Classification supervisée d'images d'observation de la Terre à haute résolution par utilisation de méthodes markoviennes

Aurélie Voisin

UNIVERSITÉ DE NICE-SOPHIA ANTIPOLIS - UFR Sciences

ÉCOLE DOCTORALE STIC
SCIENCES ET TECHNOLOGIES DE L'INFORMATION ET DE LA COMMUNICATION

T H È S E

pour obtenir le titre de

Docteur en Sciences

de l'Université de Nice - Sophia Antipolis
Specialisée en : Automatique, Traitement du Signal et des Images

Présentée et soutenue par

Aurélie VOISIN

Classification supervisée d'images d'observation de la Terre à haute résolution par utilisation de méthodes markoviennes

Thèse dirigée par Josiane ZERUBIA

préparée à INRIA Sophia Antipolis Méditerranée, Équipe AYIN

soutenue le 17 octobre 2012

Jury :

Président :	Marc BERTHOD	-	INRIA
Rapporteurs :	Paolo GAMBA	-	Université de Pavie
	Grégoire MERCIER	-	Telecom Bretagne
Directrice :	Josiane ZERUBIA	-	INRIA (Ayin)
Examinateurs :	Roger FJÖRTOFT	-	CNES
	Véronique SERFATY	-	DGA

Remerciements

Je tiens en premier lieu à remercier ma directrice de thèse Josiane Zerubia, le président du jury Pr Marc Berthod, les rapporteurs Pr Grégoire Mercier et Pr Paolo Gamba pour leurs corrections et leurs conseils avisés, ainsi que les autres membres du jury, Dr Roger Fjortoft et Dr Véronique Serfaty.

Je remercie également la DGA pour son soutien financier, ainsi que l'INRIA pour le soutien à la fois financier et matériel.

Je remercie beaucoup Gabriele Moser et Pr S. Serpico pour leur aide précieuse au cours de ces trois années, pour le temps qu'ils m'ont consacré, et pour leur présence pendant la soutenance. Je remercie Zoltan Kato pour son aide concernant les graph-cuts, et Dr Hervé Yesou pour ses commentaires sur mon manuscrit. Et bien sûr je n'oublie pas Vladimir, qui m'a tellement aidée ! Merci pour tout, pour ces voyages inoubliables, sans oublier les épagneuls, Cloclo et cette bonne vieille Bonnie.

Remontons le temps et partons aux origines de cette thèse, car un peu plus de trois ans en arrière, qui aurait imaginé que mon stage me conduirait à ce manuscrit ? Guillaume, grâce à toi, je suis désormais titulaire d'un doctorat !

Une pensée maintenant pour tous mes collègues, partis ou encore présents, qu'ils soient arianautes, ayiniens, morphemiens, ou même geometricaciens : Saima, Christine, Guillaume, Sylvain, Alexis, Adrien, Ikhlef, Yuliya, Xavier (pour m'avoir à moitié adoptée chez Morpheme), Florent (et ses meubles Ikea), Yannick, Ahmed, Aymen, et tous les autres :). En repensant à ces trois années écoulées, je me remémore toutes les personnes que j'ai pu rencontrer au cours des écoles d'été et des Doctoriales, et qui m'ont fait passer de très bons moments. Bien sûr, je me dois de remercier mes chères coquilles Saint Jacques pour avoir occupé mes midis et mes 21 juin : Maxime, Marco, Luc, Anny et Fred. Et le meilleur de tous, Clément, à qui je ne vais peut-être pas consacrer 6 lignes, car ça risque de faire un peu long. Longue vie à Akustrix/Elektrix (ou pas).

Merci à ma cousine Sophie et à Marie-Sophie d'avoir été présentes, et une petite pensée pour les toulousaings Thomas, Luc, Inès, Jérôme et Katia, et pour le JB orléannais Nayaz. Sans oublier de remercier ma maman et ma mamie préférées.

Au risque de me répéter sur les remerciements de certaines personnes, je souhaite du courage à tous ceux qui n'ont pas encore soutenu leur thèse, et qui arrivent au bout, doucement mais sûrement !

Enfin, le koala remercie sa branche d'eucalyptus qu'il a trouvée sur son chemin, et qu'il ne compte pas lâcher.

Table des matières

Table des figures

Liste des tableaux

Liste des Abréviations

ΓG Gamma généralisée

ACP Analyse en Composante Principale

ASI Agence Spatiale Italienne

CEM Classification Espérance-Maximisation

CM Champ de Markov

CNES Centre National d'Études Spatiales

CSK COSMO-SkyMed

DLR Deutsches Zentrum für Luft- und Raumfahrt (Agence Spatiale Allemande)

FDP Fonction de Densité de Probabilité

ICM Modes Conditionnels Itérés (*Iterated Conditional Modes*)

MAP Maximum A Posteriori

MCNG Matrices de Co-occurrence de Niveaux de Gris

MMD Dynamique de Metropolis Modifiée

MoLC Méthode des Log-Cumulants

MPM Mode de la Marginale a Posteriori (*Marginal Posterior Mode*)

MV Maximum de Vraisemblance

PPV Plus Proches Voisins

RADAR RAdio Detection And Ranging

RSO Radar à Synthèse d'Ouverture

RVB Rouge Vert Bleu

SEM algorithme Stochastique d'Espérance-Maximisation

SVM Séparateur à Vaste Marge

TSX TerraSAR-X

Chapitre 1

Introduction

Motivations et objectifs

Les premières photographies satellites de la Terre ont été réalisées le 14 août 1959 par le satellite américain Explorer 6 (voir Fig. 1.1). Depuis lors, les avancées dans le domaine satellitaire n'ont cessé, de sorte qu'il existe à ce jour un nombre incalculable de données de télédétection, en constante augmentation. Rien que pour le programme ORFEO, plus de 2000 images d'observation de la Terre sont acquises chaque jour, incluant 450 images acquises par un seul satellite de la constellation Pléiades [CNE 2012], et 1800 images acquises par la constellation COSMO-SkyMed [CSK 2007].

Fig. 1.1 – Première image satellite acquise en 1959 par le satellite américain Explorer 6. [Source : *http ://grin.hq.nasa.gov/*].

La difficulté principale du traitement d'image est d'arriver à informatiser ce que notre oeil est capable de voir naturellement, voire de détecter des choses que l'oeil ne peut percevoir. Le but est de pouvoir extraire des informations souhaitées dans les diverses acquisitions de manière automatique, afin de pouvoir traiter la multitude d'images disponibles sans systématiquement recourir à une expertise humaine. Les applications sont multiples, mais celle qui nous intéresse davantage est la cartographie. Selon les résolutions des acquisitions, les informations à extraire sont diverses, et peuvent aller de la détection de zones vastes (forêts, océans...), à des zones plus petites (routes, bâtiments...). De telles informations sont particulièrement utiles dans les différentes phases de traitement d'une catastrophe naturelle : de la prévention à la reconstruction, en passant par l'assistance durant l'évènement. Au cours de cette thèse, une attention toute particulière a été portée sur les zones urbaines, qui sont actuellement les zones les plus critiques face à des catastrophes naturelles, car elles regroupent une majeure partie de la population.

Parmi les observations possibles de la surface terrestre, les images radars sont particulièrement adaptées aux zones urbaines, car elles présentent un fort coefficient de rétrodiffusion au niveau de ces zones, qui apparaissent donc de manière assez lumineuse sur les images. Cela permet d'établir, par exemple, une carte d'occupation des sols assez précise. De telles images nous ont été fournies par l'agence spatiale

italienne (ASI) dans le cadre d'un projet dans lequel l'équipe Ariana/Ayin du centre INRIA-SAM était partenaire.

Les images radars présentent aussi de bonnes propriétés pour les catastrophes naturelles, car elles permettent, par exemple, de détecter des zones inondées, qui apparaissent de manière plus sombre sur les images. Le SERTIT (Service Régional de Traitement d'Image et de Télédétection) a exploité de telles propriétés pour établir les localisations des inondations dans la vallée de la Moselle en France en octobre 2006 (Fig. 1.2). Les zones inondées sont plus difficiles à détecter sur les images optiques, car l'eau peut être plus ou moins boueuse, et le courant plus ou moins fort, ce qui donne des colorations et des textures variables, contrairement à l'imagerie radar.

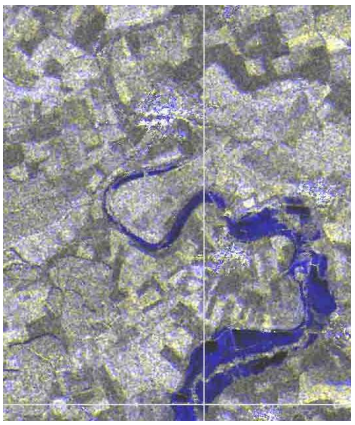


FIG. 1.2 – Carte de zones inondées dans le nord-est de la France (©SERTIT) établie sur une image RADARSAT composite acquise en octobre 2006. [Source : *sertit.u-strasbg.fr*].

La très haute résolution est particulièrement utile dans le cadre de tremblements de terre, car elle permet de détecter les zones urbaines endommagées, comme cela a été fait, par exemple, par le SERTIT sur la ville de Port-au-Prince (Haïti), suite à l'important séisme de janvier 2010 (figure 1.3). Un tel traitement relève davantage de la considération de données multitemporelles, et d'un problème de détection de changements, en particulier pour des acquisitions radars. Au cours de la thèse, nous avons travaillé sur des acquisitions de Port-au-Prince faites après le séisme, en établissant des cartes d'occupation des sols, étant donné la résolution des images disponibles ($2,5$ mètres) et le manque de vérité de terrain en notre possession. Une telle démarche entre dans une problématique de surveillance post-sismique, et peut déjà constituer une source d'information, de la même façon que le projet Kal-Haïti (*http ://kal-haiti.kalimsat.fr*), initié par le centre national d'études spatiales (CNES), vise à exploiter des données de télédétection afin de fournir une multitude d'informations utiles à la reconstruction haïtienne.

Nous avons aussi voulu travailler sur la fusion de différents types de données afin d'améliorer les cartes de classification qui peuvent être parfois imprécises. Une telle considération permet de prendre en compte simultanément un nombre varié de données de différents types à différentes résolutions afin de tirer parti de la multitude d'informations disponibles – pas toujours mises à disposition malheureusement.

Organisation du manuscrit

La thèse s'articule autour de la problématique de la classification d'images radars et optiques acquises à haute résolution.

Dans le chapitre 2, nous nous efforçons de présenter de la manière la plus didac-

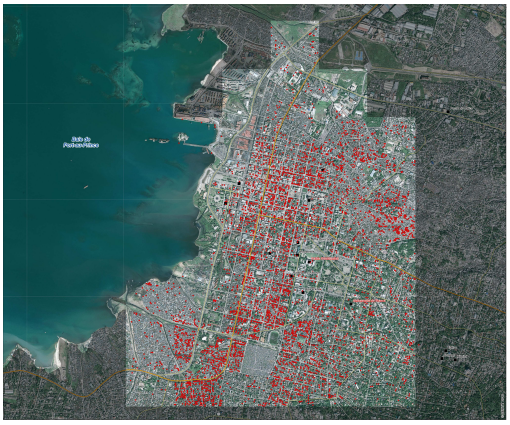


FIG. 1.3 – Carte des bâtiments endommagés de Port-au-Prince (©SERTIT) établie sur une image GeoEye-1 (résolution de 50 cm) acquise le 13 janvier 2010. [Source : *sertit.u-strasbg.fr*].

tique possible la façon dont sont formées les images que nous traitons. Ces images sont des photographies de la surface de la Terre réalisées par des capteurs actifs (RSO) et passifs (optiques). Nous présentons également en détail les satellites qui ont acquis les images réelles sur lesquelles nous avons appliqué les algorithmes de classification que nous avons développés.

Ces algorithmes, reposant sur la théorie bayésienne, ont pour point commun de requérir un apprentissage des statistiques des niveaux de gris pour chacune des classes. Il faut donc trouver la meilleure modélisation possible des images présentées dans le chapitre 2. Les acquisitions à haute résolution constituent un avantage dans la reconnaissance de petits objets. Cependant, elles augmentent la complexité du traitement par l'apparition de zones extrêmement hétérogènes, d'autant plus difficiles à modéliser statistiquement. Comme nous le verrons dans le chapitre 3, cette hétérogénéité intrinsèque requiert une modélisation statistique elle-même hétérogène, choisie comme un mélange pondéré de fonctions statistiques adaptées, dont les paramètres sont estimés automatiquement à partir d'une vérité de terrain. Cette modélisation de la vraisemblance statistique est ensuite intégrée dans des modèles de classification bayésienne markoviens.

Dans le chapitre 4, nous traitons d'images essentiellement monorésolution. Nous utilisons, pour ce faire, un champ de Markov. Il s'agit de trouver le maximum a posteriori connaissant la modélisation statistique de la vraisemblance, sous hypothèse d'un a priori markovien. Le modèle contextuel est développé pour faire face au bruit de chatoiement du radar, mais nous avons montré qu'il s'adapte très bien à des données optiques dans le cadre d'autres applications. La classification des images radars, qui sont, dans notre cas, des images monobandes, est améliorée, grâce à l'introduction d'une image texturée. Cette nouvelle entrée, multibande, est fusionnée statistiquement grâce à l'outil mathématique des copules, qui modélise la

dépendance statistique entre les observations.

Dans le chapitre 5, nous traitons d'images essentiellement multirésolution, acquises avec le même satellite (monocapteur) ou avec différents systèmes (multicapteur). Voulant tirer parti de la multirésolution inhérente à ces acquisitions, nous avons pris en compte un modèle markovien hiérarchique, dans lequel les images sont intégrées dans les différents niveaux du graphe. Nous avons opté pour une représentation en quad-arbre, afin de bénéficier des bonnes propriétés qui lui sont relatives, en particulier sa causalité, qui permet d'appliquer un algorithme d'estimation des étiquettes non itératif. En outre, le contexte est, dans ce cas, en échelle, ce qui permet de tirer parti simultanément des données détaillées (haute résolution) souvent bruitées, et de données à moins bonne résolution, mais en général moins affectées par le bruit.

Pour terminer ce manuscrit, nous donnons, dans le chapitre 6, des perspectives possibles, qui permettraient d'améliorer potentiellement ce travail de thèse.

Contributions

Pour finir cette introduction, voici une liste de nos principales contributions dans cette thèse :

1. Modélisation statistique des différentes classes, adaptée à différents types d'images acquises à haute résolution (Chap. 3) ;
2. Modélisation des statistiques jointes pour chacune des classes par le biais de copules multivariées, choisies comme extension de copules bivariées (Chap. 3) ;
3. Introduction d'un attribut de texture pour améliorer la classification des zones urbaines : choix et modélisation (Chap. 3) ;
4. Intégration de ces diverses statistiques dans un champ de Markov, et étude du paramètre du champ (Chap. 4) ;
5. Intégration des différentes statistiques dans un champ de Markov hiérarchique, fondé sur un modèle en quad-arbre, et intégration d'une mise à jour spécifique pour augmenter la robustesse au bruit de l'algorithme de classification (Chap. 5) ;
6. Tests des différentes méthodes sur différents jeux de données réels mono/multirésolution, et mono/multicapteur (Chaps. 4 et 5).

Chapitre 2

L'imagerie satellitaire

Sommaire

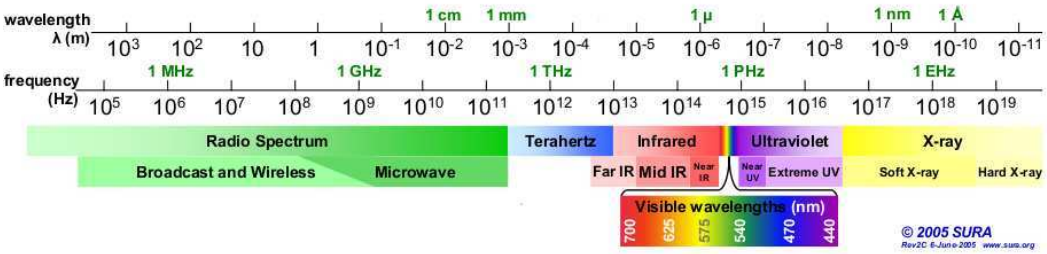


FIG. 2.1 – Spectre électromagnétique [d'après *www.sura.org*]. Les bandes d'intérêt pour nos applications sont situées autour du spectre visible, ainsi que dans les bandes micro-onde (*micro-wave*).

Dans ce chapitre, nous présentons les systèmes satellitaires d'acquisitions d'images terrestres, avec une focalisation particulière sur ceux ayant donné naissance aux images que nous avons traitées. N'ayant pas pour but de fournir une liste exhaustive de l'intégralité du matériel gravitant autour de nos têtes terriennes, nous excluons dès à présent la catégorie des satellites non-défilants. Les satellites défilants sont appelés ainsi car ceux-ci effectuent le tour de la Terre en une centaine de minutes, et comme la Terre elle-même est mobile (rotation d'environ 25 degrés pendant cette durée), à l'issue de ce tour, le satellite ne repasse pas au-dessus d'un même lieu. La période de passage sur un même endroit, appelée *temps de revisite*, varie selon les satellites mais atteint, en général, plusieurs jours. Pour les satellites les plus récents, cette période peut être inférieure à un jour, en particulier dans le cas de constellations de satellites. Les satellites défilants évoluent sur des orbites qualifiées de "basses", de l'ordre de 600 à 800 km en moyenne, altitude compatible avec une durée de vie qui ne soit pas trop affectée par les frottements atmosphériques. En outre, cette altitude relativement peu élevée permet de conférer aux imageurs, en particulier aux imageurs optiques, une bonne aptitude à distinguer des détails de la surface terrestre.

Les instruments embarqués à bord de ces satellites opèrent dans différents domaines de longueur d'onde (Fig. 2.1). En général, les imageurs de type optique utilisent des ondes concentrées autour du spectre visible, tandis que les autres plages de fréquences sont plus couramment utilisées pour des applications plus spécifiques qui nécessitent de détecter des objets qui ne sont pas forcément visibles, tels que la vitesse des vents, les fluctuations de température, etc. Cela confère un nombre considérable d'acquisitions possibles et donc d'applications possibles. En général, les satellites de télédétection sont utilisés en météorologie, en climatologie et en cartographie. Comme nous l'avons déjà évoqué dans l'introduction générale (Chap. 1), et comme le sous-entend le titre même de cette thèse, les applications cartographiques nous intéressent davantage, ce qui nous restreint à certains satellites bien spécifiques, incluant notamment les capteurs optiques [Dowman 2012] et radars de type radar à synthèse d'ouverture [Oliver 2004].

Une des caractéristiques importantes liée aux acquisitions d'images est la notion générale de *résolution*. Il y a quatre types de résolutions possibles lorsque nous traitons d'images de télédétection : spectrale, temporelle, radiométrique et spatiale [Campbell 2002]. La résolution spectrale correspond au nombre de bandes spectrales acquises par le capteur, lorsque celui-ci est passif, et dépend de la largeur des bandes. La résolution temporelle est le nombre de jours écoulés entre deux acquisitions d'une même zone. Par défaut, il s'agit du temps de revisite, mais certains satellites suffisamment mobiles peuvent être orientés depuis des plateformes de commande et, de ce fait, le temps écoulé entre deux acquisitions dépend de la mobilité (agilité) du satellite. Par ailleurs, la résolution radiométrique est le nombre de niveaux de gris des images acquises par le capteur, mesurée souvent en nombre de bits. Par exemple, 8 bits équivaut à 256 niveaux de gris. Enfin, la résolution spatiale, à laquelle nous nous réfèrerons ultérieurement comme "résolution" de l'acquisition, est la taille du plus petit élément observable sur l'image.

Maintenant que nous avons détaillé les deux termes importants que sont les notions de spectre électromagnétique et de résolution, nous pouvons nous focaliser davantage sur le contenu des différentes parties de ce chapitre. Les images traitées dans cette thèse sont essentiellement de deux types : des images optiques et des images radars. Pour ce dernier type d'images, une technologie spécifique, la synthèse d'ouverture, est employée. Nous allons donc nous focaliser dans ce chapitre sur la description des imageries optiques et radars à synthèse d'ouverture (RSO), aussi couramment caractérisé par l'acronyme anglo-saxon SAR pour *Synthetic Aperture Radar*. En particulier, nous donnons les caractéristiques des satellites COSMO-SkyMed, TerraSAR-X, GeoEye et Pléiades, qui correspondent aux images que nous avons utilisées au cours de cette thèse.

2.1 L'imagerie RSO

Le premier système imageur considéré est un système radar. Nous allons donc expliquer ce qu'est une acquisition radar, ainsi que définir des termes spécifiques que nous utilisons de manière courante dans les autres chapitres du manuscrit. Nous nous focalisons ensuite sur la synthèse d'ouverture, puisqu'il s'agit de la technologie utilisée par les récents satellites COSMO-SkyMed et TerraSAR-X, dont nous traitons les images. Nous avons fait la synthèse, pour cette partie, de plusieurs références bibliographiques [Oliver 2004, Polidori 1998, Dupont 1994, Richards 2006, Cheney 2009], ainsi que de divers cours d'écoles d'ingénieurs françaises (ENSEEIHT Toulouse, Telecom Bretagne, ISAE-SUPAERO Toulouse).

2.1.1 Le radar

Le radar, de l'anglais *RAdio Detection And Ranging*, est un système actif, c'est-à-dire qu'il émet et reçoit les ondes rétro-diffusées par les obstacles. Ces ondes sont généralement émises dans des bandes hyperfréquences (entre 1 GHz et 100 GHz).

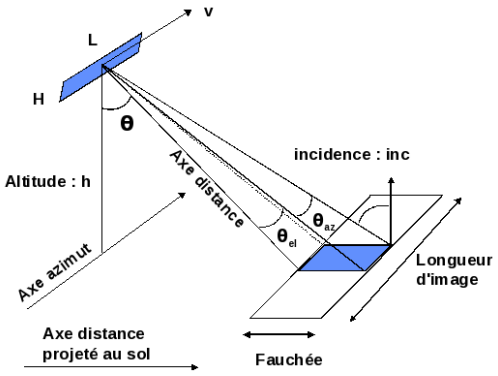


FIG. 2.2 – Système d'acquisition radar, ici à visée latérale [d'après le cours ENSEEIHT de Thierry Amiot disponible sur la plateforme Moodle]. Nous appelons l'axe en distance l'axe perpendiculaire au vecteur vitesse, l'axe azimut l'axe parallèle au vecteur vitesse du porteur et dirigé dans le même sens, et l'axe d'altitude l'axe dirigé vers le haut. L'antenne du radar concentre l'onde émise dans une zone donnée.

Dans le cas de satellites radars d'observation de la Terre – notre principal intérêt – les obstacles peuvent avoir plusieurs origines, selon la bande de fréquence choisie :

- obstacle de type interface air-surface, par exemple pour les plans d'eau ;
- obstacle sous une surface, pour des observations dans un milieu donné (type océan ou désert de sable) ;
- sommet végétal (canopée) ou obstacle présent dans le volume de la végétation (houppier).

Un gros avantage est la quasi-insensibilité aux conditions atmosphériques et à l'illumination solaire, ce qui permet, contrairement aux images optiques, de pouvoir faire des acquisitions de jour comme de nuit, et avec tout type de couverture nuageuse.

Le principe du radar est assez simple : l'émetteur lance des impulsions d'une certaine fréquence à intervalles réguliers, par exemple toutes les milli-secondes. L'antenne du radar, qui agit comme un projecteur, concentre l'émission dans une zone très étroite de l'espace (voir Fig. 2.2). C'est ainsi que sont illuminées les cibles situées dans le champ de l'antenne, et ce, d'autant plus faiblement qu'elles sont situées loin. La trace au sol du lobe d'antenne décrit une bande appelée *fauchée*. Les cibles réfléchissent les signaux reçus et l'antenne capte ces signaux réfléchis, appelés *échos*, avec un décalage d'autant plus grand que les cibles sont lointaines.

La résolution en distance R, aussi dénommée dans l'introduction de ce chapitre "résolution géométrique", correspond à la capacité du radar à distinguer deux points très proches. En théorie, un système radar devrait être capable de distinguer des cibles espacées d'un temps égal à une demi largeur d'impulsion :

$$R = \frac{cT}{2}, \tag{2.1}$$

où c est la célérité et T la durée de l'impulsion. Une impulsion la plus brève possible permet donc d'avoir la meilleure résolution possible. Mais pour avoir un signal correct, avec un rapport signal sur bruit raisonnable (pas trop petit), cette impulsion ne doit pas être trop courte. Un compromis est possible mais limitant. Il est cependant possible de recourir à une astuce du traitement du signal, qui consiste à moduler le signal émis via la technique de compression d'impulsion [Klauder 1960], ce qui permet d'augmenter la résolution en distance de la mesure ainsi que le rapport signal sur bruit. Pour réaliser la compression d'impulsion, un signal de type "chirp" est émis, signal pseudo-périodique modulé à la fois en amplitude et en fréquence sur une bande de fréquence B autour d'une fréquence porteuse. Le signal reçu par le radar est une copie retardée, atténuée et éventuellement déphasée (par effet Doppler) du signal émis. Pour détecter le signal reçu, un filtrage adapté est utilisé, équivalent à un calcul d'intercorrélation. Pour B bien choisi, le temps de l'impulsion sera de $1/B$ (choisi comme inférieur à T), et la résolution en distance sera de

$$R = \frac{c}{2B}. \tag{2.2}$$

La formation des images est la traduction en terme de pixels du signal reçu par l'antenne du radar. Le radar mesure l'*amplitude complexe* du champ électrique rétrodiffusé E_r par la surface illuminée S. On exprime l'image complexe de chaque pixel par $A \cdot \exp(i\Phi)$, où Φ est la *phase* et A l'*amplitude*. $A\cos(\phi)$ est mesurée par le canal en phase du récepteur, et $A\sin(\phi)$ par le canal en quadrature.

Cette amplitude complexe dépend d'un nombre de paramètres assez conséquent, à commencer par les propriétés physiques intrinsèques des objets observés telles que leur forme, ou encore leur(s) matériau(x) constitutif(s), qui influent sur la valeur de la permittivité diélectrique. Il est donc avantageux d'avoir quelques informations a priori sur l'observation pour traiter l'image, et éviter de potentielles indéterminations. L'amplitude dépend aussi de l'angle d'incidence du rayonnement, de la pente locale de la surface observée et de la longueur d'onde d'émission. Typiquement, si cette longueur d'onde est beaucoup plus petite que la taille des éléments de la zone observée (surface lisse à l'échelle de la longueur d'onde), comme par exemple dans le cas d'une étendue de sable, d'eau ou de glace, l'onde émise est alors totalement réfléchie. À l'inverse, si la longueur d'onde est beaucoup plus grande que la taille des éléments observés, les atomes de ceux-ci sont polarisés, c'est-à-dire que les charges négatives et positives dans les matériaux sont séparées. Ceci est décrit par le modèle de la diffusion de Rayleigh [Strutt 1899], qui prédit notamment le bleu du ciel et le rouge d'un coucher de soleil. Enfin, quand les deux longueurs sont comparables, il peut se produire des résonances entre les atomes de la cible et la réflexion se comporte selon la théorie de Mie [Stratton 1941], rendant le diagramme de ré-émission très variable.

2.1.1.1 Un radar spécifique : le radar à synthèse d'ouverture

Les images que nous avons traitées (voir chapitres ultérieurs) sont des images acquises par des capteurs COSMO-SkyMed (©ASI) et TerraSAR-X (©DLR),

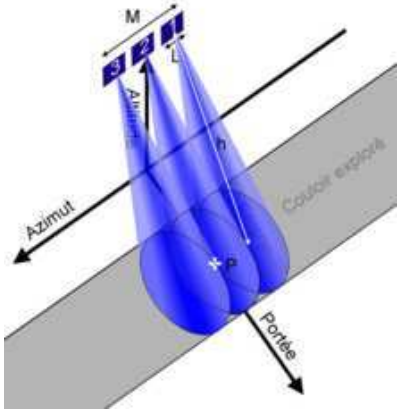


FIG. 2.3 – Schéma d'imageur RSO (©Wikipedia). Le satellite se déplace selon une trajectoire parallèle à l'axe azimutal, permettant ainsi à un point P d'être illuminé plusieurs fois par le faisceau du radar. La transformée de Fourier inverse de la sommation cohérente des ondes reçues pour un même point P affine les résultats. L'acquisition après post-traitement est mieux résolue grâce à la synthèse d'ouverture.

dont les capteurs radars sont à synthèse d'ouverture [Oliver 2004, Polidori 1998, Massonnet 2008]. Ce sont des radars à visée latérale, c'est-à-dire, fixés sur la face latérale d'un porteur (avion ou satellite). De ce fait, l'antenne embarquée émet perpendiculairement au vecteur vitesse. La synthèse d'ouverture est une technique qui s'applique spécifiquement à ce type de radars afin d'améliorer la résolution géométrique en azimut, axe horizontal projeté à la surface de la Terre et parallèle à la trajectoire du satellite (Fig. 2.3). Concrètement, une image ayant une meilleure résolution est obtenue en simulant une antenne large tout en conservant une taille physique d'antenne raisonnable. Ce traitement de synthèse d'ouverture, qui repose sur une transformée de Fourier, est appliqué en fin de chaîne (post-traitement) au signal brut reçu par le radar. En utilisant une antenne placée sur un porteur en mouvement, la sommation cohérente du signal reçu correspondant à un même point de l'espace est réalisée, sur plusieurs instants successifs, en s'arrangeant pour que l'objet reste dans le lobe principal de l'antenne sur cette durée. Il repose donc sur le fait que le radar se déplace physiquement avec le porteur, ce qui implique que le même point est illuminé plusieurs fois (Fig. 2.3). La transformée de Fourier inverse de cette sommation cohérente aboutit au calcul d'un nouveau point, virtuellement acquis par une grande antenne, donnant ainsi une image finale de meilleure résolution. Le résultat est dépendant de deux hypothèses, assez faciles à obtenir pour des porteurs de type satellite. La première est que le vol du porteur est parfaitement stable : vitesse constante, altitude constante, etc. La deuxième est une parfaite stabilité des oscillateurs de démodulation du signal (nous avons vu que le signal radar émis est un signal modulé en fréquence), afin d'assurer une relation de phase correcte entre tous les signaux reçus pendant le passage sur une zone.

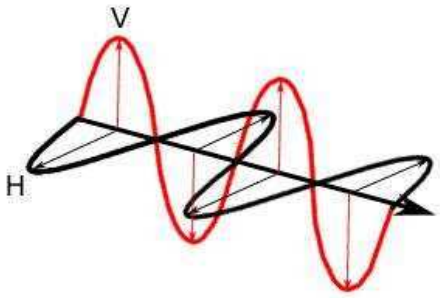


FIG. 2.4 – Polarisations verticale et horizontale des ondes émises et reçues par l'antenne du radar.

Malgré la présence du bruit de chatoiement (sous-partie 2.1.1.4), qui dégrade visuellement les acquisitions RSO, ces systèmes radars possèdent de nombreuses qualités, notamment liées au fait que les acquisitions se font en bandes hyperfréquences (ou de manière équivalente micro-ondes). Dans certaines applications, la capacité de pénétration des ondes à travers des arbres, des murs, ou des surfaces de sable par exemple, permet un élargissement du champ d'observations. En outre, la polarimétrie apporte de nombreuses informations complémentaires, comme nous allons voir dans la sous-partie suivante.

2.1.1.2 Polarisation

Les champs électriques incident E_i et réfléchi E_r sont polarisés, c'est-à-dire que les ondes se propagent dans des directions bien définies. Les polarisations sont généralement choisies comme verticales et horizontales, et toujours perpendiculaires à la direction de propagation de l'onde (Fig. 2.4). On parle alors de radar *polarimétrique* [Lee 2009]. En pratique, jusqu'à quatre images simultanées peuvent être générées suivant les différentes combinaisons de polarisations (horizontale et verticale) en émission et en réception.

La polarisation traduit la structure ou le volume d'une scène et permet de mettre en exergue les différentes propriétés polarisantes des cibles observées. Il est ainsi possible de rehausser les contrastes de certains détails non visibles sur des images classiques (non polarimétriques), ou de déduire des propriétés de la cible telles que le type de végétation. Typiquement, l'observation d'une structure horizontale est plus adéquate avec une émission/réception HH, c'est-à-dire, champs E_i et E_r horizontaux. Par exemple, elle permet de distinguer l'eau de la glace. Pour les structures verticales, de type arbres ou cultures hautes (blé...), une polarisation VV est plus adaptée. La polarisation HH passe au travers de telles cultures "verticales", et permet par exemple de mesurer la teneur en eau du sol. Les polarisations HV ou VH mettent en évidence la rétrodiffusion de volume ou des surfaces rugueuses, utile dans des domaines tels que la culture, la géologie, ou la glaciologie.

2.1.1.3 Autres types de radars

Nous considérons dans cette thèse uniquement les capteurs radars de type polarimétrique, car ils sont bien adaptés à l'utilisation que nous souhaitons en faire, à savoir établir des cartographies de zones. Cependant, d'autres types de données radars peuvent être générées pour d'autres types d'applications, comme l'établissement de modèles numériques de terrain [Noferini 2007], la cartographie de déplacements de plaques tectoniques [Madsen 2001] ou bien la modélisation de prévisions météorologiques [Meischner 2010].

Le radar interférométrique (InSAR) estime la différence de phase pour une même cible observée par deux radars à synthèse d'ouverture ou bien un même radar utilisé à des instants différents [Ketelaar 2010]. C'est une technologie très utilisée dans la construction de modèles d'élévation de terrain. Par exemple, en zone urbaine, les modèles d'élévation sont utilisés en vue de déterminer des hauteurs des bâtiments [Soergel 2010]. La tomographie RSO [Reigber 2000], permet de gagner une précision supplémentaire par rapport à l'InSAR car elle peut être utilisée pour la séparation et la localisation directe de diffuseurs au sein d'une même cellule de résolution. De ce fait, il est possible de gagner une précision au niveau de la résolution en altitude. La tomographie RSO équivaut donc à une modélisation $3D$ effectuée grâce à une synthèse d'ouverture en direction de l'altitude.

Les impulsions émises par un radar peuvent être modifiées (durée d'émission, temps entre les impulsions) afin de donner naissance à des données nécessaires pour certaines applications spécifiques telles que la radiométrie ou l'altimétrie radar. Ces diverses applications ont largement été exploitées dans le cadre de la mission ERS (European Remote Sensing Satellite) [Prata 1990, Eymard 1994].

2.1.1.4 Bruit de chatoiement

En pratique, les images radars sont affectées par un bruit à texture, dite "poivre et sel", omniprésent sur l'acquisition, et qui, de fait, dégrade la qualité de celle-ci. Il s'agit de bruit produit par un système cohérent que l'on appelle phénomène de *chatoiement* et qui dépend des paramètres du système radar et de la nature de la surface imagée. Le modèle du chatoiement classique présume de la présence d'un grand nombre de réflecteurs ponctuels indépendants répartis aléatoirement et ayant des caractéristiques de diffusion semblables à l'intérieur de la cellule de résolution. Lorsqu'illuminée par le radar, chaque cible rétrodiffuse une partie de l'énergie qui, en plus des changements de phase et de puissance, est additionnée de façon cohérente pour tous les diffuseurs. Cette variation ou incertitude statistique (la variance) est associée à la brillance de chaque pixel de l'image radar. Par exemple, sans l'effet de chatoiement une cible homogène (comme une grande étendue de gazon) apparaîtrait en tons plus clairs. Mais la réflexion provenant de chaque brin d'herbe à l'intérieur de chaque cellule de résolution produit des pixels plus clairs et d'autres plus sombres que la moyenne, de sorte que le champ apparaît tacheté. En effet, selon leur emplacement, les contributions s'additionnent majoritairement de manière constructive (points brillants) ou destructive (points sombres).

Les procédures d'analyse et de traitements d'images classiques considèrent habituellement le chatoiement comme un élément indésirable contenant peu d'information. Deux techniques permettent de réduire le chatoiement : le traitement multi-vues (plus connu sous le terme anglo-saxon *multi-look*) et le filtrage spatial.

Le traitement multi-vues consiste en un découpage du spectre azimuth (ou distance) en plusieurs vues, qui sont ensuite sommées de manière incohérente. Dans certains cas particuliers, ce traitement peut aussi consister en l'acquisition de plusieurs images d'une même scène [Bruniquel 1997, Xia 1997]. Comme le suggère le nom de ce traitement, chaque visée produit sa propre image de la scène illuminée. Chacune de ces images est aussi sujette au phénomène de chatoiement, mais en faisant la moyenne de toutes les images pour obtenir une image finale, il est possible de réduire globalement les effets du bruit. Ce traitement est appliqué par défaut à certaines des acquisitions sur lesquelles nous avons travaillé (voir Annexe A).

La réduction du chatoiement par filtrage adaptatif [Lee 1994] est un post-traitement qui consiste à appliquer une fenêtre glissante de quelques pixels à chaque pixel de l'image, et à moyenner les pixels de cette fenêtre suivant une certaine pondération. L'effet de lissage obtenu réduit visuellement le chatoiement. D'autres méthodes de filtrage plus complexes mais plus adaptées (et donc moins lissantes) peuvent aussi être utilisées [Gagnon 1997, Foucher 2001, Achim 2006, Xia 1997].

Cependant, de tels traitements ont tendance à dégrader la résolution géométrique à cause du lissage, ou bien à introduire des artéfacts. Ainsi, si une image à basse résolution est souhaitée, alors la réduction du chatoiement peut être tout à fait indiquée. En revanche, si l'application requiert des détails fins et une haute résolution, comme dans nos tests où l'objectif est de classifier des images à la meilleure résolution possible, la réduction drastique du bruit de chatoiement n'est pas très appropriée car elle diminue la résolution. En outre, le bruit de chatoiement est corrélé au signal (bruit multiplicatif), et donc à haute résolution, il peut contenir des informations sur la scène observée qu'il peut être utile de conserver [Dell'Acqua 2003, Oliver 2004].

2.1.1.5 Autres spécificités

D'autres spécificités compliquent d'avantage l'analyse des images radars, et nécessitent parfois un post-traitement. Nous proposons ici d'en établir une liste non exhaustive :

- **Distorsion oblique :** elle est due au fait que les acquisitions sont faites selon un certain angle. Les éléments dans le plan de fauchée (Fig. 2.2) le plus proche dit portée proximale (*near range*) sont comprimés par rapport à ceux qui sont présents dans la portée distale (*far range*) à cause de la non-linéarité de l'échelle des distances. Ceci requiert une conversion de l'image en distance-sol, conversion qui est directement assimilable à une projection. Cette conversion peut être faite par le processeur radar lors du traitement des données, ou par transformation géométrique de l'image. Dans la plupart des cas, cette conversion ne sera qu'une estimation, à cause des complications introduites par la variation du relief et de la topographie.

- **Distorsion géométrique :** un autre type de distorsion est le phénomène de repliement lié au fait que les objets en hauteur répondent avant des objets situés au sol. Par conséquent, leur réponse est mélangée avec celle d'autres points du sol. Cela peut être contraignant, en particulier en zones urbaines, et tend à générer des confusions dans la lecture de l'acquisition.
- **Variation en intensité (teinte) à travers l'image :** une antenne radar transmet plus d'énergie dans la partie centrale du couloir balayé que dans la portée proximale ou la portée distale. Cet effet est connu sous le nom de *diagramme d'antenne* et donne une rétrodiffusion plus forte en provenance de la partie centrale du couloir que des côtés. Ces effets se combinent pour produire une image dont l'intensité (teinte) varie de la portée proximale à la portée distale. Un procédé appelé correction de diagramme d'antenne peut s'appliquer pour produire une tonalité uniforme moyenne afin de faciliter l'interprétation visuelle. En outre, les systèmes radars peuvent différencier jusqu'à plus de 10000 niveaux de gris. Puisque l'oeil humain ne peut percevoir simultanément qu'environ 40 niveaux d'intensité [Gomarasca 2009], il y a trop d'information pour l'interprétation visuelle. Ainsi donc, la plupart des données radars ont une résolution radiométrique de 16 bits (65536 niveaux d'intensité) que l'on réduit à 8 bits (256 niveaux) pour l'interprétation visuelle et l'analyse par ordinateur. Nous avons eu recours à un tel procédé pour les images traitées par la suite. Cela est, certes, discutable dans le sens où une telle normalisation crée une distorsion non linéaire (au mieux un seuil) et une quantification des distributions initiales qui peuvent dégrader les informations présentes dans les images. Cependant, nous avons étudié les histogrammes des niveaux de gris de diverses images RSO 16 bits, et avons pu constater que la majorité (99.9%) des informations relatives aux images est concentrée sur les 256 premiers niveaux de gris. Nous considérons dès lors que la portion d'information perdue par la considération d'une image à 8 bits est négligeable par rapport au temps de calcul qui sera 2^8 plus conséquent si l'image considérée en entrée est de 16 bits.
- **Mauvais étalonnage du radar :** l'étalonnage est un procédé qui assure que le système radar et le signal qu'il mesure soient aussi consistants et précis que possible [Freeman 1992]. La plupart des images radars requièrent un étalonnage relatif avant d'être analysées. Cet étalonnage corrige les variations connues de l'antenne et la réponse du système, et assure que des mesures uniformes et répétitives sont possibles. Cette opération permet d'effectuer en toute confiance des comparaisons relatives entre la réponse des éléments dans une même image, et entre d'autres images. Cependant, si des mesures quantitatives précises sont requises, représentant l'énergie vraiment retournée par les différentes structures ou cibles pour des fins de comparaison, alors un étalonnage absolu est nécessaire. L'étalonnage absolu est un procédé beaucoup plus laborieux que l'étalonnage relatif. Il tente de relier la magnitude du signal enregistré avec la vraie valeur de l'énergie rétrodiffusée à partir de chaque cellule de résolution. Pour ce faire, des mesures détaillées des propriétés du système

radar sont requises, ainsi que des mesures quantitatives au sol faites à l'aide de cibles calibrées. Aussi, des appareils appelés transpondeurs peuvent être placés au sol avant l'acquisition des données pour calibrer l'image. Ces appareils reçoivent le signal radar, l'amplifient et le retournent vers le radar avec une intensité déterminée. En connaissant l'intensité de ce signal dans l'image, la réponse des autres structures peut être trouvée par extrapolation.

