, **67**, 5, 786-803.

Harris J. R., R. Murray, 1990, IHS transform for the integration of radar imagery with other remotely sensed data, *Photogrammetric Engineering & Remote Sensing*, **56**, 12, 1631-1641.

Hayes-Roth B., 1983, The blackboard architecture: a general framework for problem solving ?, Technical Report HPP-83-30, Stanford University, USA.

He D. C., L. Wang, 1990, Texture unit, texture spectrum, and texture analysis, *IEEE Transactions on Geoscience and Remote Sensing*, **28**, 4, 509-512.

Houzelle S., G. Giraudon, 1994, Contribution to multisensor fusion formalization, *Robotics and Autononous Systems*, **13**, 69-85.

Iverson A. E., J. R. Lersch, 1994, Adaptive image sharpening using multiresolution representations, In Proceedings of the *SPIE's Int. Symp. on Optical Engineering in Aerospace Sensing*, Orlando (Florida), USA, April 4-8 1994, **2231**, 72-83.

Kam M.,W. Chang, Q. Zhu, 1991, Hardware complexity of binary distributed detection systems with isolated local bayesian detectors, *IEEE Transactions on Systems, Man, and Cybernetics*, **21**, 3, 565-571.

King R. W., V. H. Kaupp, W. P. Waite, H. C. MacDonald, 1984, Digital color space transformations, In Proceedings of IGARSS'84, Strasbourg, France, ESA SP-215, 649-654.

Kong X. N., D. Vidal-Madjar, 1988, Effet de la résolution sur les propriétés statistiques des images satellites: une étude de cas, *International Journal of Remote Sensing*, **9**, 1315-1328.

Lee J. S., 1985, Speckle suppression and analysis for synthetic aperture radar images, In Proceedings of the *SPIE's 1985 Int. Conf. on speckle*, **556**, 170-179.

Lerman I. C., 1981, *Classification et analyse ordinale des données*, Editeur Dunod, Paris, ISBN n° 2-04-015405-1, 740 p.

Lippmann R. P., 1987, An introduction to computing with neural nets, *IEEE ASSP Magazine*, April 1987, 4-22.

Lopes A., E. Nezry, R. Touzi, H. Laur, 1990, Maximum a posteriori speckle filtering and first order texture models in SAR images, In Proceedings of IGARSS'90, **3**, 2409-2412, IEEE catalog number 90CH2825-8.

Mallat S., 1989, A theory for multiresolution signal decomposition: the wavelet representation, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **11**, 7, 674-693.

Mangolini M., T. Ranchin, L. Wald, 1992, Procédé et dispositif pour augmenter la résolution spatiale d'images à partir d'autres images de meilleure résolution spatiale, brevet n° 92-13961, 20 novembre 1992.

Mangolini M., T. Ranchin, L. Wald, 1993-a, Fusion d'images SPOT multispectrales (XS) et panchromatique (P), et d'images radar", *De l'optique au radar, les applications de SPOT et ERS*, 199-209. Cépaduès-Editions, 111 rue Vauquelin, Toulouse, France, 574 p.

Mangolini M., T. Ranchin, L. Wald, 1993-b, A new fusion method of multispectral multiresolution images : application to SPOT data, Communication to IGARSS'93, August 18-21, Tokyo, Japan.

Mc Culloch W. S., W. Pitts, 1943, A logical calculus of the ideas imminent in nervous activity, *Bulletin of Mathematical biophysics*, **5**, 115-133.

Milgram M., 1993, *Reconnaissance des formes - Méthodes numériques et connexionistes*, Armand Colin Editeur, Paris, 176 p.

Minsky M., S. Papert, 1969, *Perceptrons: an introduction to computational geometry*, MIT Press, Cambridge (MA), USA, Expanded edition 1988.

Munechika C. K., J. S. Warnick, C. Salvaggio, J. R. Schott, 1993, Resolution enhancement of multispectral image data to improve classification accuracy, *Photogrammetric Engineering & Remote Sensing*, **59**, 1, 67-72.

Pau L. F., 1989, Knowledge representation approaches in sensor fusion, *Automatica*, **25**, 2, 207-214.

Pellemans A. H. J. M., R. W. L. Jordans, R. Allewijn, 1993, Merging multispectral and panchromatic SPOT images with respect to the radiometric properties of the sensor, *Photogrammetric Engineering & Remote Sensing*, **59**, 1, 81-87.

Pradines D., 1986, Improving SPOT images size and multispectral resolution, In Proceedings of the *SPIE's Earth Remote Sensing Using the Landsat Thematic Mapper and SPOT Sensor Systems*, **660**, 98-102.

Pratt W. K., 1978, *Digital image processing*, Wiley, New York, USA, 749 p.

Raffy M., 1992, Change of scale in models of remote sensing: a general method for spatialization of models, *Remote Sensing of Environment*, **40**, 101-112.

Raffy M., 1993, Remotely-sensed quantification of covered areas and spatial resolution, *International Journal of Remote Sensing*, **14**, 1, 135-159.

Ramstein G., M. Raffy, 1989, Analysis of the structure of radiometric remotely-sensed images, *International Journal of Remote Sensing*, **10**, 6, 1049-1073.

Ranchin T., 1993, Applications de la transformée en ondelettes et de l'analyse multirésolution au traitement des images de télédétection, Thèse de Doctorat en Sciences, Université de Nice-Sophia Antipolis, France, 110 p.

Ranchin T., L. Wald, 1993, The wavelet transform for the analysis of remotely sensed images, *International Journal of Remote Sensing*, **14**, 3, 615-619.

Ranchin T., L. Wald, M. Mangolini, 1993-a, Application de la transformée en ondelettes à la simulation d'images SPOT multispectrales de résolution 10 m, Comptes-rendus du 14ème colloque GRETSI, Juan-les-Pins, France, 1387-1390.

Ranchin T., L. Wald, M. Mangolini, 1993-b, Efficient data fusion using wavelet transform : the case of SPOT satellite images, In Proceedings of the *SPIE's 1993 Int. Symp. on Optics, Imaging and Instrumentation*, San Diego (California), USA, July 11-16 1993, **2034**.

Robertson P. K., J. F. O'Callaghan, 1988, The application of perceptual color spaces to the display of remotely sensed imagery, *IEEE Transactions on Geoscience and Remote Sensing*, **26**, 1, 49-59.

Schowengerdt R. A., 1983, *Techniques for Image Processing and Classification in Remote Sensing*, Academic Press, ISBN n° 0-12-628980-8, 272 p.

Shafer G., 1976, *A mathematical theory of evidence*, Princeton University Press, Princeton (NJ), USA.

Shettigara V. K., 1992, A generalized component substitution technique for spatial enhancement of multispectral images using a higher resolution data set, *Photogrammetric Engineering & Remote Sensing*, **58**, 5, 561-567.

Shynk J. J., S. Roy, 1990, Convergence properties and stationary points of a perceptron learning algorithm, In Proceedings of the *IEEE*, **78**, 10, 1599-1604.

Smith A. R., 1978, Color gamut transformation pairs, *Computer Graphics*, **12**, 12-19.

Tanré D., C. Deroo, P. Duhaut, M. Herman, J. J. Morcrette, J. Perbos, P. Y. Deschamps, 1990, Description of a computer code to simulate the satellite in the solar spectrum : the 5S code, *International Journal of Remote Sensing*, **11**, 4, 659-668.

Thomopoulos S. C. A., R. Viswanathan, D. C. Bougoulias, 1987, Optimal decision fusion in multiple sensor systems, *IEEE Transactions on Aerospace and Electronic Systems*, **23**, 5, 644-653.

Toutin T., Y. Carbonneau, L. St-Laurent, 1992, An integrated method to rectify airborne radar imagery using DEM, *Photogrammetric Engineering & Remote Sensing*, **58**, 4, 417-422.

Travis D., 1991, *Effective color displays*, Academic Press, ISBN 0-12-697690-2, 301 p.

Tsitsiklis J. N., M. Athans, 1985, On the complexity of decentralised decision making and detection problems, *IEEE Transactions on Automatic Control*, **AC-30**, 5, 440-446.

Ulaby F. T., R. K. Moore, A K. Fung, 1982, *Microwave remote sensing*, vol. 2, Addison-Wesley Pub., Reading (MA) USA.

Welch R. M., M. S. Navar, S. K. Sengupta, 1989, The effect of spatial resolution upon the texture-based cloud field classifications, *Journal of Geophysical Research*, **94**, D12, 14767-14781.

Widrow B., M. A. Lehr, 1990, 30 years of adaptive neural networks: perceptron, madaline, and backpropagation, In Proceedings of the IEEE, **78**, 9, 1415-1442.

Woodcock C. E., A. H. Strahler, 1987, The factor of scale in remote sensing, *Remote Sensing of Environment*, **21**, 311-325.

Woodcock C. E., A. H. Strahler, D. L. Jupp, 1988, The use of variograms in remote sensing: II. Real digital images, *Remote Sensing of Environment*, **25**, 1, 349-379.

Yair E., A. Gersho, 1990, Maximum a posteriori decision and evaluation of class probabilities by Boltzmann perceptron classifiers, In Proceedings of the *IEEE*, **78**, 10, 1620-1628.

Yésou H., Y. Besnus, J. Rolet, J. C. Pion, A. Aing, 1993, Merging Seasat and SPOT imagery for the study of geological structures in a temperate agricultural region, *Remote Sensing of Environment*, **43**, 265-279.

ANNEXES

Annexe A

METHODES DE FUSION D'IMAGES MULTISPECTRALES MULTIRESOLUTIONS

Nous décrivons dans cette annexe les méthodes existantes visant à améliorer la résolution spatiale d'images multispectrales, à l'aide d'une image de meilleure résolution spatiale. Ces méthodes, qui s'appliquent à des images préalablement recalées, peuvent être classées en deux catégories : les premières combinent les radiométries (ou les luminances) des différentes bandes, alors que les secondes dissocient information spectrale et information spatiale. Nous montrons ici pourquoi ces méthodes altèrent le contenu spectral des images initiales, ou nécessitent des hypothèses très sévères qui en limitent le champ d'application.