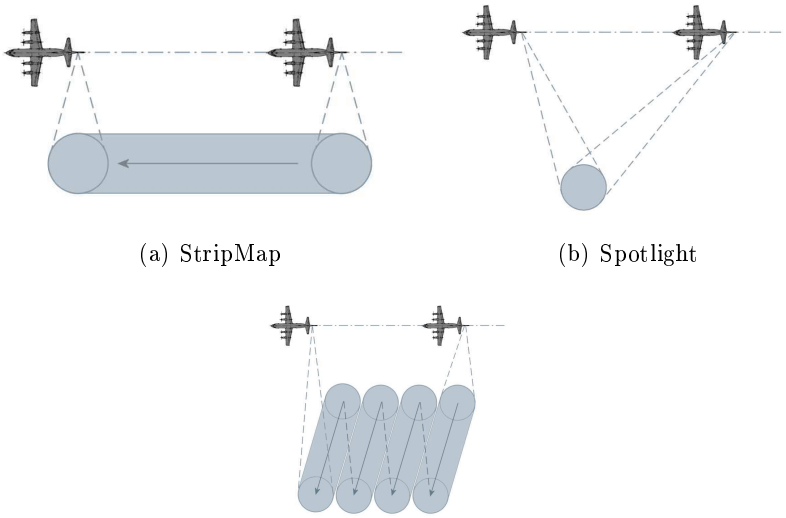
- **Erreur de démodulation :** les imageurs radars sont aussi affectés par des erreurs de démodulation temporelles au niveau du récepteur radar, dues à des retards indéterminés dans le signal reçu. Cela génère une image radar affectée par des erreurs multiplicatives au niveau de la phase [Morrison 2006].
- **Présence d'échos parasites :** des échos parasites, aussi appelés fantômes, ou "clutter", peuvent apparaître à cause des réflexions multiples. Typiquement, ils proviennent du sol, de l'eau, des turbulences atmosphériques, et peuvent sérieusement affecter les performances des systèmes radars, notamment en occultant des point-cibles. Ces effets peuvent être évités grâce à des polarisations bien choisies [Unal 2009], ou bien par filtrage numérique [Banjanin 1991]. Il peut être intéressant de supprimer les échos parasites naturels tout en conservant les échos parasites relatifs aux constructions de l'humain, qui peuvent contenir des informations intéressantes (par exemple, les échos ponctuels provenant des immeubles). Une telle application est explorée dans [Fogler 1994] en projetant les images dans le domaine log-magnitude.
- **Mouvement de cibles :** le mouvement des cibles, telles que les voitures ou les bateaux, peuvent modifier les observations [Raney 1971].
- **Ombres :** le phénomène d'ombres peut se comprendre facilement puisqu'il apparaît aussi dans le cas courant de l'imagerie optique. Il dépend notamment de l'angle d'illumination de l'antenne du radar. En général, les régions ombragées apparaissent plus sombres sur une image, car il n'y a pas d'énergie disponible pour la rétrodiffusion, et ne contiennent que très peu d'information sur la zone observée. Cependant, ces ombres peuvent être exploitées à bon escient, et s'avérer très utiles dans les zones urbaines, soit pour aider à la détermination de la hauteur d'un bâtiment, soit pour détecter les bâtiments par utilisation d'images radars interférométriques [Tison 2003b].

De nombreux autres éléments peuvent affecter les données à pleine résolution (compression, bruit de capteur...), mais nous avons listé ici ceux qui, à notre sens, sont les plus importants étant donné le problème considéré dans ce manuscrit de thèse.

2.1.1.6 Les images RSO en pratique

Dans cette brève sous-partie, nous allons décrire les allures des acquisitions au vu des diverses zones observées dans les images RSO que nous avons traitées.

La zone la plus importante pour nous est la zone urbaine, sur laquelle nous focalisons nos expérimentations. Les observations relatives à ces zones sont assimilables à des points brillants générés par des éléments comme des bâtiments qui agissent



(a) StripMap (b) Spotlight

(c) Balayage

FIG. 2.5 – Modes d'acquisition d'un radar RSO, ici illustré comme radar aéroporté [*www.radartutorial.eu*].

comme des coins réflecteurs. Ainsi, l'allure générale d'une zone urbaine est un ensemble de points très lumineux assez irréguliers, car entrecoupés, en général, de zones d'ombres (cf. 2.1.1.5). On parlera par la suite d'*hétérogénéité* de ces zones.

Dans un contexte de catastrophes naturelles, les images RSO ont leur importance [Boni 2007]. En effet, dans le cas d'une inondation, ou même plus généralement de présence d'eau, le coefficient de rétrodiffusion, en général, diminue, ce qui se traduit sur l'image par des zones sombres. Ainsi, il est plus aisé de relever des zones humides grâce à de telles caractéristiques physiques, plutôt que sur des images optiques, pour lesquelles nous pouvons avoir de sérieuses indéterminations. En effet, visuellement, dans les images optiques, la couleur de l'eau lors d'inondations peut varier entre plusieurs bleus, du vert ou du marron si celle-ci est boueuse. Dans le cas de feux de forêts, le coefficient de rétrodiffusion augmente, générant des zones plus claires que la moyenne.

Enfin, les zones de végétation, que nous avons précédemment évoquées en illustration des effets du chatoiement (voir 2.1.1.4), sont des zones tachetées dans des niveaux de gris intermédiaires entre les zones d'eau sombres et les zones urbaines assez claires.

2.1.1.7 Les modes d'acquisition

Trois modes d'acquisition sont généralement possibles (Fig. 2.5) par les radars à synthèse d'ouverture. Jusqu'à présent, nous avons parlé du mode en bande, aussi connu par son nom anglophone de *Strip mapping*, qui est le mode classique d'utilisation : le radar est pointé perpendiculairement à la trajectoire du porteur, avion

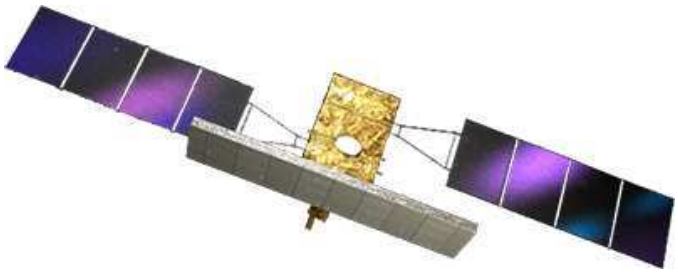


FIG. 2.6 – Vue d'artiste d'un satellite COSMO-SkyMed [*www.mitsubishicorp.com*].

ou satellite, et à un angle fixe d'incidence vers le sol. Il balaie ainsi un couloir ayant comme largeur celle du faisceau et comme longueur le trajet effectué par le porteur.

Le mode de saisie hyperfine, désigné par le terme anglais de *Spotlight*, permet d'obtenir une haute résolution (dans la direction azimutale) comme son nom l'indique. Il s'agit d'orienter le radar toujours vers la même zone lors du déplacement du porteur. Dans ce mode, une antenne réseau à commande de phase dont le faisceau est orienté grâce à un logiciel est utilisée. Cela permet d'obtenir un plus grand nombre de balayages de la zone d'intérêt que dans le mode classique et donc plus d'informations, ce qui a pour effet d'augmenter la longueur de l'ouverture synthétique. Le tout se fait cependant aux dépends de la couverture spatiale [Wahl 1996].

Le dernier mode est le balayage, connu aussi sous le nom de *ScanSAR*, pour lequel le faisceau radar effectue un balayage angulaire entre le point sous le porteur (le nadir) et un angle donné d'incidence au sol. Comme le porteur, avion ou satellite, se déplace, le couloir sondé prendra la forme d'une série de bande en zigzags si l'angle varie linéairement du nadir vers l'extérieur, puis l'inverse.

2.1.2 Le satellite COSMO-SkyMed

COSMO-SkyMed (COnstellation of small Satellites for Mediterranean basin Observation) (CSK) est un ensemble de satellites d'observation de la Terre italien, et est exploité conjointement pour des applications civiles et pour la défense. Cette constellation est formée de quatre satellites (Fig. 2.6) à orbite basse, eux-mêmes baptisés CSK, chacun étant équipé d'un système RSO multimode opérant en bande X (9.6 gigahertz). Une telle combinaison permet une couverture globale de la Terre, et un temps de revisite faible. Ce dernier est de 12 heures, ce qui signifie qu'une même zone peut être balayée deux fois dans la même journée. Ceci est un avantage considérable dans le cadre des risques naturels, pour lesquels une observation multitemporelle rapprochée est nécessaire, typiquement pour la coordination des secours. Les deux premiers satellites ont été lancés en 2007, le troisième en 2008 et le dernier module en 2010.

Comme illustré dans la Figure 2.7, les acquisitions peuvent se faire selon trois modes (voir la sous-partie 2.1.1.7), selon la résolution souhaitée :

– Mode Spotlight pour des résolutions de l'ordre de 1 mètre sur des petites zones

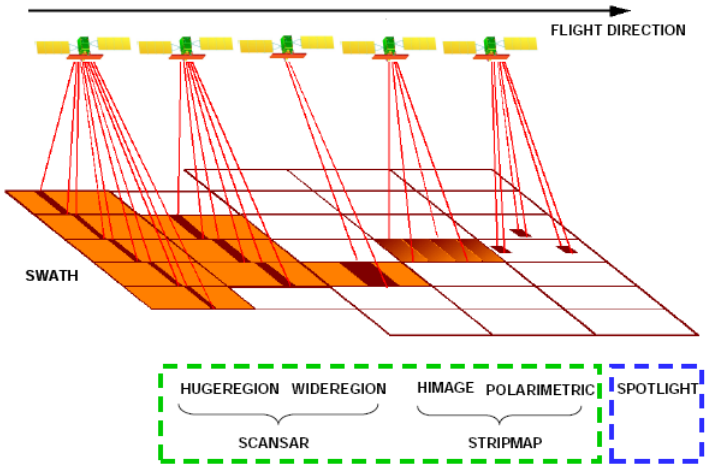


FIG. 2.7 – Les trois modes d'acquisition du capteur COSMO-SkyMed [*www.eurimage.com/products/pdf/csk−user_guide.pdf*], du moins résolu (à gauche) au mieux résolu (à droite).

(10×10 km).

- Modes StripMap, pour des résolutions de plusieurs mètres permettant l'acquisition de zones de plusieurs dizaines de kilomètres (autour de 30×30 km) : un mode HImage à polarisation simple de résolution 3 mètres, et un mode Ping-Pong de résolution 15 mètres, qui permet d'acquérir des images à polarisation multiple (combinaisons HH, VV, HV et VH, voir la sous-partie 2.1.1.2).
- Modes ScanSAR pour des résolutions moyennes à faibles (30 mètres à 100 mètres) permettant l'observation de zones très étendues (zones de plus de $100\times$ 100 km). Ils sont désignés par WideRegion et HugeRegion. Nous n'entrerons pas dans les détails concernant ce mode d'acquisition étant donné que les résolutions sont beaucoup trop faibles, et que nous cherchons à travailler avec des images à haute résolution.

Le mode StripMap HImage, étant le mode par défaut, nous avons travaillé essentiellement sur ce type d'acquisitions. Dans les sous-parties suivantes, ce mode sera simplement désigné comme *mode StripMap*. Le cas multipolarisation sera désigné comme *mode PingPong*.

Outre les acquisitions polarimétriques, qui sont parfois utilisées dans cette thèse, la constellation permet l'acquisition de données interférométriques (sous-partie 2.1.1.3), appelé mode "Tandem".

La plupart des images-tests présentées dans les chapitre suivants sont des images obtenues grâce à l'agence spatiale italienne (ASI) dans le cadre du projet "Développement et validation de méthodes d'analyse d'images multitemporelles pour la surveillance multirisques de structures critiques et d'infrastructures (2010-2012)". Ainsi, nous avons pu obtenir un grand nombre de jeux de données qui ont permis de valider nos méthodes.

De plus amples informations concernant la mission COSMO-SkyMed sont disponibles sur le site web *www.cosmo-skymed.it*.

2.1.3 Le satellite TerraSAR-X

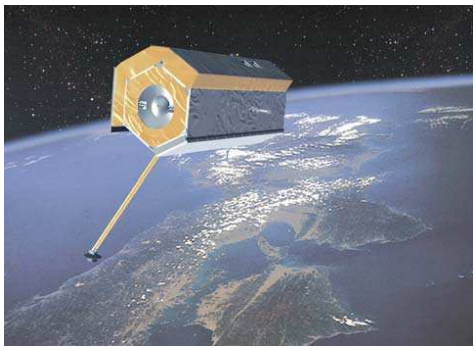


FIG. 2.8 – Vue d'artiste du satellite TerraSAR-X (©EADS-Astrium).

TerraSAR-X (TSX) est une famille de satellites radars d'observation de la Terre allemands développée dans le cadre d'un partenariat public-privé entre l'agence spatiale allemande (DLR) et EADS Astrium Allemagne. Le premier satellite TerraSAR-X a été lancé en 2007. Il a été rejoint en 2010 par un satellite jumeau TanDEM-X (TerraSAR-X add-on for Digital Elevation Measurements) en vue d'obtenir des images interférométriques. Les caractéristiques du satellite TSX (Fig. 2.8) sont similaires à celles du satellite CSK. En particulier, comme son nom l'indique, la bande d'acquisition est la bande X. Quatre modes d'acquisition sont possibles :

- Mode Spotlight à haute résolution, pour des résolutions de $1,1$ mètre en polarisation simple et $2,2$ mètres en multipolarisation, sur des petites zones (5×10 km).
- Mode Spotlight pour des résolutions de $1,7$ mètre en polarisation simple et $3,4$ mètres en multipolarisation sur des petites zones (10×10 km).
- Mode StripMap, pour des résolutions de plusieurs mètres (autour de 3 mètres en polarisation simple, 6 mètres en mode PingPong) permettant l'acquisition de zones de plusieurs dizaines de kilomètres (autour de 30×30 km).
- Mode ScanSAR pour des résolutions de $18,5$ mètres, permettant l'observation de zones très étendues (zones de plus de 100×150 km). Un fois encore, nous ne détaillerons pas ce mode d'acquisition étant donné que les résolutions sont beaucoup trop petites, et que nous cherchons à travailler avec des images à haute résolution.

Le mode par défaut est le mode StripMap. Une des différences notables avec le satellite CSK est la possibilité d'acquérir des images SpotLight à polarisation multiple, là où CSK ne se restreint qu'à une seule polarisation.

Etant donné que les deux satellites (TSX et TanDEM-X) sont physiquement proches, et non dispersés comme dans la constellation CSK, afin de permettre des

acquisitions les plus simultanées possibles (condition sine qua non pour établir les modèles d'élévation avec les données interférométriques), le temps de revisite est plus long, à savoir de 1 à 3 jours pour des zones spécifiques, jusqu'à 11 jours pour des zones situées à l'équateur.

2.2 L'imagerie optique

Un système d'acquisition optique est un système dit passif, car il capte les ondes émises naturellement par le milieu observé. Ce type d'images est moins affecté par le bruit que les images radars car l'illumination incohérente ne génère pas de bruit de chatoiement. Les bruits les plus caractéristiques sont le bruit thermique, le bruit de capteur et le bruit atmosphérique, qui influent sur les acquisitions. Ces deux derniers peuvent être compensés, et ne dégradent que peu l'image. En revanche, les conditions météorologiques ont un fort impact, ce qui donne en général des images partiellement cachées par des nuages. En outre, les prises de vue de nuit ne sont pas exploitables en général.

Le fonctionnement du système d'acquisition est fondé sur un télescope embarqué à bord d'un satellite. L'objectif du télescope est composé de deux miroirs, un miroir dit primaire et un miroir secondaire. La lumière émise par la zone observée atteint le miroir primaire, qui focalise celle-ci en un point appelé foyer image. Ce faisceau convergent peut être renvoyé vers un oculaire à l'aide du miroir secondaire. L'oculaire est la partie de l'instrument qui permet d'agrandir l'image produite par l'objectif au niveau du foyer-image, en somme il s'agit d'une loupe. Nous ne détaillerons pas davantage les principes des télescopes spatiaux, ceux-ci n'étant pas un point clé dans cette thèse. Nous proposons de nous focaliser sur les satellites Pléiades et GeoEye, qui sont deux satellites optiques haute-résolution spécifiques dont les acquisitions nous ont permis d'établir des tests de classification sur des images multicapteur. D'autres missions de ce type sont en cours, notamment, QuickBird, IKONOS et WorldView, que nous ne présentons pas dans cette thèse.

2.2.1 Le satellite GeoEye

Le satellite que nous dénommons de manière abusive GeoEye est le satellite américain GeoEye-1, qui a été lancé en 2008. Il est notamment un des fournisseurs d'images pour les sites Google Maps et Google Earth. Son homologue GeoEye-2 est prévu pour 2013. Ce satellite possède deux modes, un mode *panchromatique* (niveaux de gris) pour lequel la résolution est de 41 centimètres et un mode *multispectral* pour lequel la résolution spectrale est meilleure au détriment d'une moins bonne résolution (1,65 mètre). En pratique, le gouvernement américain impose que les images panchromatiques soient ré-échantillonnées à 50 centimètres en ayant recours à une convolution bi-cubique pour la vente en dehors du territoire américain. Les différentes bandes multispectrales sont les bandes bleue, verte, rouge et proche infra-rouge (PIR). Le temps de revisite est d'environ 3 jours.

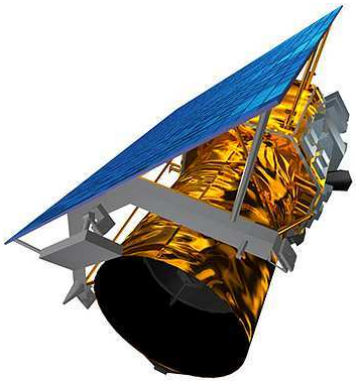


FIG. 2.9 – Vue d'artiste du satellite GeoEye [*www.satimagingcorp.com*].

2.2.2 Le satellite Pléiades

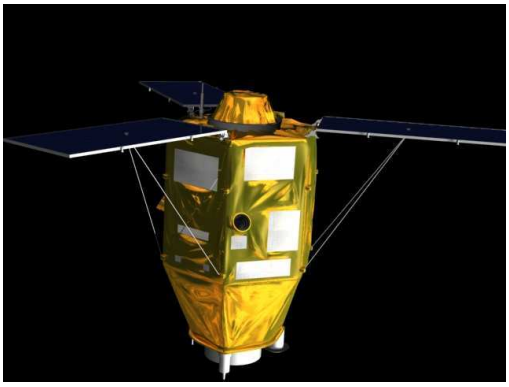


FIG. 2.10 – Vue d'artiste du satellite Pléiades (©CNES).

Pléiades est un couple de deux satellites optiques d'observation de la Terre développé notamment par l'agence spatiale française (CNES) dans un but à la fois civil et militaire (dual). Pléiades a été mis au point conjointement à la constellation COSMO-SkyMed sous l'égide du programme ORFEO (Optical and Radar Federated Earth Observation) en vue d'une coopération spatiale franco-italienne. Pléiades-1 a été lancé fin 2011, et Pléiades-2 sera lancé d'ici 2013. Par abus de langage, nous parlerons du module satellitaire Pléiades-1 comme étant le satellite Pléiades. Tout comme GeoEye, ce satellite possède deux modes, un mode panchromatique et un mode multispectral. Ces deux modes sont possibles grâce à deux détecteurs embarqués, l'un permettant la détection panchromatique et l'autre plus petit en terme de nombre de pixels, mais plus évolué (assemblage de lignes photosensibles), qui permet d'acquérir quatre bandes spectrales (bleu, vert, rouge, proche infrarouge). L'objectif du programme Pléiades est de fournir une nouvelle génération d'images de résolution 70 centimètres en mode panchromatique et de 2,8 m en mode multis-

pectral. En outre, après traitement par des stations au sol de l'agence spatiale, la taille des pixels est de 50 centimètres en mode panchromatique et 2 mètres en mode multispectral. Le champ de vue en visée verticale peut s'étendre jusqu'à 20 km. En combinant mathématiquement des mosaïques d'images, il est possible d'élargir le champ de vue à 100×100 km. Selon l'angle de la prise de vue, le temps de revisite est inférieur à une journée, mais peut s'étendre à plusieurs jours.

2.3 Conclusion

Nous avons présenté dans ce chapitre la technologie RSO, qui est une technologie radar évoluée, munie à la fois d'une compression d'impulsion et d'une synthèse d'ouverture qui visent à apporter une bonne résolution en distance et en azimut tout en minimisant le bruit thermique inhérent à ce type d'images. Ces systèmes permettent d'obtenir des images d'une zone quelles que soient l'illumination solaire et les conditions météorologiques. Cependant, le bruit de chatoiement est omniprésent, en particulier dans les zones à fort coefficient de rétrodiffusion, et rend le traitement de ces images plus ardu. Ce chatoiement est un phénomène physique qui ne peut être réduit sans dégrader la résolution initiale de l'image. Nous avons, ensuite, brièvement évoqué les imageurs optiques, dont les acquisitions sont moins bruitées et mieux résolues, mais beaucoup plus sensibles aux conditions extérieures.

Dans les chapitres suivants (Chaps. 4 et 5), nous cherchons à établir des cartes de classification d'images RSO et également sur des images multicapteur. Les premiers tests ont été réalisés sur des images RSO à polarisation simple, avec une possible extension à des images multipolarisées. Puis, nous avons cherché à étendre notre contribution à l'étude de données multirésolution en ayant recours à des modèles markoviens hiérarchiques. Enfin, reflétant l'intérêt de programmes tels que le programme ORFEO, nous avons pu souligner une certaine complémentarité entre les acquisitions optiques et radar. Ainsi, il est important, dans un contexte de classification, de pouvoir trouver des méthodes qui sont suffisamment générales pour traiter indépendamment ou de manière simultanée ces types d'acquisition. L'exploitation de la complémentarité peut avoir un impact positif sur les résultats finaux attendus. Nous avons donc pour but final de cette thèse la mise en place de méthodes de classification qui se veulent relativement générales pour permettre de traiter un panel le plus large possible d'acquisitions satellitaires réelles.

Chapitre 3

Modélisation statistique des images

Sommaire

Le chapitre précédent a démontré que les données de chaque pixel d'une image numérique sont physiquement assimilables à une valeur d'intensité, valeur notamment utilisée afin de connaître les propriétés de la scène observée. Il en résulte un lien direct entre la connaissance du monde (structure de l'observation elle-même) et une donnée physique (intensité au niveau d'un capteur), similaire à la notion même de classification. Ce lien est établi dans un contexte supervisé, donc avec apprentissage, afin de garantir la précision des résultats. Tout au long de ce manuscrit, les méthodes de classification présentées – objectif même de la thèse – sont essentiellement fondées sur la théorie bayésienne [Barnard 1958]. Nous allons donc présenter, dans ce chapitre, l'estimation fondamentale commune à toutes nos méthodes : l'estimation de la vraisemblance. Nous nous plaçons dans un formalisme purement statistique des niveaux de gris de(s) image(s) traitée(s). D'une manière générale, la vraisemblance est la probabilité d'une observation étant donné son appartenance à une classe précise (étiquette) notée m et comprise dans l'ensemble $[1; M]$. L'ensemble des observations est noté Y et l'ensemble des étiquettes X, nous garderons ces mêmes notations tout au long du manuscrit. Cette vraisemblance est calculée à partir d'une *vérité de terrain*, carte de référence dans laquelle sont dénotées manuellement les diverses classes, d'où un contexte entièrement supervisé.

Dans le cas où les observations sont contenues dans une seule bande, la vraisemblance s'exprime comme étant la modélisation des statistiques de niveaux de gris de cette image pour chacune des classes considérées. Typiquement, il peut s'agir, par exemple, d'une image d'amplitude RSO monorésolution et monopolarisée, puisque nous rappelons que nous traitons uniquement de l'amplitude de l'image RSO et non de son signal complexe. Cela détermine alors ce que nous nommons la *probabilité marginale*. La détermination de telles probabilités est traitée dans la partie 3.1.

Cependant, les observations peuvent former un vecteur. Il peut s'agir d'un vecteur-image d'acquisitions optiques de type multibandes (RVB) ou encore de données RSO acquises selon plusieurs polarisations. Il peut aussi s'agir d'une combinaison d'une image RSO avec un attribut de texture. En effet, comme nous le verrons ultérieurement (Chap. 4), une image RSO à simple polarisation contient une information assez limitée qui peut donner des résultats de classification légèrement erronés. Ainsi, l'introduction d'une information supplémentaire, typiquement sous la forme d'une texture, peut améliorer la classification, comme nous le verrons dans le chapitre 4. Ces attributs de texture sont étudiés dans la partie 3.2.

Les différentes images en entrée sont qualifiées de *bandes*. Lorsque les observations forment un vecteur de bandes, nous ne cherchons alors plus à estimer une probabilité marginale, mais une vraisemblance sous la forme d'une probabilité jointe. Nous proposons d'estimer ces probabilités jointes en ayant recours au modèle mathématique des copules [Nelsen 2006] : la probabilité jointe est ainsi modélisée par le produit des densités de probabilités marginales de chacune des bandes multipliées par un terme de densité – densité de copules – qui modélise la dépendance entre les différentes bandes. Cette théorie des copules est explicitée dans la partie 3.3.

Nous orientons ce chapitre dans un but final de classification (voir chapitres 4 et 5). Cependant, cette modélisation peut être également utilisée dans d'autres

applications comme, par exemple, la détection de changements [Cianci 2012, Serpico 2012].

3.1 Modélisation statistique des probabilités marginales

Les probabilités marginales peuvent être estimées de diverses manières. Une première possibilité est une approche paramétrique, pour laquelle l'estimation de la vraisemblance se résume à une estimation de paramètres, sous hypothèse de la connaissance de la loi de probabilité. Par opposition, les probabilités marginales peuvent être obtenues via une approche non paramétrique. Parmi les modèles non-paramétriques existants, nous pouvons citer les estimateurs de fenêtres de Parzen standards [Parzen 1962, Duda 2000], le recours aux réseaux de neurones artificiels [Bishop 1996] ou bien les machines à vecteurs de support, également appelées séparateurs à vastes marges [Mantero 2003]. Nous avons opté pour un modèle paramétrique au détriment de ces modèles non paramétriques, car celui-ci qui est moins général et plus spécifique que, par exemple, un modèle de Parzen. En outre, le modèle paramétrique modélise de manière plus optimale les vraisemblances recherchées en utilisant des modèles de probabilité adaptés.

Dans un contexte d'imagerie essentiellement satellitaire, nos acquisitions sont de deux types bien définis : d'un côté, les images optiques, de l'autre, les images radars. Dans cette partie, nous traitons séparément les deux types d'acquisition, et pour chaque type, nous traitons indépendamment les différents canaux (i.e. différentes bandes), qui peuvent être les différentes bandes colorimétriques de l'image optique, ou bien les différentes polarisations des images radars. La modélisation statistique des différentes images est fondée essentiellement sur une représentation des niveaux de gris par modèle de mélanges finis [Figueiredo 2002] estimés en ayant recours à un algorithme stochastique d'espérance-maximisation [Celeux 1996]. Les distributions choisies, en revanche, varient selon la nature de l'image.

3.1.1 Modélisation statistique des images optiques

Un modèle naturel de représentation des images optiques est la loi normale, aussi appelée loi gaussienne. Pour chaque classe m,

$$p_{mi}(y|\theta_{mi}) = \frac{1}{\sqrt{2\pi\sigma_{mi}^2}} \exp\left[-\frac{(y-\mu_{mi})^2}{2\sigma_{mi}^2}\right], \tag{3.1}$$

avec $\theta_{mi} = \{\mu_{mi}, \sigma_{mi}^2\}$, où la moyenne μ_{mi} et la variance σ_{mi}^2 sont estimées par un algorithme stochastique d'espérance-maximisation (SEM) [Celeux 1996], rappelé dans la sous-partie 3.1.3.1. D'autres algorithmes d'estimations des paramètres existent, mais nous avons opté pour l'algorithme SEM par homogénéité avec l'estimation des paramètres dans le cas d'images RSO (voir 3.1.3.2).

Au vu des résultats satisfaisants obtenus en considérant cette loi gaussienne, somme toute assez simple (voir sous-partie 3.1.3.3), nous n'avons pas jugé nécessaire d'exploiter la version généralisée de cette loi gaussienne.

3.1.2 Modélisation statistique des images RSO

Comme évoqué dans le chapitre 2, une partie de l'information radar est mélangée au bruit. Ainsi, en filtrant ce bruit, nous prenons le risque de perdre de l'information. Pour cela, nous n'avons procédé à aucun pré-traitement de cette image autre que ceux effectués par les distributeurs de celles-ci.

De nombreuses représentations statistiques ont été proposées, et peuvent être divisées en deux catégories : les fonctions de densités de probabilité (FDP) théoriques et les FDP heuristiques. La multitude de modèles proposés provient du fait que les statistiques de telles images sont corrompues par le bruit omniprésent.

3.1.2.1 Modèles théoriques

Nous avons vu dans le chapitre précédent (sous-partie 2.1.1) que l'acquisition radar est une acquisition d'amplitude complexe $A\exp(i\Phi)$, et nous cherchons, dans notre cas particulier, à modéliser les statistiques de l'amplitude A. L'intensité, carré de l'amplitude, est aussi couramment modélisée, cependant nous avons opté pour les amplitudes car, d'une part, ce sont les données qui nous ont été fournies, et d'autre part, d'après [Lee 1989], les amplitudes conduisent à de meilleurs résultats de segmentation (notre principal objectif).

Des hypothèses sont avancées quant à la rétrodiffusion, notamment que le nombre de cibles à l'intérieur d'une cellule de résolution est grand, et que ces cibles ont des dimensions similaires, supérieures à la longueur d'onde émise et sont distribuées de manière aléatoire. Dans ce cas, les surfaces homogènes apparaissent comme des champs stationnaires. Sous ces hypothèses, le phénomène de rétro-diffusion des ondes peut être modélisé par un modèle gaussien circulaire complexe, donc les parties réelles et imaginaires de l'amplitude complexe suivent un tel modèle. L'amplitude des images RSO à observation simple suit alors la loi de Rayleigh [Ulaby 1989] :

$$p_{mi}(y|\theta_{mi}) = p_{mi}(y|\sigma_{mi}) = \frac{y}{\sigma_{mi}^2} e^{-y^2/2\sigma_{mi}^2}.$$

Cette modélisation, parfaitement exacte pour du chatoiement pleinement développé, est limitée, par exemple dans le cas de la modélisation de fonctions de densité de probabilité à "queue lourde" (*heavy tailed*) qui ne sont clairement pas des distributions de Rayleigh, et qui sont généralement présentes dans les zones urbaines. En effet, les zones extrêmement texturées ont tendance à contenir des réflecteurs prédominants par rapport aux autres et donc à influer sur l'histogramme de telle sorte qu'il apparaisse à "queue lourde". Un tel histogramme est souvent difficile à estimer.

Pour les observations à "multi-visée" (voir sous-partie 2.1.1.4), la loi de Rayleigh se généralise à la distribution Nakagami-Gamma pour les amplitudes [Oliver 2004] de la forme :

$$p_{mi}(y|\theta_{mi}) = p_{mi}(y|L_{mi},\lambda_{mi}) = \frac{2}{\Gamma(L_{mi})} \left(\lambda_{mi} L_{mi}\right)^{L_{mi}} y^{2L_{mi}-1} \exp\left(-\lambda_{mi} L_{mi} y^2\right),$$

où $\Gamma(.)$ est la fonction Gamma :

$$\Gamma : z \mapsto \int_0^{+\infty} t^{z-1}\, e^{-t}\, \mathrm{d}t.$$

Pour la même raison que celle évoquée plus haut, dans le cas "simple visée", cette distribution est acceptable dans le cas où l'image est peu texturée [Tison 2003a]. En pratique, L_{mi} est un réel positif, et est interprété comme le nombre d'observations équivalent (*equivalent number of looks*), qui est le nombre effectif d'images à "simple visée", qui, moyennées, aboutissent à la formation de l'image "multi-visée". Ces images à "simple visée" pouvant être corrélées, le nombre effectif d'images à "simple visée" L_{mi} est un peu inférieur au nombre réel d'images qui ont été moyennées.

La rétrodiffusion a aussi été représentée par une distribution symétrique α-stable (SαS) par généralisation du théorème central limite [Kuruoglu 2003], ce qui est possible grâce à des hypothèses sur le bruit de chatoiement, fondées sur la supposition que ce chatoiement est le résultat d'un très grand nombre de réflecteurs ponctuels indépendants. Les FDP qui en découlent peuvent être définies comme une distribution SαS de Rayleigh généralisée, appelée ainsi car elle généralise la distribution de Rayleigh à une distribution non nécessairement de Rayleigh [Kuruoglu 2004] ; cette distribution est donnée par :

$$p_{mi}(y|\theta_{mi}) = p_{mi}(y|\gamma_{mi}, \alpha_{mi}) = y \int_0^{\infty} t \exp(-\gamma_{mi} t_{mi}^{\alpha}) J_0(yt) dt,$$

où $J_0(.)$ est la fonction de Bessel de première espèce à l'ordre 0 [Sneddon 1972]. Ces FDP conduisent, d'après [Kuruoglu 2003], à de bons résultats.

Lorsque nous considérons que le nombre de cibles à l'intérieur d'une cellule de résolution suit un processus de naissance et mort (*birth-and-death*), alors ce nombre suit une loi binomiale, et les statistiques des intensités des images RSO sont modélisées par la distribution K [Oliver 2004]. Les amplitudes suivent donc une loi dénommée "K-root" :

$$p_{mi}(y|\theta_{mi}) = p_{mi}(y|L_{mi}, M_{mi}, \mu_{mi})$$

$$= \frac{4}{\Gamma(L_{mi})\Gamma(M_{mi})} \left(\frac{L_{mi} M_{mi}}{\mu_{mi}}\right)^{(L_{mi}+M_{mi})/2} y^{L_{mi}+M_{mi}-1} K_{M_{mi}-L_{mi}} \left[2y \left(\frac{L_{mi} M_{mi}}{\mu_{mi}}\right)^{1/2}\right],$$

où $\Gamma(.)$ est la fonction Gamma et $K_\nu(.)$ est la fonction de Bessel de deuxième espèce à l'ordre ν [Sneddon 1972]. Cependant, cette distribution n'est pas très simple analytiquement et peut être difficile à manier. Ainsi, quand le nombre d'observations L des images "multi-visée" est suffisamment grand, cette distribution K peut être approximée par une loi log-normale [Oliver 2004] :

$$p_{mi}(y|\theta_{mi}) = p_{mi}(y|\sigma_{mi}, \mu_{mi}) = \frac{1}{\sigma_{mi} y \sqrt{2\pi}} \exp\left[-\frac{(\ln y - \mu_{mi})^2}{2\sigma_{mi}^2}\right].$$

Cette distribution s'avère adaptée dans la modélisation de zones urbaines, car elle permet de modéliser des zones extrêmement texturées [Oliver 2004] avec des histogrammes dits à "queue lourde", que nous avons évoqués auparavant.

Le modèle de distribution de Rayleigh gaussienne généralisée prend en compte une modélisation du bruit de chatoiement par gaussienne circulaire généralisée, sous hypothèse que les parties réelles et imaginaires du signal rétrodiffusé soient des gaussiennes généralisées de moyenne nulle [Moser 2006b]. Le modèle qui en résulte pour l'amplitude RSO est :

$$p_{mi}(y|\theta_{mi}) = p_{mi}(y|c_{mi},\gamma_{mi})$$

$$= \frac{\gamma_{mi}^2 c_{mi}^2 y}{\Gamma^2(1/c_{mi})} \int_0^{\pi/2} \exp[-(\gamma_{mi} y)^{c_{mi}} (|\cos(\theta)|^{c_{mi}} + |\sin(\theta)|^{c_{mi}})] d\theta.$$

3.1.2.2 Modèles heuristiques

La considération de modèles heuristiques est liée aux récents progrès en terme de technologie RSO, qui ne permettent plus de valider les hypothèses évoquées au début de la sous-partie précédente (voir 3.1.2.1). Par exemple, en pratique, la taille de la longueur d'onde émise n'est pas négligeable par rapport à celle de la cellule de résolution. En outre, des observations de zones urbaines ne vérifient pas forcément les hypothèses selon lesquelles les réflecteurs ponctuels par cellule de résolution sont semblables et nombreux, car certains réflecteurs peuvent être prédominants.

Les distributions de type Weibull ont été utilisées pour modéliser tant les amplitudes que les intensités, et représentent un bon modèle pour les zones agricoles et les zones des végétation [Oliver 2004]. Leur modèle de FDP est donné par :

$$p_{mi}(y|\theta_{mi}) = p_{mi}(y|\eta_{mi},\mu_{mi}) = \frac{\eta_{mi}}{\mu_{mi}^{\eta_{mi}}} y^{\eta_{mi}-1} \exp\left[-\left(\frac{y}{\mu_{mi}}\right)^{\eta_{mi}}\right].$$

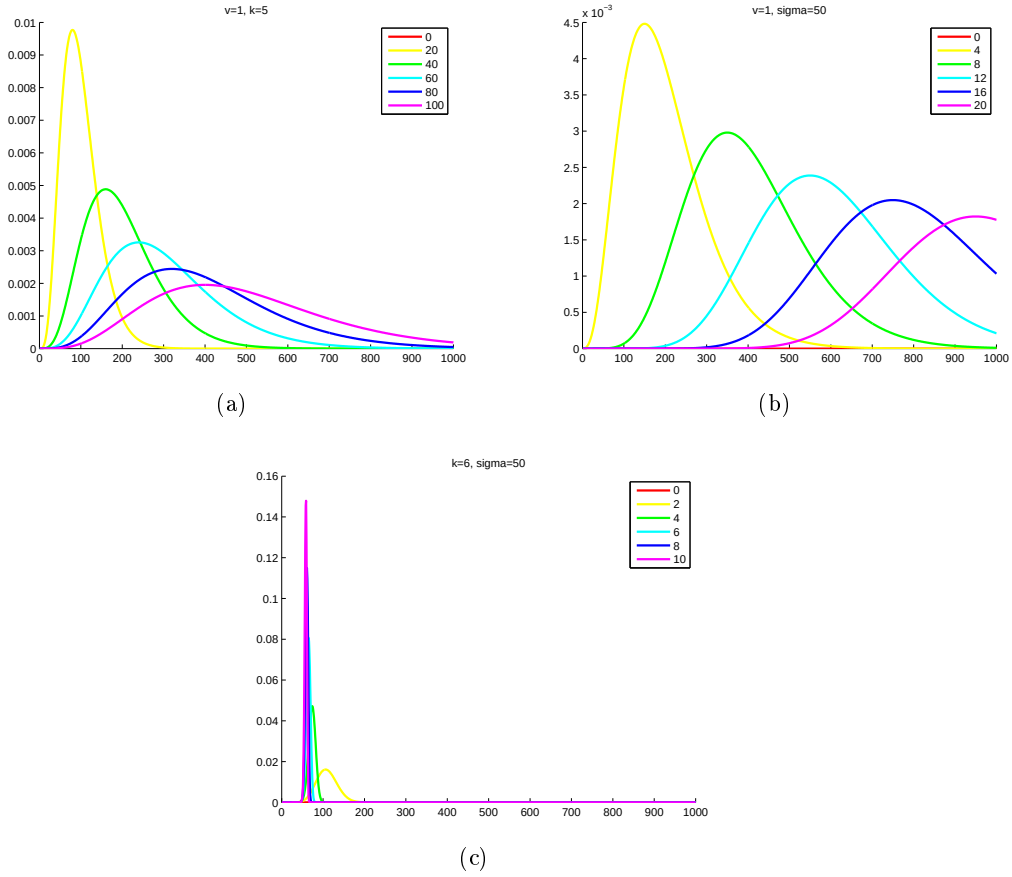
Pour les zones urbaines, la distribution de Fisher a été adoptée comme un modèle empirique pour les statistiques RSO des zones urbaines [Tison 2004], particulièrement difficiles à modéliser pour des images RSO à haute résolution et à simple-visée. L'avantage de ce modèle est la flexibilité au niveau de la modélisation de la queue des histogrammes à "queue lourde" [Tison 2004]. Sa distribution s'écrit :

$$p_{mi}(y|\theta_{mi}) = p_{mi}(y|L_{mi},M_{mi},\mu_{mi}) = \frac{\Gamma(L_{mi}+M_{mi})}{\Gamma(L_{mi})\Gamma(M_{mi})} \frac{L_{mi}}{M_{mi}\mu_{mi}} \frac{\left(\frac{L_{mi}y}{M_{mi}\mu_{mi}}\right)^{L_{mi}-1}}{\left(1+\frac{L_{mi}y}{M_{mi}\mu_{mi}}\right)^{L_{mi}+M_{mi}}},$$

où $\Gamma(\cdot)$ est la fonction Gamma [Sneddon 1972].

Plus récemment, la distribution Gamma généralisée a été proposée comme modèle empirique pour la modélisation des images RSO [Li 2011] :

$$p_{mi}(y|\theta_{mi}) = p_{mi}(y|\nu_{mi},\sigma_{mi},\kappa_{mi}) = \frac{\nu_{mi}}{\sigma_{mi}\Gamma(\kappa_{mi})} \left(\frac{y}{\sigma_{mi}}\right)^{\kappa_{mi}\nu_{mi}-1} \exp\left\{-\left(\frac{y}{\sigma_{mi}}\right)^{\nu_{mi}}\right\}.$$



(a) (b)

(c)

FIG. 3.1 – Évolution de la densité de probabilité de la distribution Gamma généralisée pour des variations de : σ_{mi} (a), κ_{mi} (b), ν_{mi} (c).

Nous remarquons qu'un tel modèle est une généralisation d'un large panel de familles de FDP comprenant : Rayleigh ($\nu_{mi} = 2$, $\kappa_{mi} = 1$), Nakagami ($\nu_{mi} = 1$), log-normal ($\kappa_{mi} \to \infty$) et Weibull ($\kappa_{mi} = 1$). Ainsi, ce modèle est assez flexible pour permettre de modéliser un grand nombre de scènes différentes. Cette flexibilité est due, avant tout, à la présence d'un degré de liberté supplémentaire (3 variables) par rapport aux distributions évoquées précédemment, même si, cependant, il n'y a pas toujours d'explication physique à ce degré de liberté supplémentaire. L'évolution de la densité de probabilité de la distribution Gamma généralisée en fonction de chacune des variables, les autres étant fixées, est donnée en figure 3.1. En pratique, un tel modèle a fait ses preuves, et aboutit à de meilleurs résultats de modélisation d'histogrammes que les lois Weibull, Nakagami, Fisher, Rayleigh gaussienne généralisée et K pour un assez grand nombre d'images [Li 2011].

Une multitude d'autres modèles ont également été pris en compte, tels que la distribution de Nakagami-Rice, la distribution gaussienne inverse [Tison 2004], ou encore la distribution de Pearson [Quelle 1993]. Cependant, ces modèles ne sont

pas vraiment adaptés dans notre cas. En effet, la distribution de Nakagami-Rice est surtout utilisée pour la détection de cibles. La distribution gaussienne inverse a montré de moins bons résultats que la distribution de Fisher pour la modélisation d'histogrammes à queue lourde. La distribution de Pearson a vu son intérêt croitre pour la modélisation de la surface de l'eau, ce qui n'est pas nécessairement notre intérêt principal. En outre, la modélisation de statistiques complexes telles que nous allons le voir par la suite n'est pas nécessairement aisée à appliquer à des systèmes de Pearson, de par leur complexité calculatoire.

3.1.3 Modèles de mélanges finis

En haute résolution, une zone peut être très hétérogène. Typiquement, une zone de végétation peut être constituée de diverses cultures. Les zones urbaines sont davantage hétérogènes, dans le sens où elles sont composées de matériaux aussi divers que du béton, des métaux, du bois, etc. Nous avons déjà évoqué cette hétérogénéité dans le chapitre précédent (voir 2.1.1.1). Ainsi, au lieu de modéliser ce type de zones par une seule FDP, nous proposons d'utiliser un mélange de diverses FDP, reflétant les contributions des divers matériaux présents dans une même classe m.

En outre, nous avons déjà pu évoquer le fait que des familles de FDP sont plus ou moins adaptées selon la nature de la classe observée [Oliver 2004]. Par exemple, la distribution de Weibull a empiriquement fait ses preuves pour la modélisation des océans ou des glaces. Celle de Rayleigh est adaptée pour la modélisation de régions naturelles uniformes, alors que la loi log-normale est plus adéquate dans les régions construites.

Pour chaque image (bande) initiale, nous cherchons à modéliser les distributions de chacune des classes ω_m considérées pour la classification, $m \in [1; M]$, étant donnée une base d'apprentissage. En effet, nous verrons ultérieurement que les modèles de classification employés sont suffisamment flexibles pour permettre de considérer en entrée plusieurs images, pouvant correspondre à diverses bandes d'une même acquisition, ou bien à diverses résolutions. Pour chaque classe, la FDP $p_m(y|\omega_m)$, où y est une observation, est modélisée par des mélanges finis [Figueiredo 2002] de distributions de niveaux de gris indépendantes :

$$p_m(y|\omega_m) = \sum_{i=1}^{K} P_{mi} p_{mi}(y|\theta_{mi}), \tag{3.2}$$

où les P_{mi} sont les proportions telles que, pour une classe m donnée, $\sum_{i=1}^{K} P_{mi} = 1$ avec $0 \leq P_{mi} \leq 1$. θ_{mi} est l'ensemble des paramètres de la i^e composante du mélange attribué à la classe m. Au lieu de travailler sur des observations correspondant à chacun des pixels de l'image, nous optons pour une modélisation des statistiques des niveaux de gris. Ainsi, cela permet de réduire la complexité calculatoire en considérant non plus le nombre de pixels N d'une image (dans notre cas, $N = 10^6$ pour une image 1000×1000 pixels), mais en considérant le nombre de niveaux de

gris (e.g., $Z = 256$). Les densités sont choisies parmi les familles précédemment considérées.

Nous présentons maintenant la façon dont sont estimés les paramètres de ces mélanges finis.

3.1.3.1 Algorithme stochastique d'espérance-maximisation

Il est important de rappeler les étapes d'un tel algorithme, puisqu'il est utilisé, dans ce manuscrit, pour l'estimation des paramètres moyenne et variance relatifs à l'image optique, et exploité dans l'estimation des paramètres des mélanges finis dans le cas RSO.

Cet algorithme vise à trouver le maximum de vraisemblance des paramètres de modèles probabilistes lorsque le modèle dépend de variables latentes non observables. Dans notre cas, les variables latentes sont les étiquettes d'appartenance des pixels r_k à une parmi les K composantes du mélange, et la log-vraisemblance à maximiser s'exprime :

$$L = \sum_{k=1}^{N} log\left(p_m(r_k|\omega_m)\right) = \sum_{z=0}^{Z-1} h_m(z) log\left(p_m(z|\omega_m)\right), \tag{3.3}$$

où $h_m(.)$ est l'histogramme correspondant à la classe m de l'image

Un simple estimateur du maximum de vraisemblance, obtenu en général par dérivée de la vraisemblance, n'est pas nécessairement aisé à estimer. Nous faisons donc appel à un algorithme d'espérance-maximisation [Dempster 1977]. Cependant, nous lui préfèrerons sa version stochastique [Celeux 1996], pour ses aptitudes à s'approcher du maximum global, et à être formulable pour la plupart des FDP des images RSO. En outre, cette version stochastique est parfaitement adaptée [Celeux 1996] lorsque nous traitons de données incomplètes. En effet, dans notre cas, nous ne savons pas à quelle composante K chaque donnée y observée correspond. Nous introduisons donc l'ensemble des étiquettes inconnues $s(y)$ telles que les données complètes sont représentées par le couple $\{(y,s), y \in [0; Z-1], s \in [1; K]\}$. L'algorithme SEM est un processus itératif. Pour chaque classe m, à chaque itération t :

- **Étape E :** pour chaque observation y et pour la i^e composante, calcul de la probabilité a posteriori correspondant à la FDP courante, c'est-à-dire, pour $y \in [0; Z-1], \quad i \in [1; K]$:

$$\tau_i^t(y) = \frac{P_{mi}^t p_{mi}^t(y|\theta_{mi})}{\sum_{j=1}^{K} P_{mj}^t p_{mj}^t(y|\theta_{mj})},$$

  où $p_{mi}^t(.)$ est la FDP estimée courante de la i^e composante ;
- **Étape S :** attribution d'une étiquette $s^t(y)$ à chaque niveau de gris y selon la probabilité a posteriori précédemment estimée $\{\tau_i^t(y) : i \in [1; K]\}$, $y \in [0; Z-1]$;

– **Étape M :** pour chaque composante i du mélange, nous estimons les paramètres de chaque composante i pour chacune des classes m : les proportions sont données par :

$$P_{mi}^{t+1} = \frac{\sum_{y \in Q_{mi}^t} h_m(y)}{\sum_{y=0}^{Z-1} h_m(y)},$$

et les paramètres estimés avec :

$$\theta_{mi}^{t+1} = \arg\max_{\theta_{mi}} L_{mi}^t = \arg\max_{\theta_{mi}} \sum_{y \in Q_{mi}^t} h_m(y) \ln p_{mi}^t(y|\theta_{mi}^t)$$

où Q_{mi}^t est l'ensemble des niveaux de gris attribués à la i^e composante dans l'étape S.

L'étape stochastique de cet algorithme garantit la convergence en loi vers une distribution stationnaire concentrée autour d'un maximum global de la log-vraisemblance L. Ainsi, à la stabilité de l'algorithme, nous obtenons non pas une seule partition mais une classe de partitions statistiquement admissibles pour les estimations des paramètres du mélange. Les estimations sont précises et asymptotiquement sans biais. Pour déterminer la meilleure des solutions dans cet ensemble stationnaire, nous procédons comme dans [Biernacki 2003] et calculons à chaque itération de l'algorithme la log-vraisemblance globale L. La plus grande vraisemblance obtenue correspond alors au meilleur jeu de paramètres pour le mélange.

Nous ne disposons actuellement pas de procédure statistique afin de déterminer le nombre minimum d'itérations à partir duquel nous pouvons considérer que la suite des paramètres acquiert son comportement stationnaire, donc pour assurer cette stationnarité, nous proposons de fixer un nombre d'itérations t_{max} moyennement élevé (dans notre cas, $t_{max} = 200$), et de faire tourner l'algorithme sur ces t_{max} itérations.

L'attribution des étiquettes s à l'initialisation de l'algorithme est faite aléatoirement. Elle est plus coûteuse calculatoirement, mais contourne une étape d'estimation préalable du mélange fini.

Dans le cas optique, où les FDP suivent par hypothèse des distributions gaussiennes, les paramètres peuvent être estimés (étape M) en ayant recours à des estimateurs de ceux-ci. L'estimateur de la moyenne est donné par :

$$\widehat{moy}_{mi} = \frac{\sum_{y \in Q_{mi}^t} h_m(y) \times y}{\sum_{y \in Q_{mi}^t} h_m(y)},$$

et celui de la variance par :

$$\widehat{var}_{mi} = \frac{\sum_{y \in Q_{mi}^t} h_m(y) \times y^2}{\sum_{y \in Q_{mi}^t} h_m(y)} - \widehat{moy}_{mi}^2.$$

D'autres stratégies ont été proposées pour garantir l'obtention du maximum de vraisemblance le plus élevé, notamment en couplant plusieurs algorithmes de type EM (par exemple, SEM-EM) [Biernacki 2001]. De nombreux autres algorithmes ont

été utilisés comme alternative à l'algorithme SEM, tel que l'algorithme d'estimation itérative conditionnelle (ICE) [Delignon 1997] qui fait appel à l'espérance conditionnelle plutôt qu'au maximum de vraisemblance.

3.1.3.2 Estimation des mélanges finis

Dans le cas où l'image à modéliser est de type radar, le mélange fini est un peu plus complexe que dans le cas optique, car les familles considérées ne suivent physiquement plus des lois normales (sous-partie 3.1.2). Nous considérons que les lois peuvent être librement choisies dans un dictionnaire prédéfini. Dans ce cas-là, l'estimation du mélange fini revient à estimer le nombre de paramètres de la somme K et les proportions P_{mi}, et à sélectionner le modèle de FDP optimal $p_{mi}(.)$ pour chaque élément du mélange, et estimer les paramètres correspondants θ_{mi}.

Nous avons pu constater qu'il existe une variété étendue de familles pour modéliser les différentes classes des images RSO (voir 3.1.2). Cependant, l'estimation du maximum de vraisemblance (MV), utile dans la dernière étape de l'algorithme SEM, n'est pas toujours possible, comme par exemple pour la distribution de Nakagami. Il en est de même pour l'estimation de moments avec la distribution de Fisher [Tison 2004]. Nous utilisons comme alternative la méthode des Log-cumulants [Moser 2006a, Tison 2004] (MoLC), qui a montré de bonnes capacités d'estimation par rapport à l'utilisation du MV ou de la méthode des moments.

Par analogie avec la transformée de Laplace pour la fonction génératrice des moments [Nicolas 2000], la méthode MoLC est fondée sur la transformée de Mellin [Sneddon 1972], et sur la généralisation des concepts de moments et de cumulants qui en découlent. Les cumulants sont définis grâce à la fonction génératrice des cumulants :

$$g(t) = log(E\{\exp(t.u)\}) = \sum_{n=1}^{\infty} \kappa_n \frac{t^n}{n!},$$

où u est une variable aléatoire. Cela permet d'établir un ensemble d'équations dépendant des paramètres inconnus d'un modèle paramétrique donné via un ou plusieurs log-cumulants, selon le nombre de paramètres à estimer :

$$\begin{cases} \kappa_1 = g'(0) = E\{\ln u\} \\ \kappa_2 = g''(0) = \mathrm{Var}\{\ln u\} \\ \kappa_3 = g^{(3)}(0) = E\{(\ln u - \kappa_1)^3\} \end{cases}, \tag{3.4}$$

où $E\{.\}$ est l'espérance mathématique et $Var\{.\}$ la variance. En pratique, pour une famille de distributions donnée, nous estimons empiriquement les i premiers moments correspondant aux i paramètres de la distribution, puis calculons les paramètres par inversion des équations (3.4), aussi données de manière plus spécifique dans le tableau 3.1.

La méthode MoLC est applicable à la majorité des distributions définies dans la sous-partie 3.1.2, et permet de définir un unique jeu de paramètres solutions des

TAB. 3.1 – Équations MoLC pour les familles paramétriques du dictionnaire. $\Psi(\cdot)$ est la fonction Digamma [Sneddon 1972] et $\Psi(\nu,\cdot)$ est la fonction polygamma du ν^e ordre [Sneddon 1972].

Famille	Équations MoLC
Gamma Généralisée	$\kappa_1 = \Psi(\kappa)/\nu + \ln\sigma$ $\kappa_2 = \Psi(1,\kappa)/\nu^2$ $\kappa_3 = \Psi(2,\kappa)/\nu^3$
Log-normale	$\kappa_1 = m$ $\kappa_2 = \sigma^2$
Weibull	$\kappa_1 = \ln\mu + \Psi(1)\eta^{-1}$ $\kappa_2 = \Psi(1,1)\eta^{-2}$
Nakagami	$2\kappa_1 = \Psi(L) - \ln\lambda L$ $4\kappa_2 = \Psi(1,L)$

équations. Sont exclues de cette propriété générale les distributions K-root et Rayleigh gaussienne généralisée, pour lesquelles certaines combinaisons de paramètres aboutissent à un système d'équations sans solutions [Moser 2006b, Jakeman 1976].

Parmi les densités évoquées en 3.1.2, une en particulier a retenu notre attention : la distribution Gamma généralisée (ΓG). En effet, un tel modèle est la généralisation d'un large panel de familles de FDP (Nakagami, Weibull...), et a montré de bons résultats expérimentaux (voir 3.1.2.2). Ainsi, nous favorisons celle-ci, mais aussi, dans les cas où la condition d'applicabilité n'est pas respectée, nous avons opté pour un dictionnaire contenant trois FDP spécifiques aux images RSO : Log-Normal, Weibull et Nakagami. La condition d'applicabilité de la distribution ΓG, qui garantit la consistance des estimations, est [Krylov 2011b] :

$$\kappa^t_{2mi} \geqslant 0.63(\kappa^t_{3mi})^{(2/3)}.$$

Ce choix de mélange de ΓG est motivé par le fait que les performances d'un simple mélange de distributions ΓG sont similaires à celles obtenues avec un dictionnaire plus complet [Krylov 2010]. En outre, le temps de calcul est plus avantageux dans le cas où une seule famille est considérée, puisqu'il n'y a pas d'étape de sélection du modèle (voir plus bas).

La sélection de seulement quatre familles de FDP dans le dictionnaire est motivée par l'étude menée par Krylov et al. dans [Krylov 2011a]. En effet, l'étude comparative démontre que le choix de ces quatre familles spécifiques n'affecte pas notablement les résultats obtenus en comparaison avec l'utilisation d'un dictionnaire plus complet, incluant notamment les lois de Fisher, K-root et Rayleigh.

Les paramètres de chacune de ces FDP du dictionnaire peuvent être estimés en ayant recours à la méthode MoLC, notamment grâce aux équations listées dans le tableau 3.1.

Nous reprenons alors l'algorithme SEM présenté dans la sous-partie 3.1.3.1, en utilisant les mêmes notations, et intégrons l'étape MoLC d'estimation des para-

mètres. Une étape d'estimation du nombre de composantes K de la somme est aussi intégrée. A chaque itération, si la proportion d'un des éléments de la somme P_{mi} est négligeable (typiquement, inférieur à 10^{-6}), alors cet élément est supprimé et K est décrémenté. Nous fixons donc préalablement une borne supérieure au nombre de composantes, soit $K_{MAX} = 7$ choisi de manière heuristique. Ce choix n'est pas critique au regard de l'algorithme SEM. En outre, il s'agit d'une sur-estimation du nombre de composantes, et cette sur-estimation ne doit pas être trop élevée afin de permettre à l'algorithme de converger plus rapidement. D'autres méthodes ont également été utilisées pour déterminer le nombre optimal de composantes, et peuvent se diviser selon deux familles [Figueiredo 2002] : les méthodes déterministes et les méthodes stochastiques. Parmi les méthodes déterministes, nous retrouvons des approches bayésiennes, comme par exemple le critère d'inférence de Schwarz [Schwarz 1978], ou bien des méthodes fondées sur la théorie de l'information, telles que la longueur de description minimale [Rissanen 1978]. Pour plus de détails concernant ces méthodes déterministes, voir [McLachlan 2000]. Parmi les méthodes stochastiques, qui sont coûteuses calculatoirement, les méthodes de Monte-Carlo par chaînes de Markov (MCMC) peuvent être employées [Figueiredo 2002].

Pour chaque classe m, à chaque itération t, l'algorithme se déroule comme suit :

- **Étape E :** pour chaque observation y et pour la i^e composante, calcul de la probabilité a posteriori correspondant à la FDP courante, c'est-à-dire, pour $y \in [0; Z-1], \quad i \in [1; K^t]$:

$$\tau_i^t(y) = \frac{P_{mi}^t p_{mi}^t(y|\theta_{mi})}{\sum_{j=1}^{K^t} P_{mj}^t p_{mj}^t(y|\theta_{mj})},$$

  où $p_{mi}^t(.)$ est la FDP estimée courante de la i^e composante ;
- **Étape S :** attribution d'une étiquette $s^t(y)$ à chaque niveau de gris y selon la probabilité a posteriori précédemment estimée $\{\tau_i^t(y) : i \in [1; K^t]\}$, $y \in [0; Z-1]$;
- **Étape MoLC :** pour chaque composante i du mélange, nous estimons les paramètres (proportions P_{mi} et paramètres θ_{mi}) de chaque composante i pour chacune des classes m : les proportions sont données par :

$$P_{mi}^{t+1} = \frac{\sum_{y \in Q_{mi}^t} h_m(y)}{\sum_{y=0}^{Z-1} h_m(y)},$$

  et les paramètres estimés en ayant recours aux équations MoLC :

$$\kappa_{1mi}^t = \frac{\sum_{y \in Q_{mi}^t} h_m(y) \ln y}{\sum_{y \in Q_{mi}^t} h_m(y)}, \quad \kappa_{bmi}^t = \frac{\sum_{y \in Q_{mi}^t} h_m(y)(\ln y - \kappa_{1mi}^t)^b}{\sum_{y \in Q_{mi}^t} h_m(y)},$$

  où Q_{mi}^t est l'ensemble des niveaux de gris attribués à la i^e composante dans l'étape S, $b = 2$ ou $b = 3$, et $h_m(y)$ est l'histogramme correspondant à la classe m de l'image. Les moments étant calculés, nous utilisons ensuite les relations du tableau 3.1 pour estimer les paramètres correspondants de chacune des familles de FDP du dictionnaire ;

- **Étape K :** pour chaque composante i du mélange, si la proportion P_{mi}^{t+1} est inférieure à un certain seuil (typiquement, 10^{-4}), alors la composante en question est éliminée, et K^{t+1} est décrémenté ;
- **Étape de sélection du modèle :** pour chaque composante i du mélange et pour chaque famille du dictionnaire, nous estimons la log-vraisemblance

$$L_{mi} = \sum_{y \in Q_{mi}^t} h_m(y) \ln p_{mi}^t(y|\theta_{mi}^t),$$

et définissons $p_{mi}^{t+1}(.)$ comme étant la famille pour laquelle la log-vraisemblance obtenue est la plus élevée. Dans le cas où la condition d'applicabilité de MoLC pour la distribution Gamma généralisée est respectée, à savoir $\kappa_{2mi}^t \geqslant 0.63(\kappa_{3mi}^t)^{(2/3)}$, cette étape n'est pas exécutée.

Tout comme précédemment (sous-partie 3.1.3.1), le meilleur modèle de mélange est obtenu après les t_{max} itérations en relevant les paramètres correspondant à la plus grande vraisemblance globale.

3.1.3.3 Validations expérimentales des estimations par mélanges finis

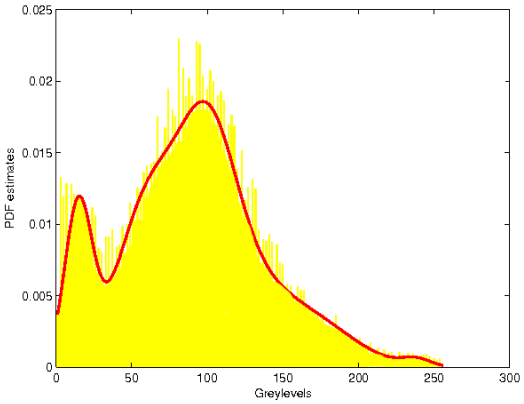
Nous proposons maintenant de valider expérimentalement l'efficacité des méthodes d'estimation proposées (voir 3.1.3.1 et 3.1.3.2), en comparant les histogrammes de niveaux de gris de diverses classes et les estimations de ces histogrammes obtenues par des mélanges finis de gaussiennes pour les images optiques et de distributions Gamma généralisées pour les images radars.

La première validation expérimentale concerne l'algorithme SEM appliqué aux images optiques, et plus particulièrement à l'image GeoEye du quai de Port-au-Prince (Haïti), dont les caractéristiques techniques sont données en Annexe A. Les résultats sont donnés dans la figure 3.2 pour l'estimation des statistiques des zones urbaines et des zones d'eau de cette image à partir d'une base d'apprentissage représentant approximativement 5% de l'image complète.

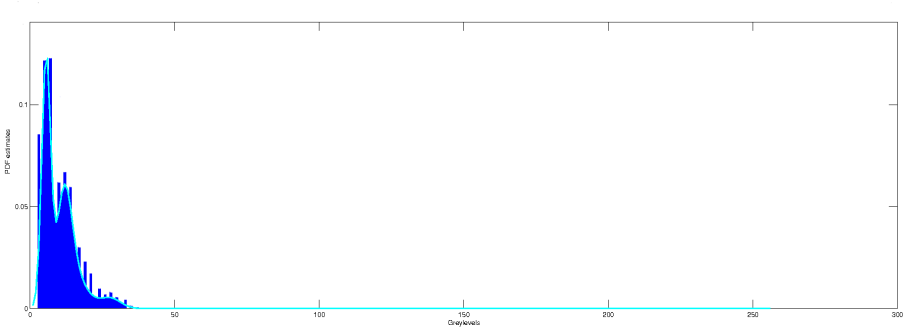
La deuxième validation expérimentale concerne l'algorithme SEM modifié (intégrant l'estimation des paramètres par MoLC) appliqué à des images RSO COSMO-SkyMed (©ASI) de deux zones : Port-au-Prince (Haïti) et Amiens (France). Les caractéristiques techniques de ces images sont données, en détail, dans l'Annexe A. Les résultats sont donnés dans les figures 3.3 et 3.4 pour l'estimation des statistiques de diverses classes à partir d'une base d'apprentissage représentant approximativement 5% de l'image complète dans chaque cas.

Dans le cas optique tout comme dans le cas RSO, la modélisation des statistiques des classes par des mélanges de distribution bien choisies est satisfaisante. L'algorithme – quasi-automatique – est capable d'estimer le bon mélange afin de trouver la meilleure modélisation possible. En outre, cet algorithme donne de bons résultats, peu importe le niveau d'hétérogénéité de la classe considérée.

(a)



(b)

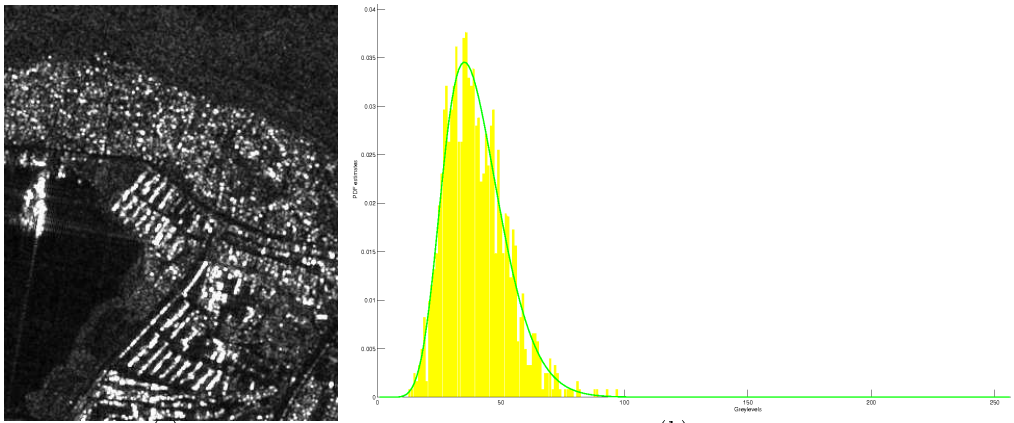


(c)

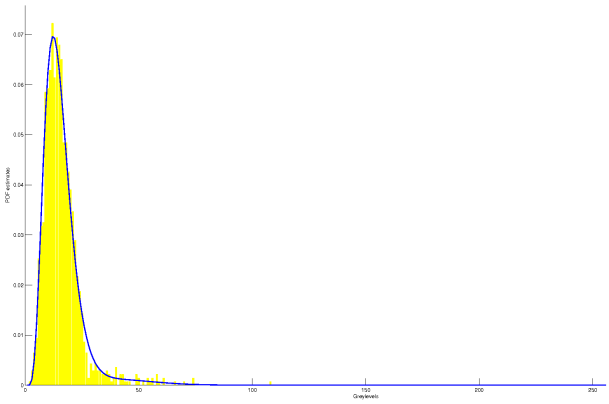
FIG. 3.2 – (a) : Image optique de Port-au-Prince (Haïti) (©GeoEye) ; (b) : Histogramme de niveaux de gris (en jaune) et son approximation par la méthode proposée pour la zone urbaine (en rouge) ; (c) : Histogramme de niveaux de gris (en bleu) et son approximation par la méthode proposée pour les zones d'eau (en cyan).

3.2 Les limites des images monorésolution à polarisation simple et introduction des attributs de texture

Les images sur lesquelles nous travaillons sont des images à polarisation simple, ce qui rend d'autant plus difficile leur analyse. En effet, nous avons en notre possession seulement des acquisitions monobandes 8-bits, donc pour toute indétermination éventuelle, il nous a paru nécessaire d'extraire une information complémentaire : un attribut de texture. De précédentes publications [Dekker 2003, Wen 2009] ont montré que l'extraction de données supplémentaires permet d'obtenir une meilleure classification, en particulier pour les zones urbaines. Cela est appuyé par nos propres résultats expérimentaux (cf. chapitres 4 et 5), qui ont été publiés en 2010 à la confé-



(a) (b)



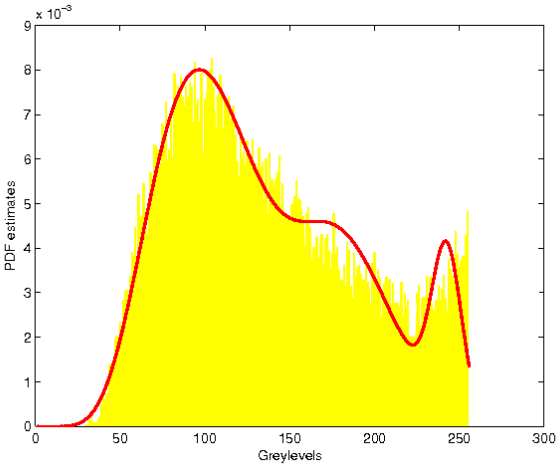
(c)

FIG. 3.3 – (a) : Image RSO originale de Port-au-Prince (Haïti) (COSMO-SkyMed,©ASI) ; (b) : Histogramme de niveaux de gris (en jaune) et son approximation par la méthode proposée pour la végétation (en vert) ; (c) : Histogramme de niveaux de gris (en jaune) et son approximation par la méthode proposée pour les zones d'eau (en bleu).

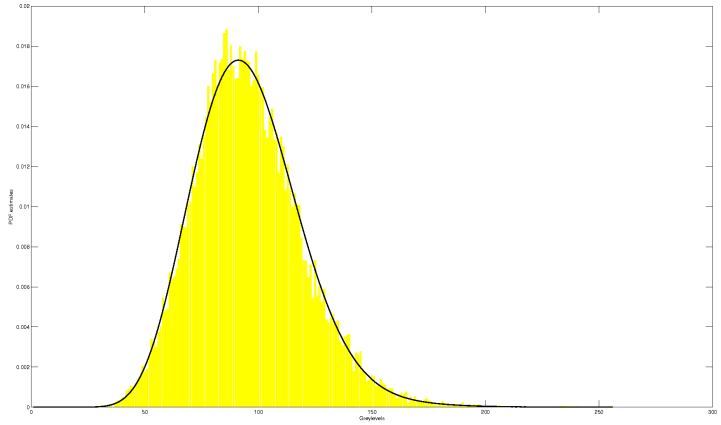
rence SPIE Remote Sensing (Annexe E.1).

La notion de texture en elle-même est difficile à décrire car il s'agit d'une notion abstraite. Il n'y a pas de modèles mathématiques généraux, mais plutôt diverses possibilités de les caractériser, qui sont plus ou moins adaptées selon les propriétés intrinsèques de la texture : périodicité, homogénéité, etc. Pour éviter toute dispersion autour de la notion de texture, qui pourrait constituer en elle-même un chapitre entier, nous nous focalisons sur la prise en compte des textures comme information utile à des fins de classification d'images RSO. Il s'agit d'une application bien spécifique, qui a largement été étudiée dans la littérature. La texture elle-même peut être

(a)



(b)



(c)

FIG. 3.4 – (a) : Image RSO originale de Amiens (France) (COSMO-SkyMed,©ASI) ; (b) : Histogramme de niveaux de gris (en jaune) et son approximation par la méthode proposée pour les zones urbaines (en rouge) (c) : Histogramme de niveaux de gris (en jaune) et son approximation par la méthode proposée pour les arbres (en noir).

considérée comme une partie intégrante de l'image RSO. Ainsi, dans [Greco 2007], l'amplitude de l'image est prise comme produit d'un bruit de chatoiement et d'une texture, cette dernière étant choisie comme moyenne de l'intensité dans une fenêtre donnée. Les auteurs prennent ici le parti de séparer totalement le bruit de la texture, et de modéliser ceux-ci de manière indépendante.

Plus généralement, les textures sont estimées mathématiquement afin de caractériser les diverses zones (donc les diverses textures) de l'image à traiter.

Les *textures du premier ordre* correspondent à des valeurs originales de l'image initiale, telles que la moyenne ou l'écart-type, souvent estimées à partir de l'histogramme de niveaux de gris. La correspondance entre diverses classes est effectuée par des calculs de distances entre les histogrammes par exemple [Dobson 1996]. Il s'agit typiquement d'une application assez simple, des applications plus évoluées ont aussi vu le jour.

Lors de l'utilisation des champs de Markov gaussiens [Dong 1999], les textures, estimées après un pré-filtrage de l'image initiale, jouent un rôle dans la détermination des étiquettes pour la classification et sont caractérisées en terme de paramètres représentant les interactions spatiales entre les pixels voisins. Cette méthodologie a été étendue à une approche non supervisée dans [Du 2002]. Ces champs de Markov ont montré leur potentialité à extraire assez spécifiquement des zones urbaines [Corbane 2009]. Cependant, cette texture extraite dépend fortement des paramètres physiques des images RSO traitées, à savoir de la polarisation, de l'angle d'incidence, de la résolution, etc. De ce fait, elle ne nous semble pas assez générale par rapport à l'utilisation que nous souhaitons en faire.

C'est sur la notion même d'interaction spatiale qu'est fondée l'utilisation de textures calculées à partir de statistiques du *deuxième ordre*, qui correspondent à l'estimation de valeurs relatives à deux (groupes de) pixels de l'image. Ainsi, en pratique, nous prenons deux pixels séparés d'une distance "offset" h et selon une direction θ, et nous leur attribuons une certaine mesure. La plus classique est le calcul de textures de type Haralick utilisant les matrices de co-occurrence de niveaux de gris. Dans la même lignée de ce type de matrices, le calcul de "run lengths" sur les niveaux de gris permet aussi d'aboutir à des textures du même type [Galloway 1975]. Les fractales ont aussi été considérées [Dellepiane 1991, Pentland 1984] D'autres types de textures ont été développés plus récemment, notamment les semivariogrammes [Chen 2004].

Dans notre cas, nous avons plutôt considéré l'extraction d'attributs de texture comme étant l'extraction d'une image complémentaire intégrée comme une image d'entrée supplémentaire. À chaque pixel de l'image est attribuée une nouvelle valeur, ou un vecteur de valeurs selon le nombre d'attributs choisi. Le modèle mathématique reste identique aux modèles précédemment évoqués, seule la façon d'être ensuite traité change, puisque nous avons cherché à générer une "image texture".

La texture peut aussi être vue comme une répétition d'éléments à une certaine fréquence. Ainsi, l'image RSO originale peut être filtrée en ayant recours à des méthodes spectrales, comme par exemple l'utilisation d'ondelettes [Zhang 2005]. Divers filtres ont aussi déjà été proposés [Buades 2005], comme par exemple le filtre de Lee [Lee 1980], ou des filtres plus spécifiques aux images RSO dans le sens où ils prennent en compte les formulations statistiques théoriques du bruit [Achim 2002]. Les filtres de Gabor [Jain 1991], qui sont le produit de gaussiennes par des sinus ou des cosinus, ont aussi largement été exploités. Ce type de filtrage est équivalent à un "débruitage" des images initiales, et aboutit à une image texturée, comme nous le souhaitons. Cependant, les images obtenues sont trop corrélées par rapport aux

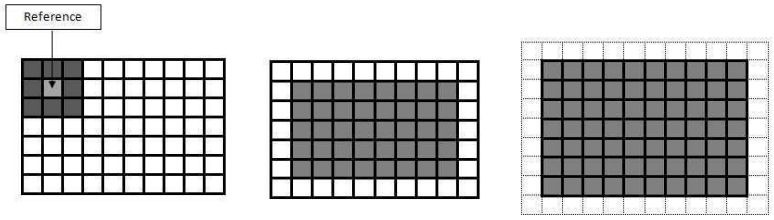


FIG. 3.5 – Illustration de la fenêtre glissante pour l'estimation des attributs de texture choisis. (a) : Fenêtre de taille $w = 3$ utilisée pour estimer le paramètre de variance du pixel de référence ; (b) et (c) : En gris : pixels pour lesquels la variance est estimée ($w = 3$) respectivement sans duplication des pixels sur les bords (b) et avec duplication (c).

images initiales pour nous permettre d'obtenir une réelle discrimination et une réelle amélioration des résultats de classification.

3.2.1 Attributs de texture

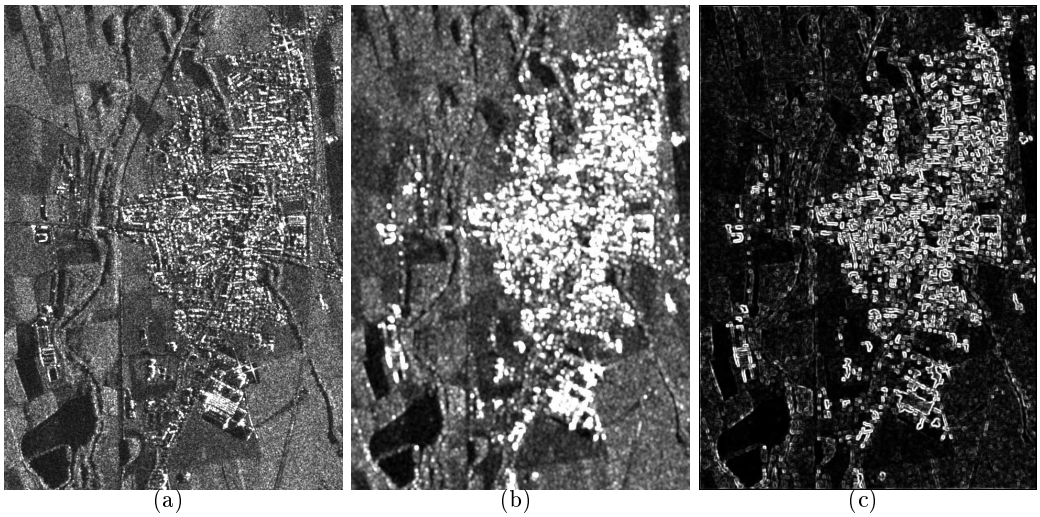
Motivés par l'étude bibliographique réalisée en introduction de cette partie, nous nous sommes intéressés à deux types d'attributs de texture en particulier :

- Méthode utilisant des *matrices de co-occurrence de niveaux de gris* (MCNG) [Haralick 1973], qui permet l'estimation des statistiques du deuxième ordre de l'image.
- Les *semivariogrammes* [Curran 1988, Chen 2004].

Nous n'avons opté que pour ces deux modèles pour diverses raisons. Ces méthodes sont faciles à manipuler, elles ont des propriétés intéressantes pour les zones urbaines, elles ont été largement utilisées et ont fait leurs preuves, comme expliqué dans de précédentes publications [Wen 2009].

Les deux extractions d'attributs de texture de l'image originale sont fondées sur un principe de fenêtre glissante de taille $w \times w$. Chaque pixel de l'image devient tour à tour pixel de référence, et sa valeur est remplacée par la variance calculée dans la fenêtre (Fig. 3.5(a)). Pour calculer les variances des pixels situés en bordure de l'image, nous avons dupliqué ces pixels, en vue de générer une image qui a la même taille que l'image originale (Fig. 3.5(c)). Une alternative aurait été, par exemple, le zero-padding, c'est-à-dire ne pas recourir à une duplication mais à une mise à 0 des pixels externes.

Nous remarquons que de telles textures peuvent être rapprochées des informations contextuelles dans le sens où chaque pixel de ces attributs renferme une certaine information concernant les pixels environnants. Une deuxième remarque concerne le fait qu'il y a une perte de résolution de l'image obtenue proportionnelle à la taille w de la fenêtre utilisée. Cependant, cette perte de résolution n'est pas critique dans le sens où, en pratique, les classes considérées sont majoritairement thématiques.



(a) (b) (c)

FIG. 3.6 – (a) : Image RSO originale de Cavallermaggiore (Italie) (COSMO-SkyMed,©ASI) ; (b) : Attribut de texture extrait par semivariogramme ; (c) : Attribut de texture extrait avec MCNG. Nous pouvons constater que les zones urbaines sont bien discriminées. La taille de la fenêtre utilisée est $w = 5$.

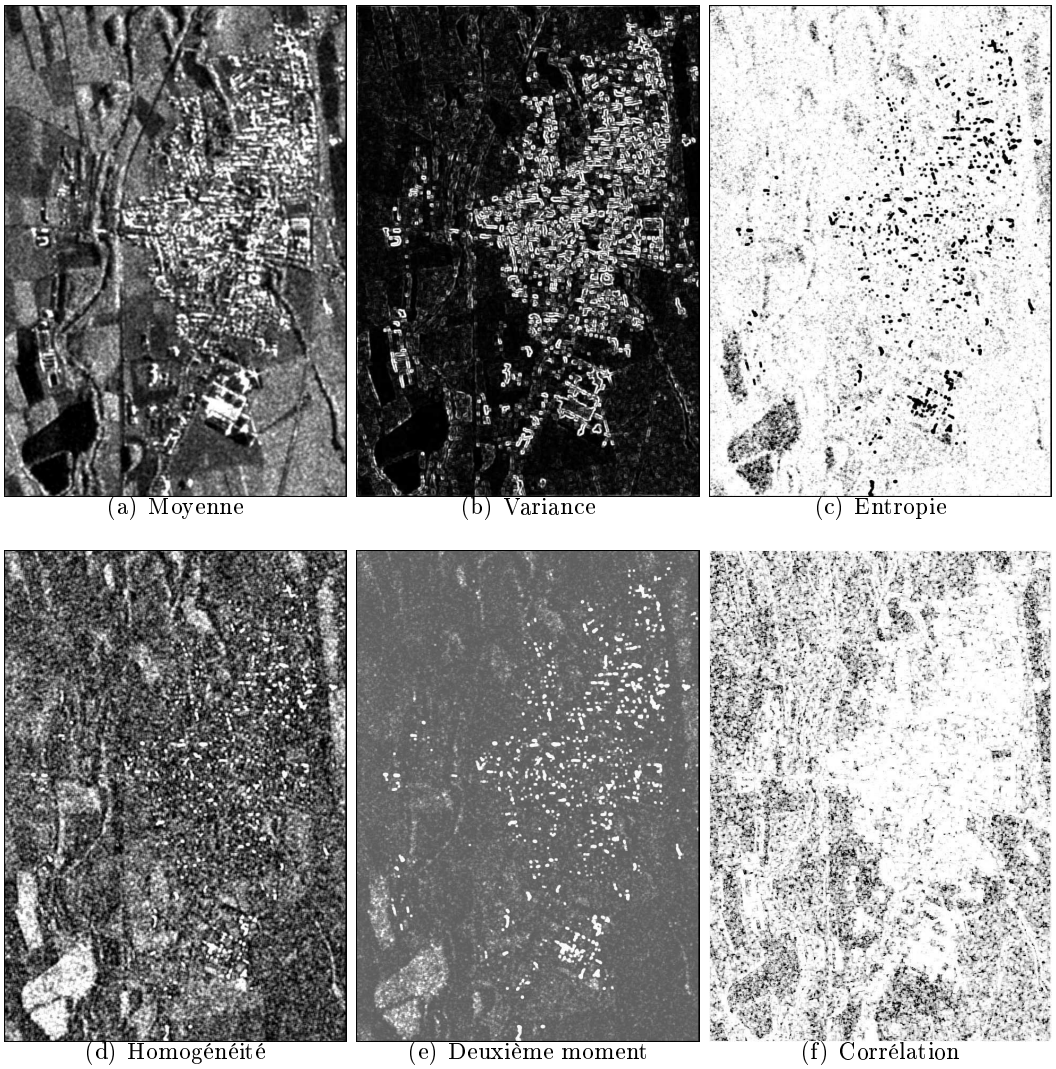
TAB. 3.2 – Paramètres d'Haralick calculés à partir de la matrice de co-occurrence afin de décrire la texture de l'image.

Paramètres	Expressions
Moyenne	$moy_i = \sum_i \sum_j i \times p_{Mcng}(i,j)$
Variance	$var_i = \sum_i \sum_j (i - moy_i)^2 \times p_{Mcng}(i,j)$
Homogénéité	$hom = \sum_i \sum_j \frac{p_{Mcng}(i,j)}{1+(i-j)^2}$
Entropie	$ent = \sum_i \sum_j p_{Mcng}(i,j) \times (-\ln p_{Mcng}(i,j))$
Deuxième moment	$mom = \sum_i \sum_j (p_{Mcng}(i,j))^2$
Corrélation	$cor = \sum_i \sum_j \frac{p_{Mcng}(i,j) \times (i-moy_i)(j-moy_j)}{var_i \times var_j}$

Les attributs de texture extraits sont représentés en Fig. 3.6. Ils montrent effectivement une bonne habilité à discriminer les zones urbaines. Nous verrons en pratique les bonnes aptitudes à améliorer les résultats de classification dans le chapitre dédié aux images monorésolution (Chap. 4).