Certaines méthodes sont spécifiques à un capteur particulier, comme la méthode "P+XS" du CNES qui s'applique à SPOT, mais la plupart sont généralisables à d'autres capteurs. Nous attachons une importance particulière à la méthode "P+XS", car elle a été utilisée comme référence pour l'évaluation des performances de la méthode ARSIS (cf. chap. 3). On adopte, pour les capteurs autres que SPOT, les notations suivantes concernant les images à fusionner : HR désigne l'image de meilleure résolution spatiale, et BRi (i=1,..., n) les images multispectrales de moins bonne résolution spatiale (correspondant par exemple à P et XSi respectivement pour la fusion SPOT P et XS). Les images multispectrales haute résolution synthétisées sont notées HRi.

A.1 Méthodes combinant les radiométries

Les méthodes combinant les radiométries (ou les luminances) des différentes bandes modifient, par définition, le contenu spectral des images initiales. Parmi ces méthodes, on décrit tout d'abord la méthode "P+XS" du CNES, qui repose sur le recouvrement spectral des bandes XS1, XS2, et P de SPOT. Nous illustrons l'erreur

radiométrique d'une part numériquement, à l'aide de simulations réalisées sur trois terrains particuliers, d'autre part visuellement sur des images de Toulouse. Les autres méthodes existantes sont ensuite brièvement exposées et critiquées.

A.1.1 Méthode P+XS du CNES

Cette méthode vise à produire, à partir des images SPOT panchromatique (P) de résolution 10 m, et multispectrales (XSi) de résolution 20 m, des images multispectrales de résolution 10 m notées XPi. Elle repose sur l'approximation suivant laquelle la luminance reçue dans la bande P est la demi-somme des luminances reçues dans les bandes XS1 et XS2 (Anonyme 1986). Cette hypothèse est intuitivement suggérée par l'examen des bandes spectrales de SPOT (cf. fig. A.1). Nous allons évaluer numériquement l'erreur qui en découle.

Soit L_k la luminance équivalente SPOT dans la bande k (k désignant P, XS1, XS2, ou XS3), définie par :

$$L_k = \frac{\int L(\lambda).S_k(\lambda)\, d\lambda}{\int S_k(\lambda)\, d\lambda} \qquad \text{(A-1)}$$

avec : $L(\lambda)$: densité spectrale de luminance (W m^{-2} sr^{-1} μm^{-1}),
$S_k(\lambda)$: sensibilité spectrale de l'instrument HRV pour la bande k (unités arbitraires). Les unités de $S_k(\lambda)$ ont été choisies de telle sorte que $\int S_k(\lambda)\, d\lambda$ ait la même valeur dans toutes les bandes.

Notons L_{XP1} et L_{XP2} les luminances simulées pour réaliser les images XP1 et XP2 à la résolution de 10 m. L'hypothèse spectrale précédente se traduit, sur les images à 10 m, par :

$$L_P = \frac{L_{XP1} + L_{XP2}}{2} \qquad \text{(A-2)}$$

Il est nécessaire d'effectuer une seconde hypothèse, qui exprime comment répartir la luminance de l'image P entre les images XP1 et XP2. La méthode P+XS du CNES propose de conserver le rapport des luminances entre les bandes 1 et 2, suivant l'équation (A-3). Cette seconde hypothèse n'est entièrement justifiée que dans le cas où

le pixel 20 m peut être considéré comme composé d'objets d'échelle 10 m, possédant tous la même réponse spectrale.

$$\frac{L_{XP1}}{L_{XP2}} = \frac{L_{XS1}}{L_{XS2}} \quad \text{(A-3)}$$

On obtient finalement les formules de calcul suivantes :

$$L_{XP1} = 2\,L_P \cdot \frac{L_{XS1}}{L_{XS1} + L_{XS2}} \qquad L_{XP2} = 2\,L_P \cdot \frac{L_{XS2}}{L_{XS1} + L_{XS2}} \quad \text{(A-4)}$$

Ces formules ne permettent pas de créer une image haute résolution XP3, du fait du recouvrement quasi nul des bandes P et XS3. La solution retenue par le CNES, pour obtenir cette troisième bande au pas de 10 m, est une simple duplication de pixels.

La méthode P+XS présente deux inconvénients majeurs perceptibles à l'examen de ces équations. Premièrement, l'ensemble des structures présentes dans l'image P et dans l'image XS2 se retrouvent dans XP1, quelles que soient leurs échelles caractéristiques. L'un des effets est de fortement modifier le contenu spectral de XP1 par rapport à XS1. Le même inconvénient se produit pour XP2. Il s'ensuit en particulier que l'image XP1 (ou XP2) ramenée à 20 m diffère fortement de l'image originale XS1 (ou XS2). Deuxièmement, l'approximation de base (éq. A-2) est assez grossière, comme le montre la comparaison des filtres spectraux P et $\left(\frac{XS1 + XS2}{2}\right)$ sur la figure A.1.

Afin d'évaluer l'erreur en luminance, nous avons effectué des simulations sur différents terrains homogènes typiques - terrain de réflectance égale à 0,2, végétation, et sable - dont les courbes de réflectance sont représentées sur la figure A.2. Ces simulations prennent en compte un modèle atmosphérique qui permet d'obtenir la luminance observée par le satellite (cf. fig. A.3) à partir de la réflectance au sol. Les densités spectrales de luminance satellite ont été calculées pour un modèle atmosphérique standard à l'aide du logiciel "5S" (Tanré *et al.* 1990), et l'on a vérifié la faible sensibilité du résultat final aux différents paramètres atmosphériques. Le produit de la luminance satellite simulée $L(\lambda)$ par le filtre $\left(\frac{XS1 + XS2}{2} - P\right)$ donne la densité spectrale d'erreur en luminance, liée à la première hypothèse de la méthode P+XS (cf. éq. A-2).