3.2.1.1 Matrice de co-occurrence de niveau de gris

Suggérées par Haralick [Haralick 1973], les MCNG sont encore très largement employées. A chaque pixel est attribuée une matrice $Mcng$, calculée dans la fenêtre



(a) Moyenne (b) Variance (c) Entropie

(d) Homogénéité (e) Deuxième moment (f) Corrélation

FIG. 3.7 – Attributs de texture de l'image RSO de Cavallermaggiore donnée Fig. 3.6(a), calculés pour différents paramètres.

de taille $w \times w$ incluant le-dit pixel et son voisinage. La matrice $Mcng$ est une matrice carrée de taille $Z \times Z$, Z étant le nombre de niveaux de gris de l'image originale, qui décrit les statistiques jointes des niveaux de gris des différents pixels comme une fonction de leur localisation mutuelle. En d'autres termes, l'élément (i, j) de la matrice est la probabilité $p_{Mcng}(i, j)$ représentant la fréquence d'occurrence de deux pixels ayant respectivement les valeurs i et j et espacés d'une distance h. Nous avons choisi une distance par défaut de 1, ce qui nous amène à considérer des pixels adjacents. L'adjacence peut être définie dans chacune des directions possibles (horizontale, verticale ou en diagonale). Dans notre cas, nous avons considéré une adjacence horizontale, c'est-à-dire que la matrice $Mcng$ est remplie en considérant

un pixel de référence et le pixel situé à sa droite à l'intérieur de la fenêtre de taille $w \times w$. Nous avons pu constater que l'adjacence n'a, en pratique et dans notre cas spécifique, que peu d'influence sur les résultats de classification, car peu d'influence sur les formes de textures obtenues.

Haralick a dérivé de cette matrice quatorze paramètres, dont le contraste, la corrélation ou encore la variance. Nous en listons quelques-uns à titre indicatif dans le tableau 3.2, et nous montrons les images de textures obtenues pour chacun de ces paramètres dans la Fig. 3.7, textures extraites à partir de l'image de Cavallermaggiore (Fig. 3.6(a) et Annexe A). Nous cherchons de manière empirique l'attribut qui met le mieux en exergue un type de zones donné (par exemple, les zones urbaines) et qui donne une image de textures un peu différente de l'image initiale. Typiquement, l'image texturée obtenue avec la moyenne est un peu trop similaire à notre image de départ. La variance est, à notre sens, la plus convenable car elle est assez contrastée avec une bonne mise en évidence de la zone urbaine, tout en étant peu bruitée.

3.2.1.2 Semivariogramme

Le semivariogramme, qui décrit les propriétés spatiales de l'image, est une mesure de la variance d'une variable en fonction de la distance. Soit s et t deux pixels adjacents séparés d'une distance h, le semivariogramme est défini comme l'espérance du carré de la différence de niveaux de gris entre les deux pixels :

$$\gamma(h) = \frac{E[|z_s - z_t|^2]}{2} \tag{3.5}$$

Empiriquement, le semivariogramme s'estime grâce à :

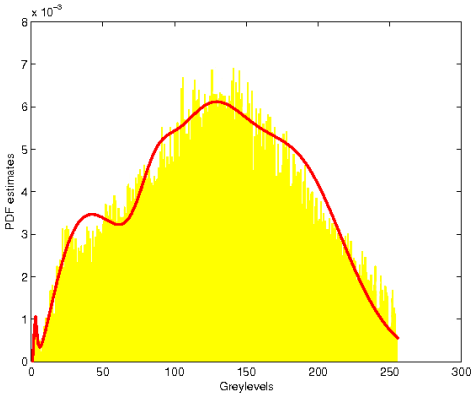
$$\hat{\gamma}(h) = \frac{\sum_{(i,j)\in N(h)} |z_i - z_j|^2}{|N(h)|} \tag{3.6}$$

où i, j sont des pixels adjacents séparés d'une distance h, z_i, z_j leurs niveaux de gris respectifs et $N(h)$ est l'ensemble des paires d'observation. Par fenêtre glissante, chaque pixel est remplacé par sa valeur de semivariogramme correspondante. Un exemple est donné en Fig. 3.6.

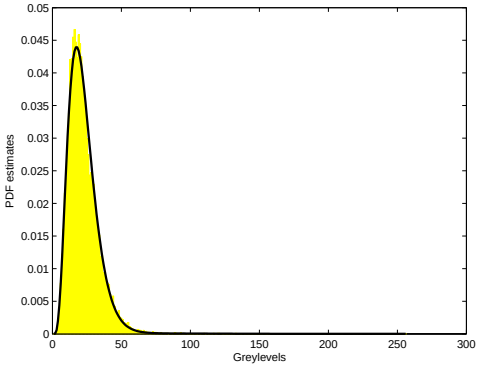
3.2.2 Modélisation statistique des attributs de texture

Il s'agit maintenant d'intégrer l'information de texture dans notre modèle. Comme précédemment (cf. 3.1), nous cherchons à modéliser les statistiques des niveaux de gris afin d'obtenir des informations initiales similaires, ici sous forme de FDP. Peu de modèles statistiques de FDP existent actuellement pour modéliser les images de textures. Nous proposons alors d'utiliser le modèle de mélange fini de distributions Gamma généralisées comme modèle pour les attributs de texture car il s'agit d'un mélange de FDP heuristiques assez flexible pour permettre la modélisation de diverses classes et donc de divers types d'images. En outre, comme les

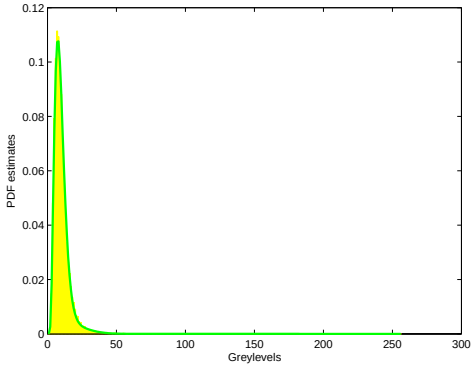
(a)



(b)



(c)



(d)

FIG. 3.8 – (a) : Image de texture extraite de l'image RSO d'Amiens donnée Fig. 3.4(a) par utilisation de MCNG ; (b) : Histogramme de niveaux de gris (en jaune) et son approximation par la méthode proposée pour les zones urbaines (en rouge) ; (c) : Histogramme de niveaux de gris (en jaune) et son approximation par la méthode proposée pour les arbres (en noir) ; (d) : Histogramme de niveaux de gris (en jaune) et son approximation par la méthode proposée pour la végétation (en vert).

attributs de texture sont extraits d'images modélisables statistiquement par des dis-

tributions ΓG, nous avons supposé, par extension, que ces attributs sont eux-mêmes modélisables par ΓG.

Les modélisations des histogrammes de niveaux de gris sont données en Fig. 3.8, et montrent que le mélange de distributions de ΓG est adéquat car la modélisation est satisfaisante. Nous avons illustré ici la modélisation des statistiques des images texturées extraites par MCNG, car comme nous le verrons dans la sous-partie 4.5.1 du chapitre 4, nous avons privilégié de tels attributs de texture dans notre travail de thèse.

3.3 Modélisation des statistiques jointes par utilisation de copules

Nous avons présenté dans la partie 3.1 comment les FDP marginales de chacune des images initiales sont estimées grâce à des mélanges finis. Cependant, comme nous le verrons par la suite, nous sommes amenés le plus souvent à considérer simultanément un jeu d'images en entrée à une même résolution. Typiquement, une image RSO et son attribut de texture, ou encore un jeu d'images RSO/optique. Il faut donc construire un modèle de FDP jointe prenant en compte de manière optimale les d FDP marginales, où d est le nombre de bandes disponibles en entrée.

Une première possibilité est de considérer des lois multivariées. Pour plusieurs bandes optiques, les distributions gaussiennes multivariées sont couramment utilisées [Hardie 2004]. Dans des contextes d'images RSO polarimétriques, des lois plus spécifiques, de préférence non gaussiennes, ont été utilisées telles que la distribution de KummerU [Bombrun 2011] par exemple.

Peu de modèles statistiques existent pour modéliser conjointement les données optiques et RSO. Cependant, quelques-uns ont été proposés, notamment dans [Storvik 2003], où la FDP conjointe est égale au produit des FDP marginales des bandes RSO et optiques multipliées par un terme de corrélation, fonction des statistiques des images préalablement projetées dans un espace bien choisi. Une FDP jointe a également été proposée dans [Pellizzeri 2002], sous hypothèse que les intensités des images radars et optiques soient non corrélées. Dans ce modèle, les intensités des deux capteurs sont modélisées par une loi log-normale, et la FDP jointe est modélisée comme étant le produit de ces lois marginales. Ce modèle peut être discuté, d'une part car les données étant acquises sur le même site conservent intrinsèquement une certaine dépendance, et d'autre part par le manque de généralité du modèle des FDP marginales dans ce cas. Dans [Lombardo 2003], la FDP jointe est considérée comme suivant une loi gaussienne multivariée, sous condition que les logarithmes des intensités des images RSO soient pris comme observations, partant du principe que les images RSO sont modélisables par des lois log-normales. Dans [Chabert 2011] est introduite une modélisation par système de Pearson bivarié, qui permet d'estimer la probabilité jointe relative à une image RSO et une image optique. Des modèles non paramétriques ont également été proposés, tel que l'utilisation d'arbres de dépendance [Datcu 2002], qui modélise une FDP définie dans une

dimension d comme un produit de $d-1$ FDP de dimension 2.

Pour combiner les marginales, nous proposons d'utiliser la théorie des copules [Nelsen 2006] qui est une théorie largement utilisée dans le milieu de la finance, et qui, dans notre cas, va permettre de modéliser la dépendance existant entre les différentes images de départ.

Définition 1 *Une copule à d dimensions est une distribution jointe multivariée définie sur* $[0;1]^d$ *telle que les distributions marginales soient uniformément distribuées sur* $[0;1]$. *Plus spécifiquement, il s'agit d'une fonction* $C : [0;1]^d \rightarrow [0;1]$, *qui satisfait les propriétés suivantes :*

1. $C(u_1, ..., u_d)$ *est croissante pour chaque composante* u_i ;
2. $C(1, 1, ...u_i, 1, ..., 1) = u_i$ *pour tout* $i = 1, ..., d$, $u_i \in [0;1]$;
3. *pour tout* $(a_1, ..., a_d)$, $(b_1, ..., b_d) \in [0;1]^d$, *avec* $a_i \leq b_i$, $\forall i \in [1;d]$, *alors*

$$\sum_{i_1=1}^{2} ... \sum_{i_d=1}^{2} (-1)^{i_1+...+i_d} C(u_{1_{i_1}}, ...u_{d_{i_d}}) \geq 0$$

où $u_{j1} = a_j$ *et* $u_{j2} = b_j$ *pour tout* $j = 1, ..., d$.

Elles ont été introduites par A. Sklar [Nelsen 2006] en 1959, dont émane le théorème :

Théorème 1 *Soient les variables aléatoires* $Y^1, ..., Y^d$ *dont la fonction de répartition jointe est* H *et les fonctions de répartition marginales sont respectivement* $F^1, ..., F^d$. *Alors il existe une fonction* C, *dite d-copule, telle que :*

$$H(y^1, ..., y^d) = C(F^1(y^1), ..., F^d(y^d)), \tag{3.7}$$

pour tout $y^1, ..., y^d$ *dans* $\mathbb{R}$. *Si les* $F^j, j \in [1;d]$ *sont continues, alors* C *est unique.*

Le but de l'utilisation des copules est de pallier le manque de modèles de probabilités jointes pour les images RSO (par exemple, pour modéliser des images multipolarisées), alors que de nombreux modèles ont déjà été considérés pour des FDP marginales (partie 3.1). Ainsi, un des principaux avantages résulte dans le fait que les marginales peuvent être modélisées librement. De ce fait, le modèle est suffisamment général pour pouvoir être appliqué à tous les types d'images, attributs de texture (partie 3.2) inclus. Nous illustrons cette généralité à la fin du chapitre 5. En outre, aucun modèle n'a jusqu'à présent été considéré pour modéliser des statistiques jointes d'amplitudes d'images RSO et de leur attribut de texture.

Un modèle alternatif a été proposé par Storvik et al. dans [Storvik 2009]. Les auteurs proposent de modéliser indépendamment les marginales des diverses images d'entrées et, ensuite, de recourir à un modèle normal multivarié. Pour ce faire, les FDP non gaussiennes sont transformées en FDP gaussiennes jointes, puis la loi multivariée est générée avant de faire la transformation inverse pour se replacer dans l'espace de départ. Dans ce cas, le modèle est apparenté à la considération bien spécifique des copules gaussiennes.

Si nous dérivons (3.7) par rapport aux d variables continues y^j de FDP f^j, nous obtenons la distribution jointe :

$$h(y^1, ..., y^d) = \prod_{j=1}^{d} f^j(y^j) \times c(F^1(y^1), ..., F^d(y^d)), \tag{3.8}$$

En pratique, les $\{f^1, ..., f^d\}$ sont les FDP marginales estimées préalablement pour la j^e image en entrée et pour la m^e classe, à savoir que, à chaque classe de la classification, correspondra une copule. $\{F^1, ..., F^d\}$ sont les fonctions de répartition correspondantes, qui ont les mêmes paramètres que les FDP respectives puisque par définition, la fonction de densité de probabilité F^j est l'intégrale sur $]-\infty; y^j]$ de sa FDP correspondante f^j. c est la densité de la copule C. Ainsi, pour chaque classe, l'estimation de la densité jointe repose entièrement sur la détermination de la famille de la copule C d'après l'équation (3.8).

L'étude des copules bivariées ($d = 2$) a été largement traitée dans la littérature. En revanche, lorsque la dimension est supérieure, les références sont beaucoup moins étendues. Nous avons traité essentiellement des copules archimédiennes [Savu 2008], qui sont un bon compromis entre une souplesse analytique et une large possibilité de modélisation [Nelsen 2006]. En outre, pour des dimensions multiples, peu de modèles non-archimédiens ont été étudiés, voire même la bibliographie dans ce domaine est quasi-inexistante. Parmi ces copules archimédiennes, nous avons sélectionné un dictionnaire de trois familles à un paramètre θ : Clayton, Ali-Mikhail-Haq (AMH) [Kumar 2010] et Gumbel. Ce choix de copules permet la modélisation d'un grand nombre de structures de dépendance [Huard 2006] et a montré une bonne capacité à modéliser les données RSO [Krylov 2011c, Mercier 2008]. Notamment, les copules de Gumbel et de Clayton ne sont pas symétriques, ce qui permet de toucher un plus grand nombre de dépendances. En outre, l'étude dans [Krylov 2011c] montre que parmi plus d'une dizaine de candidats, les copules de Clayton et de Gumbel sont les plus performantes dans le sens où elles sont un bon compromis entre complexité calculatoire et précision des résultats.

L'expression analytique de ces densités-copules, qui peut être obtenue à partir des fonctions génératrices Φ_θ, ne contient qu'un seul paramètre à estimer.

Définition 2 *Soit* $\Phi_\theta : [0; 1] \mapsto [0; \infty[$ *une fonction continue, strictement décroissante et convexe telle que* $\Phi_\theta(1) = 0$ *et* $\Phi_\theta(0) = \infty$. *Cette fonction a pour inverse* Φ_θ^{-1}. *Alors la copule* C *est archimédienne si et seulement si* Φ_θ^{-1} *est monotone sur* $[0; \infty[$. *A ce moment-là,*

$$C(u_1, ..., u_d) = \Phi_\theta^{-1}(\Phi_\theta(u_1) + ... + \Phi_\theta(u_d)). \tag{3.9}$$

Ayant connaissance de la fonction génératrice des différentes familles, il est alors possible de calculer l'expression analytique des copules multivariées (première colonne du tableau 3.3).

Cependant, l'expression analytique qui nous intéresse davantage est l'expression de la densité des copules, puisque c'est elle qui détermine la distribution jointe

TAB. 3.3 – Copules considérées, $\theta(\tau)$, et intervalles de validité pour τ et θ.

Copule	$C(u_1,...,u_d)$	$\theta(\tau)$	Intervalle de τ	Intervalle de θ
Clayton	$\left[\left(\sum_{i=1}^{d} u_i^{-\theta}\right) - d + 1\right]^{-1/\theta}$	$\theta = \frac{2\tau}{1-\tau}$	$\tau \in]0;1]$	$]0;+\infty[$
AMH	$\frac{\prod_{i=1}^{d} u_i}{1-\theta \prod_{i=1}^{d}(1-u_i)}$	$\tau = \frac{3\theta-2}{3\theta}$ - $\frac{2}{3}\left(1-\frac{1}{\theta}\right)^2 \ln(1-\theta)$	$\tau \in$ $[-0,182;\frac{1}{3}]$	$[-1;+1[$
Gumbel	$\exp\left(-\left[\sum_{i=1}^{d}(-\ln u_i)^{\theta}\right]^{1/\theta}\right)$	$\theta = \frac{1}{1-\tau}$	$\tau \in [0;1]$	$[1;\infty[$

recherchée (Eq. (3.8)). Cette dernière n'est pas toujours simple à obtenir en pratique, puisqu'il s'agit, en fait, de dériver les expressions des copules multivariées données dans le tableau 3.3 en fonction des d variables. Elles sont bien connues pour des copules bivariées, mais pas toujours immédiates à obtenir pour une dimension d non fixée. Parmi notre choix de copules, la densité la plus simple à calculer de manière explicite est la densité de la copule de Clayton :

$$c(u_1,...,u_d) = \prod_{j=1}^{d} u_j^{-(\alpha+1)} \cdot \prod_{n=0}^{d-1}(1+\alpha n) \cdot (\sum_{j=1}^{d} u_j^{-\alpha} - d + 1)^{(\frac{-1-d\alpha}{\alpha})} \quad \text{où} \quad u_j = F^j(y^j). \tag{3.10}$$

Les détails de ce calcul sont donnés en Annexe B. Pour les autres densités (AMH, Gumbel), nous avons eu recours directement à des expressions analytiques estimées par des outils informatiques tels que Matlab.

Pour estimer θ et trouver la meilleure famille de copules, nous utilisons la relation entre les copules et le τ de Kendall, $\tau \in [-1;1]$ [Nelsen 2006]. Théoriquement, le τ de Kendall est une mesure de la similarité (concordance-discordance) entre deux réalisations indépendantes (Y_1, Y_2) et $(\hat{Y_1}, \hat{Y_2})$:

$$\tau = P\{(Y_1 - \hat{Y_1})(Y_2 - \hat{Y_2}) > 0\} - P\{(Y_1 - \hat{Y_1})(Y_2 - \hat{Y_2}) < 0\}.$$

Cette relation peut être étendue à une dimension supérieure. Sa version empirique est, dans notre cas, davantage utile. Pour deux réalisations données $y_{1,l}$ et $y_{2,l}$ ($l \in [1;N]$), l'estimateur empirique du τ de Kendall est [Joe 1990] :

$$\hat{\tau} = \frac{N(N-1)}{4} \sum_{i \neq j} I[Y_{1,i} \leqslant Y_{1,j}] I[Y_{2,i} \leqslant Y_{2,j}] - 1, \tag{3.11}$$

où $I[.]$ est la fonction indicatrice, (Y_1, Y_2) sont différentes marginales bivariées, correspondant aux observations de deux canaux d'entrée différents, et N est la taille des échantillons. Lorsque nous considérons un nombre supérieur d'images en entrée (3 bandes R,V,B par exemple), nous estimons cette mesure par paire d'images, et obtenons le τ final par moyennage [Kojadinovic 2010]. Ce $\hat{\tau}$ est obtenu grâce à la

vérité de terrain qui sert de base d'apprentissage, puisque nous nous plaçons dans un contexte supervisé.

Suite à cette estimation, nous utilisons la relation générale entre le τ de Kendall et la copule C afin d'estimer pour chacune des familles du dictionnaire l'unique paramètre θ. Cette relation est donnée par le biais d'une intégrale de Lebesgue-Stieltjes [Carter 2000] :

$$\tau_{d,C} = \frac{1}{2^{d-1}-1}\left[2^d \int_{[0;1]^d} C(\mathbf{u})dC(\mathbf{u}) - 1\right]. \tag{3.12}$$

Grâce à cette relation, nous pouvons calculer l'expression du paramètre θ en fonction du $\hat{\tau}$ empirique. Cette expression s'avère indépendante de la dimension d de la copule [Nelsen 2006, Joe 1990].

Pour chacune des familles du dictionnaire (Tab. 3.3), nous regardons ensuite si le paramètre θ estimé est effectivement dans l'intervalle de définition de ce paramètre pour une famille donnée, et écartons les familles pour lesquelles ce n'est pas le cas.

Enfin, pour chaque classe m, nous choisissons la meilleure famille de copules parmi le dictionnaire prédéfini, soit celle qui conduit à la plus haute p-valeur dans le test de Pearson du χ^2 [Lehmann 2007, Krylov 2011c]. L'hypothèse nulle dans ce test est que les fréquences d'échantillonnage $C_m(F_m^1(u^1), ..., F_m^d(u^d))$, où $(u^1, ..., u^d)$ représentent les données observées, soient consistantes avec les fréquences théoriques $C_m(v^1, ..., v^d)$ pour chaque copule du dictionnaire.

Un modèle similaire à celui proposé ci-dessus a été employé dans [Mercier 2009], à savoir que la probabilité jointe est le produit des probabilités marginales par une densité de copule. Les auteurs n'utilisent pas, comme dans notre cas, une densité de copule de dimension d directement dérivée de copules multivariées, mais considèrent les inter-dépendances des bandes d'entrée deux à deux, pour former une densité de copule à d dimensions comme produit de $\frac{d(d-1)}{2}$ copules bivariées, plus faciles à manier. La principale différence entre notre parti pris et celui des auteurs de [Mercier 2009] est que dans notre cas, nous n'avons qu'un seul paramètre à estimer. En revanche, le choix des copules bivariées est plus large que celui des copules multivariées de par le fait que ces dernières n'ont pas fait l'objet d'études aussi étendues. En outre, le modèle proposé dans [Mercier 2009] permet de contourner le fait que le choix d'une construction d'une d-copule, où $d \geq 4$, est assez discutable puisqu'il consiste à rendre circulaire les dépendances entre composantes. Cependant, en pratique, nous n'avons pas été amenés à travailler sur un grand nombre de bandes d'entrée, comme cela serait le cas si nous traitions des images multi-spectrales ayant un grand nombre de bandes, qui est un autre problème que celui traité au cours de cette thèse.

3.4 Conclusion

Dans ce chapitre, nous avons établi les modèles mathématiques des images initiales, qu'elles soient RSO ou optiques. Ces modèles mathématiques forment l'étape

d'apprentissage de l'algorithme. Chaque classe de chaque bande d'entrée est modélisée par une fonction de densité de probabilité, elle-même mélange de fonctions de densité de probabilité dont la famille et les paramètres sont déterminés automatiquement par un algorithme stochastique d'espérance-maximisation (SEM) adapté. Ensuite, pour diverses bandes d'entrée à une même résolution, les FDP marginales estimées sont combinées mathématiquement en ayant recours à des copules multivariées, qui sont une modélisation de l'inter-dépendance des bandes considérées, et constituent une modélisation générale dans le sens où elles prennent en compte des cas de dépendance et d'indépendance (densité de copule égale à 1) des bandes d'entrée. Les bandes d'entrée à une même résolution peuvent être diverses bandes d'une même acquisition, typiquement les bandes RVB optiques ou encore les bandes polarimétriques RSO. Elles peuvent aussi être la combinaison d'une image RSO monopolarisée avec son attribut de texture. Dans tous les cas, ces images sont des acquisitions recalées d'une même zone.

Ces modèles mathématiques sont, en fait, la modélisation de la vraisemblance, qui va être intégrée dans nos modèles bayésiens. Cette étape est donc cruciale pour la suite des expérimentations, puisqu'elle constitue une base pour les approches méthodologiques développées dans les chapitres 4 et 5 de ce manuscrit.

Chapitre 4

Classification supervisée d'images RSO monorésolution

Sommaire

Les bonnes propriétés des images RSO jouent un rôle important dans la gestion d'aléas, en permettant d'établir des cartographies d'utilisation et/ou de couverture des sols, ou de zones endommagées par des phénomènes naturels tels que des inondations ou des tremblements de terre [Boni 2007]. Dans le cadre d'une première approche face aux risques environnementaux, nous proposons d'établir des classifications d'images RSO contenant des zones urbaines, puisqu'elles sont susceptibles d'être le plus affectées lors de catastrophes naturelles. La classification est un processus qui vise à attribuer à chaque pixel de l'image une classe d'appartenance. Les classes considérées pour la classification peuvent être regroupées selon deux grands types : les classes thématiques, et les classes physiques. Au vu des acquisitions que nous avons été amenés à traiter, nous avons opté pour une classification thématique, typiquement, végétations, zones urbaines, bien que nous ayons introduit également une classification physique en intégrant une classification des zones d'eau, de par leur potentialité à être catégorisées comme classe thématique.

Nous utilisons les notations déjà introduites dans le chapitre précédent, à savoir que chaque classe est notée m et est comprise dans l'ensemble $[1; M]$. L'ensemble des observations est noté Y et l'ensemble des étiquettes X. Connaissant les observations Y, nous cherchons à déterminer les étiquettes X en chacun des sites (ici, des pixels). Dans ce chapitre, les observations sont un ensemble d'images recalées d'une même zone à une même résolution (d'où le terme monorésolution). En général, les expérimentations seront effectuées sur une image RSO avec ou sans son attribut de texture (voir partie 3.2 du chapitre précédent), ou bien encore sur une image optique RVB.

Nous proposons d'établir, dans un premier temps, un état de l'art de la classification d'images monorésolution, préférentiellement monopolarisées, en mettant un certain accent sur les méthodes markoviennes, puisque ces méthodes nous intéressent davantage, de par leur adaptabilité avec des critères bayésiens, et aussi car la prise en compte contextuelle est un réel avantage pour des images aussi bruitées que les images RSO. L'état de l'art étant très étendu, il est malheureusement difficile d'établir une liste exhaustive des méthodes publiées. Nous avons tenté de donner la liste la plus complète possible, tout au moins pour les méthodes markoviennes. D'autres ouvrages [Richards 2006] ou papiers [Lu 2007] traitent aussi du large choix de méthodologies pour classifier des images RSO, en incluant aussi des études sur le traitement d'images de télédétection optiques.

Nous présentons, ensuite, une partie plus technique qui concerne les champs de Markov tels que nous les avons exploités durant la thèse dans un cadre monorésolution, et décrivons comment nous intégrons dans ce modèle les statistiques des images et des attributs de texture, détaillées dans le chapitre précédent (Chap. 3). Nous présentons aussi des méthodes de classification de référence, couramment étudiées dans la littérature, afin d'évaluer les résultats de classification obtenus par nos soins. Après une brève partie qui vise à lister les méthodes possibles pour évaluer les classifieurs, nous présentons des résultats obtenus sur des images réelles, et publiés.

4.1 État de l'art

Depuis plusieurs décennies maintenant, de nombreuses méthodes ont été proposées pour permettre la classification d'images RSO. Celles-ci deviennent de plus en plus spécifiques, notamment au vu de l'émergence de capteurs toujours plus performants, notamment avec des résolutions de plus en plus fines. Elles constituent une discipline à part entière de par le fait de la spécificité de ces images, qui rend la plupart des méthodes de classification valides dans le cas optique obsolètes dans le cas radar notamment à cause du bruit de chatoiement. Les méthodes de classification peuvent être séparées selon deux groupes, les méthodes supervisées et non supervisées, c'est-à-dire, avec ou sans apprentissage. Récemment, un troisième groupe intermédiaire a émergé : les méthodes semi-supervisées.

4.1.1 Les méthodes supervisées

Les méthodes supervisées font appel à des techniques très variées, plus ou moins générales dans le sens où certaines ne sont pas spécifiques au traitement d'images RSO. Commençons par quelques méthodes générales. Les réseaux de neurones, technique qui divise la communauté scientifique par son manque de validation théorique, ont été employés avec succès. Les réseaux de Hopfield [Jacob 2002, Xue 2003] ont donné de bons résultats dans le cadre de la segmentation par rapport à d'autres méthodes telles que le seuillage, car ils présentent une certaine robustesse au bruit de chatoiement. Les séparateurs à vaste marge (SVM) [Vapnik 2000, Su 2004] ont été largement exploités pour tous les types d'images de télédétection, qu'elles soient optiques, RSO, monobandes ou multibandes. Ceux-ci sont davantage détaillés dans la partie 4.3.

Un autre technique possible est le recours à une segmentation par ligne de partage des eaux (*watershed*). Par exemple, des résultats encourageants ont été obtenus en la combinant à la connaissance supposée de la statistique du bruit de chatoiement [Martins 2006].

Plus spécifiquement, des méthodes bayésiennes fondées sur des champs de Markov avec modélisation des statistiques des images sont utilisées depuis de nombreuses années. Au début des années 90, des modèles markoviens à deux niveaux [Derin 1990, Rignot 1991] ont été utilisés en vue de modéliser d'une part les régions de l'image idéale (non bruitée) et d'autre part le bruit inhérent de l'image par des statistiques différentes, ensuite intégrées dans des algorithmes de type maximum a posteriori (MAP), modes conditionnels itérés (ICM) ou autres, que nous serons amenés à détailler ultérieurement (sous-partie 4.2.2.2). A la fin des années 90, les champs de Markov gaussiens ont été largement utilisés. Les statistiques des coefficients de rétrodiffusion sont choisies comme des lois gaussiennes multivariées, et les étiquettes sont attribuées en ayant recours à des calculs de statistiques du premier et du deuxième ordre ainsi que des estimations de paramètres de textures [Dong 1999].

Un modèle de Markov structuré en arbre (*tree-structured Markov random field*) a été introduit par les chercheurs de l'université de Naples [D'Elia 2003] et par la suite

étendu dans [Poggi 2005, Liu 2009]. Un tel modèle vise à décrire la structure cachée des données (étiquettes) grâce à une séquence de champ de Markov binaires. Plus clairement, l'ensemble des classes est partitionnée en structure binaire, où chaque noeud de la structure est lié à un paramètre du champ de Markov. Par exemple, l'ensemble des classes peut être sous-divisée une première fois en zones d'eau/zones sèches, les zones sèches en sable/autre etc. Les structures binaires sont bien connues pour être plus rapides au niveau calculatoire en informatique.

Plus généraux que les modèles structurés en arbres, les champs de Markov hiérarchiques sont couramment exploités. Ils peuvent être combinés à la technique de classification fondée sur les sacs-de-mots ou bien des sacs-d'attributs (*bags-of-feature*) [Yang 2009], c'est-à-dire par apprentissage de portions d'images intégrées dans un dictionnaire. Pour de la reconnaissance de textes, cela serait lié à un apprentissage par mots, inclus ensuite dans un dictionnaire sémantique. L'image-test est subdivisée, puis chaque sous-image (ou les attributs de texture de ces imagettes) est comparée au dictionnaire puis classifiée. Un contexte markovien est ensuite pris en compte pour la recombinaison des résultats de classification de chaque sous-image en une carte de classification finale. Dans un contexte plus spécifique de détection de zones urbaines, il est aussi possible d'utiliser des sacs-de-mots sans nécessairement les prendre en compte avec un modèle markovien [Weizman 2008]. Dans ce cas, il s'agit simplement de comparer les sous-divisions de l'image avec un dictionnaire spécifiquement généré pour les zones urbaines. Les champs de Markov hiérarchiques ont aussi été combinés à la théorie de Dempster-Shafer. Par exemple, dans la méthode proposée dans [Foucher 2002], cette théorie permet de combiner les informations contenues à chaque niveau de l'arbre hiérarchique, qui correspondent à chaque niveau de décomposition de l'image initiale par ondelettes [Mallat 2008, Daubechies 1988]. Dans un contexte hiérarchique, les ondelettes sont particulièrement adaptées. En dehors du champ de Markov hiérarchique, les ondelettes ont aussi été intégrées dans un arbre de Markov caché (*Hidden Markov tree (HTM)*). Dans ce cas, le modèle est un arbre probabiliste qui prend en compte les propriétés statistiques des ondelettes [Choi 2001]. La classification est faite à chacun des niveaux, puis combinée en utilisant des techniques de fusion bayésiennes inter-échelles afin de peaufiner les mauvaises classifications du niveau le mieux résolu dues au chatoiement. Une version non supervisée de cet algorithme a été proposée dans [Qing 2004].

Des techniques de classification bien spécifiques ont aussi été envisagées, qui donnent de bons résultats dans le cadre d'applications bien précises. Les contours actifs ont été exploités pour la séparation entre zones d'eau et zones terrestres (typiquement, séparations côtières) [Bates 2000, Silveira 2009] ; un rectangle de départ est sélectionné dans la zone d'eau, puis la forme de ce rectangle évolue afin de minimiser une certaine énergie, qui caractérise la limite entre terre et mer, limite modélisée mathématiquement par une courbe. La connaissance de départ est la modélisation des statistiques des classes. La séparation entre zones de basse végétation et forêts a été étudiée dans [Fosgate 1997], zones pour lesquelles la vraisemblance est donnée en ayant recours à un modèle auto-régressif, et la séparation entre les classes se fait par règle de décision bayésienne. Les images RSO ont été largement

exploitées pour la détection et le suivi de marées noires, qui est une classification spécifique, puisqu'il s'agit de détecter une tache d'hydrocarbure dans une étendue d'eau, qui apparaît en pratique comme une zone plus foncée. Pour procéder à une telle classification, des modèles markoviens ont été exploités, souvent de manière hiérarchique [Derrode 2007, Morales 2008].

Outre le cadre spécifique des champs de Markov hiérarchiques précédemment évoqués, de nombreuses méthodes hiérarchiques plus générales ont été employées en vue de classifier des images monorésolution. Dans [Wen 2009], une représentation en quad-arbre est exploitée, pour laquelle les différents niveaux sont obtenus par moyennage cohérent des pixels de niveaux inférieurs (le niveau le plus bas étant l'image RSO). L'observation à la meilleure résolution, et de par là même la vraisemblance correspondante, est donnée par un modèle auto-régressif combinant les pixels à différents niveaux dont les paramètres sont estimés par une technique de *bootstrap*. Les pixels sont classifiés par relation entre l'estimation des paramètres et le niveau de confiance en chacune des classes [Wen 2009], ou bien en ayant recours à un champ de Markov [Zhang 2009a]. Le *bootstrap* est une technique de rééchantillonnage qui permet d'estimer l'écart entre l'erreur d'apprentissage (risque empirique) et l'erreur de généralisation (risque fonctionnel) [Efron 1994].

4.1.2 Les méthodes non supervisées

Moins approfondie que pour le mode supervisé, la bibliographie des méthodes sans apprentissage est somme toute assez conséquente, de par son côté pratique, puisqu'aucune vérité de terrain n'est nécessaire. Des méthodes variées ont été considérées. Parmi elles, nous pouvons citer le tracé de contour combiné à l'utilisation d'informations de texture et d'intensité d'images RSO monobandes [Chamundeeswari 2007]. L'image est préalablement divisée en blocs, puis chaque bloc est identifié, d'après sa moyenne et variance comme uniforme, texturé, ou bordure. Les blocs similaires sont ensuite connectés entre eux, de façon à former des régions, et enfin une classification des K-moyennes (K-*means*) est appliquée pour déterminer la classe des régions. Plus dans un objectif de segmentation que de classification, des méthodes fondées sur des grilles (*grid*) ont aussi été développées [Galland 2009]. L'image initiale est sub-divisée selon une grille uniforme, puis en déformant ce maillage (rajout de noeuds/fusion), les auteurs cherchent à minimiser une fonction énergie, qui dépend des paramètres de l'image, du maillage, mais aussi des fonctions de densité de probabilité des classes (par exemple, choisies comme des distributions de Fisher).

Plus proche de la méthode que nous allons détailler en partie 4.2, l'utilisation des mélanges finis dans un cadre non supervisé a été adoptée par Peng et al. [Peng 1995], fondé sur des mélanges de gaussiennes. Les auteurs ont ensuite adapté des algorithmes amplement utilisés, de type EM, afin d'estimer les paramètres des gaussiennes ainsi que les probabilités a priori locales. La combinaison mélanges finis / champs de Markov a déjà été proposée de manière non supervisée, dans le cadre de statistiques dans le système de Pearson, dans [Delignon 1997]. Dans l'applica-

tion spécifique de la classification non-supervisée d'images RSO, des mélanges finis de distributions K et Gamma ont été intégrés dans des champs et des chaînes de Markov cachés [Fjortoft 2003]. La chaîne est obtenue par scannage de l'image selon la méthode de Hilbert-Peano, et son utilisation est adéquate dans le sens où le paramètre de régulation est plus facile à estimer que dans le cas d'un champ de Markov. En revanche, les frontières inter-classes ne sont pas très précises dans ce cas. Les auteurs de l'article [Fjortoft 2003] ont, par ailleurs, proposé une méthode hybride, palliant les défauts des chaînes et des champs. Dans le cas multibandes, les chaînes de Markov multivariées [Brunel 2010] ont été proposées. Elles intègrent des distributions T et K, dont les paramètres sont estimés par algorithme d'espérance-maximisation (EM).

Une première application des champs de Markov est la considération d'un champ de Markov sur la paire observations/étiquettes (Y, X) [Pieczynski 2000], qui a certains avantages au niveau calculatoire et au niveau de la modélisation. Plus évolués, les champs de Markov triplet proposent une modélisation de la paire (Y, X) comme une distribution marginale du champ de Markov (Y, U, X), où U est un processus auxiliaire [Benboudjema 2005].

4.1.3 Les méthodes semi-supervisées

Depuis quelques années, les méthodes par apprentissage (*active learning*) connaissent un réel essor. Ce sont, en général, des méthodes bien formalisées fondées sur des techniques dites semi-supervisées, c'est-à-dire que l'algorithme procède à une détermination des pixels, puis les pixels les plus incertains sont donnés par l'utilisateur (apprentissage) de façon à améliorer l'algorithme. Ainsi, ces méthodes minimisent le nombre de pixels à étiqueter pour l'apprentissage en choisissant uniquement les plus incertains [Tuia 2009, Demir 2011, Tuia 2012]. Pour la prédiction des étiquettes des pixels, de nombreuses méthodes non supervisées peuvent être envisagées comme celles listées dans les sous-parties 4.1.1 et 4.1.2, par exemple via des méthodes de type SVM [Tuia 2009]. Théoriquement, la considération de méthodes semi-supervisées aboutit à de meilleurs résultats, avec un apprentissage sur un nombre inférieur de pixels. Cependant, la plupart de ces méthodes intègrent de nombreux paramètres, qui sont en général estimés empiriquement.

4.2 Méthode proposée

La méthode supervisée de classification bayésienne que nous proposons dans ce manuscrit se divise en deux étapes principales, et entre dans la lignée des méthodes supervisées fondées sur des champs de Markov, présentées dans l'état de l'art (sous-parties 4.1.1). La première étape consiste à modéliser les statistiques de l'amplitude de l'image RSO et de son attribut de texture pour chaque classe considérée pour la classification via la méthode détaillée dans le chapitre 3. La deuxième étape de la méthode utilisée consiste à générer la classification bayésienne à partir de l'apprentissage des statistiques conjointes de chacune des classes. Pour augmenter la

robustesse face au bruit de chatoiement, nous avons considéré un modèle contextuel fondé sur des champs de Markov (CM) [Besag 1974, Dubes 1989]. De tels champs intègrent un contexte spatial, c'est-à-dire que les étiquettes des pixels voisins du pixel à classifier sont prises en compte comme a priori, en considérant que le pixel à classifier a une forte probabilité d'appartenir à la même classe que ses voisins.

La nouveauté du modèle proposé par rapport à des méthodes bayésiennes fondées sur des champs de Markov détaillées dans l'état de l'art (partie 4.1), elles-aussi combinées à une modélisation statistique, est multiple :

- un choix de modélisation statistique plus souple, et particulièrement adaptée aux types d'images traitées ;
- l'introduction d'un modèle contextuel de texture ;
- une modélisation des densités jointes RSO/attribut de texture par utilisation des copules.

Cette méthode a été publiée dans la conférence SPIE Remote Sensing 2010 (Annexe E.1), et a servi de comparaison dans des papiers publiés ultérieurement.

4.2.1 Champs de Markov

Dans cette sous-partie, nous faisons quelques rappels afin de décrire les modèles markoviens appliqués à des problématiques de traitement d'images.

Un champ de Gibbs discret [Besag 1974] est un champ tel que la fonction de probabilité de masse soit définie par une fonction du type

$$p(X = x) = \frac{\exp(-H(x))}{Z}, \tag{4.1}$$

où $H(.)$ est appelée fonction énergie et Z est une constante de normalisation. Cette constante n'est, en général, pas calculable sur tout l'ensemble des réalisations X à cause du très grand nombre de configurations possibles.

C'est alors qu'entre en considération l'hypothèse markovienne, qui est fondée sur le voisinage des pixels. La probabilité conditionnelle locale en un site (pixel) s n'est fonction que de la configuration de son voisinage, ce qui se traduit formellement par

$$p(x_s = \omega_m | x_t, t \neq s) = p(x_s = \omega_m | x^{(s)}),$$

où s est le site courant ($s \in S$), x_s l'étiquette du site, $x^{(s)}$ la configuration en dehors du site s telle que $x^{(s)} = \{x_t, t \neq s, t \sim s\}$ et $t \sim s$ signifie que t et s sont des pixels voisins. X est donc de ce fait un processus markovien.

D'après le théorème de Hammersley-Clifford [Besag 1974], il est possible de définir une caractéristique locale (probabilité conditionnelle) pour chacune des classes considérées $m \in [1; M]$, donnée par :

$$p(x_s = \omega_m | x^{(s)}) = \frac{\exp(-H(x_s = \omega_m | x^{(s)}))}{\sum_{j=1}^{M} \exp(-H(x_s = \omega_j | x^{(s)}))} \tag{4.2}$$

Ce théorème est valable seulement si aucune configuration n'est interdite, soit $p(x_s = \omega_m|x^{(s)}) > 0$. Lorsqu'il est souhaitable de supprimer certaines configurations interdites, telles que des routes à l'intérieur de zone d'eau par exemple, des solutions ont été proposées [Li 1995].

Nous travaillons ici avec un voisinage isotrope du deuxième ordre, c'est-à-dire que seules les cliques de taille 2 sont considérées. Il s'agit donc de la prise en compte d'un voisinage composé de 8 pixels autour d'un pixel de référence. Une *clique* est un ensemble de sites dans lequel les paires de sites sont des voisins mutuels. À chaque clique c est associée un potentiel V_c, et la fonction énergie s'exprime alors comme :

$$H(x_s = \omega_m|x^{(s)}) = \sum_{c \in Q} V_c(x_s = \omega_m|x^{(s)}),$$

où Q représente l'ensemble des cliques associées au site s. Le potentiel est, dans notre cas, choisi comme un modèle de Potts, soit, pour des cliques de taille 2,

$$V_{c=\{s,t\}}(x_s, x_t) = -\beta\delta_{x_s=x_t}, \quad \text{avec } \beta > 0, \quad \text{où } \delta_{x_s=x_t} = \begin{cases} 1, & \text{si } x_s = x_t \\ 0, & \text{sinon} \end{cases}. \tag{4.3}$$

Ainsi, avec (4.1) et (4.3), nous pouvons définir une caractéristique locale pour chaque site :

$$p(x_s) = \frac{1}{Z}\exp(-\beta \times \sum_{t:\{s,t\}\in C} \delta_{x_s=x_t}), \tag{4.4}$$

où nous rappelons que s désigne un site, x_s l'étiquette attribuée à ce site et x_t tel que t et s appartiennent à la même clique.

Le modèle choisi est donc isotrope et homogène, car le paramètre β ne dépend ni de l'orientation des cliques, ni de leur localisation.

4.2.2 Les méthodes d'optimisation

Étant donné que l'on cherche la configuration des étiquettes x connaissant les observations y, le problème se résout à l'aide de champs de Markov cachés. Grâce au théorème de Bayes [Barnard 1958], étant donné les probabilités a priori (sous-partie 4.2.1) et les vraisemblances (Chap. 3) définies précédemment, nous pouvons formuler une probabilité a posteriori pour chaque classe m :

$$p(x|y) \propto p(x) \times p_m(y|x). \tag{4.5}$$

En supposant que les vraisemblances sont indépendantes, soit :

$$p_m(y|x) = \prod_{t \in S} p_m(y_t|x_t),$$

nous pouvons alors définir une distribution a posteriori pour x telle que

$$p(x|y) = \frac{\exp(-H(x|y,\beta))}{Z}, \tag{4.6}$$

où Z est une constante de normalisation et $H(.)$ une fonction énergie qui s'exprime en fonction d'un seul paramètre $\beta > 0$:

$$H(x|y,\beta) = \sum_{t \in S} \left[-\log p_m(y|x) - \beta \sum_{s' : \{s',t\} \in C} \delta_{x_{s'} = x_t} \right]. \tag{4.7}$$

Une telle distribution est donc une distribution de Gibbs, et, ainsi, le champ des étiquettes conditionnellement aux observations est un champ de Markov (cf. théorème d'Hammersley-Clifford). Il est alors possible de simuler ce processus, par exemple à l'aide d'un recuit simulé (détaillé ultérieurement) qui maximise la probabilité a posteriori.

Nous nous sommes limités pour l'a priori à ce modèle de Potts, somme toute assez simple, mais qui aboutit, en pratique, à des résultats satisfaisants. Des modèles plus complexes ont aussi été étudiés dans la littérature, comme par exemple dans [Celeux 2003], où est intégré un champ dit externe, qui est un système comparable à celui des proportions dans les mélanges finis. Il permet d'intégrer implicitement un poids sur les différentes classes lors de la segmentation. Le modèle d'éléments finis employé dans l'expression de la vraisemblance est suffisamment flexible pour éviter d'avoir à prendre en compte des modèles de mélanges au niveau des a priori. En effet, l'estimation d'un seul paramètre β du modèle a priori pour lequel nous avons opté est déjà assez délicate (voir 4.2.2.1). Typiquement, un tel modèle de champ externe implique l'estimation d'autres paramètres du champ (paramètres des mélanges), augmentant ainsi énormément la complexité calculatoire.

4.2.2.1 Estimation du paramètre β du champ de Markov

Nous pouvons noter la présence d'un paramètre β, appelé inverse-température, nécessaire à la détermination de l'a priori d'après l'équation (4.4). En théorie, connaissant une base d'apprentissage – par exemple, celle utilisée pour la détermination des vraisemblances statistiques – il est possible de déterminer la valeur du paramètre β de manière automatique.

Plusieurs méthodes existent, la plus commune étant l'algorithme de type Ho-Kashyap [Serpico 2006]. Le paramètre est estimé à la fois grâce à une vérité de terrain, mais aussi grâce à l'estimation préliminaire d'une carte de classification. Nous avons souhaité éviter cette estimation préalable. Une des méthodes particulièrement adaptée alors est l'exploitation du calcul de la vraisemblance, la contrainte majeure dans ce cas étant que la fonction de partition Z de l'équation (4.6) n'est pas connue de manière explicite. La vraisemblance peut être estimée par algorithme MCMC (*Monte Carlo Markov Chain*), comme dans [Descombes 1999] par exemple. Un tel algorithme est suffisamment général pour permettre d'être utilisé dans des modèles à forte dépendance. Nous avons choisi d'avoir plutôt recours à une maximisation de la pseudo-vraisemblance PL, qui est plus simple à calculer que la vraisemblance

globale. La PL est définie à partir des caractéristiques locales de l'équation (4.2) :

$$\log PL(x|\beta) = \log \left[\prod_{s \in S} p(x_s|y) \right]. \tag{4.8}$$

Pour optimiser cette fonction, un algorithme de recuit simulé s'avère efficace [Geman 1984].

Description de l'algorithme de recuit simulé pour l'estimation du paramètre β

Le recuit simulé s'appuie sur l'algorithme de Metropolis-Hastings [Hastings 1970], qui permet de décrire l'évolution d'un système thermodynamique. Par analogie avec le processus physique, la fonction à minimiser est appelée l'énergie E du système, ici telle que $E = \log PL(x|\beta)$.

Nous partons d'une solution initiale comprise dans l'espace des solutions – l'espace des β dans notre cas – puis nous cherchons à améliorer cette estimation en faisant baisser l'énergie E du système. À chaque itération de l'algorithme, une modification élémentaire du paramètre β est effectuée. Cette modification entraîne une variation ΔE de l'énergie du système. Si cette variation est négative (c'est-à-dire qu'elle fait baisser l'énergie du système), elle est appliquée à la solution courante. Sinon, elle est acceptée avec une probabilité $e^{-\frac{\Delta E}{T}}$. Cette technique d'optimisation s'appelle *dynamique de Metropolis*. L'acceptation d'une "mauvaise" solution permet d'explorer une plus grande partie de (voire tout) l'espace des solutions et tend à éviter de s'enfermer trop vite dans la recherche d'un optimum local. Nous introduisons également un paramètre fictif, la température T du système, fixée initialement à une valeur T_0 élevée, puis diminuée itérativement. Nous avons opté pour une diminution continue de celle-ci, la plus couramment utilisée étant $T_{t+1} = \lambda T_t$ avec $\lambda < 1$, tout en restant proche de 1, t étant le numéro de l'itération. La température joue un rôle important. À haute température, le système est libre de se déplacer dans l'espace des solutions en choisissant des solutions ne minimisant pas forcément l'énergie du système. À basse température, les modifications baissant l'énergie du système sont choisies, mais d'autres peuvent être acceptées, empêchant ainsi l'algorithme de tomber dans un minimum local.

L'algorithme du recuit simulé est donc un algorithme itératif qui construit la solution au fur et à mesure, dont le déroulement est le suivant :

1. choix initial de la température T_0 et du paramètre β_0 ;
2. à l'itération t, tant que le nombre maximum d'itérations n'est pas atteint et que la variation d'énergie est assez élevée, faire :
   a. modification élémentaire de l'ancien β, et calcul de la nouvelle énergie $E = \log PL(x|\beta)$ correspondante.
   b. si la variation de l'énergie ΔE entre le nouvel et l'ancien état de β est négative, ce nouvel état est accepté. Si elle est positive, alors le nouvel état est accepté avec une probabilité $e^{-\frac{\Delta E}{T}}$. Sinon, l'ancien état de β est conservé.
   c. diminution lente de la température : $T_{t+1} = \lambda T_t$;

Les principaux inconvénients du recuit simulé résident dans le choix des nombreux paramètres, tels que la température initiale, la loi de décroissance de la température λ, les critères d'arrêt, etc. Ces paramètres sont souvent choisis de manière empirique. Des études théoriques du recuit simulé ont pu montrer que sous certaines conditions, cet algorithme converge vers un optimum global. C'est l'une des raisons pour lesquelles un tel algorithme a été pris en compte.

Si nécessaire, cette procédure de recuit simulé peut être suivie de la méthode suggérée dans [Yu 2003], qui combine une méthode de Metropolis-Hastings avec une descente de gradient. Dans nos simulations, nous n'avons pas utilisé une telle procédure.

Cependant, en pratique, les cartes de vérité de terrain sont difficiles à établir de manière précise, surtout au niveau des frontières inter-classes. Les étiquettes des pixels frontaliers sont, en général, un mélange des classes qu'ils bordent, et prendre la décision de leur assigner une seule classe s'avère difficile. En outre, nous avons pu constater qu'une vérité de terrain exhaustive – ou du moins contenant un nombre suffisant de transitions inter-classes – était nécessaire à l'estimation du paramètre du champ β. Cela est cohérent avec le rôle de ce paramètre vis-à-vis de l'ajustement des probabilités des transitions spatiales inter-classes. Cependant, pour être le plus précis possible, nous n'avons établi de vérités de terrain que dans des zones homogènes, sans prise en compte de frontières, comme nous le montrons dans la Fig. 4.3. C'est pour cette raison qu'au lieu d'estimer le paramètre β en maximisant la pseudo-vraisemblance par recuit simulé, nous avons été contraints de fixer sa valeur empiriquement, en trouvant un compromis. En effet, celui-ci ne doit pas être trop élevé de manière à ce que les cartes de classification ne soient pas trop lissées. Si, au contraire, ce paramètre n'est pas assez élevé, l'a priori n'aura alors que peu d'influence, et la prise en compte contextuelle sera négligée au profit d'un classification plus "pixellisée" et donc moins robuste au bruit. Nous avons opté en pratique pour $\beta \in [1; 2]$, intervalle qui est un bon compromis.

Pour surpasser cette limitation liée au manque de précision des vérités de terrain, une première solution serait d'estimer le paramètre β de manière totalement non supervisée via des algorithmes de type EM (échantillonneurs de Gibbs-Metropolis) ou fondés sur des approximations de l'EM. Des méthodes EM semi-supervisées, qui utilisent à la fois des pixels étiquetés (apprentissage) et non étiquetés ont aussi été proposées, par exemple dans [Jackson 2001].

L'estimation du paramètre β pourrait aussi être intégrée directement dans la procédure de classification, comme cela a été proposé dans [Lakshmanan 1989], pour laquelle le paramètre β est mis à jour itérativement, puis intégré dans la procédure de classification. Cette procédure est très lourde calculatoirement, et est difficile à faire converger [Fjortoft 2003] ; elle n'a pas été choisie essentiellement pour ces raisons.

Dans nos expériences, nous avons considéré un modèle de base pour lequel le paramètre β est constant sur l'image, mais il est aussi possible de prendre un tel paramètre comme dépendant de la localisation des cliques [Descombes 1999] afin d'avoir un modèle plus adaptatif, et aussi plus complexe. Le choix d'un paramètre β

imposé constant est une manière économique et efficace de procéder pour l'obtention d'un a priori somme toute satisfaisant.

4.2.2.2 Recherche du Maximum a Posteriori

Pour trouver les étiquettes des pixels, et donc la carte de classification, nous cherchons à appliquer une méthode de type maximum a posteriori (MAP), qui vise, comme son nom l'indique, à trouver le maximum global de l'a posteriori. Pour des raisons pratiques, nous proposons plutôt de minimiser la fonction énergie H (Eq. (4.7)), ce qui est totalement équivalent au fait de trouver le MAP. Plusieurs algorithmes ont été employés à de telles fins, tels que le recuit simulé [Geman 1984] ou les modes conditionnels itérés (ICM) [Besag 1986, Serpico 2006]. Un autre estimateur, qui utilise une fonction de coût légèrement différente, est l'estimateur associé au critère du mode de la marginale a posteriori (MPM). Nous étudierons et utiliserons cet estimateur essentiellement dans le chapitre suivant. La principale différence avec le MAP est que le MPM fournit le maximum de l'a posteriori local, et pénalise une configuration proportionnellement au nombre de différences entre deux configurations.

Le recuit simulé

Nous avons déjà parlé de l'algorithme du recuit simulé dans la sous-partie précédente (4.2.2.1) dans le cadre de l'estimation du paramètre β. Pour rappel, l'algorithme du recuit simulé est dédié à la recherche d'une configuration d'énergie minimale d'un champ de Gibbs. L'idée d'intégrer un paramètre de température et de simuler un recuit a été initialement proposée en traitement d'images, de manière indépendante, par Kirkpatrick et al. en 1983 [Kirkpatrick 1983] et reprise par Geman et Geman [Geman 1984], qui ont mis en oeuvre l'algorithme itératif suivant (en notant t le numéro de l'itération) :

1. choix d'une température initiale T_0 suffisamment élevée et d'une configuration initiale quelconque $x^{(0)}$;
2. à l'itération t, tant que le nombre maximum d'itérations n'est pas atteint et que la variation d'énergie est assez élevée :
   a. simulation d'une configuration $x^{(t)}$ pour la loi de Gibbs d'énergie $\frac{H(x^{(t)})}{T_t}$ à partir de la configuration $x^{(t-1)}$; la simulation peut se faire par l'échantillonneur de Gibbs ou l'algorithme de Metropolis ; en général, un balayage complet de l'image à la température T_t est réalisé ;
   b. diminution lente de la température : $T_{t+1} = \frac{c}{log(1+t)}$;

La décroissance logarithmique de la température se fait à un rythme très lent. La constante c intervenant dans la décroissance dépend de la variation énergétique globale maximale sur l'espace des configurations. Au départ, toutes les configurations sont équiprobables, puis les minima énergétiques apparaissent et s'accentuent. Notons que contrairement aux algorithmes de l'échantillonneur de Gibbs et de Metropolis, qui échantillonnent selon la loi de Gibbs, et qui sont en mesure de donner

toutes les configurations possibles, les images obtenues par recuit simulé sont uniques et doivent en théorie correspondre aux minima globaux de l'énergie. Il existe une preuve de convergence de cet algorithme, qui repose à nouveau sur la construction d'une chaîne de Markov, mais qui est hétérogène cette fois-ci à cause de la variation du paramètre de température [Geman 1984]. Intuitivement, le recuit simulé permet d'atteindre un optimum global, car il accepte des remontées en énergie. Avec la décroissance de la température, ces sauts énergétiques sont progressivement supprimés au fur et à mesure que nous nous rapprochons de l'optimum global. La descente en température doit donc se faire suffisamment lentement pour que l'algorithme ne reste pas piégé dans un minimum local de l'énergie.

L'ICM

L'algorithme du recuit simulé peut être très lourd en temps de calcul puisqu'il demande la génération d'un grand nombre de configurations au fur et à mesure que la température décroît. Des algorithmes sous-optimaux sont donc souvent utilisés en pratique. Besag [Besag 1986] a ainsi proposé un autre algorithme, beaucoup plus rapide, qui converge vers un minimum local. Il s'agit de l'ICM, *Iterated Conditional Mode*, algorithme itératif modifiant de façon déterministe à chaque étape les valeurs x_s de l'ensemble des sites de l'image. Nous construisons donc, partant d'une configuration initiale $x^{(0)}$, une suite de configurations $x^{(t)}$, convergeant vers une approximation du MAP recherché. A chaque itération t, l'algorithme parcourt tous les sites de l'image, et en chaque site, les deux opérations suivantes sont effectuées :

1. calcul des probabilités conditionnelles locales, pour toutes les valeurs possibles des étiquettes (ici le nombre de classes) du site ;
2. la meilleure étiquette est celle qui maximise la probabilité conditionnelle locale en chaque site.

Le processus s'arrête lorsque le nombre de changements d'une étape à l'autre devient suffisamment faible (stabilité de l'algorithme). Nous pouvons montrer que l'énergie globale de la configuration x diminue à chaque itération. Cet algorithme, contrairement au recuit simulé, est très rapide puisqu'en pratique, une dizaine de balayages permettent d'arriver à la convergence et est peu coûteux en temps de calcul puisqu'il ne nécessite que le calcul des énergies conditionnelles locales. En contrepartie, ses performances dépendent très fortement de l'initialisation puisqu'il converge vers un minimum local. L'ICM s'apparente à une descente en gradient (l'énergie est diminuée à chaque itération) ou à un recuit simulé gelé à température nulle, et peut donc rester bloqué dans le minimum énergétique local le plus proche de l'initialisation. Le recuit simulé, au contraire, grâce au paramètre de température et aux remontées en énergie qu'il autorise, permet d'accéder au minimum global.

La dynamique de Metropolis modifiée

Les auteurs du papier [Kato 1992] ont proposé d'utiliser une dynamique de Metropolis modifiée (MMD), qui offre en général un bon compromis entre complexité

calculatoire et précision des résultats. En outre, il s'agit d'un bon compromis entre l'algorithme déterministe ICM, qui est rapide mais trouve un minimum local, et le recuit simulé, beaucoup plus lent mais qui fournit le minimum global. Nous avons utilisé en pratique un tel algorithme pour la recherche du maximum a posteriori, et donc pour l'estimation des étiquettes des images que nous souhaitons classifier.

La principale différence entre l'algorithme de Metropolis modifié et l'algorithme de Metropolis-Hastings [Hastings 1970] est au niveau du choix du seuil utilisé dans les dynamiques. Celui-ci est choisi aléatoirement à chaque itération dans la méthode de Metropolis-Hastings, alors que dans la version modifiée, ce seuil, noté α, est fixé entre 0 et 1 au début de l'algorithme, ce qui signifie simplement qu'un nouvel étiquetage n'est possible que sous la condition d'une augmentation non "excessive" de l'énergie. Ce seuil constant permet le contrôle de cette augmentation.

L'algorithme MMD se déroule comme suit :

1. sélection d'une configuration aléatoire $x^{(0)} = \omega^0$, avec une température initiale T_0 ;
2. à l'itération t, tant que le nombre maximum d'itérations n'est pas atteint et tant que l'algorithme n'a pas convergé ($\Delta H/H$ inférieur à un seuil bien choisi), faire :
   a. en utilisant la distribution uniforme, sélection d'un état global η, qui diffère exactement d'un élément par rapport à ω^t ;
   b. estimation de $\Delta H(x^{(t)}) = H(x^{(t)} = \eta) - H(x^{(t)} = \omega^t)$. η est accepté d'après la règle :

$$\omega^{t+1} = \begin{cases} \eta, & \text{si } \Delta H \leq 0, \\ \eta, & \text{si } \Delta H \geq 0 \text{ et } ln(\alpha) \leq -\frac{\Delta H}{T_t}, \\ \omega^t, & \text{sinon.} \end{cases}$$

   où α est un seuil constant, $\alpha \in]0;1[$, choisi au début de l'algorithme ;
   c. diminution lente de la température : $T_{t+1} = \lambda T_t$ où $\lambda \in [0,95; 0,99]$ en pratique ;

Les graph-cuts

Plus récemment, des méthodes plus rapides, fondées sur la théorie des graphes, et appelées *graph-cuts*, ont été implémentées en vue de minimiser la fonction d'énergie (et donc de maximiser l'a posteriori) [Boykov 2004, Kolmogorov 2004, Boykov 2001]. Ce problème de minimisation revient à trouver la coupe minimale dans un graphe. La complexité étant polynomiale et non pas exponentielle, ce type d'implémentation permet un gain en temps de calcul considérable. En outre, la coupe dans un graphe n'est pas itérative. En pratique, le temps de calcul dans le cadre de la recherche de l'énergie minimale est divisé par 50 ou 100 par rapport à l'utilisation d'un algorithme MMD, pour des résultats finaux similaires.

4.3 Autres méthodes utilisées pour la comparaison

Dans cette partie, nous présentons en détail quelques méthodes de classification de la littérature, qui sont utilisées dans la partie des résultats (4.5) à des fins de comparaison avec la méthode que nous avons détaillée dans la partie précédente (recherche du MAP avec minimisation par MMD ou graph-cuts).

4.3.1 Les K-plus proches voisins

Dans sa version de base, la méthode des K-plus proches voisins (PPV) vise à classifier de manière supervisée un pixel par rapport à un vote majoritaire sur les K pixels les plus proches dont la classe est connue [Shakhnarovish 2005]. La proximité est définie selon un critère de distance, par exemple la distance euclidienne ou encore la distance de Hamming.

Si K est trop grand, les résultats de classification obtenus sont trop lisses, de sorte que les frontières entre les classes sont moins distinctes. En revanche, si K est trop petit, il peut y avoir des mauvaises classifications dues aux indéterminations pour le vote majoritaire. Il faut donc trouver un bon compromis, si possible de manière automatique. De nombreux algorithmes ont été développés pour déterminer ce paramètre K, le plus connu étant la validation croisée [Kohavi 1995], qui est une estimation de fiabilité par échantillonnage. Le meilleur K est choisi parmi un intervalle de valeurs de K possibles.

En outre, pour éviter le calcul (lourd) de toutes les distances, et surtout la recherche systématique de tous les voisins pour tous les pixels à classifier, il est possible de recourir à des algorithmes rapides de recherche des plus proches voisins, qui sont en général des méthodes graphiques [Dubuisson 1990].

Dans nos expériences, nous utilisons cette méthode des plus proches voisins afin de déterminer les vraisemblances, au lieu d'estimer celles-ci par méthode statistique utilisant des mélanges finis (Chap. 3). Puis, de la même manière qu'expliqué dans la sous-partie 4.2.1, ces probabilités sont intégrées comme connaissances dans un champ de Markov. Nous nous réfèrerons à une telle méthode comme la méthode K-PPV-CM. Tout comme les copules permettent un modélisation conjointe des statistiques d'une image RSO et de son attribut de texture, il est aussi possible d'estimer cette probabilité conjointe grâce aux K-PPV.

4.3.2 L'algorithme ATML-CEM

La méthode décrite ici a été présentée à la conférence ICIP en 2011 (Annexe E.2). Elle combine, via des produits d'experts, les statistiques de l'image RSO, modélisées par une densité de Nakagami (voir sous-partie 3.1.2.1 du chapitre précédent), et celles de l'attribut de texture qui lui est associé. L'attribut de texture est modélisé comme un modèle autorégressif dans lequel un pixel est exprimé comme une combinaison linéaire des pixels voisins. L'erreur de régression suit une distribution t de Student indépendante et identiquement distribuée [Kayabol 2010]. Sous ces conditions, l'attribut de texture suit une distribution t. La vraisemblance de chaque classe

est donc modélisée par le produit d'une loi de Nakagami et d'une loi t, et appelle à estimer quatre paramètres (des lois) ainsi qu'un vecteur de paramètres, dont la taille dépend du nombre de pixels voisins pris en compte dans le modèle autorégressif de la texture. Typiquement, si nous prenons un voisinage de huit pixels, il y aura donc 12 paramètres à estimer au total.

La classification est obtenue en ayant recours à un algorithme de classification espérance-maximisation (CEM), qui estime de manière conjointe les étiquettes des pixels et les paramètres [Celeux 1992]. Il utilise un principe similaire à l'algorithme SEM détaillé en sous-partie 3.1.3.1 du chapitre 3, sauf que l'étape stochastique du SEM est remplacée par une étape de classification via la maximisation de l'a posteriori pour chaque pixel.

L'a posteriori, estimé de manière bayésienne, est donc proportionnel au produit de la vraisemblance par un a priori sur les classes. Cet a priori est pris comme un modèle multinomial logistique (MnL) [Krishnapuram 2005].

L'algorithme est désigné comme ATML-CEM, pour indiquer qu'il s'agit d'un algorithme fondé sur les densités de l'Amplitude et de la Texture combinées à un modèle MnL et intégrées dans un algorithme CEM. Étant dans un contexte supervisé, nous n'avons pas pris en compte la version non supervisée de l'algorithme également présentée dans cette publication.

4.3.3 Les Séparateurs à Vastes Marges

Nous pourrions consacrer un chapitre entier à discuter des séparateurs à vaste marge, cependant ce n'est pas l'objectif de cette thèse. Nous souhaitons juste rappeler le plus brièvement possible le principe de base d'un tel algorithme, qui va être utilisé comme méthode de comparaison à la fois dans ce chapitre et dans le chapitre suivant.

Les séparateurs à vastes marges sont des classifieurs issus de la théorie de l'apprentissage statistique, qui permettent de reformuler le problème de classification comme un problème d'optimisation quadratique [Vapnik 2000, Bishop 2006].

Pour ce faire, nous recherchons une *marge maximale*, la marge étant la distance entre la frontière de séparation et les échantillons les plus proches, appelés *vecteurs supports* (Fig. 4.1). Dans les SVM, la frontière de séparation est choisie comme celle qui maximise la marge. Elle sera appelée *hyperplan séparateur optimal*. Le problème est de trouver cette frontière séparatrice optimale à partir d'un ensemble d'apprentissage. Ceci est fait en formulant le problème comme un problème d'optimisation quadratique, pour lequel il existe des algorithmes connus.

Afin de pouvoir traiter des cas où les données ne sont pas linéairement séparables, l'espace de représentation des données d'entrée est transformé en un espace de plus grande dimension (voire de dimension infinie), dans lequel il est probable qu'il existe un hyperplan séparateur linéaire. Ceci est réalisé grâce à une fonction noyau, qui doit respecter certaines conditions, et qui a l'avantage de ne pas nécessiter la connaissance explicite de la transformation à appliquer pour le changement d'espace. Les fonctions noyau permettent de transformer un produit scalaire dans un espace de grande

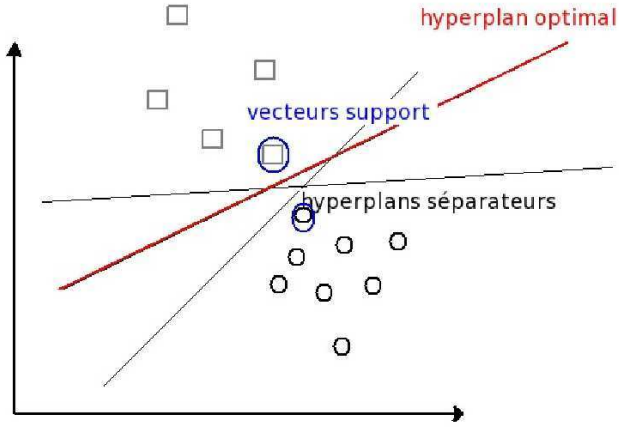


FIG. 4.1 – Échantillons d'apprentissage (ronds et carrés) correspondant à deux classes. Il existe en théorie une infinité d'hyperplans séparateurs, l'algorithme SVM doit trouver l'hyperplan séparateur optimal (en rouge), qui maximise la distance entre les vecteurs supports. Ici les échantillons, au choix, initialement linéairement séparables, ou bien représentés après projection dans un espace de représentation permettant d'établir un hyperplan séparateur linéaire.

dimension, ce qui est coûteux, en une simple évaluation ponctuelle d'une fonction. Cette technique est connue sous le nom de *kernel trick*.

Théoriquement, l'algorithme SVM résout un problème de discrimination à deux classes (discrimination binaire), c'est-à-dire $x \in \{-1, 1\}$, le vecteur d'entrée y étant dans un espace Y muni d'un produit scalaire. Nous pouvons prendre par exemple $Y = \mathbb{R}^N$, où N est le nombre d'observations. La résolution du problème de classification passe par la construction d'une fonction h, qui, à un vecteur d'entrée y, fait correspondre une sortie x :

$$x = h(y).$$

Dans le cas simple d'une fonction discriminante linéaire, h est obtenue par combinaison linéaire du vecteur d'entrée $y = (y_1, ..., y_N)^T$, avec un vecteur de poids $w = (w_1, ..., w_N)$:

$$h(y) = w^T y + w_0,$$

où T désigne la transposée et w_0 est un scalaire. La frontière de décision $h(y) = 0$ est un hyperplan appelé *hyperplan séparateur*, et il est décidé que y est de classe $x = 1$ si $h(y) \geq 0$ et de classe $x = -1$ sinon. Puisque nous sommes dans un contexte supervisé, la fonction h est déterminée par apprentissage.

La marge est la distance entre les échantillons d'apprentissage et l'hyperplan séparateur qui satisfasse la condition de séparabilité, à savoir :

$$x_z(w^T y_z + w_0) \geq 0.$$

La distance d'un échantillon y_z à l'hyperplan est donnée par sa projection orthogonale sur l'hyperplan :

$$\frac{x_z(w^T y_z + w_0)}{||w||}.$$

L'hyperplan séparateur recherché est donc donné par :

$$\arg\max_{w,w_0} \left\{ \frac{1}{||w||} \min_z \left[x_z(w^T y_z + w_0) \right] \right\}.$$

Afin de faciliter l'optimisation, nous choisissons de normaliser w et w_0, et la formulation dite primale des SVM s'exprime alors sous la forme suivante :

$$\text{Minimiser} \quad \frac{1}{2}||w||^2 \quad \text{sous les contraintes} \quad x_z(w^T y_z + w_0) \geq 1 \tag{4.9}$$

Ceci peut se résoudre par la méthode classique des multiplicateurs de Lagrange [Bertsekas 1996], que nous ne détaillons pas ici.

La notion de marge maximale et la procédure de recherche de l'hyperplan séparateur telles que présentées pour l'instant ne permettent de résoudre que des problèmes de discrimination linéairement séparables. C'est une limitation sévère qui condamne à ne pouvoir résoudre qu'un certain nombre de problèmes. Afin de remédier à l'absence de séparateur linéaire, l'idée des SVM est de reconsidérer le problème dans un espace de dimension supérieure, éventuellement de dimension infinie. Dans ce nouvel espace, il est alors probable qu'il existe une séparatrice linéaire. Une transformation non-linéaire ϕ est alors appliquée aux vecteurs d'entrée y. L'espace d'arrivée $\phi(Y)$ est appelé *espace de redescription*. Dans cet espace, nous cherchons alors l'hyperplan

$$h(y) = < w, \phi(y) > + w_0.$$

La nouvelle formulation du problème introduit un produit scalaire entre vecteurs dans l'espace de redescription, noté $< ., . >$, de dimension élevée, ce qui est coûteux en termes de calculs. Pour résoudre ce problème, nous utilisons une astuce connue sous le nom de *kernel trick*, qui consiste à utiliser une fonction noyau, qui vérifie $K(y_i, y_j) = < \phi(y_i), \phi(y_j) >$. L'utilisation d'un tel noyau permet d'éviter le calcul de la transformation ϕ, et ne nécessite que la connaissance du noyau. Le théorème de Mercer [Herbrich 2002] explicite que K doit être symétrique et semi-définie positive.

L'exemple le plus simple de fonction noyau est le noyau linéaire $K(y_i, y_j) = y_i^T \cdot y_j$, qui nous ramène au problème de classification linéaire évoqué précédemment. Les noyaux usuels employés avec les SVM sont :

- le noyau polynomial $K(y_i, y_j) = (y_i^T \cdot y_j + 1)^d$, d étant le degré du polynôme ;
- le noyau gaussien, aussi appelé RBF pour *Radial Basis Function* :
  $K(y_i, y_j) = \exp\left(-\frac{\|y_i - y_j\|^2}{2\sigma^2}\right)$.

Comme évoqué précédemment, le SVM dans sa forme de base vise à classifier de manière binaire les images. Il s'agit donc de trouver un schéma de combinaison afin d'en déduire la classification en M classes. Les deux stratégies les plus connues sont appelées *one-versus-all* et *one-versus-one* [Bishop 2006]. La méthode *one-versus-all*

consiste à construire M classifieurs binaires en attribuant l'étiquette 1 aux échantillons de l'une des classes et l'étiquette -1 à toutes les autres. En phase de test, le classifieur donnant la valeur de confiance la plus élevée remporte le vote. La méthode *one-versus-one* consiste à construire $M(M-1)/2$ classifieurs binaires en confrontant chacune des M classes. En phase de test, l'échantillon à classer est analysé par chaque classifieur et un vote majoritaire permet de déterminer sa classe.

Comme nous pouvons le remarquer, cette méthode SVM n'est pas contextuelle, et a été choisie comme modèle de comparaison pour cette raison, qui s'ajoute au fait que la technique des SVM est bien connue et largement utilisée. Cependant, pour pallier cette limitation, des champs de Markov peuvent être intégrés dans un tel processus de classification [Moser 2010]. Cette intégration repose sur une reformulation analytique de la règle d'énergie minimale markovienne, afin de l'introduire facilement dans l'expression des noyaux de l'algorithme SVM.

4.4 Analyseurs de performance des classifieurs

Cette partie vise à fournir une brève explication sur les méthodes utilisées en vue d'analyser les performances des méthodes de classification. Il y a deux manières d'évaluer les résultats : la manière visuelle (analyse qualitative) et la manière numérique (analyse quantitative).

Visuellement, nous pouvons donner un avis sur les résultats expérimentaux, par analyse directe de la carte de classification en comparant celle-ci avec l'image de départ, ou bien avec d'autres types d'acquisition de la même zone. Nous avons beaucoup utilisé les images GeoEye de Google Maps (©Google) pour valider ou infirmer visuellement les résultats de classification.

Numériquement, pour obtenir une validation de tout algorithme de classification, il est nécessaire d'être en possession d'une vérité de terrain qui sert d'image test. À partir de cette vérité de terrain, il est possible d'établir une *matrice de confusion*, qui rassemble les taux de bonne classification de chaque classe, ainsi que les taux d'erreurs de classification. À partir des taux de bonne classification, qui sont, en fait, un simple pourcentage de pixels bien classifiés, un *taux de classification moyen* peut être établi par simple moyennage. Nous pouvons aussi déterminer un *taux d'erreur*, qui est simplement l'addition des pixels i classifiés en j avec les pixels j classifiés en i. Les taux de bonne classification accompagnent de manière systématique nos résultats visuels. En revanche, la matrice de confusion n'est pas toujours indiquée, d'autant plus que, souvent, les pixels mal classifiés sont visuellement détectables sur la carte de classification, et, en général, nous accompagnons nos résultats expérimentaux d'une explication concernant la raison pour laquelle certains pixels sont mal classifiés.

Dans un contexte supervisé, nous avons donc besoin d'une vérité de terrain qui est utile à la fois pour établir la base d'apprentissage et la base de test. La question qui en découle est : comment choisir une telle base pour obtenir l'estimation de la matrice de confusion la plus cohérente possible ?

La méthode de *resubstitution* consiste à utiliser les mêmes données pour l'apprentissage et le test. Elle fournit une estimation optimiste de la probabilité d'erreur du classifieur, ce qui n'est pas forcément le critère le plus adapté et le plus proche de la réalité.

La *validation croisée* est une méthode générale qui vise à séparer les échantillons de la vérité de terrain en base de test et en base d'apprentissage [Kohavi 1995]. La méthode du *hold-out* est la plus simple des validations croisées. Elle consiste à prendre une partie (par exemple, 2/3) de la vérité de terrain pour entrainer le classifieur, et utiliser le reste (par exemple, 1/3) pour tester les résultats en établissant, par exemple, des matrices de confusion. Nous avons utilisé cette méthode pour évaluer nos classifieurs, même si elle soulève la question de la représentativité de chaque zone d'apprentissage et de test. En effet, il se pourrait que nous soyons "chanceux" au niveau de la sélection des échantillons-test, et que dans l'hypothèse où les taux de bonne classification soient corrects, cela n'est dû qu'à la prise en compte de pixels-test assez bien choisis. Nous avons, en pratique, sélectionné les pixels de la vérité de terrain dans différentes parties de l'acquisition, afin d'obtenir dans les bases de test et d'apprentissage une représentativité maximale, par exemple, diverses cultures dans la classe végétation.

Pour tenter de pallier cette limitation, il est possible de répéter le hold-out un certain nombre de fois sur des zones de test et d'apprentissage sélectionnées aléatoirement à chaque répétition. Cependant, cette procédure n'est pas très rigoureuse, car les échantillons de test peuvent se recouvrir, et d'autres être occultés.

Une manière d'améliorer la méthode du *hold-out* est d'utiliser la méthode du *k-fold*, qui consiste à diviser k fois la vérité de terrain (échantillon), puis de sélectionner un des k échantillons comme ensemble de validation ; les $(k-1)$ autres échantillons constituent alors l'ensemble d'apprentissage. Nous évaluons la classification sur une telle base. Puis nous répétons l'opération en sélectionnant un autre échantillon de validation parmi les $(k-1)$ échantillons qui n'ont pas encore été utilisés pour la validation du modèle. L'opération se répète ainsi k fois pour qu'en fin de compte, chaque sous-échantillon ait été utilisé exactement une fois comme ensemble de validation. Nous établissons, ensuite, une évaluation moyenne de la classification. Cela évite le problème de recouvrement des échantillons-test, que nous avons évoqué précédemment.

La méthode du *leave-one-out* est un cas particulier de la méthode *k-fold*. Cette fois-ci, k est égal au nombre de pixels n de la vérité de terrain. L'apprentissage est fait sur $(n-1)$ observations, et le modèle est validé sur la n^e observation. Cette opération est répétée n fois, ce qui nous a semblé long et fastidieux, surtout si nous considérons que notre vérité de terrain totale représente environ 10% de nos images. Soit $n = 10000$ pour une image de 1000×1000 pixels. Il s'agit, en outre, d'une estimation pessimiste de la performance du classifieur.

D'autres analyseurs existent, tels que le jackknife et le boostrap [Efron 1994, Shao 1995], qui sont initialement des méthodes de ré-échantillonnage adaptées à l'échantillonnage de la vérité de terrain, celui-ci étant réalisé de manière assez simple dans la validation croisée. Pour davantage de méthodes, plus élaborées, nous propo-

sons au lecteur de se référer à la littérature [Devroye 1996, Jain 2000].

Pour finir, mentionnons qu'une autre manière, bayésienne, d'évaluer le classifieur exploité, est d'estimer la probabilité d'erreur d'un classifieur associée à chaque classe de la classification [Theodoridis 2008] en partant de la connaissance de vecteurs de test dont les classes sont connues. Cette probabilité d'erreur peut s'exprimer de manière ponctuelle, ou bien en ayant recours à un intervalle de confiance. Nous n'avons pas procédé à une telle analyse dans notre cas, qui semblait assez délicate dans le sens où notre base de test n'est pas assez précise pour permettre d'évaluer un classifieur de manière totalement optimale.

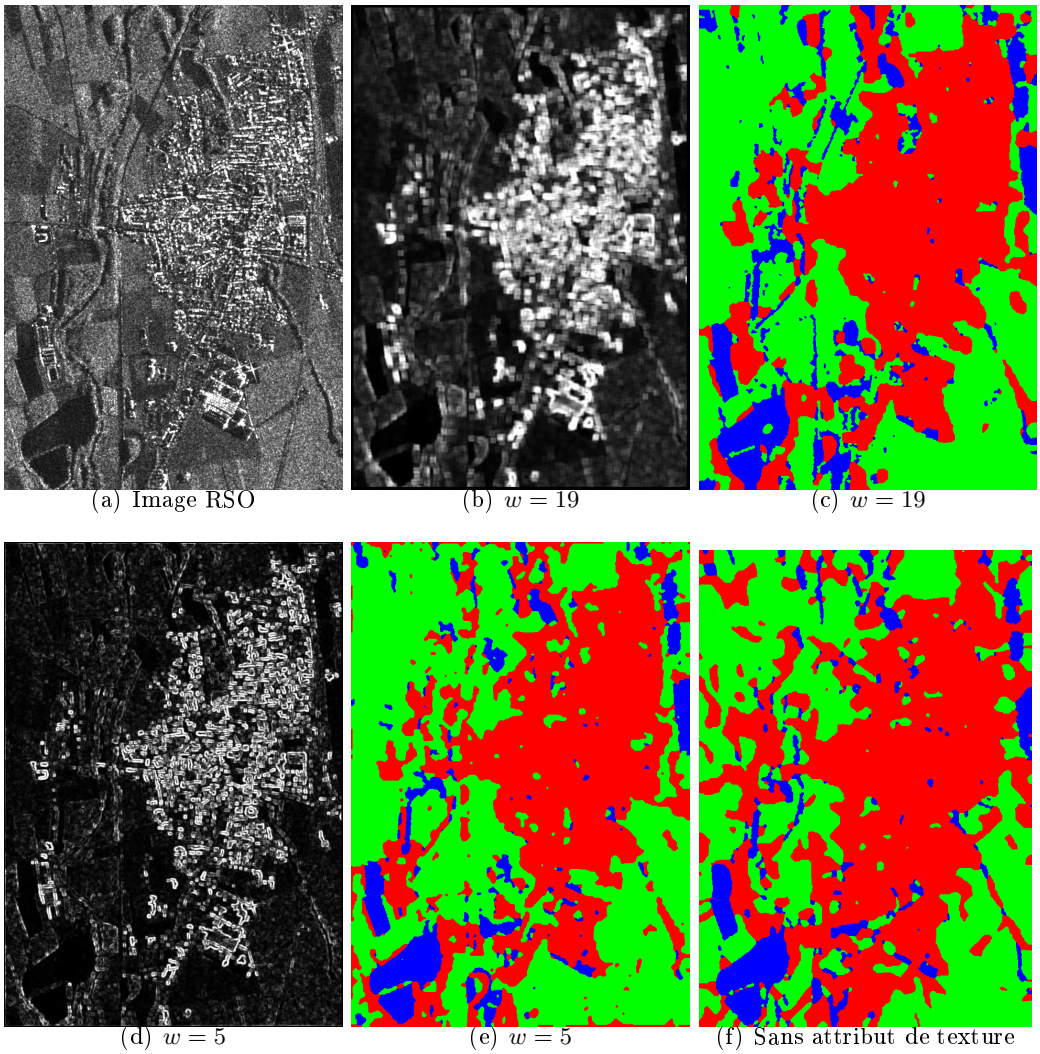
4.5 Résultats expérimentaux

Nous dédions cette partie à l'analyse expérimentale de diverses méthodes de classification, incluant avant tout celle proposée dans ce chapitre (partie 4.2). Les caractéristiques techniques des acquisitions traitées sont données en détail dans l'Annexe A. Pour l'ensemble de ces résultats, nous avons opté pour un paramètre β égal à $1,3$; qui est une valeur choisie empiriquement, mais qui est un bon compromis (voir sous-partie 4.2.2.1). Les apprentissages sont effectués sur une base contenant environ 5% de l'image, et chaque classe est apprise sur un nombre de pixels similaire à ceux des autres classes. Les images à traiter étant pour la plupart des images RSO à simple résolution, nous intégrons comme donnée initiale complémentaire un attribut de texture (voir la partie 3.2 du chapitre précédent). Nous allons tout d'abord observer les effets d'un tel attribut de texture sur la classification en travaillant avec une acquisition de la ville de Cavallermaggiore en Italie, avant de nous intéresser davantage aux méthodes de classification et à l'efficacité de celle que nous avons proposée dans ce chapitre, qui combine la modélisation statistique vue dans le chapitre 3 avec des champs de Markov. La comparaison entre les diverses méthodes est effectuée visuellement (évaluation qualitative) grâce à l'affichage de cartes de classification et à notre savoir-faire empirique, ainsi que numériquement (évaluation quantitative) via des matrices de confusion, que nous avons évoquées dans la partie 4.4.

L'acquisition de la ville de Cavallermaggiore en Italie a été également utilisée en Annexe C afin de mener une brève étude sur la robustesse de la classification par rapport à la base d'apprentissage. Nous avons pu observer qu'une modification de cette base d'apprentissage n'influe que peu sur les résultats de classification. De ce fait, nous pouvons en déduire que la méthode proposée est robuste, et qu'une certaine tolérance d'erreur est acceptée sur la base d'apprentissage, permettant à une personne non experte d'entraîner l'algorithme.