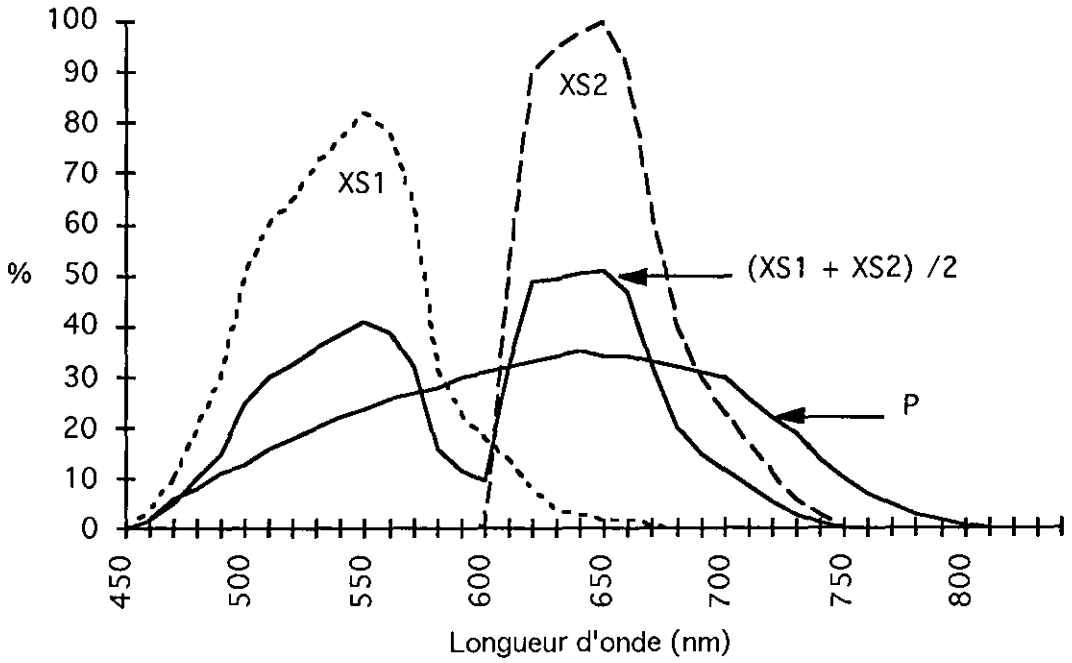


Figure A.1 : approximation spectrale de la méthode du CNES

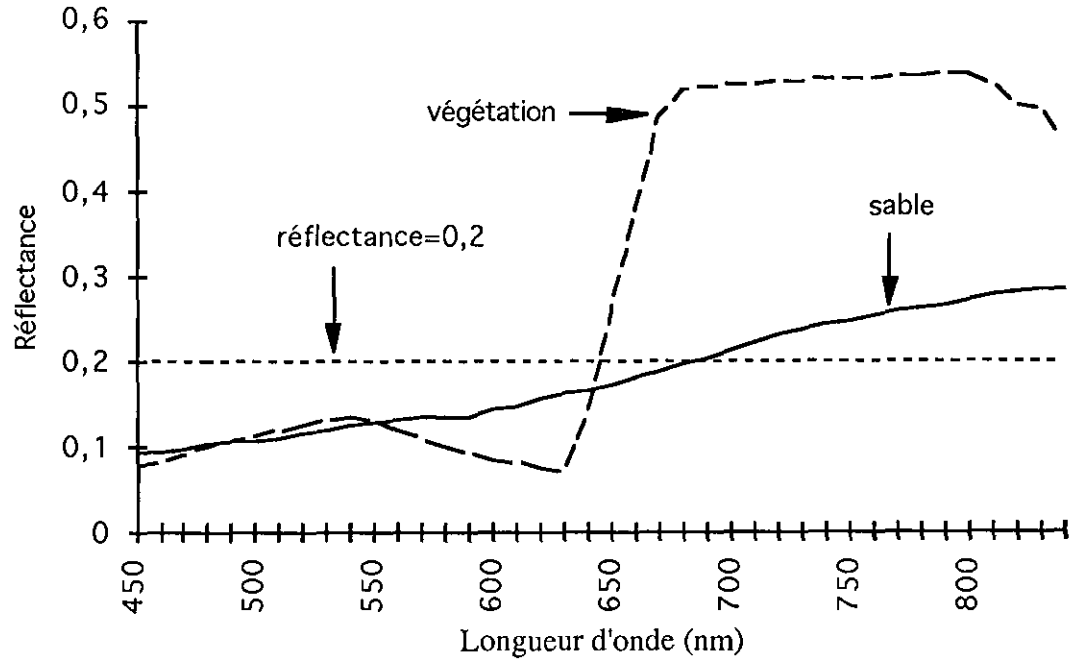


Figure A.2 : courbes de réflectance de trois types de terrains

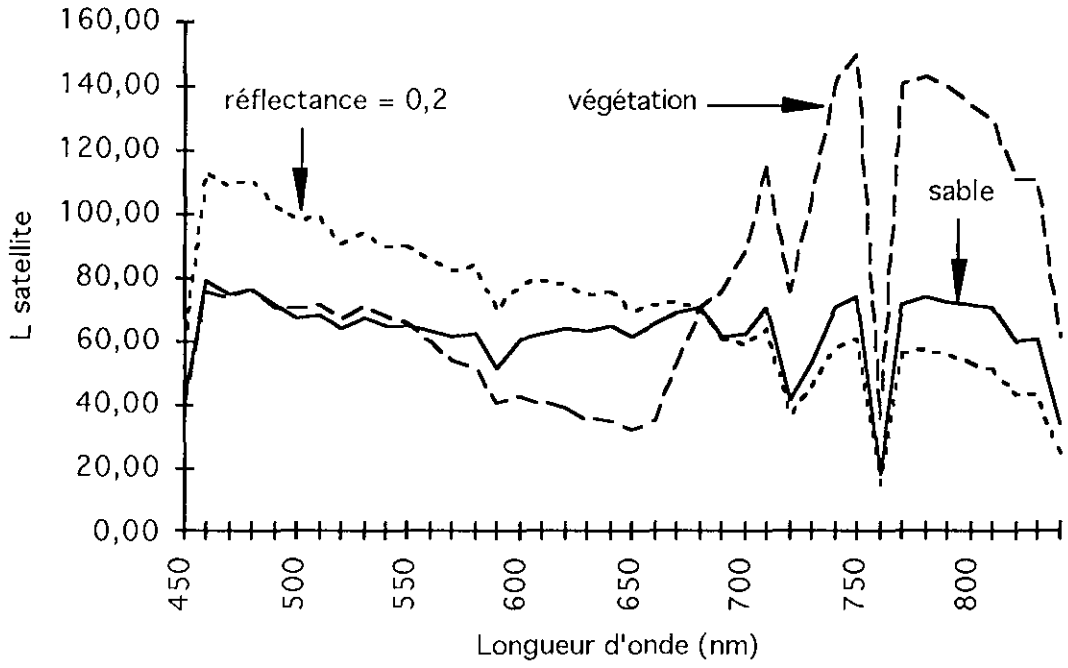


Figure A.3 : luminance observée par le satellite pour trois types de terrains (en W m^{-2} sr^{-1} μm^{-1})

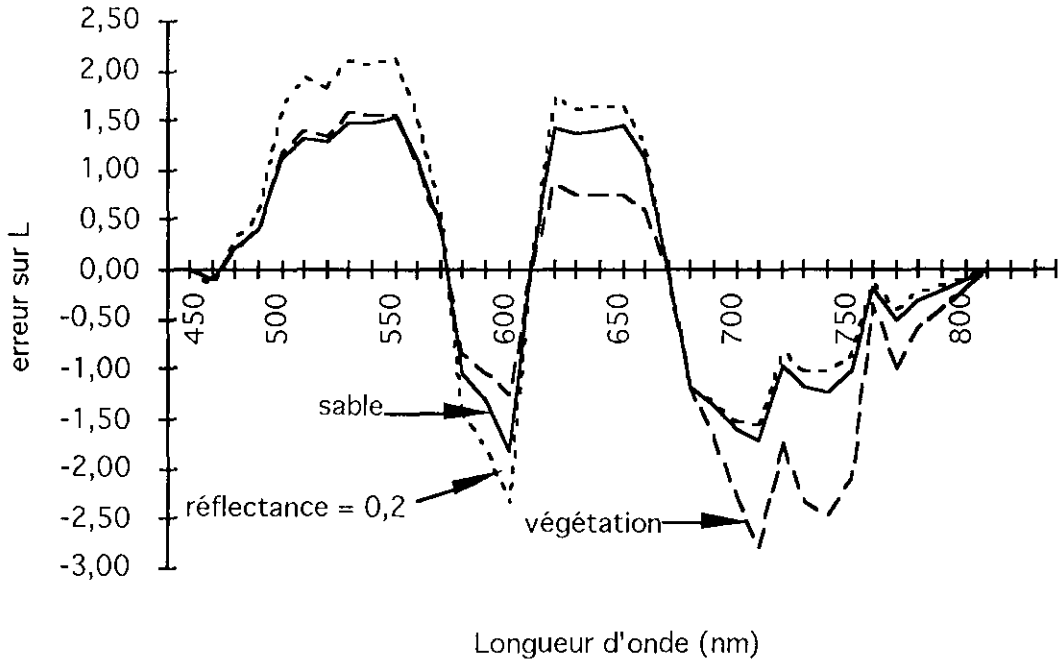


Figure A.4 : densité spectrale d'erreur en luminance de la méthode P+XS (en W m^{-2} sr^{-1} μm^{-1})

Cette densité spectrale d'erreur est reportée sur la figure A.4, qui permet de localiser les domaines du spectre où l'erreur en luminance est importante, pour chaque type de terrain envisagé. On vérifie que la densité spectrale d'erreur dépend nettement du type de terrain. L'erreur liée à la première hypothèse de la méthode P+XS s'en déduit par intégration sur tout le spectre. Les écarts de signes opposés se compensent en partie.

L'erreur radiométrique finale *Er* sur XP1 et XP2 est calculée à partir de l'équation (A-4) :

$$Er = \frac{L_{XP1} - L_{XS1}}{L_{XS1}} = \frac{L_{XP2} - L_{XS2}}{L_{XS2}} = \frac{2\,L_P}{L_{XS1} + L_{XS2}} - 1 \qquad \text{(A-5)}$$

Le tableau A.1 donne L_{XS1}, L_{XP1}, et l'erreur radiométrique *Er* pour les trois terrains homogènes retenus.

CIBLE :	Réflectance = 0,2	Végétation	Sable
L_{XS1} (W m^{-2} sr^{-1} µm^{-1})	90	63	65
L_{XP1} (W m^{-2} sr^{-1} µm^{-1})	83	72	64
Erreur radiométrique Er (%)	- 8 %	15 %	- 2 %

Tableau A.1 : erreur radiométrique finale *Er* de la méthode P+XS, sur trois types de terrains

L'erreur radiométrique, de signe variable, est assez élevée pour le terrain de réflectance constante 0,2, et très élevée pour la végétation. Elle peut être encore beaucoup plus grande pour des objets présentant un pic de réflectance autour de l'une des longueurs d'onde suivantes : 550 nm, 600 nm, 650 nm et 700-750 nm, où l'erreur de la méthode est maximale (cf. fig. A.1).

Les simulations précédentes montrent que la méthode P+XS commet une erreur radiométrique élevée. Nous allons maintenant illustrer cette erreur sur des images réelles de la ville de Toulouse (France). L'erreur radiométrique de la méthode P+XS est clairement mise en évidence sur une chute d'eau réalisée en travers de la Garonne, appelée Chaussée du Bazacle. Les forts courants induisent la formation d'écume sur une vingtaine de mètres en aval. Cette écume apparaît très brillante sur l'image panchromatique, sur laquelle on peut la confondre avec un pont (cf. fig. A.5).



Figure A.5 : image panchromatique de la ville de Toulouse.
La Chaussée du Bazacle est entourée d'un cercle.

Par contre, sa réponse est quasi-nulle sur XS1 et XS2. On vérifie donc que, pour cette cible particulière, la bande P est très différente de la moyenne de XS1 et XS2. La forte charge turbide de l'eau et de l'écume (image datée de mai) induit une forte réponse dans l'infrarouge nettement visible dans XS3 et dans P, en raison de l'étendue de sa bande spectrale (cf. fig. A.1). A ce facteur spectral, s'ajoute l'influence du changement de résolution spatiale. En effet, la largeur de la Chaussée du Bazacle est d'environ 30 m, soit 3 pixels sur l'image P. Sur les images XS de résolution 20 m, la valeur radiométrique de la Chaussée est le résultat d'une moyenne avec la Garonne et les quais environnants. Le contraste entre la Chaussée et la Garonne est donc plus faible sur XS que sur P. Les images XP1 et XP2 obtenues par la méthode P+XS font apparaître à tort la Chaussée du Bazacle. Par contre, les images XS1-HR et XS2-HR obtenues par la méthode ARSIS ne font pas apparaître la Chaussée, conformément à sa réponse spectrale. Cet exemple a de plus l'intérêt de montrer la nécessité de l'utilisation des bandes multispectrales afin de pouvoir discriminer la Chaussée du Bazacle des ponts.

A.1.2 Autres méthodes

L'hypothèse de recouvrement des bandes spectrales effectuée par la méthode P+XS peut être affinée : Munechika *et al.* (1993) proposent d'exprimer la bande P comme somme pondérée des bandes XS, les coefficients étant déterminés par régression linéaire à partir de cibles caractéristiques de spectre connu. Cette technique permet de plus d'étendre la fusion à d'autres bandes que P et XS (par exemple SPOT P et Landsat TM). Cependant, le mélange des bandes spectrales modifie encore la radiométrie et les structures des images XS originales.

Une autre famille de méthodes combinant les radiométries peut être qualifiée de "méthodes de changement de base". Le principe de ces méthodes, illustré par la figure A.6, peut être décomposé en trois étapes :

- on effectue un changement de base sur le vecteur constitué des images basse résolution {BRi}, qui aboutit à un nouveau vecteur {Vi},
- l'image haute résolution HR est subtituée à l'une des composantes de {Vi},
- on effectue enfin un changement de base inverse qui fournit l'ensemble {HRi} d'images haute résolution.