4.5.1 Influence de l'attribut de texture

L'image considérée pour cette étude de l'influence du choix de l'attribut de texture est une acquisition monobande à simple polarisation de la ville de Cavallermaggiore située dans la plaine du Pô en Italie. Nous lui avons appliqué la méthode vue



(a) Image RSO (b) $w = 19$ (c) $w = 19$

(d) $w = 5$ (e) $w = 5$ (f) Sans attribut de texture

FIG. 4.2 – (a) : Image RSO de Cavallermaggiore (Italie) (COSMO-SkyMed,©ASI) ; (b) : Attribut de texture extrait par MCNG ($w = 19$) ; (c) : Carte de classification obtenue avec la méthode proposée appliquée à l'image RSO et sa texture MCNG ($w = 19$) ; (d) : Attribut de texture extrait par MCNG ($w = 5$) ; (e) : Carte de classification obtenue avec la méthode proposée appliquée à l'image RSO et sa texture MCNG ($w = 5$) ; (f) : Carte de classification obtenue avec la méthode proposée appliquée à l'image RSO uniquement.

en partie 4.2. Trois classes sont considérées : les zones d'eau (représentées en bleu), les zones urbaines (représentées en rouge) et la végétation (représentée en vert).

La figure 4.2 montre l'influence de l'introduction d'attributs de texture, qui améliore visuellement les résultats de classification. En effet, dans la figure 4.2(f), la végétation est parfois confondue avec la zone urbaine, problème en partie résolu

TAB. 4.1 – Résultats numériques de classification obtenus pour l'image de Cavallermaggiore.

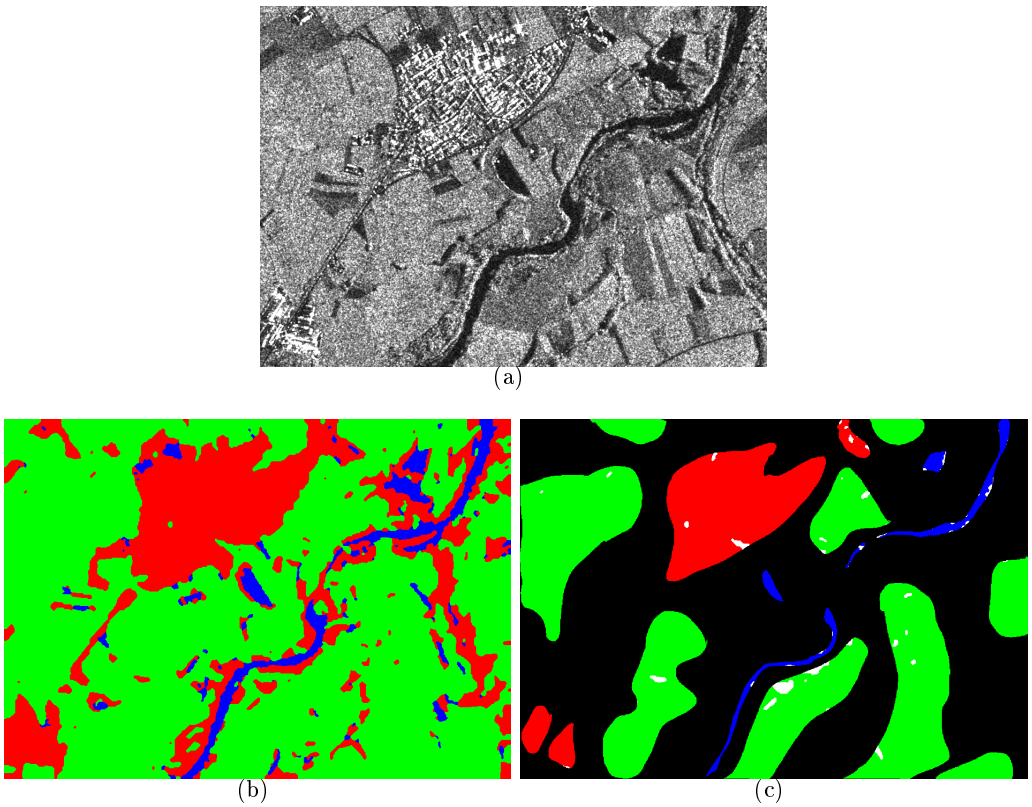
	Zone d'eau	Urbain	Végétation	**Global**
CM proposé (Semivar. $w = 5$)	$98,37\%$	$98,91\%$	100%	**99,09%**
CM proposé (MCNG $w = 5$)	$98,62\%$	$98,42\%$	100%	**99,01%**
CM proposé (MCNG $w = 11$)	$97,83\%$	$98,55\%$	100%	**98,79%**
CM proposé (MCNG $w = 19$)	$93,56\%$	$99,27\%$	100%	**97,61%**
CM proposé sans texture	$95,96\%$	$98,88\%$	$84,65\%$	**93,16%**

dans les autres cartes de classification obtenues après introduction de la donnée de texture. En outre, les attributs choisis, qu'il s'agisse de matrices MCNG ou de semivariogrammes, conduisent à des résultats tout à fait similaires (Tab. 4.2). Nous avons donc par la suite considéré un seul type de texture au détriment de l'autre. Pour faire notre choix, nous comparons visuellement ces deux attributs de texture (Fig. 3.6 dans le chapitre 3), en particulier dans les zones urbaines, qui sont nos zones d'intérêt. Les descripteurs obtenus, correspondant aux bâtiments, sont de forme rectangulaire pour la variance d'Haralick, alors qu'ils sont circulaires dans le cas du semivariogramme. Il paraît donc plus naturel de considérer la variance d'Haralick.

Nous avons ensuite effectué diverses simulations en vue de trouver empiriquement la meilleure taille de fenêtre w (sous-partie 3.2.1 du chapitre 3). Pour ce faire, nous avons lancé notre algorithme sur des attributs de texture obtenus avec des tailles de fenêtres variables, telles que $w \in [5; 19]$. D'après le tableau 4.1, il s'est avéré que $w = 5$ a conduit à l'obtention de résultats numériques meilleurs que ceux avec d'autres tailles de fenêtres. Cependant, visuellement, les résultats obtenus avec une grande fenêtre sont plus satisfaisants (Fig. 4.2), notamment dans la distinction de zones d'eau. Nous pouvons constater que l'écart de pourcentage dans les zones d'eau du tableau 4.1 réside dans l'introduction, par le classifieur, d'une zone de végétation dans la zone d'eau en bas à gauche de l'image 4.2(c). L'appréciation de la meilleure taille de fenêtre est largement discutable et peut être choisie au gré de l'utilisateur. Nous avons opté pour $w = 5$ en général.

4.5.2 Comparaison de nos résultats avec d'autres méthodes de l'état de l'art

Nous proposons dans cette partie de montrer les résultats obtenus avec la méthode proposée en partie 4.2 sur divers jeux de données, qu'elles soient de télédétection ou autres. Nous comparons cette méthode avec celles détaillées dans la partie 4.3, et soulignons l'intérêt de l'introduction d'un attribut de texture.



(a) (b) (c)

FIG. 4.3 – (a) : Image RSO de Lombriasco (COSMO-SkyMed,©ASI) ; (b) : Carte de classification obtenue avec la méthode proposée. (c) : Même carte de classification projetée sur la base de test. En blanc : pixels mal classifiés ; en noir : pixels non pris en compte dans la base de test ; dans les différentes couleurs : correspondance entre la vérité de terrain test et l'estimation des étiquettes après algorithme (rouge : zone urbaine | bleu : zones d'eau | vert : végétation).

4.5.2.1 Lombriasco, Italie

Nous montrons ici les résultats de classification obtenus pour une acquisition RSO à polarisation simple effectuée autour de la ville de Lombriasco en Italie, dont les détails techniques sont donnés en Annexe A. Trois classes sont considérées : les zones d'eau, les zones urbaines et la végétation. En général, le nombre de pixels pour les bases d'apprentissage est légèrement supérieur à ceux des bases de test afin de procéder à la validation croisée (partie 4.4). Cependant, pour cette simulation, nous avons opté pour une vérité de terrain plus étendue pour des raisons visuelles en vue d'illustrer la précision des résultats, insérés dans la vérité de terrain (Fig. 4.3).

Les taux de bonne classification sont donnés Tab. 4.2, et comparent les résultats obtenus avec notre méthode – qui combine des champs de Markov et une modélisation des statistiques par mélanges finis – avec les résultats obtenus par la méthode des K-PPV-CM (partie 4.3). Nous observons aussi les effets positifs de l'intégra-

TAB. 4.2 – Résultats numériques de classification obtenus pour l'image de Lombriasco.

	Zones d'eau	Urbain	Végétation	**Global**
CM proposé avec attribut	95, 28%	98, 67%	98, 50%	**97,48%**
CM proposé sans attribut	97, 74%	98, 90%	81, 80%	**92,82%**
K-PPV-CM	93, 86%	85, 54%	99, 91%	**93,10%**

tion de l'attribut de texture sur les résultats de classification, qui améliore ceux-ci d'environ 5%.

4.5.2.2 Rosenheim, Allemagne

Nous proposons, dans cette sous-partie, des tests sur une acquisition TerraSAR-X de la ville de Rosenheim en Allemagne, dont les caractéristiques techniques sont données en Annexe A. Nous proposons d'observer les résultats obtenus avec et sans attributs de texture, et de les comparer à des algorithmes de référence type SVM et K-PPV ainsi qu'à un algorithme développé au sein de l'équipe Ayin (ATML-CEM). Cette comparaison est effectuée visuellement (Fig. 4.4) et numériquement (Tab. 4.3).

Cette image est difficile à traiter, notamment au niveau des arbres présents au bord de l'eau, et qui, statistiquement, sont proches des zones urbaines. Cela explique en partie les erreurs au niveau des taux de classification, ainsi que les erreurs visuelles. Les deux méthodes fondées sur des champs de Markov, et utilisant un attribut de texture comme image complémentaire, donnent des résultats similaires tant numériquement que visuellement. Cela montre la robustesse et l'efficacité de l'introduction de l'élément de texture, en particulier pour l'aide à la discrimination des zones urbaines. Quant au départage entre la méthode que nous avons proposée et les K-PPV-CM, dans ce cas bien précis, cela n'est pas immédiat. En revanche, lorsque nous nous reportons à la sous-partie précédente, nous pouvons clairement en déduire que la modélisation statistique par modèle de mélanges finis est préférable, car elle semble plus robuste par rapport à divers jeux de données. Nous pouvons observer que lorsque nous ne prenons pas en compte de contexte (classifieur SVM), la carte de classification est très "pixelisée", très bruitée.

Numériquement (Tab. 4.3), la méthode ATML-CEM est meilleure que les autres. Cependant, en étudiant les cartes de classification (Fig. 4.4), nous pouvons observer

TAB. 4.3 – Résultats numériques de classification obtenus pour l'image de Rosenheim.

	Zones d'eau	Urbain	Végétation	**Global**
CM proposé avec attribut	91, 28%	98, 82%	93, 53%	**94,54%**
CM proposé sans attribut	92, 95%	98, 32%	81, 33%	**90,87%**
K-PPV-CM	90, 56%	98, 49%	94, 99%	**94,68%**
ATML-CEM	98, 05%	98, 30%	95, 87%	**97,41%**

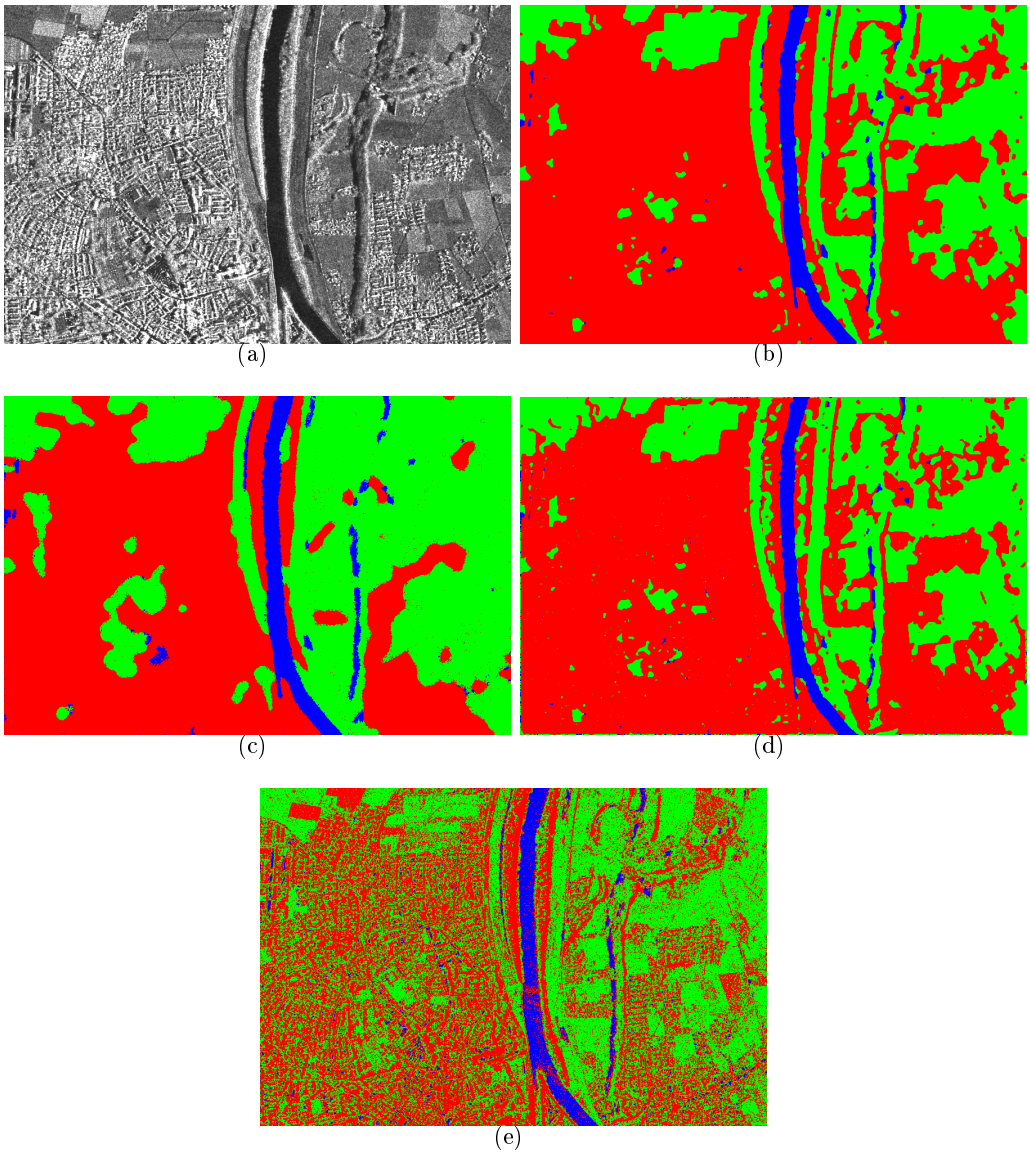
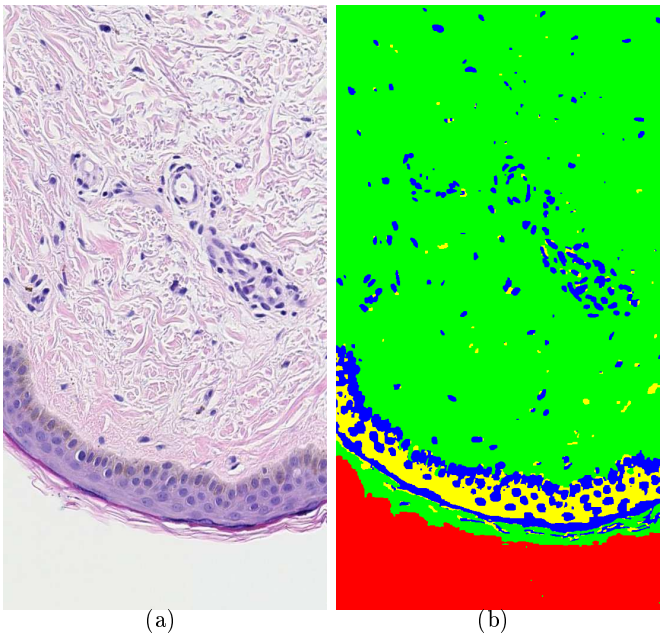


FIG. 4.4 – (a) : Image RSO de Rosenheim (TerraSAR-X,©Infoterra) ; (b) : Carte de classification obtenue avec la méthode proposée avec attribut de texture ; (c) : Carte de classification obtenue avec la méthode ATML-CEM ; (d) : Carte de classification obtenue avec la méthode des K-PPV-CM ; (e) : Carte de classification obtenue avec la méthode SVM (rouge : zone urbaine | bleu : zones d'eau | vert : végétation).

que cette différence numérique est due au fait que cette méthode induit davantage de sur-lissage que les méthodes fondées sur des champs de Markov. Cela se traduit par des zones moins bien définies. En outre, certaines zones urbaines disparaissent au profit de la végétation, ce qui n'apparaît pas dans les résultats numériques à cause de vérités de terrain qui n'ont pas été sélectionnées dans cette partie de l'image.



(a) (b)

FIG. 4.5 – (a) Image RVB d'une coupe histologique (©Galderma) (550 × 1020 pixels) et (b) classification obtenue avec la méthode proposée dans ce chapitre.

4.5.3 Généralité du modèle

Il nous paraît important de démontrer que le modèle mathématique considéré est suffisamment général pour être appliqué à divers types d'images, qui ne sont pas forcément des images radars de télédétection. L'entreprise Galderma R&D nous a gracieusement fourni des images de coupes histologiques auxquelles nous avons appliqué notre algorithme. La classification est exécutée pour 4 classes physiques : le derme en vert, le noyau fibroblaste en bleu, l'épithélium en jaune et le fond en rouge. Les images en entrée sont les bandes R, V, B de l'image initiale, qui sont combinées grâce à la méthode des copules. Chacune de ces bandes est statistiquement modélisée par des mélanges de distributions gaussiennes, et aucun attribut de texture n'est ici utilisé. Visuellement (Fig. 4.5) et numériquement (Tab. 4.4), les résultats obtenus sont satisfaisants et montrent la robustesse de l'algorithme proposé vis-à-vis de différents types d'images.

TAB. 4.4 – Taux de bonne classification pour chacune des classes de la coupe histologique.

	Noyau fibroblaste	Derme	Fond	Épithélium	**Global**
CM proposé	99,92%	99,97%	97,72%	96,65%	**98,56%**

4.5.4 Les limites du modèle

D'une manière générale, nous avons vu que les résultats obtenus sont satisfaisants, puisque les taux de bonne classification sont bons, et que visuellement, les pixels sont bien classifiés. En outre, l'algorithme est robuste au bruit de chatoiement. Les pixels mal classifiés correspondent, en général, à de petites zones, et peuvent être facilement expliqués. Par exemple, des zones d'ombre dans les villes confondues avec des zones d'eau de par leur faible niveau de gris sur l'image. Il peut aussi s'agir de pixels bien classifiés mais non pris en compte dans nos vérités de terrain quelque peu grossières (maisons isolées dans la campagne, etc.). La considération de classes supplémentaires, telle que par exemple la classe "ombres", n'est pas forcément une solution car les statistiques de cette classe sont très proches de celles de la classe "eau" en imagerie RSO, et, en pratique, cela génère des erreurs de classification plus importantes. La solution est d'utiliser l'information contenue dans une autre acquisition de la même zone (optique, ou autre bande polarimétrique) quand ce type d'information est disponible, ou bien d'intégrer des informations géométriques, soit par la connaissance d'un modèle numérique d'élévation (MNE), soit par l'extraction de propriétés géométriques des zones urbaines. Par exemple, nous pouvons supposer que la détection de bâtiments pourrait permettre de détecter les zones d'ombre et discerner celles-ci par rapport à des zones d'eau.

De manière générale, certaines notions sont discutables, comme la précision qui doit être donnée sur la vérité de terrain, ou bien la nécessité d'obtenir des avis d'experts, etc. La base d'une telle discussion est donnée dans [Madhok 2002]. Typiquement, une vérité de terrain se doit d'être plus ou moins détaillée, selon si nous cherchons à connaître parfaitement l'image, ou bien s'il s'agit uniquement de repérer une zone d'une classe donnée, par exemple pour observer l'étendue d'une zone urbaine sur une carte. En somme, cela dépend entièrement de l'utilisation finale. Mais il est assez difficile d'obtenir des vérités de terrain faites par des experts (car cela est coûteux et donc peu diffusé).

Pour en revenir à notre modèle et aux résultats qui en découlent, nous pouvons observer que les cartes de classification tendent à être très lisses, ce qui a pour conséquence une perte au niveau des détails de l'image. Diminuer le paramètre de régularisation du champ de Markov aboutit à un lissage moins important, mais aussi à une image davantage affectée par le bruit. Augmenter ce paramètre tend à sur-lisser la carte de classification. Il s'agit donc d'un compromis à faire sur la détermination de ce paramètre, et notre choix empirique concernant la valeur du paramètre β respecte ce compromis. Une modification de celui-ci n'est malheureusement pas dans notre cas la solution adéquate pour pallier les limites du modèle proposé.

4.6 Conclusion

Comme nous l'avons constaté dans ce chapitre, les cartes de classification obtenues sont globalement satisfaisantes. L'introduction de l'élément de texture, possible grâce à la flexibilité du modèle proposé, apporte une réelle amélioration sur la clas-

sification finale et en particulier sur les zones urbaines, ce qui se traduit par une amélioration du taux de classification global d'environ 5% en moyenne. Cependant, visuellement, les cartes de classification présentent des lissages aux frontières inter-classes non négligeables. Cela n'est pas surprenant, et peut être expliqué par deux raisons : (i) l'extraction de l'attribut de texture par fenêtre glissante ; (ii) le fait d'utiliser un modèle de champ de Markov de base et isotrope. Ce deuxième point est traité dans le chapitre suivant par l'intégration d'un modèle de voisinage adaptatif. La considération d'un attribut de texture différent, en revanche, ne permet pas d'améliorer de manière significative les taux de bonne classification qui sont déjà très satisfaisants pour toutes les classes. Quelques tests ont été réalisés avec d'autres attributs de texture, et corroborent cette affirmation.

Nous remarquons également un certain sur-lissage à l'intérieur des classes, généré par le champ markovien. Nous avons alors décidé de considérer une inter-dépendance en échelle par l'utilisation d'un modèle de Markov hiérarchique, qui prend en compte les interactions entre les différents niveaux du graphe hiérarchique. C'est un avantage, car cela permet de prendre en compte simultanément une image à haute résolution – donc détaillée – mais particulièrement bruitée, et son homologue à plus faible résolution, moins affectée par du bruit. En outre, le modèle hiérarchique ouvre une porte vers la généralisation de notre modèle mono-échelle (monorésolution) à une prise en compte d'images multirésolution. C'est ce que nous allons étudier en détail dans le chapitre suivant (Chap. 5).

Chapitre 5

Classification supervisée d'images multirésolution et multicapteur

Sommaire

Il existe de plus en plus de données de télédétection, qui deviennent, au cours des années, plus précises et plus variées. Ainsi, les algorithmes de traitement d'images dans ce domaine doivent répondre à des attentes de résultats toujours plus précis sur des images difficiles à manipuler. Dans la lignée directe du chapitre précédent, nous avons cherché à traiter les images multirésolution comme extension directe du cas monoresolution, puis par suite logique nous avons ensuite travaillé sur de la fusion d'images multicapteur. Le cas de la classification d'images multirésolution et multicapteur a été regroupé en un seul chapitre dans ce manuscrit de thèse, étant donné que les acquisitions sont, souvent, issues de différents capteurs (optiques et RSO), et par conséquent ont différentes résolutions. Nous cherchons donc à développer une méthode robuste aux divers types d'images que nous lui imposons en entrée. En pratique, il s'agit d'images RSO, d'attributs de texture et éventuellement d'images optiques haute résolution. Nous notons ici que les images optiques ne sont pas hyperspectrales, car c'est une étude assez spécifique, qui n'entre pas dans le cadre de cette thèse, celle-ci ayant pour objectif principal de classifier des images RSO et d'étudier, ensuite, les extensions possibles de la méthode développée. Dans un cas hyperspectral, il nous faudrait étudier des algorithmes plus spécifiques, incluant notamment des réductions de dimension de l'espace de départ [Dell'Acqua 2004].

Après une étude de l'état de l'art, nous présentons la méthode hiérarchique que nous avons développée au cours de la thèse. Nous détaillons, ensuite, la façon dont les images initiales sont intégrées dans la hiérarchie explicite, via des recalages ou des décompositions en ondelettes. Nous présentons enfin des résultats de classification que nous comparons à diverses méthodes issues de la littérature.

5.1 État de l'art

Dans un premier temps, nous présentons un état de l'art sur les méthodes adaptées à des acquisitions multirésolution et monocapteur, avant d'être étendues au cas multicapteur. Pour les mêmes motivations que celles explicitées dans le chapitre 4, nous avons mis un certain accent sur l'étude de méthodes markoviennes. En revanche, nous n'avons pas du tout étudié le cas non supervisé, qui n'est pas le plus présent dans la littérature.

5.1.1 Classification d'images multirésolution

La bibliographie sur la classification d'images multirésolution est bien moins étendue que celle actuellement disponible pour des images monorésolution. Le traitement d'images RSO est déjà ardu en monorésolution, avec la considération de problématiques telles que la polarimétrie, l'interférométrie, etc. Il est compréhensible que les images multirésolution soient encore plus délicates à traiter. En outre, le traitement de données multirésolution n'est possible que grâce à l'évolution du matériel informatique, devenu désormais suffisamment puissant pour traiter des données assez larges de manière simultanée.

Dans le cas de la classification d'images optiques multirésolution, potentiellement multibandes, Storvik et al. [Storvik 2005] ont proposé une méthode bayésienne fondée sur une maximisation de l'a posteriori via un algorithme ICM. La vraisemblance est estimée comme étant le produit des vraisemblances à chacune des résolutions considérées, les paramètres des couches inférieures dépendent donc des paramètres des couches supérieures. Nous avons réfléchi à l'extension d'un tel modèle à des observations RSO, en faisant l'hypothèse que les observations ne suivent pas une distribution gaussienne, mais une distribution ΓG, ce qui reste faisable grâce au fait qu'une combinaison linéaire de lois de Gamma est encore une loi de Gamma. En revanche, le calcul est beaucoup plus délicat lorsque l'on considère un mélange fini de ΓG comme modélisation des vraisemblances, et n'est pas faisable en conservant notre modèle général de mélange fini de fonctions issues d'un dictionnaire de FDP. En outre, l'insertion de données multicapteur s'avère encore plus discutable pour ce modèle.

Dans l'étude bibliographique menée lors de la thèse, nous avons pu observer que l'aspect multirésolution est souvent pris en charge par la considération implicite de graphes. Les méthodes hiérarchiques vues dans l'état de l'art du chapitre précédent (chap.4) semblent donc tout à fait adéquates, moyennant certaines petites modifications, pour traiter des images multirésolution. Des modèles de Markov hiérarchiques sont particulièrement adaptés, car ils permettent de prendre en compte le contexte dans un modèle de graphe hiérarchique [Li 2000].

5.1.2 Classification d'images multirésolution issues de différents capteurs

En général, la classification d'images multirésolution issues de différents capteurs se fait après recalage (voir 5.4.1) et mise à la même résolution (ré-échantillonnage) des images de départ. Suite à ce pré-traitement, trois types de méthodes sont possibles :

- les méthodes combinant divers classifieurs, ce qui revient à combiner différents résultats de classification. Ceci vise, notamment dans un contexte bayésien, à pallier le manque de modèles statistiques multivariés des données multisources [Briem 2002], difficiles à établir étant donné les différences physiques entre les systèmes d'acquisition.
- les méthodes combinant les images elles-mêmes, faisant appel à des méthodes de fusion, puis à l'intégration de ces images fusionnées dans des techniques de classification bien connues (SVM, etc.).
- les méthodes combinant des résultats intermédiaires tels que des probabilités, par exemple.

5.1.2.1 Combinaison de classifieurs

Deux questions sont sous-jacentes à une telle combinaison de classifieurs [Al-Ani 2011] : "combien de classifieurs doivent être considérés, et combien

d'images / d'attributs doivent être utilisé(e)s ?". Il s'agit donc, pour ces deux questionnements, de faire un choix, idéalement automatique, de la meilleure combinaison possible.

Les méthodes qui combinent plusieurs classifieurs reposent sur le principe de "diviser pour mieux régner", en d'autres termes, la combinaison de plusieurs classifieurs doit permettre d'améliorer la classification par rapport à un classifieur simple. La méthode de base est de classifier les images avec différentes méthodes et de combiner les résultats par vote majoritaire [Xu 1992]. Il est aussi possible d'utiliser des méthodes plus évoluées, fondées sur la théorie de Dempster-Shafer [Xu 1992, Al-Ani 2011], ou bien de combiner les classifieurs via des réseaux de neurones [Benediktsson 1997, Cho 1995]. Wolpert [Wolpert 1992] a introduit la méthode générale de reprojection pondérée (*stacked generalization*), dans laquelle les sorties des classifieurs sont combinées via des pondérations appliquées à chacune des sorties, poids fondés sur les performances individuelles des classifieurs. La théorie des ensembles flous a aussi été employée [Fauvel 2006] : la sortie de chaque classifieur est considérée comme un ensemble flou (*fuzzy set*), et le niveau de flou détermine lui-même l'efficacité de l'algorithme d'un classifieur donné. Ensuite, les ensembles flous sont eux-mêmes fusionnés afin d'obtenir une carte de classification finale.

Il est aussi possible de combiner des classifieurs "faibles" en un classifieur plus robuste via des méthodes de type boosting [Schapire 1999] ou bagging (bootstrap aggregating) [Breiman 1996].

Dans le cas où la classification est bayésienne, les résultats obtenus pour les différents classifieurs sont des probabilités a posteriori, et diverses opérations peuvent être réalisées sur ces probabilités. Elles peuvent être simplement moyennées [Xu 1992] ou bien, de manière plus générale, additionnées suivant une somme pondérée linéaire (*linear opinion pool* (LOP)), règle des consensus la plus utilisée [Briem 2002]. Une variation à cette LOP est sa version logarithmique, fondée sur le produit pondéré des probabilités a posteriori (ou somme pondérée des logarithmes) [Benediktsson 1992]. D'autres opérations peuvent être appliquées à ces probabilités, telles que le calcul de la médiane, ou des estimations de normes [Kittler 1998]. La LOP, tout comme sa version logarithmique ne sont que des exemples tirés de la théorie générale des consensus, qui vise à trouver un consensus entre les membres d'un groupe d'experts [Benediktsson 1992].

Toujours suivant cette règle des consensus, Benediksson et al. [Benediktsson 1999] ont proposé de combiner des classifieurs de manière hiérarchique, c'est-à-dire qu'une première comparaison est effectuée entre deux classifieurs, un classifieur neuronal et un classifieur statistique. Les pixels pour lesquels les classifieurs n'ont pas trouvé de consensus sont réinjectés dans un troisième classifieur.

Dans l'application spécifique de la classification d'images multicapteur optique/RSO, quelques méthodes ont été proposées. Dans [Waske 2008], par exemple, les auteurs appliquent une classification préalable de type séparateur à vaste marge (SVM) aux deux jeux de données préalablement mises à la même résolution, avant

de fusionner les SVM par utilisation d'un classifieur supplémentaire.

5.1.2.2 Fusion d'images

Une autre stratégie, plus courante, consiste à combiner les images initiales, puis à appliquer un algorithme de classification. Nous nous focalisons, dans cette partie, davantage sur les techniques de fusion possibles, au détriment des méthodes de classification elles-mêmes.

La fusion a deux utilités principales : elle peut être utilisée pour réduire la dimension de l'espace d'images de départ, ou bien pour obtenir une image à hautes résolutions spectrale et spatiale par la combinaison d'une image à haute résolution spatiale (image panchromatique à haute résolution) et d'une image à haute résolution spectrale (image couleur à basse résolution). Ce type de fusion est appelé *pan-sharpening*, et vise donc à rechercher un jeu de données précis, en vue d'obtenir la meilleure classification possible [Amarsaikhan 2004]. Deux principes généraux de pan-sharpening existent :

- Teinter les pixels de l'image panchromatique avec ceux de l'image couleur.
- Appliquer les hautes fréquences de l'image haute résolution à l'image couleur, méthode qui, en général, aboutit à de meilleurs résultats.

Parmi les algorithmes de fusion bien connus et couramment utilisés, nous pouvons citer :

- la méthode intensité-hue-saturation (IHS) [Pohl 1999, Mather 2011], dans laquelle la composante "intensité" de l'image à basse résolution est remplacée par la donnée haute résolution.
- la transformée de Brovey, qui multiplie chaque pixel multispectral, suréchantillonné à la taille de l'image à haute résolution, par le rapport de l'intensité du pixel panchromatique correspondant sur la somme de toutes les intensités multispectrales [Zhou 1998].
- l'analyse en composante principale (ACP) consiste à remplacer la première composante de l'image à basse résolution par l'image haute résolution (après ajustement des histogrammes) [Chavez 1991].

Ces méthodes ont tendance à négliger la qualité de la fusion au niveau spectral [Wang 2005]. D'autres méthodes [Metwalli 2010], un peu moins limitantes, ont aussi été largement utilisées, telles que le filtrage passe-haut, qui vise à extraire les détails spatiaux de l'image à haute résolution et à les injecter dans l'image basse résolution, la méthode de Gram-Schmidt, ou bien encore l'emploi de la transformée en ondelettes discrète. Ces méthodes peuvent aussi être combinées de manière adéquate, afin d'améliorer la fusion [Metwalli 2010]. Nous remarquons que la plupart de ces méthodes font appel à des projections dans des espaces bien choisis, tels que le domaine de Fourier [Palubinskas 2011], l'IHS, etc. Les images fusionnées sont, ensuite, obtenues par reprojection dans l'espace de départ.

D'autres algorithmes, un peu moins généralement utilisés, ont aussi été développés, comme par exemple la technique multicapteur multirésolution (*Multiresolution Multisensor Technique* (MMT)) proposée par Zhukov et al. [Zhukov 1999] et

Hu [Hu 1999], pour laquelle l'information spatiale de l'image à la meilleure résolution est utilisée pour analyser la composition des pixels de l'image basse résolution et les démélanger.

Comme nous l'avons déjà précisé, ces stratégies de fusion sont générales, et peuvent être employées à d'autres fins que celle de la classification. Par exemple, elle peuvent être utilisées pour de la détection de changement suite à une inondation [Sun 2007].

Cependant, dans notre cas, il nous paraît difficile d'envisager ce type de fusion étant donné qu'en pratique, nous avons traité de cas optique/RSO où l'image optique à haute résolution spectrale, est aussi celle à la plus haute résolution spatiale. Nous souhaitions, cependant, utiliser les informations contenues dans l'image RSO, qui peuvent aider à la discrimination des zones urbaines. Ainsi, comme alternative, puisque nous sommes dans un contexte bayésien, au lieu de fusionner les images, il est possible de fusionner des vraisemblances, et d'intégrer une telle fusion dans, par exemple, un simple champ de Markov, comme cela a été fait dans le chapitre précédent (chap. 4). La fusion des vraisemblances (probabilités marginales) peut se faire grâce à la théorie des copules (voir partie 3.3), ou bien grâce à toute autre distribution qui modélise la dépendance entre les images de départ, comme par exemple les distributions meta-gaussiennes [Storvik 2009]. La principale nouveauté de ce chapitre de thèse est la prise en compte de l'aspect multirésolution originel des images, en plus de l'aspect de fusion multicapteur.

5.2 Généralités sur la classification supervisée incluant des champs de Markov hiérarchiques

Une partie du travail de thèse a consisté à trouver un classifieur à la fois robuste au bruit de chatoiement (voir 2.1.1.4), et qui dégrade peu ou pas les résolutions des images initiales. Nous utilisons, dans cet objectif, une méthode fondée sur un modèle markovien hiérarchique. Comme nous l'avons vu dans l'état de l'art du chapitre précédent (partie 4.1), la plupart des méthodes markoviennes hiérarchiques ont été développées pour être appliquées à des images monorésolution. L'idée principale dans cette thèse est d'avoir non plus recours à une décomposition fictive de l'image via un moyennage de pixels adapté ou des ondelettes, mais de prendre en compte les résolutions effectives en intégrant ces images dans un graphe spécifique. Bien entendu, lorsque nous souhaitons appliquer notre méthode à des images monorésolution, il est alors nécessaire de recourir à une décomposition en échelle de l'image de départ. Nous avons opté pour une décomposition en ondelettes, dont nous rappelons la théorie ultérieurement (sous-partie 5.4.2).

D'une manière générale, deux types de hiérarchie sont possibles [Graffigne 1995] et peuvent être adaptées à notre problème de classification :

- une hiérarchie *induite*, regroupant les techniques de groupe de renormalisation [Gidas 1989] et les décompositions par sous-espaces contraints [Heitz 1994].

– une hiérarchie *explicite*, comprenant notamment la technique des graphes hiérarchiques, que nous avons choisie.

Nous allons maintenant donner des définitions relativement générales et présenter globalement le modèle de hiérarchie explicite que nous avons considéré (quad-arbre). Il s'agit d'un préambule à la partie suivante, qui vise à présenter des généralités, ainsi que les bonnes propriétés des graphes hiérarchiques explicites.

5.2.1 Notations et quelques définitions préalables

Commençons par établir quelques règles de notations. Nous notons s un site compris dans un ensemble S et V_s son voisinage ayant pour propriétés :

(i) $\forall s \in S, s \notin V_s$,

(ii) $\forall s_1, s_2 \subset S, s_1 \in V_{s_2} \Leftrightarrow s_2 \in V_{s_1}$.

L'ensemble des voisinages est noté V. Le couple (S, V) définit un graphe G, qui est à la fois *simple* et *non orienté* par la propriété de symétrie (ii) du voisinage. Un graphe est dit *connecté* si lorsque nous considérons deux sites s et t, nous trouverons toujours une chaîne finie pour les relier.

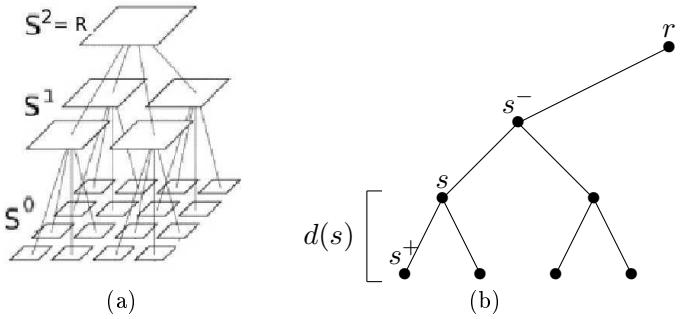
Une *clique* est un sous-espace non vide c de S de dimension supérieure ou égale à 1 tel que 2 sites distincts de c soient voisins. L'ensemble des cliques du graphe G est noté C.

Nous allons, par la suite, considérer un graphe particulier, dit *arbre*, qui est un graphe connecté sans cycle. L'ensemble des sites est alors hiérarchiquement partitionné, de telle sorte que $S = S^0 \bigcup S^1 \bigcup ... \bigcup S^R$, où R désigne la résolution la plus grossière (la *racine*), et 0 correspond au niveau de référence. Nous pouvons définir des relations de parenté entre les différents niveaux n de l'arbre, de telle sorte que si nous considérons un site s pour un n quelconque, nous pouvons lui attribuer un unique *parent* s^- et des *enfants* s^+. L'ensemble formé par s et tous ses descendants sera noté $d(s)$. Si le site s possède 4 enfants, nous considérons alors une structure de type *quad-arbre*. Notons que les sites situés sur la racine n'ont pas d'enfants ($d(s) = s$ pour $s \in S^0$), et les sites appartenant à la couche S^R n'ont pas de parents. Une telle structure est représentée Fig. 5.1. Dans un tel contexte, la relation d'ordre se fait sur les différents sites du quad-arbre, d'après une relation « parent-enfant ».

5.2.2 Graphes hiérarchiques

Il s'agit d'une technique provenant de la théorie des graphes et du traitement du signal. Les deux types les plus connus sont le graphe pyramidal $3D$ et le quad-arbre, dont la principale différence est respectivement la présence ou l'absence de voisins à un même niveau d'échelle (i.e., en spatial). Certains graphes pyramidaux ont davantage de branches du fait, par exemple, de la considération d'un nombre supérieur d'ascendants. Cela leur confère une meilleure prise en compte du voisinage, mais le graphe obtenu n'est plus un arbre.

Le problème de classification consiste à estimer un ensemble de variables cachées X à partir d'observations Y. X et Y sont des processus aléatoires indicés par les



(a) (b)

FIG. 5.1 – (a) : Structure du modèle hiérarchique : quad-arbre ; (b) : Notations utilisées sur le quad-arbre.

sommets du quad-arbre. La restriction de X (resp. Y) à la couche n est notée $X^{(n)} = \{X_s, s \in S^n\}$ (resp. $Y^{(n)} = \{Y_s, s \in S^n\}$) de réalisation $x^{(n)}$ à valeurs dans Ω. Quelques hypothèses sont nécessaires pour que X soit un *champ de Markov* sur le graphe G : $\forall s \in S$, $\forall x \in \Omega$,

(i) $p(X = x) > 0$,

(ii) $p(X_s = x_s | X_t = x_t, t \in S - \{s\}) = p(X_s = x_s | X_t = x_t, t \in V_s)$.

La propriété de Markov implique donc que la donnée des variables X_t voisines d'un site s ($t \in V_s$) suffisent à déterminer la distribution locale conditionnelle de X_s.

5.2.3 Les avantages et les inconvénients d'un tel modèle

Nous proposons de faire une liste non exhaustive des avantages liés à l'utilisation de champs de Markov hiérarchiques :

- Possibilité d'intégrer différents types de statistiques et donc d'utiliser différents types d'images comme observations (diverses acquisitions radars et/ou optiques) [Laferté 1995] ;
- Robustesse au bruit de chatoiement grâce à une prise en compte du contexte ;
- Sensibilité moindre aux conditions initiales (surtout à la probabilité a priori initiale) ;
- Diminution du risque de tomber dans un minimum local, d'où de meilleures chances de converger vers un optimum global (propriété des techniques multigrilles) ;
- Bon compromis entre résolution radiométrique (bruit) et résolution spatiale. En effet, à une faible résolution, les images sont assez peu bruitées, mais les frontières inter-classes sont floues. En revanche, à une haute résolution, le bruit est souvent plus présent, notamment pour les images RSO, mais les frontières inter-classes sont facilement localisables. La classification hiérarchique permet de prendre en compte toutes ces informations de manière simultanée.