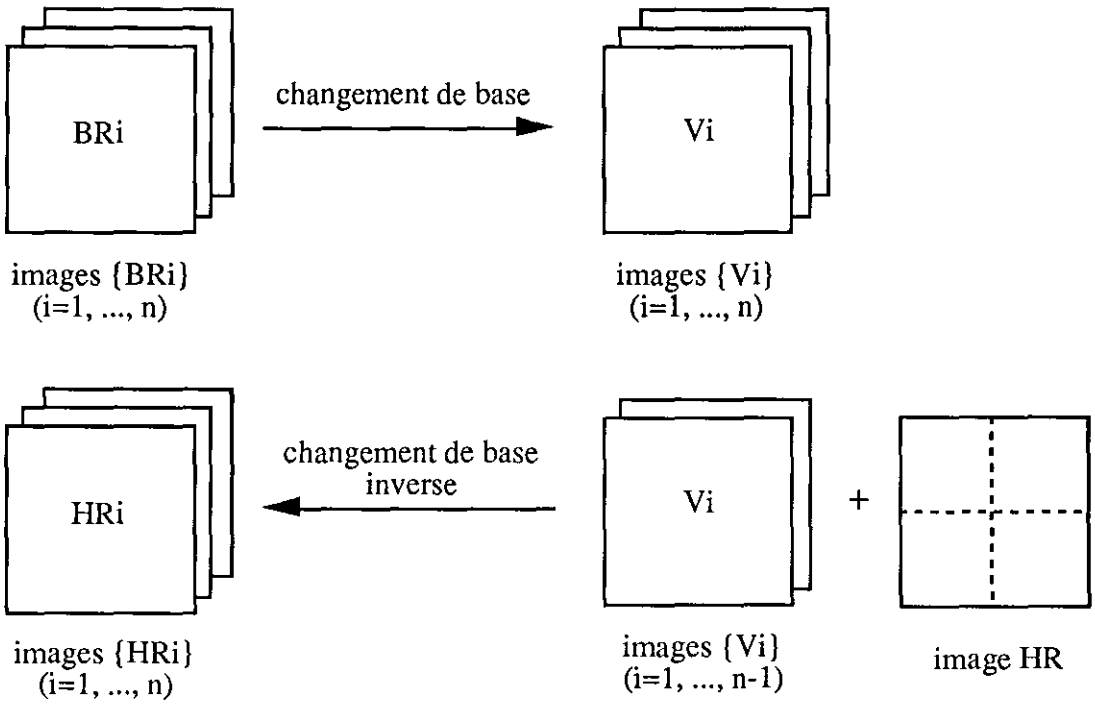


Figure A.6 : principe des méthodes de changement de base

Le changement de base est destiné à produire une image basse résolution, qui représente une information la plus proche possible de celle de l'image HR, selon un critère donné. La substitution de l'image HR à cette image basse résolution constitue la phase d'injection de la haute résolution spatiale dans les images multispectrales. Les prétraitements nécessaires sont le recalage géométrique des images {BRi} et HR, et le suréchantillonnage des images {BRi} au pas de l'image HR, qui permettent une correspondance pixel à pixel entre les images.

Les changements de base (ou transformations) les plus couramment utilisés proviennent des espaces de couleurs, de l'analyse factorielle, ou de l'étude des relations entre les bandes spectrales des images à fusionner :

- la base {luminance, teinte, saturation} (LTS ou IHS en anglais) de l'espace des couleurs se déduit de la base {rouge, vert, bleu} (RVB) utilisée en vidéo (cf. chap. 4). La luminance est approximativement la somme des composantes RVB, alors que la teinte et la saturation définissent la couleur. Les images {BRi} sont identifiées à la base {RVB}, ce qui impose l'utilisation de trois images basse résolution. Après changement de base, l'image HR, dont la dynamique est ramenée à celle de l'image de luminance, lui est substituée (Chavez *et al.* 1991, Carper *et al.* 1990). La similitude du

contenu informationnel de ces deux images n'est justifiée que si la bande spectrale de l'image HR recouvre celles des trois images BRi,

- l'analyse en composantes principales (ACP), guidée par un critère d'inertie (variance) maximale, fournit une base composée d'axes factoriels (cf. chap. 4). Selon ce critère, le premier axe factoriel est le plus représentatif de l'information contenue dans l'ensemble des images. On substitue donc l'image HR à la projection du nuage de points sur cet axe, appelée première composante principale (Chavez *et al.* 1991, Shettigara 1992),
- on peut utiliser les relations entre les bandes spectrales des images à fusionner, pour définir un changement de base linéaire. Pellemans *et al.*. (1993) synthétisent une image aussi proche que possible de l'image HR, suivant un critère spectral, par combinaison linéaire des images {BRi} : les coefficients sont déterminés par le recouvrement des bandes spectrales des images BRi et HR.

Toutes ces méthodes de changement de base présentent l'inconvénient majeur de modifier le contenu spectral des images BRi. Etant donnés les profils complexes des bandes spectrales des instruments utilisés, l'image basse résolution synthétisée à partir des images BRi, pour s'approcher de l'image HR, présente de forts écarts radiométriques avec celle-ci. Le mélange de l'image HR avec l'information provenant des images multispectrales, réalisé par le changement de base inverse, modifie donc la radiométrie des images initiales BRi.

A.2 Méthodes dissociant information spectrale et information spatiale

Les méthodes dissociant information spectrale et information spatiale extraient certaines structures géométriques de l'image haute résolution P, pour les injecter dans chaque image multispectrale BRi. Chaque image multispectrale haute résolution HRi résulte donc du mélange de l'image BRi correspondante et de l'image HR, mais pas des autres bandes spectrales. On peut espérer, grâce à ces méthodes, simuler ce que produirait réellement le capteur multispectral s'il possédait la résolution spatiale de l'image HR. Nous exposons et critiquons deux d'entre elles, désignées par "méthode de filtrage passe-haut", et "méthode de Pradines".

- Méthode de filtrage passe-haut :

L'idée est d'extraire l'information spatiale haute résolution de l'image HR en lui appliquant un filtre passe-haut qui fournit une image des hautes fréquences spatiales. Cette image est alors ajoutée, pixel à pixel, aux images multispectrales BRi (Chavez *et al.* 1991). Le principe est assez séduisant, mais les résultats de cette méthode dépendent du filtre, de sa taille, et de sa forme. En outre, à supposer que le filtrage fournisse exclusivement les structures d'échelles comprises entre les résolutions de HR et BRi, leur injection par addition des images gomme toute relation entre compte numérique et luminance. Il y a donc modification du contenu spectral des images BRi.

- Méthode de Pradines :

Cette méthode impose que la résolution des images BRi soit un multiple de celle de l'image HR. Si cette condition n'est pas remplie, il est possible de s'y ramener par rééchantillonnage des images BRi. Notons R_{HR} et R_{BR} respectivement les résolutions des images HR et BRi. On considère indépendamment chaque superpixel des images BRi, de résolution R_{BR}, couvrant $(R_{BR} / R_{HR})^2$ pixels de l'image HR, de résolution R_{HR} (cf. fig. A.7). On cherche à répartir la radiométrie de ce superpixel BRi entre les pixels correspondants qui composent l'image HRi. En notant *XBRi* la radiométrie d'un superpixel BRi, et *XHR(j)* celles des pixels correspondants de l'image HR, la radiométrie *XHRi(j)* des pixels de l'image HRi est donnée par l'équation (A-6) (Pradines 1986).

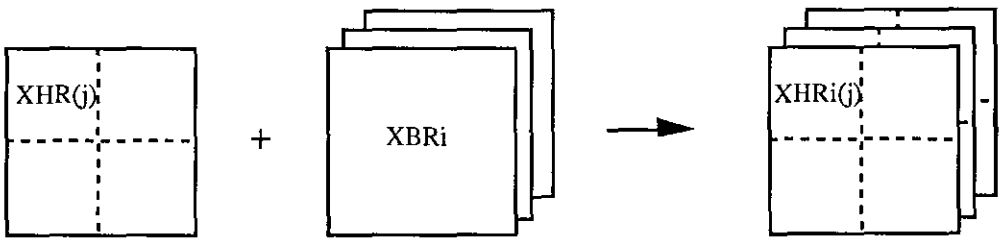


Figure A.7 : principe de la méthode de Pradines

$$XHRi(j) = XBRi \cdot \frac{XHR(j)}{\sum_{k} XHR(k)} \qquad (j=1, ..., (R_{BR} / R_{HR})^2) \qquad \text{(A-6)}$$

La somme des radiométries *XHRi(j)* est égale à la radiométrie *XBRi* du superpixel. Pour que la moyenne des radiométries *XHRi(j)* soit égale à *XBRi*, il suffit de multiplier la valeur de *XHRi(j)* dans l'équation (A-6) par le nombre de pixels $(R_{BR} / R_{HR})^2$. La méthode a été affinée par Price, comme décrit dans Munechika *et al.* (1993), mais le concept reste identique. Son inconvénient majeur est la nécessité d'une forte corrélation entre l'image HR et les images BRi. De plus, chaque superpixel étant traité indépendamment de ses voisins, l'image résultante est souvent bruitée.

Les méthodes précédentes ainsi que les méthodes dérivées disponibles dans la littérature ne satisfont pas les thématiciens pour lesquels la qualité radiométrique est primordiale. C'est pourquoi nous avons été conduits à développer la méthode ARSIS, qui utilise l'analyse multirésolution et la transformée en ondelettes (cf. chap. 3).

Annexe B

CLASSIFICATEUR K-PLUS PROCHES VOISINS

Le classificateur K-plus proches voisins (K-PPV) est un outil de classification dirigée, c'est-à-dire faisant appel à l'utilisateur pour lui fournir initialement des connaissances, par opposition à la classification non dirigée où le système réalise seul une analyse des données. Les méthodes paramétriques usuelles de classification dirigée, comme la classification bayésienne, présentent l'inconvénient majeur d'imposer des hypothèses très fortes sur les lois de probabilité des variables d'entrée. Ces lois n'étant généralement pas vérifiées sur les images de télédétection pour les classes qui nous intéressent, nous préférons utiliser des méthodes de classification non-paramétriques. Parmi ces méthodes, nous avons sélectionné le classificateur K-PPV pour sa simplicité alliée à des performances généralement satisfaisantes.

Pour effectuer la classification de n images superposables, on peut considérer chaque pixel comme un point de l'espace $\Re^n$, dont les composantes sont ses radiométries dans les différentes images. Suite à l'apprentissage, on dispose d'un ensemble de pixels, représenté par un nuage de points, dont la classe d'appartenance est connue. On cherche à affecter une classe à chaque pixel inconnu, à l'aide d'un critère de distance appliqué à son point représentatif dans $\Re^n$ et au nuage de points issu de l'apprentissage. Le déroulement de la méthode, illustré en dimension 2 par la figure B.1, s'articule en deux étapes (Milgram 1993) :

- pour chaque point X de $\Re^n$ à classer, on fait croître le rayon r d'une boule B(X, r) centrée sur X, jusqu'à ce qu'elle contienne exactement K échantillons d'apprentissage,
- la classe affectée à X est celle qui est la plus représentée parmi les K échantillons contenus dans la boule.