La considération d'un graphe de type quad-arbre allonge la liste de ces avantages, notamment en incluant la propriété de causalité en échelle, induite par la structure

propre du quad-arbre. L'intérêt de la propriété de causalité est de simplifier le modèle statistique, par considération d'un simple voisinage, et donc de pouvoir construire des algorithmes simples, efficaces et non itératifs pour les champs de Markov. Si nous avions choisi une hiérarchie induite, l'algorithme aurait été itératif. Visuellement, il est possible de d'identifier certains modèles de graphes causaux. La famille des graphes triangulés en fait partie [Whittaker 1990]. Un graphe triangulé est un graphe n'ayant pas de cycle de longueur supérieure ou égale à 4 sans corde. Les chaînes de Markov et le quad-arbre sont des cas particuliers de graphes triangulés.

L'utilisation de modèles de graphes hiérarchiques a aussi quelques inconvénients, notamment les effets de blocs [Laferté 2000], qui sont particulièrement visibles pour des structures circulaires [Laferté 1996]. La structure en quad-arbre est elle-même quelque peu contraignante, car elle impose que les données soient structurées de manière dyadique. Nous étudions plus en détail ces inconvénients par la suite.

5.3 Méthode markovienne hiérarchique avec mise à jour de l'a priori

De nombreux algorithmes ont été proposés afin d'estimer les étiquettes sur des graphes hiérarchiques. Le choix de ces algorithmes se fait notamment en fonction des caractéristiques du modèle envisagé. En général, nous procédons à une minimisation de l'énergie globale via des algorithmes de relaxation itératifs [Zerubia 1994]. L'avantage du quad-arbre est de pouvoir utiliser ses propriétés (en particulier la causalité) afin d'appliquer des algorithmes non itératifs [Fabre 1994]. Une première option consiste à rechercher un estimateur exact du Maximum A Posteriori (MAP) par filtrage linéaire de Kalman [Luettgen 1994] ou par algorithme non-linéaire de type Viterbi [Bouman 1994, Laferté 1995, Laferté 1996]. Mais ce critère nous amène à travailler avec des probabilités très petites et entraine parfois des problèmes de dépassement inférieur de capacité (« underflow »). Nous avons, de ce fait, opté pour un estimateur exact du *mode de la marginale a posteriori* (MPM) [Marroquin 1987, Laferté 2000]. La fonction de coût associée à cet estimateur a pour avantage de pénaliser les erreurs en fonction de leur nombre et de l'échelle à laquelle elles se produisent : une erreur à l'échelle grossière est plus pénalisée qu'une erreur à l'échelle la plus fine, ce qui est normal puisqu'un site au niveau de la racine équivaut à 4^R pixels de l'échelle la plus fine.

Pour estimer la probabilité a posteriori, étant dans un cadre bayésien, nous avons besoin de la connaissance préalable de certaines probabilités : la vraisemblance, la probabilité a priori et la probabilité de transition en chaque site s du quad-arbre. Le calcul de la vraisemblance a été présenté en détail dans le Chapitre 3. Nous en rappelons brièvement le principe : à chaque résolution $n \in [0; N]$ et pour chaque bande $j \in [1; d]$, nous cherchons à modéliser les distributions statistiques de chaque classe m considérée pour la classification, $m \in [1; M]$, étant donnée une base

d'apprentissage. Pour chaque classe, les FDP $p_m(y_j|\omega_m)$ sont telles que :

$$p_m(y_j|\omega_m) = \sum_{i=1}^{K} P_{mi} p_{mi}(y_j|\theta_{mi}),$$

où y_j désigne la j^e observation (bande) à un niveau de l'arbre (résolution) donné. Les FDP $p_{mi}(y_j|\theta_{mi})$ du mélange sont automatiquement choisies dans un dictionnaire prédéfini, avec une certaine préférence pour la loi gaussienne dans le cas où la bande est optique, et Gamma généralisée dans le cas où la bande est une acquisition radar. Puis, à chaque résolution, par le biais de copules (théorème de Sklar), les probabilités marginales des différentes bandes sont rassemblées en une FDP jointe $p(y|\omega_m)$, où y contient les informations des différentes bandes (optique/RSO, RSO/texture, etc.). C'est cette FDP jointe qui est intégrée dans le modèle mathématique du quad-arbre.

Nous allons voir en détail dans cette partie la définition mathématique des probabilités de transition et des probabilités a priori. Ensuite, nous nous focalisons sur la maximisation de l'a posteriori, et introduisons dans l'algorithme non itératif une mise à jour de la probabilité a priori, ce qui, en pratique, augmente la robustesse de la méthode hiérarchique par rapport au bruit inhérent des images, notamment le bruit de chatoiement des images RSO.

5.3.1 Les probabilités de transition

L'hypothèse fondamentale est de considérer le processus aléatoire X comme markovien en échelle, $p(x^{(n)}|x^{(k)}, k > n) = p(x^{(n)}|x^{(n+1)})$, où n et k désignent des échelles. Ce sont ces probabilités de transition entre les différents niveaux $p(x_s|x_{s^-})$ qui spécifient le champ de Markov, car elles traduisent la causalité des interactions statistiques entre les différentes échelles du quad-arbre. Il est donc important de bien les définir. Nous avons opté pour la forme de la probabilité de transition adoptée dans [Bouman 1994], soit pour tout site $s \in S$ et toute échelle $n \in [0; R-1]$:

$$p(x_s = \omega_m | x_{s^-} = \omega_k) = \begin{cases} \theta_n, & \text{si } \omega_m = \omega_k \\ \frac{1-\theta_n}{M-1}, & \text{sinon} \end{cases} \tag{5.1}$$

où ω_m et ω_k représentent respectivement les classes m et k, $m, k \in [1; M]$ où M désigne le nombre de classes considérées pour la classification finale (fixé dans l'apprentissage). Ce modèle permet de favoriser un étiquetage parent-enfant identique.

Le paramètre inconnu θ_n, qui détermine totalement la probabilité de transition, dépend, théoriquement, du niveau n. En pratique, il est choisi comme indépendant du niveau de l'arbre n, la raison majeure étant qu'il est difficile d'estimer de manière automatique ce paramètre, et que l'estimer en chacun des niveaux est une tâche fastidieuse, et pas forcément nécessaire.

Notre choix s'est porté sur une telle probabilité de transition car malgré son expression assez simple, les résultats obtenus (voir 5.5) sont satisfaisants, et son expression permet un emploi assez aisé et flexible de ces probabilités.

5.3.2 Les probabilités a priori

La probabilité a priori $p(x_s)$ est la probabilité que l'étiquette du site s soit x_s, où x_s prend ses valeurs dans $[1; M]$. La distribution a priori d'un niveau n dans $[0; R-1]$ est telle que :

$$p(x_s^{(n)}) = \sum_{x_{s^-}^{(n)} \in \Omega = \{\omega_i, i=1,\ldots,M\}} p(x_s^{(n)} | x_{s^-}^{(n)}) p(x_{s^-}^{(n)}), \tag{5.2}$$

où $p(x_s^{(n)}|x_{s^-}^{(n)})$ sont les probabilités de transition, vues dans la sous-partie 5.3.1. La seule connaissance de la probabilité a priori au niveau le plus grossier R permet de déduire toutes les autres probabilités a priori aux niveaux inférieurs, puisque les probabilités de transition sont bien connues.

Comme nous le verrons dans la description de l'algorithme MPM (sous-partie 5.3.3), il est nécessaire, en étape préliminaire à la classification, d'avoir une estimation de la probabilité a priori au niveau le plus grossier R. Pour ce faire, nous utilisons des estimations de probabilités a priori d'après des classifications effectuées préalablement, par exemple, en exploitant les résultats obtenus avec un algorithme des K plus proches voisins.

5.3.3 Les probabilités a posteriori et leur estimation - Méthode de Laferté et al. [Laferté 2000]

Du fait de l'absence de cycle sur le quad-arbre, les étiquettes sont estimées exactement et non itérativement par critère MPM grâce à un algorithme de type « forward-backward », comparable à l'algorithme classique de Baum pour une chaîne de Markov [Baum 1970]. Il vise à maximiser, en chaque site s, la marginale a posteriori

$$\hat{x_s} = \arg\max p(x_s|y). \tag{5.3}$$

Une telle méthode a été introduite par Laferté et al. dans [Laferté 2000], et nous l'avons, dans un premier temps, combinée avec la modélisation de statistiques par mélanges finis proposée en chapitre 3. Une telle combinaison a été exploitée pour la classification d'images RSO dans la publication du Gretsi 2011 (Annexe E.3).

Le principal objectif de la procédure est l'estimation de l'a posteriori au niveau 0 en fonction des observations de tous les niveaux $y = \{y_s, \forall s \in S, \forall n \in [0; R]\}$, en vue de maximiser cet a posteriori. Cette estimation est réalisée en deux passes, dites montante (« forward ») et descendante (« backward »), tel qu'illustré dans la figure 5.2.

Passe montante

Il s'agit de calculer, pour tous les sites $s \in S$, les marginales a posteriori partielles $p(x_s|y_{d(s)})$ et $p(x_s, x_{s^-}|y_{d(s)})$, qui seront nécessaires au calcul des probabilités a posteriori complètes $p(x_s|y)$. Par définition,

$$p(x_s, x_{s^-}|y_{d(s)}) = p(x_{s^-}|x_s) p(x_s|y_{d(s)}), \tag{5.4}$$

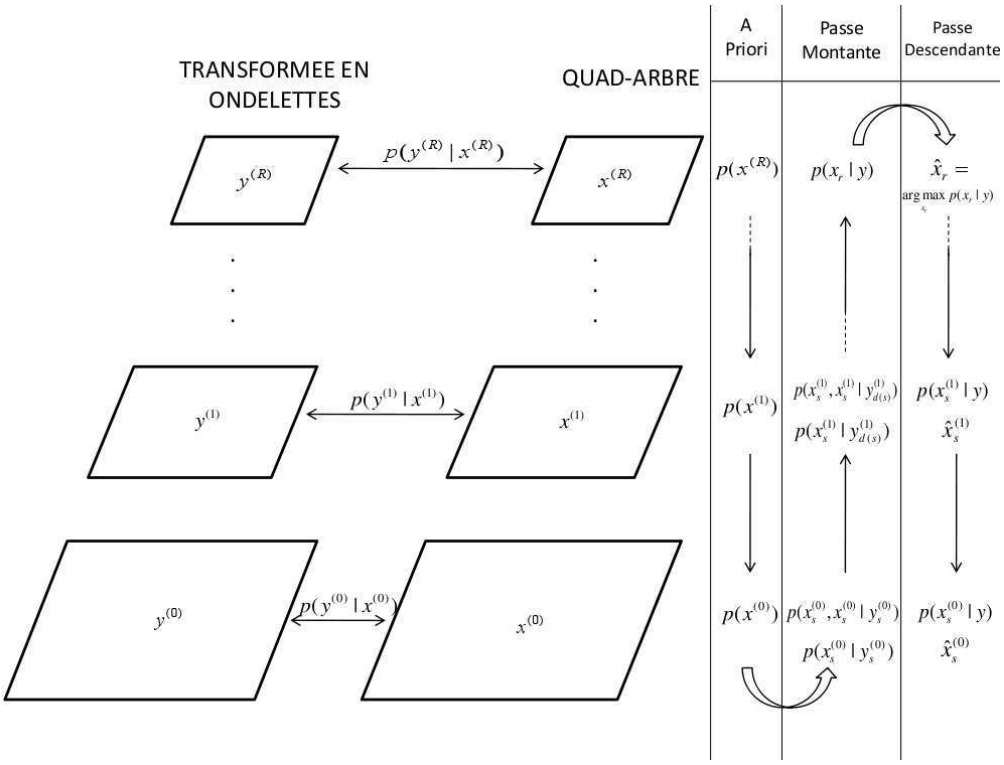


FIG. 5.2 – Modèle hiérarchique générique du quad-arbre et schéma de l'algorithme de recherche du critère MPM proposé par Laferté.

avec

$$p(x_{s^-}|x_s) = \frac{p(x_s|x_{s^-})p(x_{s^-})}{p(x_s)}. \tag{5.5}$$

Les probabilités a priori et les probabilités de transition pour tous les sites s sont déterminées préalablement, comme nous l'avons vu dans les sous-parties 5.3.1 et 5.3.2. Nous estimons tout d'abord $p(x_s|y_{d(s)})$, sachant que [Laferté 2000] :

$$p(x_s|y_{d(s)}) = \frac{1}{Z} p(y_s|x_s)p(x_s) \prod_{t \in s^+} \sum_{x_t} \left[\frac{p(x_t|y_{d(t)})}{p(x_t)} p(x_t|x_s) \right]. \tag{5.6}$$

D'après la relation (5.4), la connaissance de $p(x_s|y_{d(s)})$ permet de calculer $p(x_s, x_{s^-}|y_{d(s)})$. Nous allons donc procéder à une récurrence, qui consiste à estimer $p(x_s|y_{d(s)})$ puis $p(x_s, x_{s^-}|y_{d(s)})$ pour tous les niveaux, en commençant aux feuilles (niveau 0) et en terminant à la racine (niveau R) de l'arbre. Les estimations des $p(x_s|y_{d(s)})$ à un niveau n sont utilisées pour estimer les $p(x_s|y_{d(s)})$ du niveau directement supérieur $n+1$. Les vraisemblances $p(y_s|x_s)$ sont connues, puisque déterminées par apprentissage.

Les sites du niveau 0 n'ayant pas d'enfants, les équations, pour ce niveau, se

simplifient :

$$p(x_s|y_{d(s)}) = p(x_s|y_s) = \frac{1}{Z}p(y_s|x_s)p(x_s)$$

et

$$p(x_s, x_{s^-}|y_{d(s)}) = p(x_s, x_{s^-}|y_s) = \frac{p(x_s|x_{s^-})p(x_s|y_s)p(x_{s^-})}{p(x_s)}.$$

Passe descendante

Pour chaque niveau n, du plus grossier au plus fin, cette passe vise à estimer les probabilités a posteriori complètes en fonction des probabilités a posteriori partielles déterminées dans la passe montante. Pour un site $s \in \{S^0, ..., S^R - 1\}$, d'après le théorème de Bayes,

$$\begin{aligned} p(x_s|y) &= \sum_{x_{s^-}} p(x_s|x_{s^-}, y)p(x_{s^-}|y) \\ &= \sum_{x_{s^-}} p(x_s|x_{s^-}, y_{d(s)})p(x_{s^-}|y) \\ &= \sum_{x_{s^-}} \frac{p(x_s, x_{s^-}|y_{d(s)})}{\sum_{x_s} p(x_s, x_{s^-}|y_{d(s)})} p(x_{s^-}|y) \end{aligned} \tag{5.7}$$

Pour le niveau le plus grossier R, $p(x_s|y) = p(x_s|y_{d(s)})$ pour des sites $s \in S^R$. De ce fait, la carte de classification à ce niveau est obtenue en estimant directement $\hat{x_s} = \arg\max p(x_s|y)$ pour $s \in S^R$. Pour les autres niveaux, il suffit d'utiliser l'équation (5.7), les $p(x_s, x_{s^-}|y_{d(s)})$ ayant été estimés dans la passe montante. Les étiquettes sont déterminées par $\hat{x_s} = \arg\max p(x_s|y)$ en chacun des niveaux.

5.3.4 Les probabilités a posteriori et leur estimation - Méthode avec mise à jour de l'a priori

Nous avons cherché à améliorer l'estimation de la probabilité a priori, en partant du principe que celle-ci n'est pas estimée de manière précise dans l'étape préliminaire. En effet, les probabilités a priori ne sont estimées de manière précise qu'au niveau le plus grossier (5.3.2), celles aux autres niveaux étant déduites par la relation (5.2). Pour cette raison, nous avons proposé une mise à jour itérative des probabilités a priori, qui vise à améliorer leur estimation, et donc les résultats de classification. Cette méthode a été présentée à la conférence IS&T/SPIE Electronic Imaging en janvier 2012 (Annexe E.4) ainsi qu'à un séminaire italien (Annexe E.5). Elle a aussi été publiée dans le journal IEEE Geoscience and Remote Sensing Letters (GRSL) en 2012 (Annexe E.8). Dans ce papier de journal, ainsi que dans l'article relatif au séminaire italien, cette méthode hiérarchique est utilisée dans le cas spécifique de la classification d'images RSO monorésolution, mais elle est suffisamment générale pour être appliquée à divers jeux de données (multicapteur, multirésolution, etc.).

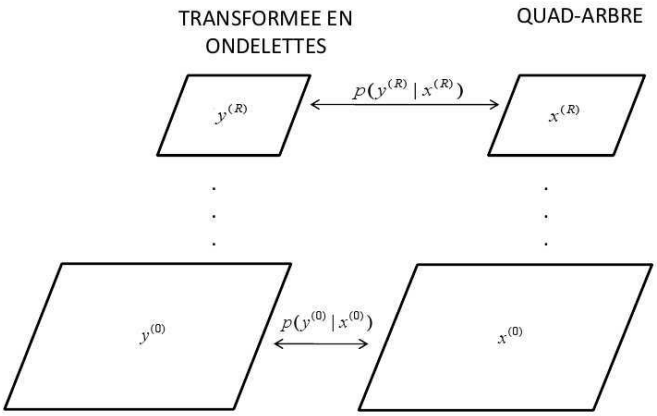


FIG. 5.3 – Modèle hiérarchique générique du quad-arbre.

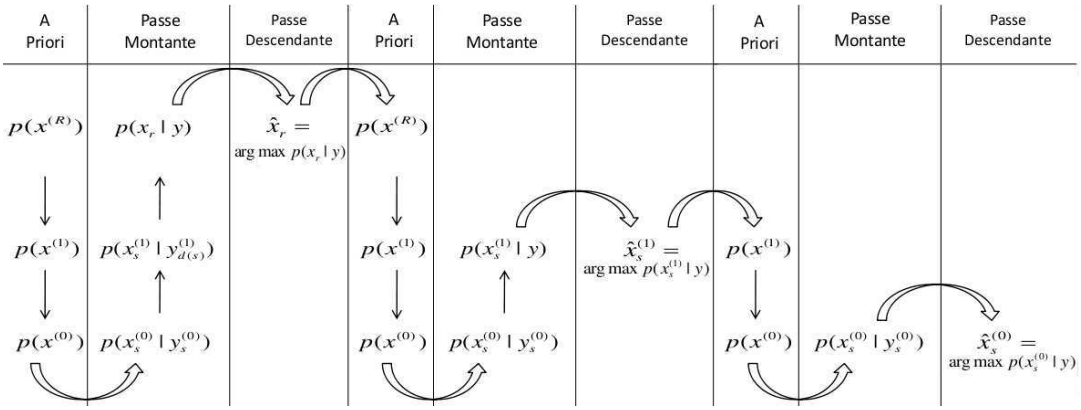


FIG. 5.4 – Estimation du critère MPM proposée sur le quad-arbre donné en Fig. 5.3. Dans cette représentation, $R = 2$.

Dans la nouvelle version de l'algorithme MPM, la passe descendante est tronquée, et seule la marginale a posteriori au niveau de la racine est estimée (voir Figs. 5.3 et 5.4). La carte de classification obtenue est alors utilisée pour mettre à jour les probabilités a priori à ce niveau-là. Pour ce faire, la probabilité a priori $P(x_s)$ est remplacée par la probabilité conditionnelle locale de l'équation (4.4) du chapitre 4. Ce choix donne, en général, une estimation de la probabilité a priori biaisée, mais favorise l'adaptabilité spatiale, ce qui est une propriété souhaitée quand il s'agit de traiter des images à très haute résolution, dans lesquelles les détails spatiaux sont, en général, apparents.

Comme vu dans la sous-partie 5.3.2, les probabilités a priori des niveaux inférieurs sont données par la relation (5.2). Ensuite, comme illustré sur la Fig. 5.4, un nouvel algorithme MPM est lancé, sur un quad-arbre étêté de son niveau supérieur

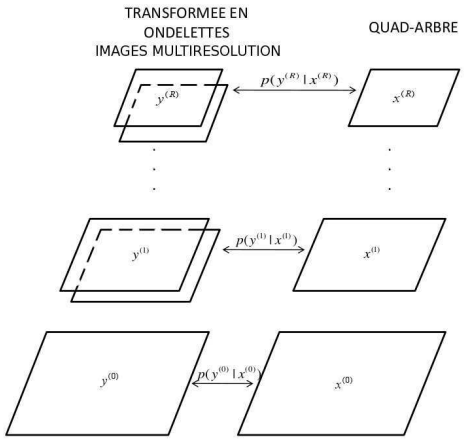


FIG. 5.5 – Modèle hiérarchique générique du quad-arbre - Nouvelle version, plus souple au niveau des observations grâce à l'insertion d'une flexibilité sur le nombre de bandes à chaque résolution. Les bandes peuvent être acquises par différents capteurs. Le modèle statistique considéré (mélanges finis et copules) est suffisamment souple pour permettre cette double flexibilité sur les bandes d'entrée.

(i.e., tronqué). Et ainsi de suite, jusqu'à atteindre le niveau 0.

L'initialisation de l'algorithme est effectuée en considérant une équiprobabilité des a priori, ce qui est possible grâce à la robustesse de la méthode proposée par rapport aux conditions initiales. Ainsi, nous pouvons, en plus, nous affranchir de l'étape préalable d'estimation des a priori (voir 5.3.3).

5.3.5 Les probabilités a posteriori et leur estimation - Version finale de la méthode proposée

La méthode telle que décrite dans la sous-partie précédente ne présente pas la possibilité d'avoir différents nombres de bandes selon les niveaux. En effet, si l'on considère une image optique à 3 bandes à un niveau 0 et une image RSO monopolarisée (et donc monobande) à un autre niveau de l'arbre, ce modèle n'est pas adapté pour les observations. Étant donné que le modèle de vraisemblance utilisé est suffisamment souple (grâce aux copules), il est donc possible d'avoir un nombre de bandes différent à chaque niveau du graphe hiérarchique, avec des bandes elles-mêmes issues de différents capteurs. Nous proposons d'introduire un paramètre adaptatif pour le nombre de bandes en entrée à chacun des niveaux. Le modèle générique du graphe pour les observations est représenté dans la figure 5.5. L'algorithme de la recherche du critère MPM, quant à lui, ne change pas (Fig. 5.4).

Cette version finale de la méthode a été présentée à la conférence EUSIPCO 2012 (Annexe E.6). Elle a aussi été soumise dans le journal IEEE Transactions on Geoscience and Remote Sensing (TGRS) en 2012 (Annexe E.9), dans lequel les résultats expérimentaux sont illustrés sur des jeux de données multicapteur.

5.4 L'intégration des données dans le quad-arbre

Les observations que nous traitons nécessitent parfois des pré-traitements. Il peut s'agir, par exemple, de décomposer une image monorésolution en image multirésolution, de recaler deux images d'une même zone, ou encore de fusionner certaines données.

5.4.1 Le recalage d'images

Comme nous l'avons vu dans le chapitre 2, les satellites sont défilants, et, de ce fait, deux acquisitions d'une même zone faites par un ou plusieurs satellites peuvent être géométriquement différentes, selon, par exemple, l'angle de prise de vue. Le *recalage* est l'opération qui consiste à la mise en correspondance des deux acquisitions. Dans notre cas, il s'agit d'une mise en correspondance géographique et géométrique. En général, le recalage est fait sur l'image géolocalisée, et il consiste en la projection des autres images dans le plan de la première image appelée image-source. C'est une étape nécessaire lorsque nous cherchons à comparer, fusionner, en un mot, traiter, plusieurs données simultanément. Il existe de très nombreux algorithmes de recalage, dont la plupart sont rappelés dans [Goshtasby 2005, LeMoigne 2011].

Nous avons utilisé l'outil de recalage du logiciel ENVI. Après la sélection de points de correspondance entre les deux images à recaler, typiquement des correspondances de routes et de bâtiments, le logiciel calcule la transformation à appliquer à l'image-cible pour qu'elle corresponde géométriquement au mieux à l'image-source. La transformation est généralement choisie comme une combinaison de transformations élémentaires telles que la rotation, la translation et la mise à l'échelle. Il est aussi possible de recourir à des transformations mathématiques polynomiales, ou bien des triangulations de Delaunay. Pour plus d'informations, il est conseillé de se référer à [Richards 2006].

5.4.2 Décomposition en ondelettes

Un modèle relativement naturel en traitement d'images pour générer une décomposition hiérarchique est l'utilisation de la décomposition en ondelettes [Mallat 2008]. Nous aurions tout aussi bien pu considérer un filtrage médian encapsulé [Arias-Castro 2009], ou d'autres méthodes de type filtrage/décimation. Cependant, de par leur variété et leur large champ d'application, nous avons préféré opter pour des ondelettes. En outre, le choix du type de décomposition hiérarchique des données n'a, en pratique, pas une influence importante sur les résultats de classification.

Nous n'allons pas détailler l'intégralité de la théorie des ondelettes dans cette partie, mais nous donnons simplement les bases de la décomposition en ondelettes discrète à 2 dimensions ($2D$) qui a conduit, typiquement, à la décomposition d'une image RSO originale comme montré sur la Fig. 5.6, utile pour intégrer une image monorésolue dans un graphe hiérarchique explicite. Cette décomposition peut aussi être utile afin d'augmenter le nombre d'informations disponibles, et ainsi éviter

FIG. 5.6 – Exemple de décomposition en ondelettes d'une image RSO à résolution simple, obtenue par utilisation d'ondelettes de Daubechies-10.

d'éventuels niveaux "vides" dans l'arbre, dus au manque d'information à certains niveaux. En effet, si, par exemple, nous considérons une image optique ayant une résolution de 1 mètre, et une acquisition RSO recalée de la même zone à 4 mètres de résolution, alors l'image à la résolution la plus fine est intégrée au niveau le plus bas du quad-arbre, et l'image RSO au niveau $R = 2$. Le premier niveau ($n = 1$) serait donc vide de toute information ; c'est pour cette raison que nous décomposons l'image à la résolution la plus fine selon 2 niveaux (ou plus), et intégrons les images obtenues comme observations dans l'arbre. Ainsi, en bas et au premier étage, nous avons l'information optique uniquement, alors qu'au deuxième niveau (de résolution 4 mètres), nous avons l'image RSO ainsi que le deuxième niveau de décomposition de l'image optique. Il est aussi possible d'augmenter la taille de l'arbre en introduisant, en plus, une décomposition en ondelettes de l'image RSO.

La décomposition en ondelettes $2D$ est la décomposition d'une image représentée mathématiquement par l'ensemble $\{y(s,t)\}$ – où (s,t) désigne les coordonnées – sur une base orthogonale de $L^2(\mathbb{R}^2)$ constituée de fonctions d'ondelettes $\{\Psi^{LH}, \Psi^{HL}, \Psi^{HH}\}$ et de fonctions d'échelle Φ^{LL}. L'espace que nous considérons est muni de la translation et de la dilatation de paramètres respectifs $\tau = \{\tau_n\}, n \in \mathbb{Z}$ et $a = a_m, m \in \mathbb{Z}$. L'analyse multirésolution dans le cadre de laquelle nous nous plaçons ici est un cas particulier où $\tau_n = 2^m n$ et $a_m = 2^m$. Nous construisons la base d'ondelettes associée dans le cas $2D$, pour laquelle la translation est doublement paramétrée (horizontalement et verticalement). Nous avons alors $\{\Psi^B_{n,q,m}(s,t)\} = \{2^{-m/2}\Psi^B(2^{-m}s-n, 2^{-m}t-q)\}$ où $B = \{LH, HL, HH\}$. De même, nous formons la base $\{\Phi^{LL}_{n,q,m}(s,t)\} = \{2^{-m/2}\Phi^{LL}(2^{-m}s-n, 2^{-m}t-q)\}$.

En considérant une échelle de décomposition J, l'image se décompose alors en tout couple (s,t) :

$$y(s,t) = \sum_{n,q} a_{n,q,J}\Phi^{LL}_{n,q,J}(s,t) + \sum_{B}\sum_{j=1}^{J}\sum_{n,q} d_{n,q,j}\Phi^{LL}_{n,q,j}(s,t) \qquad (5.8)$$

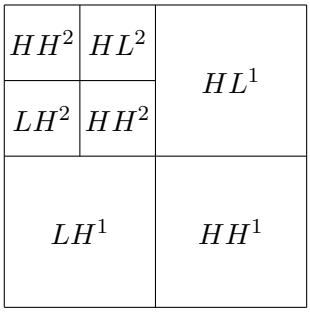


FIG. 5.7 – Schéma représentatif de la décomposition en ondelettes pour une échelle de 2.

où les coefficients $a_{n,q,J}$ et $d_{n,q,j}$ sont appelés respectivement les coefficients d'approximation et de détail.

Le coefficient d'approximation à l'échelle 0 correspond à l'image originale que nous voulons décomposer. En pratique, par filtrage et décimation de cette image, nous obtenons pour l'échelle 1 :

- les coefficients d'approximation par filtrage passe-bas sur les lignes et les colonnes, d'où l'emploi du sigle LL (en se référant au terme anglophone « Low-Low ») ;
- les coefficients de détail par filtrage passe-bas sur les lignes et passe-haut sur les colonnes (LH pour « Low-High ») , passe-haut sur les lignes et passe-bas (HL) ou passe-haut (HH) sur les colonnes.

De manière identique, les coefficients d'approximation de l'échelle j sont décomposés par filtrage et décimation pour obtenir l'ensemble des coefficients de l'échelle $j+1$. Un schéma bien connu pour illustrer cela est donné Fig. 5.7.

Les coefficients que nous conservons à chaque échelle sont les coefficients d'approximation. Le facteur d'échelle étant en puissance de 2, nous nous retrouvons bien dans une configuration de type quad-arbre, où à un pixel à l'échelle $j \neq 0$ correspond 4 enfants et un unique parent.

5.4.3 La fusion optique

Lorsque des images optiques sont considérées à une résolution donnée, celles-ci sont composées, en général, de 3 ou 4 bandes, qui correspondent aux bandes visibles R,V,B et éventuellement à la bande proche infra-rouge. Or, si nous avons en entrée 3 bandes optiques et une seule bande RSO, il est logique de constater que l'image optique aura un poids plus fort sur la probabilité jointe, ce qui n'est pas souhaité. De plus, la complexité calculatoire sur quatre bandes pour le calcul des densités de copules est telle qu'il est difficile de travailler sur des images très grandes. En effet, il faut stocker 256^4 probabilités conjointes possibles, pour des résultats qui demeurent moins satisfaisants, en pratique, que ceux obtenus après fusion des bandes optiques.

En outre, une forte dépendance entre les bandes d'entrée, probable si l'on prend

des bandes issues d'un même capteur, n'est pas toujours modélisable à l'aide des copules, car celles-ci peuvent être limitées par certaines conditions d'applicabilité. Cette forte dépendance se traduit par un τ de Kendall dont la valeur est proche de 1 ou de -1. Une des conditions d'applicabilité des copules est que ce τ doit être compris dans l'intervalle de validité du τ pour les copules (voir Tab. 3.3 du chapitre 3).

Nous avons déjà discuté de la fusion dans l'état de l'art (voir 5.1.2.2), bien qu'il s'agissait plutôt de fusionner des images RSO avec des images optiques. Cependant, les méthodes sont générales, et donc utilisables pour de la fusion de bandes RVB en une bande en niveaux de gris. Nous avons eu recours à la méthode intensité-hue-saturation [Mather 2011] fondée sur la décomposition dans le domaine IHS de l'image RVB. L'image fusionnée est la composante intensité de l'image RVB dans ce nouvel espace.

Lorsque l'image optique considérée possède, en plus de ses composantes RVB, une composante proche-infrarouge (IR), il est possible d'utiliser une décomposition IHS plus générale, qui prend en compte cette composante proche-IR dans l'expression de l'intensité [Tu 2004].

5.5 Résultats expérimentaux

Nous proposons de valider expérimentalement la méthode hiérarchique proposée en sous-partie 5.3.5. Nous allons, tout d'abord, étudier l'influence des paramètres fixés de manière expérimentale, avant de comparer les résultats avec ceux obtenus via des méthodes de référence, dans différentes configurations, à savoir monorésolution, multirésolution et même multicapteur.

Pour les simulations concernant les diverses influences des différents paramètres, nous travaillons avec l'acquisition RSO (CSK) de Cavallermaggiore (Italie), dont les caractéristiques techniques sont données en Annexe A. Étant donné que cette acquisition est monorésolue, et que nous ne possédons pas de données supplémentaires sur cette zone, qu'elles soient optiques ou radars, nous avons donc recours à une décomposition en ondelettes (voir 5.4.2) pour remplir les différents niveaux de l'arbre.

Les paramètres à estimer sont le paramètre du champ de Markov β (sous-parties 5.3.4 et 4.2.1), et le paramètre de la probabilité de transition $\theta_n = \theta$ (sous-partie 5.3.1). Nous devons aussi choisir la famille adéquate pour la décomposition en ondelettes, quand celle-ci est nécessaire (voir 5.4.2). Pour étudier les influences de ces paramètres, nous cherchons la valeur expérimentale du paramètre souhaité pour laquelle la classification est la meilleure, par compromis visuel et numérique, en fixant les valeurs des autres variables. Par défaut, les paramètres sont fixés à $\theta_n = 0,85$, $\beta = 5$ et la décomposition en ondelettes se fait en ayant recours à des ondelettes de type Daubechies [Daubechies 1988]. La méthode utilisée est celle décrite en sous-partie 5.3.5.

Notons enfin, que sauf mention contraire, la décomposition en ondelettes est en

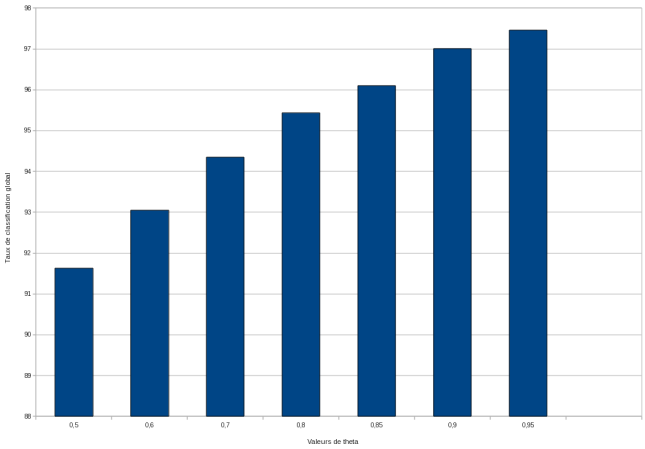
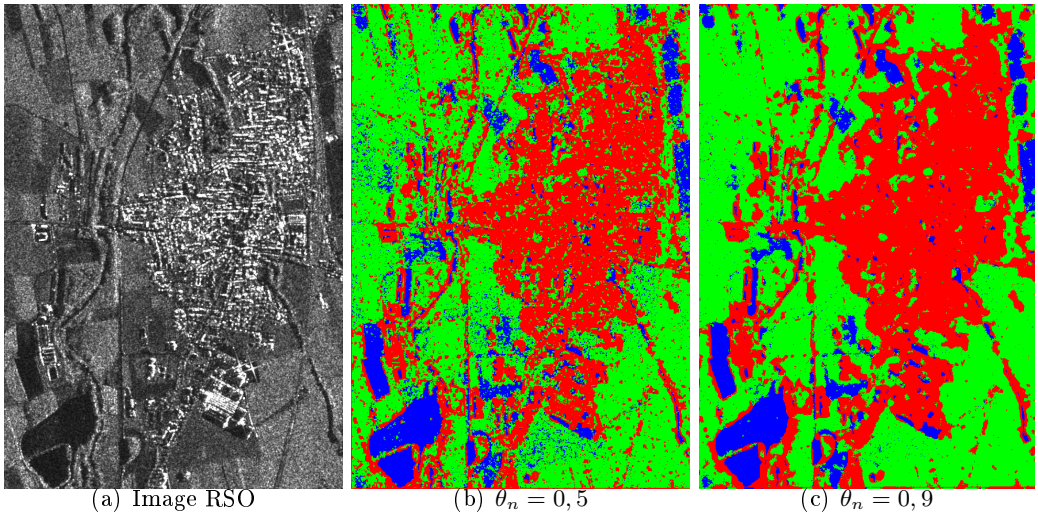


FIG. 5.8 – Taux globaux de bonne classification en fonction de différentes valeurs de θ_n, obtenus pour la classification de l'image de Cavallermaggiore.



(a) Image RSO (b) $\theta_n = 0,5$ (c) $\theta_n = 0,9$

FIG. 5.9 – Influence de θ_n sur la carte de classification obtenue pour l'image de Cavallermaggiore donnée en (a).

général effectuée sur $R = 2$ niveaux. En effet, étant donné la taille maximale des acquisitions sur lesquelles nous travaillons, soit 1000×1000 pixels, la décomposition au niveau 2 est de taille 250×250 pixels. Un niveau supérieur donne, à notre sens, une image "vignette", trop lisse, et dans laquelle les routes, les rivières, n'apparaissent plus.

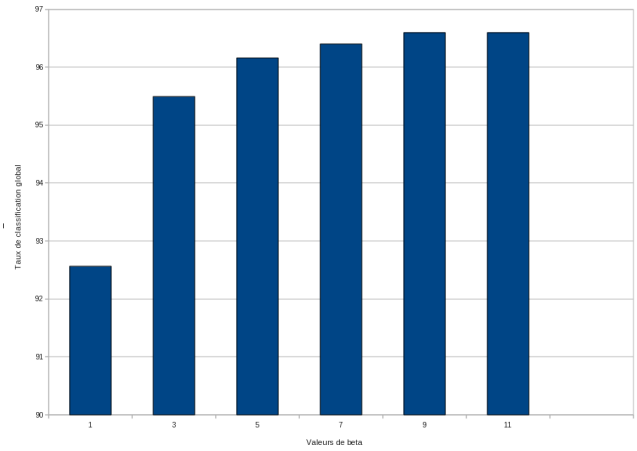


FIG. 5.10 – Taux globaux de bonne classification en fonction de différentes valeurs de β, obtenus pour la classification de l'image de Cavallermaggiore.

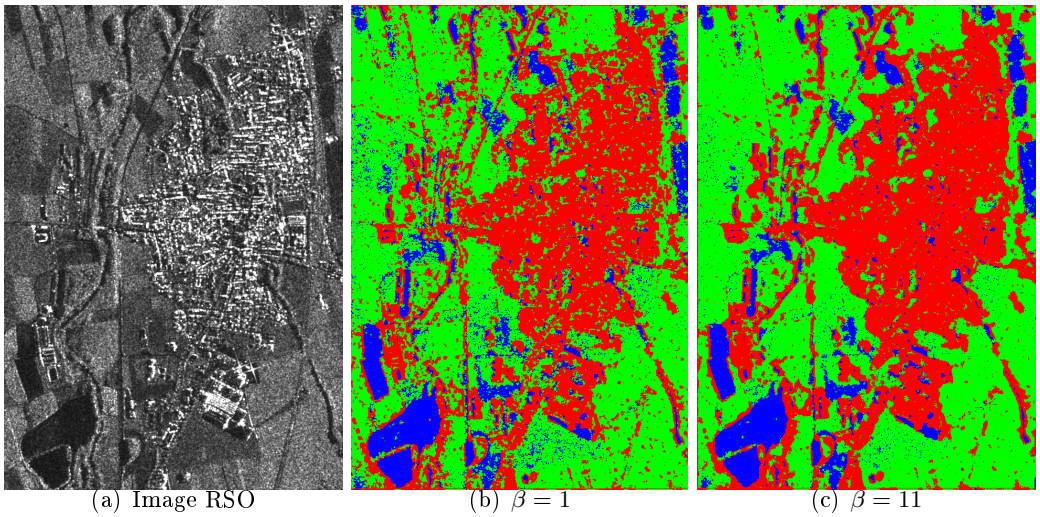
5.5.1 Influence du paramètre θ_n des probabilités de transition

Pour des raisons de simplicité visuelle, nous proposons de montrer graphiquement (Fig. 5.8) l'évolution du taux global de bonne classification obtenu pour l'acquisition de Cavallermaggiore en fonction des valeurs de θ_n. Ce graphique montre que l'augmentation du paramètre θ_n génère une augmentation du taux global de classification. Visuellement, cela se traduit par une carte de classification de plus en plus lisse (Fig. 5.9). Cependant, comme nous l'avons déjà expliqué, plus la carte de classification est lisse, plus les détails sont atténués. Nous avons donc choisi de ne pas utiliser un paramètre θ_n de plus de $0,9$ (contrainte forte sur la probabilité de transition). En pratique, nous avons choisi de fixer $\theta_n = 0,85$, ce qui implique en d'autres termes qu'un site s situé à une échelle n a une probabilité d'environ 85% d'appartenir à la même classe que son ascendant s^-.

5.5.2 Influence du paramètre β des probabilités a priori

De manière similaire à la sous-partie précédente, nous proposons de montrer graphiquement (Fig. 5.10) l'évolution du taux de classification global obtenu pour l'acquisition de Cavallermaggiore en fonction de différentes valeurs de β.

L'influence du paramètre β sur la carte de classification est un peu moins marquée que celle du paramètre θ_n. Cela est clairement visible Fig. 5.11, dans laquelle nous pouvons, certes, observer une carte de classification plus lisse pour un β plus élevé, mais cela n'est pas aussi important que dans le cas où le paramètre θ_n varie (Fig. 5.9). Nous remarquons, d'après le graphique 5.10, qu'à partir d'une certaine valeur de β, située aux alentours de $\beta = 5$, le taux de classification reste à peu près constant. Nous avons donc choisi de fixer $\beta = 5$. En outre, nous n'avons pas opté pour une valeur de ce paramètre plus élevée afin d'éviter que la carte de classification



(a) Image RSO (b) $\beta = 1$ (c) $\beta = 11$

FIG. 5.11 – Influence de β sur la carte de classification obtenue pour l'image de Cavallermaggiore donnée en (a).