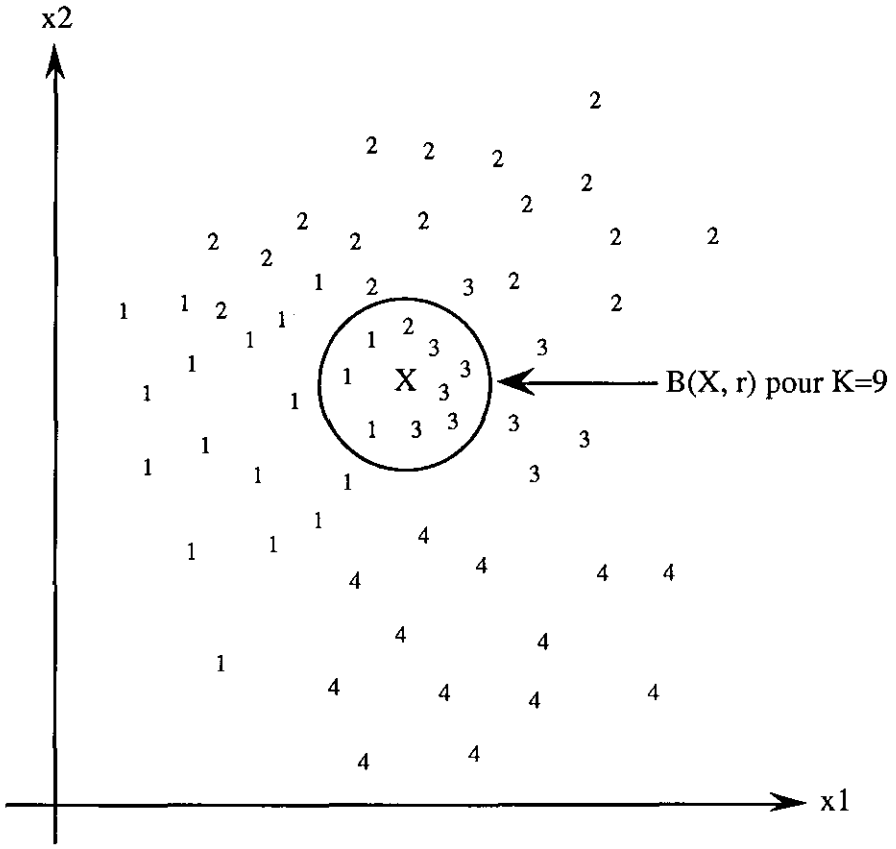


Figure B.1 : illustration de la classification K-PPV en dimension 2. Les échantillons d'apprentissage sont représentés par le numéro de leur classe d'appartenance (4 classes)

Le choix du paramètre K influe directement sur le résultat de la classification. Une faible valeur de K permet d'obtenir une bonne résolution, c'est-à-dire des frontières complexes entre les classes, mais le classificateur est alors sensible au bruit sur les échantillons d'apprentissage et sur le pixel à classer. Inversement, une forte valeur de K lisse les frontières mais améliore la robustesse vis-à-vis du bruit. Le choix de K résulte donc d'un compromis. Nous avons observé de manière empirique sur les images traitées qu'une valeur de K de l'ordre de 10 donnait de bons résultats. D'autre part, il est nécessaire de choisir une distance dans $\Re^n$; nous avons choisi la distance euclidienne qui est parmi les plus usuelles pour le classificateur K-PPV.

Il est possible d'associer à la décision de la méthode K-PPV pour chaque pixel, une valeur de confiance de la classe choisie. Nous utiliserons comme paramètre de confiance le rapport du nombre d'échantillons appartenant à la classe sélectionnée, au nombre total K d'échantillons contenus dans la boule.

Annexe C

RESEAUX DE NEURONES APPLIQUES A LA CLASSIFICATION

C.1 Introduction

Nous utilisons ici les réseaux de neurones comme outil de classification dirigée, de manière analogue au classificateur K-PPV décrit en annexe B. Les réseaux de neurones constituent une alternative connexionniste, qui s'adapte aux propriétés stochastiques de chaque problème spécifique (méthode non paramétrique), par le biais d'une phase d'apprentissage (Yair *et al.* 1990).

Sous le terme de "réseaux de neurones", on regroupe un certain nombre de modèles dont le but est d'imiter le fonctionnement du cerveau humain. Ces modèles ont en commun l'utilisation d'une entité élémentaire qui cherche à modéliser la cellule nerveuse, et diffèrent par le schéma d'interconnexion entre ces neurones. Les plus connus sont le perceptron, le modèle de Hopfield, et celui de Kohonen (Davalo, Naïm 1990). Nous avons choisi de coder un réseau de type perceptron multicouches utilisant l'algorithme de rétropropagation (Widrow *et al.* 1990), en raison de sa simplicité et de ses performances reconnues en matière de classification multisources (Bischof *et al.* 1992). Ses limitations résident dans l'absence de critères garantissant la convergence, et dans la non-unicité théorique de la solution trouvée (Shynk *et al.* 1990).

C.2 Le perceptron multicouches

Avant d'aborder le perceptron multicouches, il est nécessaire de présenter la modélisation des neurones élémentaires qui le composent, connue sous le nom de neurone formel, qui a été proposée dès les années quarante par Mc Culloch et Pitts (1943). Le neurone formel reçoit, à l'instant k, des entrées (x_0, x_{1k}, ..., x_{nk}) constituant un vecteur X_k. Il en réalise une somme pondérée par les coefficients (w_{0k},

w_{1k}, ..., w_{nk}) formant le vecteur $\mathbf{W}_k$. La sortie linéaire $s_k = \mathbf{X}_k^t \mathbf{W}_k$ résultante est filtrée par une fonction de seuillage qui fournit la valeur de sortie y_k correspondant à l'activation du neurone. Ce principe est illustré par la figure C.1, tirée de Widrow *et al.* (1990).

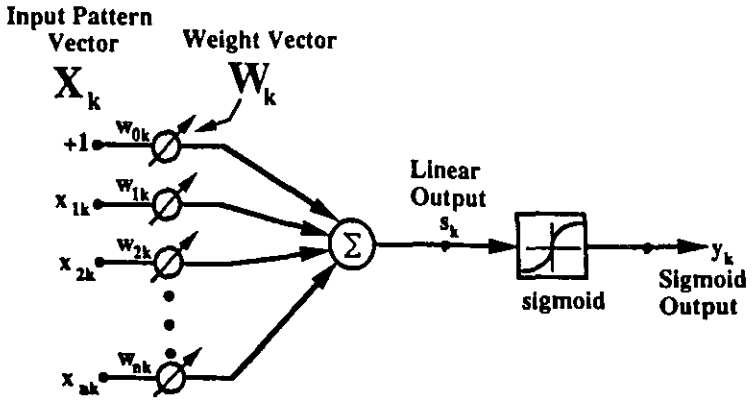


Figure C.1 : modélisation du neurone formel

Les composantes de $\mathbf{X}_k$ sont des valeurs continues ou binaires ; les poids $\mathbf{W}_k$ sont des valeurs continues positives ou négatives. La valeur $x_0 = 1$, constante dans le temps, détermine le seuil d'activation du neurone. La fonction de seuillage utilisée ici est une fonction sigmoïde définie par : $\mathrm{sgm}(x) = \dfrac{1}{1+\exp(-x)}$.

Le perceptron simple utilise le neurone formel précédent comme organe de décision. Afin que celui-ci puisse prendre une décision, il est nécessaire de lui fournir des connaissances par l'intermédiaire d'une phase dite d'apprentissage. Cette phase consiste à présenter au neurone un certain nombre d'exemples définis par le vecteur d'entrée $\mathbf{X}_k$ et la décision correspondante d_k. Les poids $\mathbf{W}_k$, initialisés aléatoirement, sont alors ajustés de telle sorte que la décision y_k coïncide pour l'entrée $\mathbf{X}_k$ avec la valeur désirée d_k, à l'erreur $\varepsilon_k = d_k - y_k$ près. La méthode utilisée pour ajuster les poids $\mathbf{W}_k$, appelée rétropropagation du gradient, vise à minimiser $(\varepsilon_k)^2$. Elle aboutit à définir les poids à l'instant k+1 par la relation (Widrow *et al.* 1990) :

$$W_{k+1} = W_k + 2\,\mu\,\varepsilon_k\,\mathrm{sgm}'(s_k)\,X_k$$

où : sgm' est la dérivée de la fonction sgm
μ est le pas du gradient

Cette relation est appliquée de manière itérative, en recalculant à chaque itération la sortie s_k et l'erreur ε_k, jusqu'à ce que cette erreur devienne inférieure à un seuil donné (cf. fig. C.2). La phase d'apprentissage est alors terminée, et le neurone est capable de fournir une décision y_k pour tout nouveau vecteur d'entrée X_k. C'est la phase dite de relaxation du réseau, dont les performances sont très liées au choix des exemples présentés à l'apprentissage.

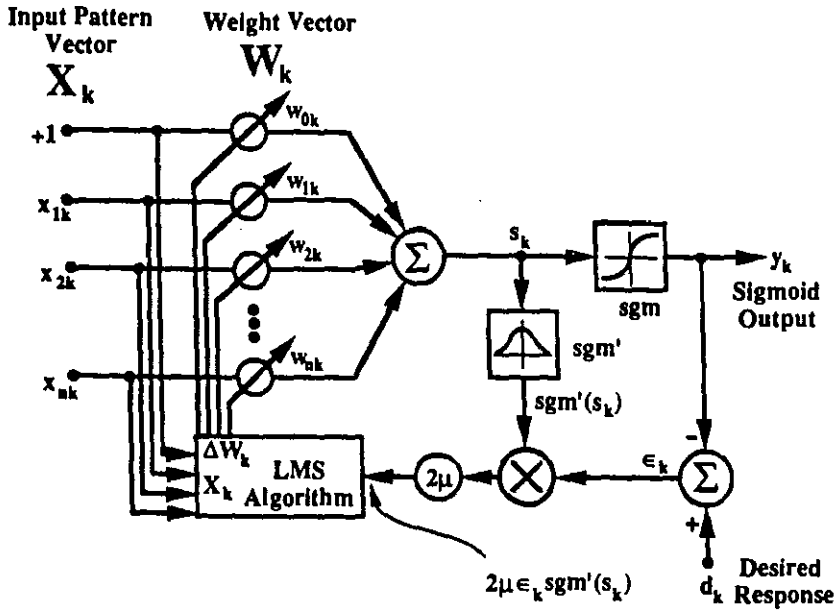


Figure C.2 : le perceptron simple (tiré de Widrow *et al.* 1990)

Le perceptron simple est cependant très limité dans ces capacités décisionnelles : il est notamment incapable de résoudre un problème non linéaire (Minski Papert 1969). L'idée a donc été proposée d'associer des perceptrons simples en réseaux. La généralisation du principe de modification des poids W_k de chaque neurone impose une structure de connexions relativement simple. La plus usitée est la structure multicouches, telle que deux neurones d'une même couche soient indépendents, et que tout neurone d'une couche donnée soit connecté seulement aux neurones de la couche

précédente et de la couche suivante. La dernière couche est appelée "couche de sortie", et les autres "couches cachées". L'exemple d'un réseau à 2 couches est présenté sur la figure C.3. La première couche reçoit les entrées **X**, et transmet ses résultats à la deuxième couche qui est ici la couche de sortie fournissant le vecteur résultat **Y**. Sur cette figure, l'exposant entre parenthèses indique le numéro de la couche et l'indice la position du neurone dans cette couche.

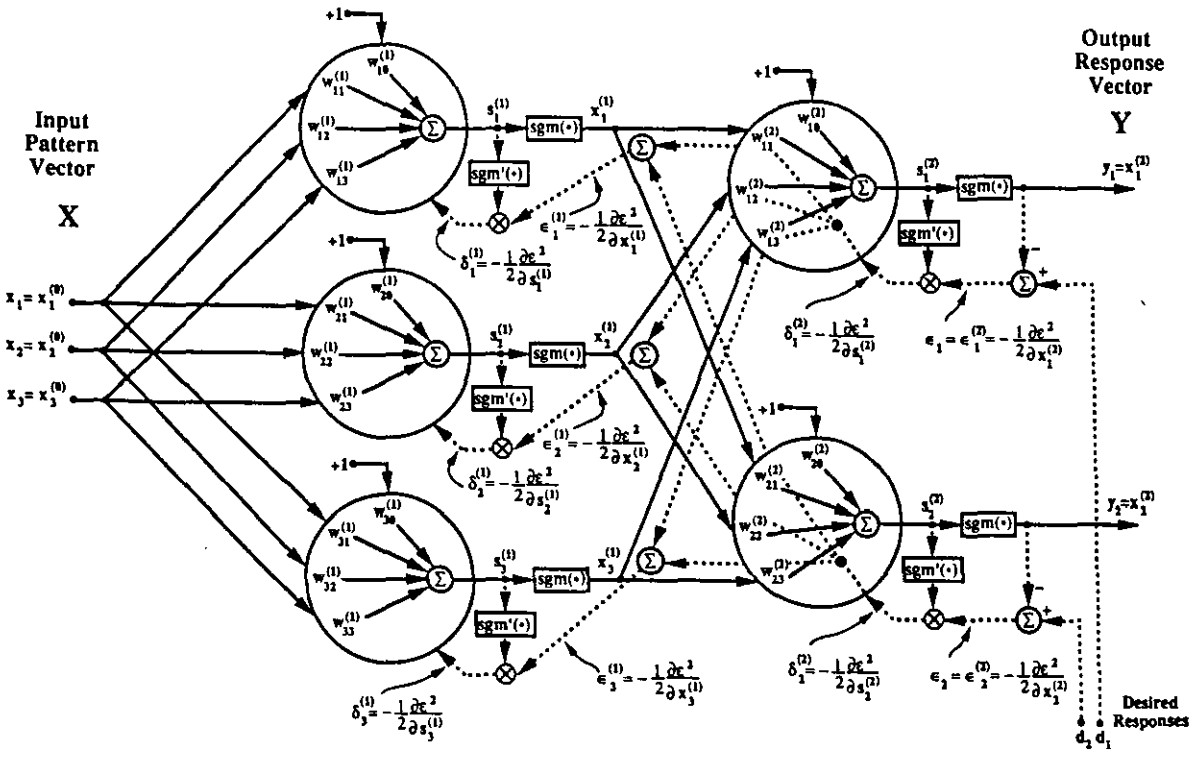


Figure C.3 : le perceptron multicouches (tiré de Widrow *et al.* 1990)

Le perceptron multicouches généralise la méthode de rétropropagation du gradient en minimisant l'erreur quadratique ε^2 définie à partir du vecteur de sortie désiré $\mathbf{D} = (d_1, ..., d_m)$ et du vecteur de sortie réel $\mathbf{Y} = (y_1, ..., y_m)$ par :

$$\varepsilon^2 = \sum_i (d_i^{\,2} - y_i^{\,2})$$

Considérons l'un quelconque des neurones du réseau, possédant à l'instant k les poids $\mathbf{W}_k$, un vecteur d'entrée $\mathbf{X}_k$, une sortie linéaire $s_k = \mathbf{X}_k^{\,t}\,\mathbf{W}_k$, et une erreur ε_k. La formule de rétropropagation du gradient s'écrit alors :

$$\mathbf{W}_{k+1} = \mathbf{W}_k + 2\,\mu\,\delta_k\,\mathbf{X}_k$$

où : μ est le pas du gradient

$$\delta_k = -\frac{1}{2}\frac{\partial\,\varepsilon^2}{\partial\,s_k}$$

δ_k traduit la sensibilité de l'erreur quadratique ε^2 du réseau à la sortie linéaire s_k du neurone.

- si le neurone appartient à la couche de sortie : $\delta_k = \varepsilon_k\ \mathrm{sgm}'(s_k)$
- si le neurone appartient à une couche cachée, δ_k s'obtient à partir des δ_k et des poids $\mathbf{W}_k$ de la couche suivante, comme indiqué par les flèches pointillées sur la figure C.3. On multiplie le δ_k de chaque neurone de la couche suivante par le poids qui lie ce neurone au neurone courant, et l'on somme ces δ_k pondérés. Le résultat est multiplié par $\mathrm{sgm}'(s_k)$ pour donner la valeur δ_k du neurone courant.

La phase d'apprentissage du perceptron multicouches se termine lorsque l'erreur quadratique ε^2 devient inférieure à un seuil prédéterminé. Le réseau est alors capable d'associer à tout vecteur d'entrée **X** un vecteur de sortie **Y** correspondant à une décision. Nous allons voir maintenant comment adapter un tel réseau à un problème de classification d'images.

C.3 Classification multicapteurs

Pour les besoins du chapitre 5, nous souhaitons réaliser une classification à partir d'un jeu de *n* images recalées géométriquement et ramenées au même pas. Tout pixel de la scène sera donc représenté par un vecteur de dimension *n*, composé des radiométries du pixel dans les différentes images. C'est ce vecteur que nous allons utiliser en entrée du réseau de neurones. Le but est de réaliser une classification en *m* classes, c'est-à-dire d'attribuer à chaque pixel de l'image une classe prédéterminée. Le réseau de neurones possédera autant de neurones sur sa couche de sortie qu'il y a de classes. Le nombre de couches cachées et le nombre de neurones qui les composent peut varier. Nous utilisons deux couches cachées, qui permettent de former toutes les régions de décision possibles avec une précision donnée, pourvu que le nombre de neurones sur ces couches soit correctement choisi (Lippmann 1987, Gibson *et al.* 1990). Il est donc inutile de développer un perceptron multicouches possédant plus de

trois couches, ce qui augmenterait la complexité des calculs sans en étendre les capacités décisionnelles. La structure du réseau est représentée sur la figure C.4.

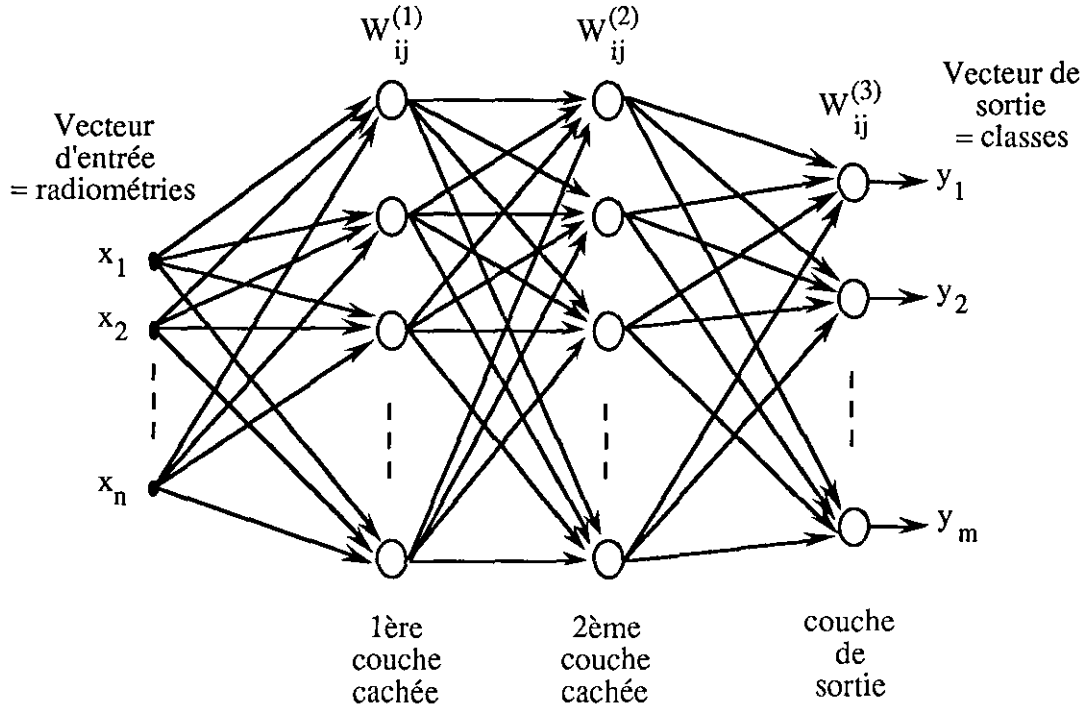


Figure C.4 : structure du réseau de classification

La phase d'apprentissage consiste à présenter en entrée du réseau un certain nombre de pixels présélectionnés, et en sortie un vecteur caractéristique de la classe correspondante. Le vecteur de sortie a toutes ses composantes nulles sauf une, égale à 1, dont la position détermine la classe. Ce codage correspond à l'activation d'un unique neurone sur la couche de sortie, tout en respectant la dynamique de la fonction sigmoïde définie plus haut.

La phase de relaxation, qui réalise la classification, consiste à présenter successivement au réseau l'ensemble des pixels de l'image et à analyser les vecteurs de sortie correspondants. La classe sélectionnée est celle dont le neurone de sortie possède la plus grande valeur d'activation, conformément au codage adopté lors de l'apprentissage. L'image de classification est constituée d'autant de codes présélectionnés qu'il y a de classes. Nous calculons de plus, pour chaque pixel, un paramètre de confiance de la classe sélectionnée, qui peut être défini de différentes façons. L'un des paramètres possibles est la valeur de la plus forte activation en sortie, qui croît jusqu'à la valeur 1 lorsque la confiance augmente. Cependant, d'autres

neurones de sortie peuvent posséder des activations proches de la valeur la plus élevée, rendant le choix de la classe incertain. Un autre paramètre a donc été proposé pour intégrer ce critère : il est défini comme la différence entre les deux plus fortes activations des neurones de sortie. L'image de confiance ainsi créée peut servir d'aide à l'interprétation visuelle de la classification.

Annexe D

RESULTATS COMPLEMENTAIRES SUR LES ARCHITECTURES DE FUSION

Afin de ne privilégier ni le site, ni le classificateur utilisé, et de tirer des conclusions aussi générales que possible sur les architectures, celles-ci ont été mises en œuvre sur deux sites (Fos et Marignane) avec deux classificateurs différents (K-PPV et réseau de neurones). Dans le chapitre 5, les matrices de confusion ne sont données que pour le site de Fos et le classificateur réseau de neurones. Les résultats complets issus des autres cas traités sont rassemblés dans cette annexe.

Deux règles de fusion ont été définies pour les architectures décentralisée et mixte : l'une est basée sur les "confiances pixel" et l'autre sur les "fiabilités statistiques" (cf.§ 5.3.3). Appliquées à l'architecture décentralisée, ces deux règles donnent des taux moyens de classification voisins mais des matrices de confusion sensiblement différentes. En ce qui concerne l'architecture mixte, les taux moyens de bonne classification sont quasiment toujours égaux pour les deux règles, et les matrices de confusion sont très proches. Pour alléger la présentation, seuls sont présentés les résultats fournis par la règle de "confiances pixel".

Au chapitre 5, la comparaison des architectures de fusion a été effectuée sur différentes combinaisons de capteurs, pour évaluer si la complémentarité des capteurs exploités, relativement aux classes souhaitées, privilégie une architecture particulière. Pour limiter le volume de cette annexe à des proportions raisonnables, on ne donne les matrices de confusion que dans le cas où les trois capteurs SPOT, ERS-1, et Landsat sont exploités conjointement.

D.1 SITE DE FOS

D.1.1 CLASSIFICATEUR RESEAU DE NEURONES

D.1.1.1 Classifications monocapteurs

référence \ classé	eau	boisé	non boisé	urbain
eau	44 %	18 %	37 %	0 %
boisé	0 %	89 %	7 %	4 %
non boisé	5 %	83 %	9 %	3 %
urbain	2 %	41 %	3 %	54 %

Taux moyen de bonne classification : 43 %

Tableau D.1 : matrice de confusion sur les données ERS
(site de Fos - classificateur réseau de neurones)

référence \ classé	eau	boisé	non boisé	urbain
eau	50 %	27 %	10 %	13 %
boisé	3 %	84 %	0 %	12 %
non boisé	0 %	6 %	72 %	22 %
urbain	0 %	0 %	0 %	100 %

Taux moyen de bonne classification : 69 %

Tableau D.2 : matrice de confusion sur les données SPOT
(site de Fos - classificateur réseau de neurones)

référence \ classé	eau	boisé	non boisé	urbain
eau	100 %	0 %	0 %	0 %
boisé	0 %	94 %	6 %	0 %
non boisé	0 %	11 %	83 %	6 %
urbain	0 %	1 %	46 %	53 %

Taux moyen de bonne classification : 86 %

Tableau D.3 : matrice de confusion sur les données Landsat
(site de Fos - classificateur réseau de neurones)

D.1.1.2 Architectures appliquées à SPOT, ERS, et Landsat

référence \ classé	eau	boisé	non boisé	urbain
eau	100 %	0 %	0 %	0 %
boisé	0 %	95 %	5 %	0 %
non boisé	0 %	11 %	87 %	2 %
urbain	0 %	1 %	19 %	80 %

Taux moyen de bonne classification : 93 %

Tableau D.4 : matrice de confusion de l'architecture centralisée (site de Fos - classificateur réseau de neurones)

référence \ classé	eau	boisé	non boisé	urbain
eau	94 %	6 %	0 %	0 %
boisé	0 %	94 %	5 %	1 %
non boisé	0 %	10 %	83 %	7 %
urbain	0 %	1 %	33 %	66 %

Taux moyen de bonne classification : 86 %

Tableau D.5 : matrice de confusion de l'architecture décentralisée utilisant les confiances pixel (site de Fos - classificateur réseau de neurones)

référence \ classé	eau	boisé	non boisé	urbain	indécision
eau	99 %	0 %	0 %	0 %	1 %
boisé	0 %	91 %	2 %	0 %	7 %
non boisé	0 %	15 %	73 %	8 %	4 %
urbain	0 %	1 %	0 %	97 %	2 %

Taux moyen de bonne classification : 92 %

Tableau D.6 : matrice de confusion de l'architecture mixte utilisant les confiances pixel (site de Fos - classificateur réseau de neurones)

D.1.2 CLASSIFICATEUR K-PPV

D.1.2.1 Classifications monocapteurs

référence \ classé	eau	boisé	non boisé	urbain
eau	35 %	48 %	16 %	0 %
boisé	0 %	98 %	2 %	0 %
non boisé	4 %	93 %	3 %	1 %
urbain	2 %	57 %	9 %	33 %

Taux moyen de bonne classification : 35 %

Tableau D.7 : matrice de confusion sur les données ERS
(site de Fos - classificateur K-PPV)

référence \ classé	eau	boisé	non boisé	urbain
eau	49 %	31 %	17 %	3 %
boisé	3 %	80 %	1 %	16 %
non boisé	3 %	7 %	73 %	16 %
urbain	0 %	0 %	6 %	94 %

Taux moyen de bonne classification : 67 %

Tableau D.8 : matrice de confusion sur les données SPOT
(site de Fos - classificateur K-PPV)

référence \ classé	eau	boisé	non boisé	urbain
eau	96 %	0 %	0 %	4 %
boisé	0 %	93 %	7 %	0 %
non boisé	0 %	10 %	84 %	6 %
urbain	0 %	1 %	36 %	64 %

Taux moyen de bonne classification : 87 %

Tableau D.9 : matrice de confusion sur les données Landsat
(site de Fos - classificateur K-PPV)

D.1.2.2 Architectures appliquées à SPOT, ERS, et Landsat

référence \ classé	eau	boisé	non boisé	urbain
eau	100 %	0 %	0 %	0 %
boisé	0 %	95 %	5 %	0 %
non boisé	0 %	9 %	88 %	3 %
urbain	0 %	0 %	17 %	83 %

Taux moyen de bonne classification : 93 %

Tableau D.10 : matrice de confusion de l'architecture centralisée (site de Fos - classificateur K-PPV)

référence \ classé	eau	boisé	non boisé	urbain
eau	58 %	25 %	15 %	2 %
boisé	0 %	88 %	2 %	9 %
non boisé	1 %	7 %	80 %	11 %
urbain	0 %	0 %	13 %	87 %

Taux moyen de bonne classification : 72 %

Tableau D.11 : matrice de confusion de l'architecture décentralisée utilisant les confiances pixel (site de Fos - classificateur K-PPV)

référence \ classé	eau	boisé	non boisé	urbain	indécision
eau	99 %	0 %	0 %	0 %	1 %
boisé	0 %	95 %	5 %	0 %	0 %
non boisé	0 %	9 %	87 %	3 %	1 %
urbain	0 %	0 %	14 %	84 %	2 %

Taux moyen de bonne classification : 93 %

Tableau D.12 : matrice de confusion de l'architecture mixte utilisant les confiances pixel (site de Fos - classificateur K-PPV)

D.2 SITE DE MARIGNANE

D.2.1 CLASSIFICATEUR RESEAU DE NEURONES

D.2.1.1 Classifications monocapteurs

référence \ classé	eau	boisé	non boisé	urbain
eau	57 %	9 %	32 %	2 %
boisé	2 %	70 %	12 %	16 %
non boisé	10 %	33 %	38 %	19 %
urbain	5 %	19 %	7 %	68 %

Taux moyen de bonne classification : 60 %

Tableau D.13 : matrice de confusion sur les données ERS
(site de Marignane - classificateur réseau de neurones)

référence \ classé	eau	boisé	non boisé	urbain
eau	67 %	33 %	0 %	0 %
boisé	8 %	76 %	13 %	3 %
non boisé	0 %	0 %	77 %	22 %
urbain	0 %	0 %	1 %	99 %

Taux moyen de bonne classification : 82 %

Tableau D.14 : matrice de confusion sur les données SPOT
(site de Marignane - classificateur réseau de neurones)

référence \ classé	eau	boisé	non boisé	urbain
eau	100 %	0 %	0 %	0 %
boisé	0 %	81 %	2 %	17 %
non boisé	0 %	0 %	79 %	21 %
urbain	0 %	0 %	10 %	90 %

Taux moyen de bonne classification : 88 %

Tableau D.15 : matrice de confusion sur les données Landsat
(site de Marignane - classificateur réseau de neurones)

D.2.1.2 Architectures appliquées à SPOT, ERS, et Landsat

référence \ classé	eau	boisé	non boisé	urbain
eau	100 %	0 %	0 %	0 %
boisé	0 %	96 %	2 %	1 %
non boisé	0 %	0 %	95 %	5 %
urbain	0 %	0 %	4 %	96 %

Taux moyen de bonne classification : 97 %

Tableau D.16 : matrice de confusion de l'architecture centralisée (site de Marignane - classificateur réseau de neurones)

référence \ classé	eau	boisé	non boisé	urbain
eau	100 %	0 %	0 %	0 %
boisé	1 %	80 %	1 %	17 %
non boisé	0 %	0 %	90 %	10 %
urbain	0 %	0 %	2 %	98 %

Taux moyen de bonne classification : 93 %

Tableau D.17 : matrice de confusion de l'architecture décentralisée utilisant les confiances pixel (site de Marignane - classificateur réseau de neurones)

référence \ classé	eau	boisé	non boisé	urbain	indécision
eau	100 %	0 %	0 %	0 %	0 %
boisé	0 %	96 %	0 %	0 %	4 %
non boisé	0 %	0 %	96 %	2 %	2 %
urbain	0 %	1 %	7 %	90 %	2 %

Taux moyen de bonne classification : 95 %

Tableau D.18 : matrice de confusion de l'architecture mixte utilisant les confiances pixel (site de Marignane - classificateur réseau de neurones)

D.2.2 CLASSIFICATEUR K-PPV

D.2.2.1 Classifications monocapteurs

référence \ classé	eau	boisé	non boisé	urbain
eau	45 %	12 %	42 %	0 %
boisé	2 %	81 %	11 %	6 %
non boisé	4 %	48 %	42 %	6 %
urbain	4 %	37 %	9 %	51 %

Taux moyen de bonne classification : 55 %

Tableau D.19 : matrice de confusion sur les données ERS
(site de Marignane - classificateur K-PPV)

référence \ classé	eau	boisé	non boisé	urbain
eau	73 %	27 %	0 %	0 %
boisé	5 %	79 %	16 %	0 %
non boisé	1 %	0 %	83 %	16 %
urbain	0 %	0 %	2 %	98 %

Taux moyen de bonne classification : 85 %

Tableau D.20 : matrice de confusion sur les données SPOT
(site de Marignane - classificateur K-PPV)

référence \ classé	eau	boisé	non boisé	urbain
eau	100 %	0 %	0 %	0 %
boisé	0 %	94 %	1 %	4 %
non boisé	0 %	0 %	32 %	68 %
urbain	0 %	2 %	22 %	76 %

Taux moyen de bonne classification : 77 %

Tableau D.21 : matrice de confusion sur les données Landsat
(site de Marignane - classificateur K-PPV)

D.2.2.2 Architectures appliquées à SPOT, ERS, et Landsat

classé / référence	eau	boisé	non boisé	urbain
eau	100 %	0 %	0 %	0 %
boisé	0 %	97 %	3 %	0 %
non boisé	0 %	0 %	94 %	6 %
urbain	0 %	0 %	4 %	96 %

Taux moyen de bonne classification : 97 %

Tableau D.22 : matrice de confusion de l'architecture centralisée (site de Marignane - classificateur K-PPV)

classé / référence	eau	boisé	non boisé	urbain
eau	80 %	20 %	0 %	0 %
boisé	2 %	82 %	13 %	3 %
non boisé	1 %	0 %	79 %	20 %
urbain	0 %	0 %	4 %	96 %

Taux moyen de bonne classification : 86 %

Tableau D.23 : matrice de confusion de l'architecture décentralisée utilisant les confiances pixel (site de Marignane - classificateur K-PPV)

classé / référence	eau	boisé	non boisé	urbain	indécision
eau	100 %	0 %	0 %	0 %	0 %
boisé	0 %	97 %	3 %	0 %	0 %
non boisé	0 %	0 %	90 %	7 %	3 %
urbain	0 %	0 %	3 %	96 %	1 %

Taux moyen de bonne classification : 96 %

Tableau D.24 : matrice de confusion de l'architecture mixte utilisant les confiances pixel (site de Marignane - classificateur K-PPV)

LISTE DES FIGURES

Figure 2.1: caractéristiques synthétiques de SPOT 1, 2, et 3 13
Figure 2.2: caractéristiques synthétiques de Landsat 5 14
Figure 2.3: caractéristiques synthétiques de ERS-1 15
Figure 2.4: trois types complémentaires de fusion d'images 18
Figure 2.5: cellule de fusion 20
Figure 2.6: exemple d'architecture de fusion 21

Figure 3.1: pyramides multirésolutions 27
Figure 3.2: méthode d'identification des coefficients des images HR et BR 28
Figure 3.3: schéma général de la méthode ARSIS 29
Figure 3.4: schéma de la méthode ARSIS limitée à trois résolutions 30
Figure 3.5: bandes spectrales de SPOT 36
Figure 3.6: extrait de la scène panchromatique sur la ville de Barcelone, et emplacements des 3 vignettes utilisées ultérieurement 37
Figure 3.7: image panchromatique du port de Barcelone (vignette n°2). 43
Figure 3.8: composition colorée XS sur le port de Barcelone (vignette n°2). 45
Figure 3.9: composition colorée XS-HR sur le port de Barcelone (vignette n°2). 47

Figure 4.1: composition colorée RVB sur Landsat TM 3-2-1 63
Figure 4.2: composition colorée RVB sur ERS, TM5, et P 65
Figure 4.3: composition colorée LTS sur P, TM5, et ERS 67
Figure 4.4: principe de la méthode de substitution dans le repère LTS 70
Figure 4.5 : composition colorée TM 3-2-1 avec transformation LTS et substitution de l'image ERS à la luminance 73
Figure 4.6: première composante principale calculée sur P, ERS, et TM4 75
Figure 4.7: troisième composante principale calculée sur P, ERS, et TM4 77
Figure 4.8: schéma d'intégration de la méthode de fusion optique-radar 80
Figure 4.9: image SPOT P de l'aéroport de Marignane 83
Figure 4.10: image ERS-1 de l'aéroport de Marignane 85
Figure 4.11: image fusionnée SPOT P + ERS-1 de l'aéroport de Marignane 87

Figure 5.1: localisation des sites traités (France) 93
Figure 5.2: image panchromatique SPOT de Fos-sur-Mer (France) 95
Figure 5.3: image panchromatique SPOT de Marignane (France) 97
Figure 5.4: architecture centralisée 108

Figure 5.5: architecture décentralisée utilisant les confiances pixel 109
Figure 5.6: architecture décentralisée utilisant les fiabilités statistiques 110
Figure 5.7: architecture mixte utilisant les confiances pixel 112
Figure 5.8: architecture mixte utilisant les fiabilités statistiques 113

Figure A.1: approximation spectrale de la méthode du CNES 142
Figure A.2: courbes de réflectance de trois types de terrains 142
Figure A.3: luminance observée par le satellite pour trois types de terrains 143
Figure A.4: densité spectrale d'erreur en luminance de la méthode P+XS 143
Figure A.5: image panchromatique de la ville de Toulouse. 145
Figure A.6: principe des méthodes de changement de base 148
Figure A.7: principe de la méthode de Pradines 150

Figure B.1: illustration de la classification K-PPV en dimension 2 154

Figure C.1: modélisation du neurone formel 156
Figure C.2: le perceptron simple (tiré de Widrow et al. 1990) 157
Figure C.3: le perceptron multicouches (tiré de Widrow et al. 1990) 158
Figure C.4: structure du réseau de classification 160

Résumé :

Le nombre croissant d'images de la Terre en provenance de capteurs satellitaires pose aux utilisateurs un problème d'exploitation, auquel la fusion de données semble pouvoir apporter des solutions. Ce travail illustre l'apport de la fusion de données en télédétection et en photo-interprétation, sur des données issues des capteurs SPOT, ERS-1, et Landsat, qui sont parmi les plus utilisés dans le domaine. Les spécificités et complémentarités de ces capteurs, de type optique (visible ou infrarouge) ou radar, ont été prises en compte pour réaliser une fusion dite hétérogène. Face à la grande diversité des méthodes de fusion de données, nous avons choisi de nous restreindre à la fusion au niveau pixel, c'est-à-dire proche du signal physique bidimensionnel, par opposition au niveau sémantique. Trois développements méthodologiques majeurs ont été réalisés.

L'exploitation thématique des images satellitaires multispectrales se heurte souvent à une limitation de leur résolution spatiale. Afin d'y remédier, nous proposons une méthode, nommée ARSIS, qui utilise une image de meilleure résolution spatiale pour améliorer celle des images multispectrales, sans en modifier l'information spectrale. Les outils mathématiques mis en œuvre sont l'analyse multirésolution, la transformée en ondelettes, et la modélisation des similitudes spectrales entre les images à fusionner. Des critères d'estimation de la qualité image des produits synthétisés ont été développés, pour pouvoir comparer la méthode ARSIS aux méthodes préexistantes, sur plusieurs scènes SPOT et Landsat. L'intérêt de la méthode ARSIS tient à sa richesse conceptuelle et à un réel gain en qualité spectrale par rapport aux autres méthodes.

Nous avons également étudié quelques méthodes de représentation visuelle du volume de données, qui sont destinées à faciliter le travail de photo-interprétation et à améliorer la qualité de l'information extraite, mais aussi à concentrer l'important volume de données en un minimum d'images. Les méthodes de changement d'espace de représentation des données et l'utilisation de la couleur ont été expérimentées. Les critiques des photo-interprètes sur les produits réalisés ont mis en évidence la nécessité de prendre en compte les spécificités des capteurs pour la fusion de données hétérogènes. Nous avons donc développé une méthode de fusion hétérogène optique-radar, qui consiste à extraire les objets ponctuels à fort écho radar pour les représenter dans un contexte optique.

Le troisième développement a porté sur la comparaison de différentes architectures de fusion, dont le rôle est de définir l'enchaînement des algorithmes qui constituent une méthode de fusion. L'application retenue est la classification automatique, très utilisée en télédétection, qui permet de comparer numériquement les performances des architectures grâce à l'utilisation d'une vérité terrain. Des architectures centralisée, décentralisée, et une architecture "mixte" sont présentées. Leurs avantages et inconvénients en termes de contraintes opérationnelles, de souplesse d'utilisation, et de qualité de la classification sont discutés. Ces derniers sont illustrés par la comparaison des résultats de classification des différentes architectures, pour deux classificateurs et deux sites différents. Enfin, ce travail met clairement en évidence l'apport de la fusion en classification par rapport aux traitements monocapteurs.

Mots-clés : Fusion de données, images, télédétection, photo-interprétation, analyse multirésolution, transformée en ondelettes, multispectral, classification, architectures, réseaux de neurones.