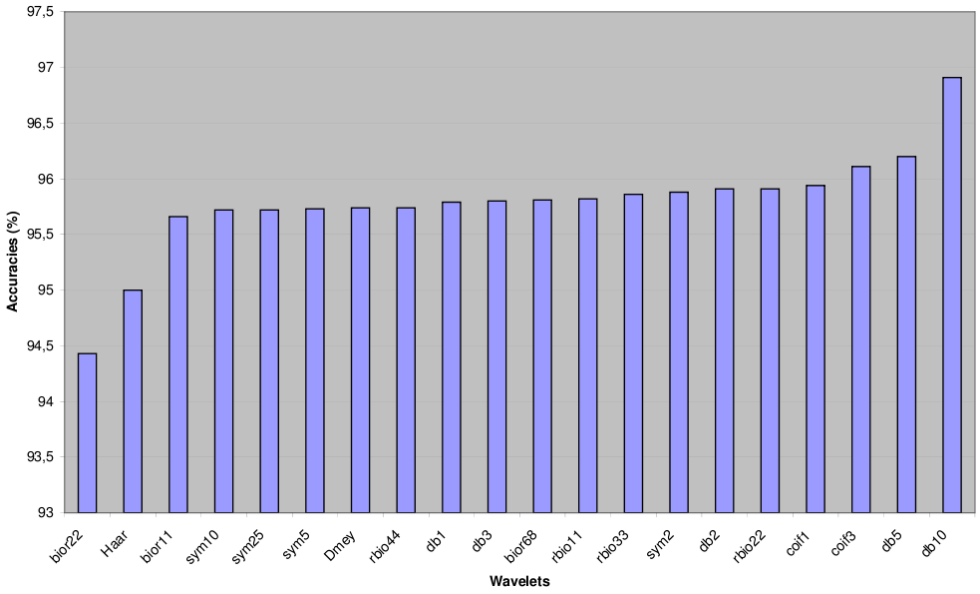


FIG. 5.12 – Taux de classification globaux en fonction de différents types d'ondelettes obtenus pour la classification de l'image de Cavallermaggiore.

ne soit trop lisse.

5.5.3 Influence de la décomposition en ondelettes

Un nombre important de fonctions d'ondelettes existe, et nous avons décomposé notre image originale en utilisant un large choix de familles, afin d'observer les différentes influences de celles-ci sur la classification. Notre choix de familles d'ondelettes est varié : Daubechies [Daubechies 1988], orthogonale, biorthogonale, et quelques autres [Mallat 2008, Vetterli 2010].

Nous avons effectué une comparaison au niveau des taux de classification globaux en fonction de l'ondelette utilisée. Ceci est donné en Fig. 5.12, figure dans laquelle nous remarquons que le meilleur taux de classification est obtenu en utilisant les ondelettes de Daubechies-10.

Une comparaison visuelle a aussi été faite (non montrée ici, étant donné le large choix des familles), et nous avons pu remarquer que les cartes de classification obtenues sont similaires, la principale différence étant au niveau du lissage des cartes de classification finales.

Nous avons donc opté pour une décomposition des images RSO par utilisation d'ondelettes de Daubechies-10. En revanche, pour la décomposition des images optiques, nous avons simplement utilisé des ondelettes de type Haar, qui donnent souvent de bons résultats pour ce type d'images.

5.5.4 Résultats pour différents jeux de données

Nous comparons les résultats obtenus, sur différents jeux de données, pour diverses méthodes de classification :

1. la méthode hiérarchique proposée dans la sous-partie 5.3.5 ;
2. la méthode hiérarchique proposée par Laferté et al. [Laferté 2000], rappelée dans la sous-partie 5.3.3 ;
3. la méthode SVM [Vapnik 2000], rappelée dans la sous-partie 4.3.3 du chapitre 4 ;
4. la méthode proposée dans le chapitre 4 fondée sur des CM. Puisqu'il s'agit d'une méthode non hiérarchique, celle-ci n'est appliquée qu'à l'image du niveau le mieux résolu de l'arbre (niveau 0).

D'après l'état de l'art donné en début de ce chapitre, peu de méthodes – voire même à notre connaissance aucune – utilisent les observations multicapteur à la résolution initiale. Il est donc difficile de pouvoir comparer notre méthode avec des méthodes de l'état de l'art qui ne "dégradent" pas les résolutions des observations, d'où le nombre restreint de tests comparatifs. Nous avons tenté d'appliquer des principes de pan-sharpening inversé, par introduction des informations contenues dans l'image RSO dans les images optiques, ce qui nous ramenait à la classification d'images multibande monorésolution étudiée dans le chapitre 4. Cependant, cela n'a pas abouti à des résultats de classification très probants. Nous avons donc décidé de ne recourir à aucune fusion, excepté celle des bandes optiques, explicitée dans la sous-partie 5.4.3.

TAB. 5.1 – Résultats numériques de classification obtenus pour l'image de Cavallermaggiore. Comparaison entre les CM hiérarchique et non hiérarchique.

	Zone d'eau	Urbain	Végétation	**Global**
CM hiérarchique avec texture	98,20%	91,80%	98,35%	**96,12%**
CM hiérarchique sans texture	98,37%	70,05%	97,99%	**88,80%**
Méthode de Laferté sans texture	94,84%	63,24%	94,60%	**84,23%**
CM mono-échelle avec texture	98,62%	98,42%	100%	**99,01%**

Les caractéristiques des images de télédétection utilisées sont données en Annexe A. Les transformations diverses appliquées aux images (fusion de bandes, décomposition en ondelettes, etc.) sont décrites au cas par cas dans les sous-parties.

Les résultats de classification sont donnés de manière visuelle (carte de classification) et numérique, grâce à une estimation des taux de bonne classification par méthode de validation croisée (voir 4.4 du Chap. 4). Comme précédemment, les vérités de terrain sont sélectionnées manuellement dans des zones homogènes, c'est-à-dire qu'aucune frontière inter-classe n'est prise en compte.

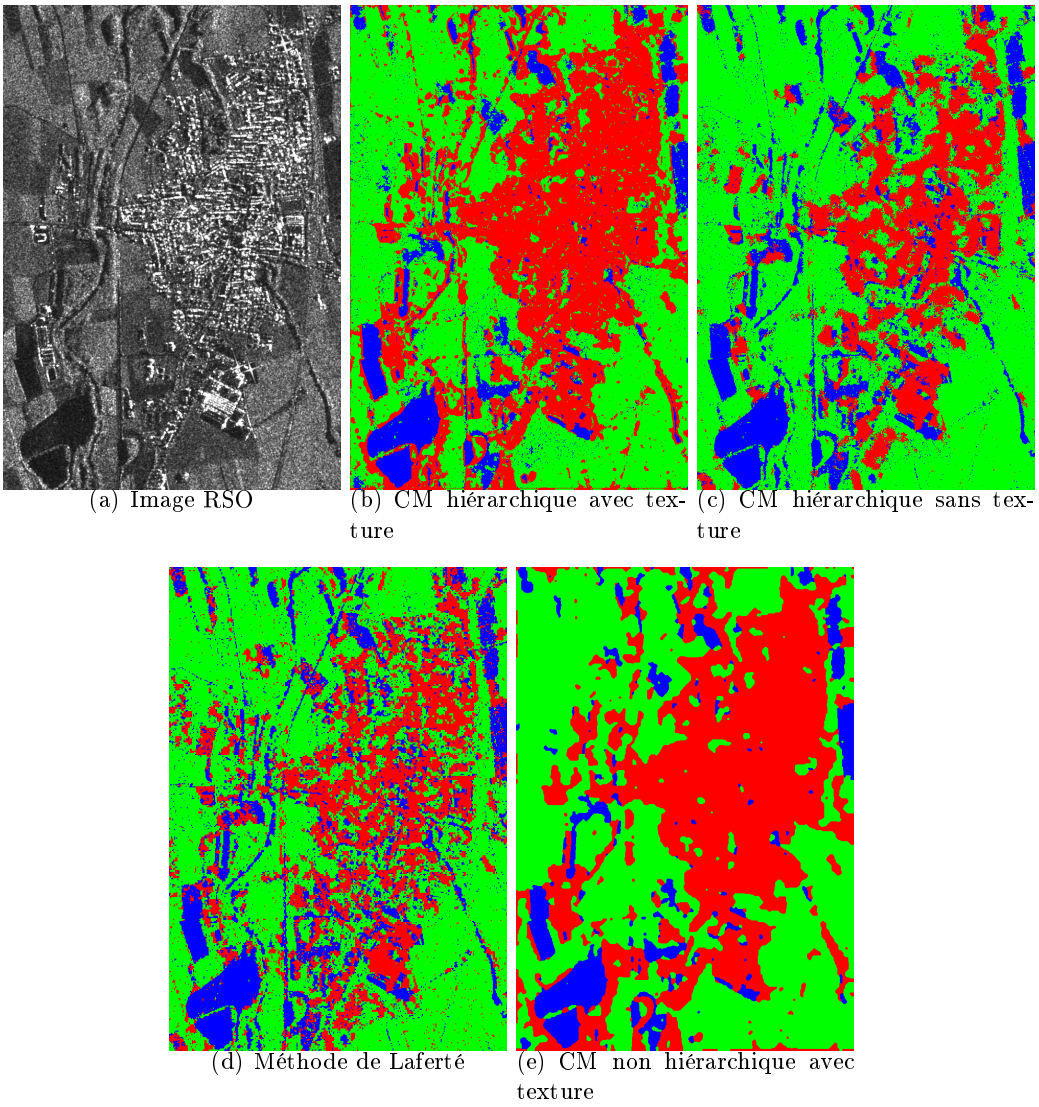
5.5.4.1 Classification monorésolution et monocapteur : Cavallermaggiore (Italie)

Comme nous l'avons déjà expliqué au début de cette partie, l'image de Cavallermaggiore est une acquisition RSO monorésolution. Nous avons voulu montrer, dans cette sous-partie, que la méthode hiérarchique proposée se veut suffisamment générale pour traiter d'images multicapteur, tout en pouvant être intégrée dans l'état de l'art sur les méthodes markoviennes hiérarchiques de classification d'images monorésolution détaillé dans la partie 4.1 du chapitre 4. Trois classes sont considérées ici : les zones d'eau (représentées en bleu), les zones urbaines (représentées en rouge) et la végétation (représentée en vert).

Pour pouvoir intégrer cette acquisition dans la méthode hiérarchique, nous avons procédé à une décomposition selon $R = 2$ niveaux de l'image grâce à des ondelettes de Daubechies-10. Nous avons, à chaque niveau, introduit un attribut de texture MCNG (cf. partie 3.2 du chapitre 3), combiné à l'image (ou sa décomposition) via les copules.

En comparant les résultats obtenus (Fig. 5.13), nous remarquons que l'introduction de l'attribut de texture permet une meilleure discrimination des zones urbaines, et que la méthode proposée est plus robuste au bruit de chatoiement que celle de Laferté, grâce à la mise à jour des probabilités a priori. Ceci est visible si nous comparons les zones d'eau des figures 5.13(c) et 5.13(d). En outre, cela n'est peut-être pas immédiat lorsque nous regardons la Fig 5.13(d), car celle-ci a été compressée pour l'affichage, mais la méthode de Laferté introduit des effets blocs bien connus [Laferté 2000] dans la classification, ce qui n'est pas le cas avec la méthode que nous avons proposée.

Enfin, les résultats numériques (Tab. 5.1) corroborent les résultats visuels, sauf

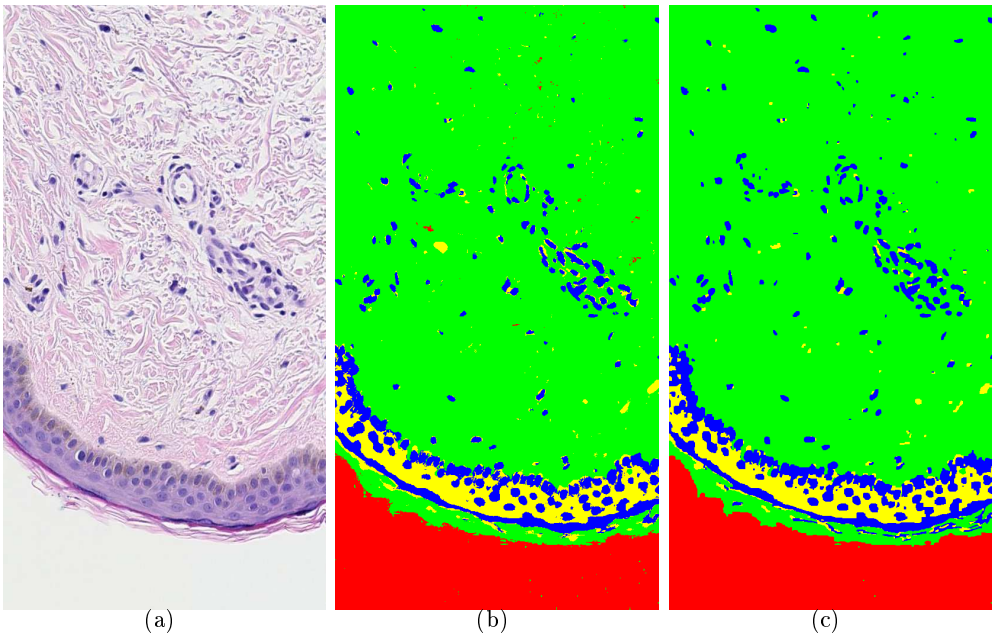


(a) Image RSO (b) CM hiérarchique avec texture (c) CM hiérarchique sans texture

(d) Méthode de Laferté (e) CM non hiérarchique avec texture

FIG. 5.13 – (a) : Image RSO de Cavallermaggiore (Italie) (COSMO-SkyMed,©ASI) ; (b) : Carte de classification obtenue avec la méthode hiérarchique proposée, appliquée à l'image RSO et sa texture MCNG ($w = 5$) ; (c) : Carte de classification obtenue avec la méthode hiérarchique proposée, appliquée à l'image RSO uniquement ; (d) : Carte de classification obtenue avec la méthode hiérarchique de Laferté, appliquée à l'image RSO uniquement ; (e) : Carte de classification obtenue avec la méthode non hiérarchique appliquée à l'image RSO et sa texture MCNG ($w = 5$).

que le taux de classification le plus élevé est obtenu avec la méthode non hiérarchique du chapitre 4. Cela peut s'expliquer par le fait que nous avons pris en compte une vérité de terrain assez grossière, dans laquelle les détails citadins (zones de végétation en ville, par exemple) ne sont pas considérés. Il est donc possible que les détails de



(a) (b) (c)

FIG. 5.14 – (a) Image RVB d'une coupe histologique (©Galderma) (550 × 1020 pixels), (b) classification obtenue avec la méthode hiérarchique proposée dans ce chapitre et (c) classification obtenue avec la méthode non hiérarchique proposée dans le chapitre précédent.

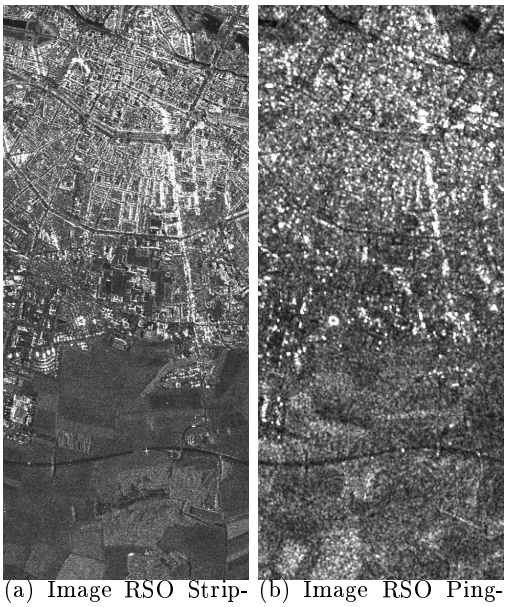
TAB. 5.2 – Taux de bonne classification pour chacune des classes de la coupe histologique. Comparaison entre les CM hiérarchique et non hiérarchique.

	Noyau fibroblaste	Derme	Fond	Épith.	**Global**
CM hiérarchique proposé	97,08%	99,87%	97,71%	97,13%	**97,95%**
CM non hiérarchique	99,92%	99,97%	97,72%	96,65%	**98,56%**

la carte de classification hiérarchique soient mal pris en compte dans les résultats. En outre, pour obtenir des résultats moins détaillés, et plus proches numériquement de ceux obtenus avec le CM mono-échelle, nous pourrions opter pour un choix de paramètres ayant des valeurs plus élevées (cf. études sur les paramètres β et θ_n dans les sous-parties précédentes).

5.5.4.2 Classification monorésolution et monocapteur : coupe histologique

Comme dans le chapitre 4, nous cherchons à montrer que les applications de télédétection ne sont pas les seules applications possibles. Nous traitons une image de coupe histologique R, V, B fournie par l'entreprise Galderma. Cette image étant



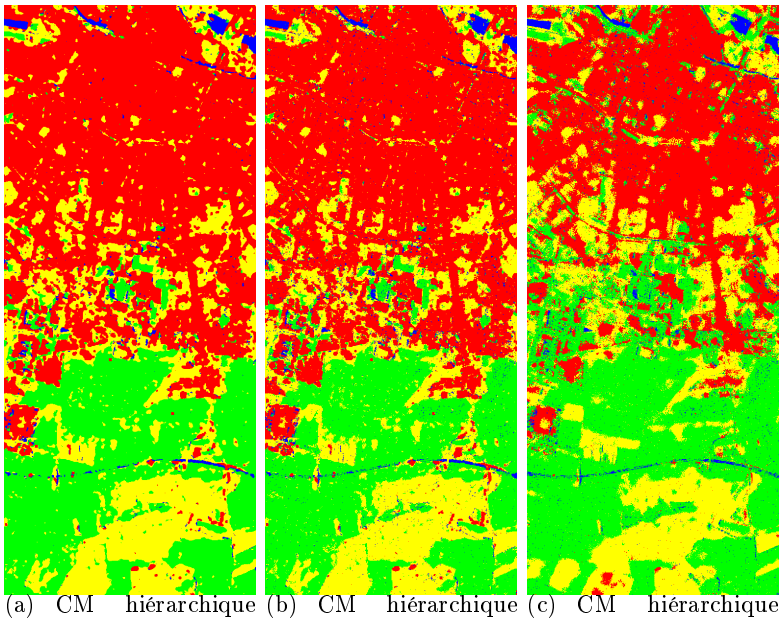
(a) Image RSO Strip-Map (©ASI, 2011) (b) Image RSO Ping-Pong (©ASI, 2011)

FIG. 5.15 – Images RSO originales d'Amiens.

monorésolue, nous avons appliqué une décomposition en ondelettes de Haar sur 2 niveaux pour chacune des bandes. À chaque niveau, ces bandes sont combinées par des copules. Chacune de ces bandes est statistiquement modélisée par des mélanges de distributions gaussiennes, et aucun attribut de texture n'est ici utilisé. La classification est exécutée pour 4 classes physiques : le derme en vert, le noyau fibroblaste en bleu, l'épithélium en jaune et le fond en rouge. Les résultats de classification sont donnés visuellement (Fig. 5.14) et numériquement (Tab. 5.2). Nous remarquons que pour ce cas spécifique, les deux méthodes markoviennes donnent des résultats similaires, ce qui montre la validité de ces méthodes pour des applications autres que la télédétection. Nous aurions tendance à opter pour la méthode la plus rapide d'exécution, qui est la méthode non hiérarchique. En effet, les temps de calcul sont de 3 minutes pour les CM, contre 8 minutes pour les CM hiérarchiques. Les expériences ont été menées sur des processeurs Intel Xeon quad-core (2.40GHz, 12MB cache), 18Gb RAM, sous Linux 64-bits.

5.5.4.3 Classification multirésolution : Amiens (France)

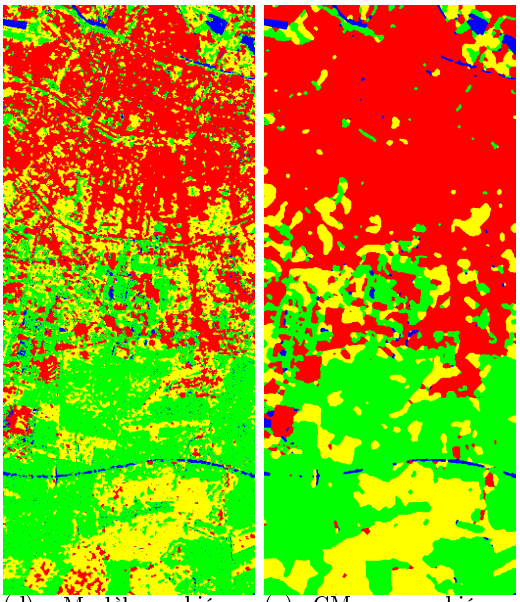
Nous traitons maintenant deux images RSO, acquises au cours de la même année, à des résolutions différentes, présentées en Fig. 5.15. Il s'agit donc d'un jeu de données monocapteur et multirésolution ($2,5$ et 5 mètres de résolution). Nous souhaitons classifier ces images suivant 4 classes : zones urbaines, zones d'eau, végétation basse et végétation haute (arbre). Pour la même raison que celle évoquée dans le chapitre 4, il est judicieux d'introduire un attribut de texture comme bande



(a) CM hiérarchique proposé, avec attribut de texture et post-traitement. $R = 2$

(b) CM hiérarchique proposé, avec attribut de texture. $R = 2$

(c) CM hiérarchique proposé, sans attribut de texture. $R = 2$



(d) Modèle hiérarchique proposé par Laferté. $R = 1$

(e) CM non hiérarchique proposée, sans attribut de texture

FIG. 5.16 – Images RSO originales d'Amiens et classifications obtenues avec les CM hiérarchique et non hiérarchique. Différentes configurations sont testées, à savoir avec ou sans attributs de texture, et avec ou sans ondelettes (resp. $R = 2$ ou $R = 1$).

supplémentaire à chaque résolution. Ce jeu de données étant initialement multirésolution, nous intégrons directement les acquisitions comme observations dans l'arbre hiérarchique. Ainsi, nous avons, de manière naturelle, un arbre de $R = 1$ niveau. Nous proposons de décomposer l'image du niveau $R = 1$ (RSO PingPong) avec des ondelettes afin d'avoir un arbre avec $R = 2$ niveaux.

Nous comparons les diverses combinaisons possibles, à savoir avec ou sans recours aux ondelettes, et avec ou sans recours aux attributs de texture. Les cartes de classification obtenues sont données Fig. 5.16. Visuellement, les résultats sont satisfaisants, et la méthode proposée (Fig. 5.16(b)) semble être un bon compromis entre la carte bruitée obtenue avec le CM hiérarchique proposé par Laferté (Fig. 5.16(d)), et la carte lisse obtenue avec le CM mono-échelle (Fig. 5.16(e)). Dans le Tab. 5.3, nous comparons les taux de bonne classification pour la méthode hiérarchique proposée, qui combine les images d'amplitudes RSO avec leur décomposition ($R = 2$), et la même méthode appliquée seulement aux images RSO d'amplitude ($R = 1$). Cela permet de souligner l'amélioration du taux de classification global de 3% relatif à la considération simultanée d'un arbre plus haut, et des ondelettes. Nous comparons ensuite la méthode proposée à celle de Laferté, toutes les deux considérées sur $R = 1$ (donc sans recours aux ondelettes). Nous pouvons alors voir les avantages de l'introduction de la mise à jour de l'a priori : numériquement, une augmentation de plus de 2% du taux de classification global est observée.

Toujours en nous référant au Tab. 5.3, nous remarquons que, pour ce jeu de données, les résultats numériques sont meilleurs avec l'approche contextuelle non hiérarchique (Fig. 5.16(e)). Cependant, cela est obtenu aux dépens d'un effet de lissage important au niveau des frontières inter-classes. Cela n'affecte pas les résultats numériques du Tab. 5.3 car la vérité de terrain de référence n'inclut pas de zones de transition (nous avions déjà évoqué ce problème dans le chapitre précédent). De plus, l'approche hiérarchique est préférable pour la classification d'images incluant des zones urbaines puisqu'elle permet d'extraire plus de détails que dans le cas du modèle de CM non-hiérarchique.

Le taux de classification des zones de basse végétation est un peu inférieur, et ce, pour toutes les méthodes de classification considérées. Cela est dû aux erreurs de classification de cette classe en végétation haute (arbres), qui apparait en bas de l'image. Cependant, par simple observation de l'image RSO, il est très difficile de détecter que cette zone est effectivement une zone de basse végétation : sa clarté au niveau des niveaux de gris nous induit en erreur. Il faut savoir que ces deux zones ont été distinguées sur la vérité de terrain, car cette dernière a été construite par interprétation visuelle d'images optiques.

Si, au lieu d'utiliser simplement la bande HH de l'acquisition RSO PingPong, nous avions utilisé à la fois les bandes HH et VV, nous n'aurions pas amélioré pour autant de manière significative les résultats de classification. Cela est dû au fait que ces deux bandes contiennent des observations hautement corrélées, et les copules prises en compte considèrent intrinsèquement le même type et le même degré de dépendance entre les différents canaux. Elles sont, de ce fait, beaucoup plus adaptées aux cas multicapteur et multibandes, bien qu'un modèle encore plus adapté serait,

TAB. 5.3 – Taux de bonne classification pour chacune des classes de l'image d'Amiens. Comparaison entre les CM hiérarchique et non hiérarchique.

	Global
CM hiérarchique, avec attribut de texture, $R = 2$ et post-traitement [a]	**93,87%**
CM hiérarchique proposé, avec attribut de texture, $R = 2$ [b]	**93,12%**
CM hiérarchique proposé, avec attribut de texture, $R = 1$ [c]	**91,96%**
CM hiérarchique proposé, sans attribut de texture, $R = 2$ [d]	**91,38%**
CM hiérarchique proposé, sans attribut de texture, $R = 1$ [e]	**88,59%**
Modèle proposé par Laferté, sans attribut de texture, $R = 1$ [f]	**86,32%**
CM non hiérarchique, sans attribut de texture [g]	**93,55%**

	Zones d'eau	Zones urbaines	Basse végétation	Arbres
[a]	97,42%	97,09%	86,20%	94,77%
[b]	96,31%	96,50%	85,31%	94,35%
[c]	95,09%	96,32%	83,34%	93,07%
[d]	96,99%	92,41%	84,91%	91,21%
[e]	96,78%	88,77%	81,63%	87,15%
[f]	96,52%	84,06%	79,13%	85,55%
[g]	98,31%	99,07%	83,95%	92,88%

peut-être, la considération de structure de copules hiérarchiques [Nelsen 2006].

Le post-traitement, détaillé dans la sous-partie 5.5.6, permet d'améliorer la classification globale d'environ 1% en "éliminant" les pixels isolés (Fig. 5.16(a)). Comme nous le verrons ultérieurement, son emploi est discutable, mais, selon l'application, peut être utile.

5.5.4.4 Classification multirésolution et multicapteur : Port-au-Prince (Haïti)

Nous appliquons maintenant l'algorithme de classification hiérarchique proposé à un jeu de données multicapteur composé de deux images recalées du port de Port-au-Prince : une image RSO CSK à 2,5 mètres de résolution, et une image optique GeoEye de résolution initiale 0,5 m, données en Fig. 5.17(a),(b). Les caractéristiques techniques des images considérées sont données en Annexe A. Cinq classes sont considérées : les zones d'eau, les zones urbaines, la végétation basse, le sable et les conteneurs.

Pour la raison évoquée dans la sous-partie 5.4.3, nous avons fusionné les bandes optiques par "pan-sharpening". Aucun attribut de texture n'a été introduit. Pour respecter la décomposition dyadique imposée par la structure même du quad-arbre, nous avons légèrement ré-échantillonné, par interpolation bilinéaire [Jain 1988, Gonzalez 2008], l'image optique pour obtenir une résolution de 0.625 m, qui correspond au quart de la résolution de l'image RSO. Dans ce cas

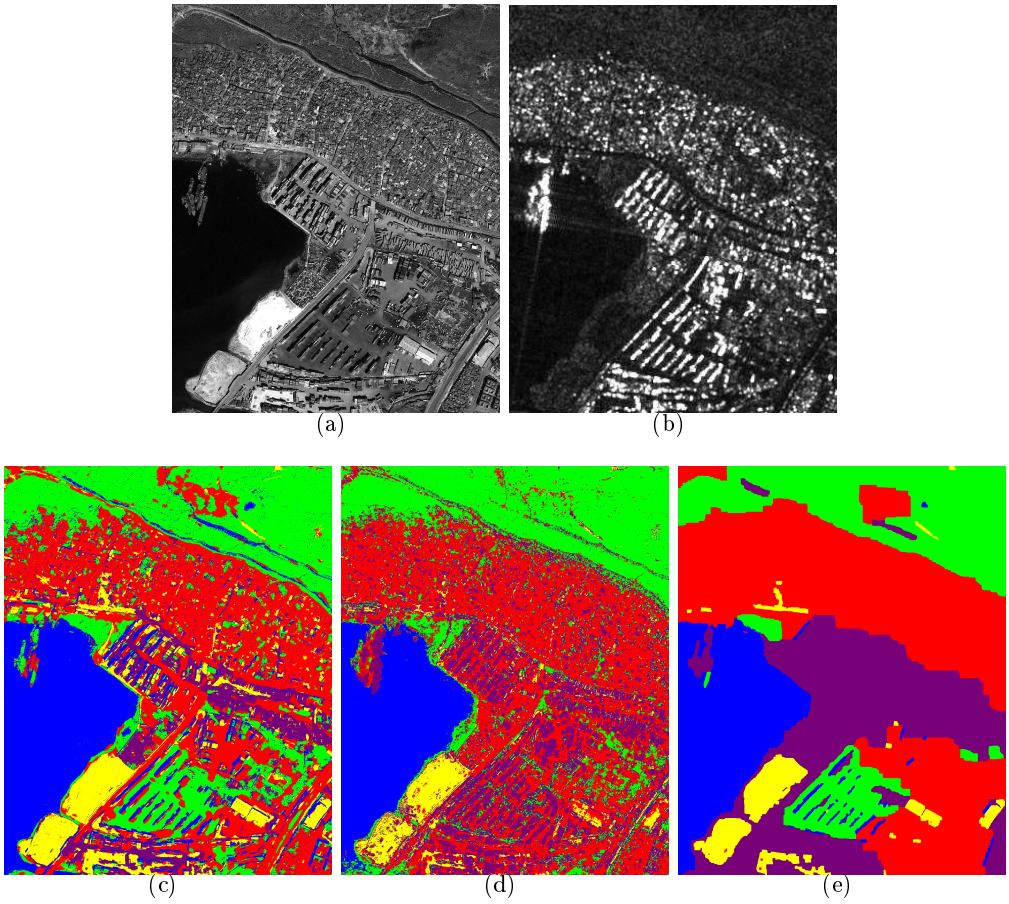


FIG. 5.17 – (a) Image optique (©GeoEye, 2010) et (b) image RSO (CSK, ©ASI, 2010), et différents résultats de classification obtenus avec : (c) la méthode hiérarchique proposée appliquée au jeu de données multicapteur, (d) la méthode SVM appliquée à l'image optique RVB et à l'image RSO sur-échantillonnée, et (e) la méthode non hiérarchique fondée sur des CM appliquée à l'image optique, la mieux résolue. Légende : zones d'eau (bleu), zones urbaines (rouge), végétation (vert), sable (jaune), et conteneurs (violet).

spécifique, l'image optique est intégrée en bas de l'arbre, et l'image RSO à l'échelle $R = 2$. Pour cette raison, et en vue d'éviter tout étage "vide" dans l'arbre, nous intégrons une information additionnelle en introduisant une décomposition hiérarchique de l'image optique originale par recours aux ondelettes de Haar. Ainsi, dans les deux premiers niveaux de l'arbre, nous avons une seule observation optique, au troisième niveau ($R = 2$), nous avons deux observations (optique et RSO). Il est possible d'augmenter la taille de l'arbre en décomposant aussi l'image RSO avec des ondelettes de Daubechies-10. Cependant, étant donnée la taille de l'image initiale, $R = 2$ apparaît comme un choix raisonnable : un plus grand niveau de décomposi-

TAB. 5.4 – Taux de bonne classification pour chacune des classes du jeu de données de Port-au-Prince. Comparaison entre les CM hiérarchique et non hiérarchique.

	Eau	Urb.	Vég.	Sable	Cont.	**Global**
CM hiérarchique	100%	75,24%	87,16%	98,89%	49,31%	**82,12%**
CM hiérarc. (Opt.)	100%	67,12%	86,89%	98,83%	41,90%	**78,95%**
CM non hiérarchique	100%	100%	81,42%	99,94%	59,62%	**88,20%**

tion pourrait supprimer intrinsèquement des détails importants de l'image tels que des routes.

Nous comparons les résultats de classification obtenus avec la méthode hiérarchique proposée, appliquée au jeu de données multicapteur ainsi qu'à l'image optique uniquement. Comme nous l'avions évoqué en préambule de cette partie, il n'y a, à notre connaissance, pas de méthode de classification de l'état de l'art permettant d'inclure les données à leur résolution initiale.

La carte de classification (Fig. 5.17) obtenue avec des données multicapteur en entrée montre des résultats satisfaisants car la classification est relativement bien détaillée (Fig. 5.17(c)). Les résultats numériques (Tab. 5.4) soulignent l'amélioration relative à l'introduction de la donnée RSO par rapport à la prise en compte de l'image optique uniquement, en particulier dans les zones urbaines, dans lesquelles les données RSO semblent d'une grande aide, puisque la séparation entre les conteneurs et les zones urbaines est améliorée. Les principales erreurs de classification relatives à la méthode proposée sont situées dans les zones de conteneurs, dans lesquelles l'asphalte est classifié comme végétation basse.

La méthode SVM, appliquée aux 3 bandes RVB optiques et à l'image RSO sur-échantillonnée, donne visuellement (Fig. 5.17) des résultats globalement similaires à ceux obtenus avec la méthode hiérarchique, bien qu'un peu plus bruités (par exemple, dans la zone de sable). Son principal désavantage est qu'elle requiert le sur-échantillonnage de l'image RSO, ce que nous évitons avec la méthode proposée.

Le tableau 5.4 indique que la méthode non hiérarchique fondée sur des champs de Markov mono-échelle donne de meilleurs résultats numériques. Cependant, la carte de classification obtenue en utilisant une telle méthode est très lisse (Fig. 5.17(e)), et cela n'affecte les résultats numériques que marginalement, les zones de test étant situées dans des régions homogènes (pas de frontières). Visuellement, la carte de classification obtenue avec la méthode hiérarchique est plus détaillée. Pour cette raison, nous préférons cette méthode.

5.5.5 Dépendance des données et validité des copules

Nous avons vu théoriquement que les copules permettent de modéliser tout type de dépendance, des observations indépendantes aux données les plus dépendantes.

Nous avons voulu mener une étude sur cette notion de dépendance, qui pourrait être utilisée comme travail préparatoire à une étude ultérieure, plus poussée, sur la théorie des copules. Il y a plusieurs moyens pour déterminer le niveau de dépendance

des données, nous en avons sélectionné deux, l'un fondé sur le tracé du χ (*chi-plot*) [Genest 2003], et l'autre sur le τ de Kendall [Nelsen 2006].

Nous avons déjà établi la façon dont peut être calculé le τ de Kendall dans la partie 3.3. Il existe un test ad hoc qui permet d'infirmer ou non l'hypothèse d'indépendance, mais, comme ce coefficient est une variable aléatoire qui suit une loi normale sous des conditions très peu restrictives, nous procédons plutôt à un test paramétrique tel que : sous l'hypothèse nulle (H_0), les observations sont indépendantes, et sous l'hypothèse H_1, elles sont corrélées. En effet, ce τ est une variable aléatoire d'espérance nulle et de variance égale à $Var(\tau) = \frac{2(2n+5)}{9n(n-1)}$, où n est le nombre d'observations. Ainsi, τ tend assez rapidement vers une loi normale, et $z = \frac{\tau}{Var(\tau)}$ suit une loi normale $N(0,1)$. Il suffit donc simplement d'une table de la loi normale pour déterminer quelle hypothèse est la plus raisonnable.

Le χ-plot est un graphique proposé par Fisher et Switzer [Fisher 2001] sur lequel sont portées les paires (λ_i, χ_i). Le nuage de points obtenu est alors analysé : s'il est situé à l'intérieur d'une certaine bande de contrôle (en rouge sur les figures 5.18 et 5.19), alors les variables aléatoires X et Y sont indépendantes. Dans le cas contraire, nous pouvons conclure à la dépendance. Pour tout couple (X_i, Y_i), tel que $i \in [1; n]$, où n est le nombre d'échantillons, nous définissons :

$$\chi_i = \frac{H_i - F_i G_i}{\sqrt{F_i(1-F_i)G_i(1-G_i)}},$$

$$\lambda_i = 4 sign(\tilde{F}_i \tilde{G}_i) max(\tilde{F}_i^{\,2}, \tilde{G}_i^{\,2}),$$

avec

$$H_i = \frac{1}{n-1} \# \{ j \neq i : X_j \leq X_i, Y_j \leq Y_i \},$$

$$F_i = \frac{1}{n-1} \# \{ j \neq i : X_j \leq X_i \},$$

$$G_i = \frac{1}{n-1} \# \{ j \neq i : Y_j \leq Y_i \},$$

où $\tilde{F}_i = F_i - 1/2$ et $\tilde{G}_i = G_i - 1/2$. Sous hypothèse d'indépendance, $H_i = F_i \times G_i$. λ_i représente la distance du couple (X_i, Y_i) par rapport au centre du nuage de points, et $\sqrt{n}\chi_i$ est la racine carrée (positive ou négative) de la statistique du χ^2, utilisée pour tester l'indépendance du tableau de contingence (X, Y). La largeur de la bande de contrôle est définie par $\chi = \pm c_p/\sqrt{n}$, qui correspond à un pourcentage des (λ_i, χ_i) se trouvant à l'intérieur de la bande. Typiquement, dans les figures présentées (5.18 et 5.19), $c_p = 2,18$, ce qui correspond à une proportion $p = 0,99$: pour garantir l'indépendance des observations, 99% du nuage de points doit être contenu dans la bande de contrôle. La correspondance entre les proportions et les c_p est déterminée par la méthode de Monte-Carlo [Fisher 2001].

Reprenons les jeux de données précédemment utilisés, et étudions les dépendances des bandes.

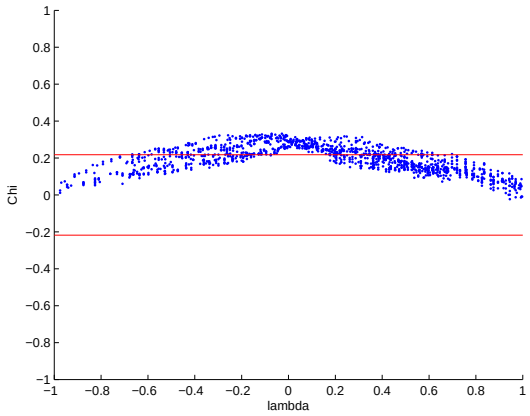


FIG. 5.18 – χ-plot pour la classe "zone urbaine" de l'image de Cavallermaggiore.

TAB. 5.5 – Valeur du τ de Kendall obtenue expérimentalement pour chacune des classes de l'image de Cavallermaggiore (Italie).

	τ de Kendall
Zones d'eau	$0,143$
Zones urbaines	$0,281$
Végétation	$0,276$

Cavallermaggiore

Étant donné que l'image de texture est directement extraite de l'image RSO, nous pouvons pressentir que ces deux bandes sont dépendantes. Les valeurs des τ de Kendall, données dans le tableau 5.5 pour chacune des classes, montrent la véracité de cette intuition. Nous avons aussi tracé le χ-*plot* pour la zone urbaine, les courbes obtenues étant similaires pour les autres classes, en figure 5.18. Les points n'étant pas tous à l'intérieur de la bande de contrôle (représentée en rouge), nous pouvons en déduire que les données sont dépendantes. L'hypothèse d'indépendance n'est, dans ce cas, pas vérifiée, il est donc nécessaire de recourir à l'utilisation de copules comme modélisation de cette inter-dépendance.

Port-au-Prince

Dans ce cas, nous cherchons à combiner des images acquises par différents capteurs. Même s'il s'agit d'images de la même zone, il est difficile de connaître le niveau de dépendance des données optiques et RSO.

En pratique, pour chacune des classes, le τ de Kendall estimé est proche de 0. Nous avons donc affaire à des données quasi-indépendantes, et cela est confirmé par le tracé du χ-plot (Fig. 5.19). L'hypothèse d'indépendance est dans ce cas vérifiée, nous avons donc cherché à comparer les résultats de classification obtenus en modé-

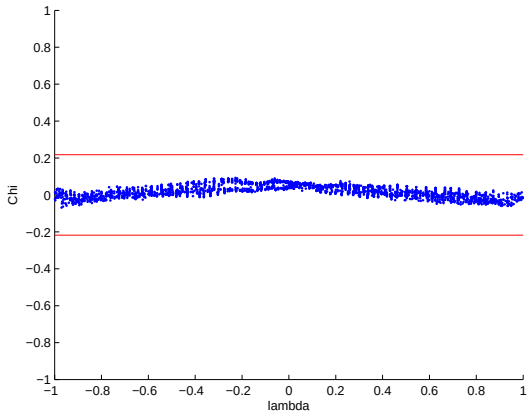


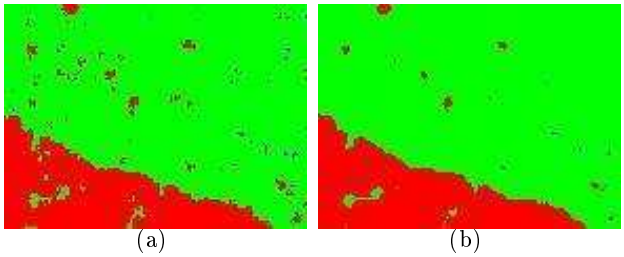
FIG. 5.19 – χ-plot pour la classe "zone urbaine" de l'image de Port-au-Prince.

lisant la probabilité jointe par simple produit de densités marginales, par rapport à l'utilisation effective des copules. Numériquement et visuellement, les résultats sont identiques, ce qui signifie que la densité de copule estimée est égale à 1 (ou proche de 1), cela étant (en grande partie) dû au bruit de chatoiement. Nous aurions donc pu utiliser, dans ce cas spécifique, un simple produit de densités marginales. Cependant, l'utilisation de densités de copules, même si celles-ci sont égales à 1, permet d'établir un modèle général, qui s'adapte à tous les types de bandes, dépendantes comme indépendantes.

5.5.6 Post-traitement

Toutes les méthodes de classification proposées peuvent subir un post-traitement via, par exemple, l'application d'un vote majoritaire [Dobson 1996]. Cette procédure, générale, est appliquée en dehors de l'algorithme de classification. Nous avons remarqué qu'il permet de lisser la carte de classification finale, permettant d'atténuer les effets (par exemple, pixels isolés) du bruit de chatoiement. Typiquement, dans la sous-partie 5.5.4.3, un tel post-traitement améliore le taux global de bonne classification de quasiment 1%. Pour voir les effets d'un tel traitement de manière plus précise, nous zoomons sur un bout de carte de classification ; cela est donné en Fig. 5.20.

Bien entendu, étant donné que cette procédure est appliquée en dehors l'algorithme de classification, elle est donc totalement optionnelle, et peut être utilisée au bon vouloir de l'utilisateur, si celui-ci veut prendre le risque d'obtenir une carte de classification plus lisse au détriment de certains détails. De fait, son utilisation dépend entièrement de l'application finale.



(a) (b)

FIG. 5.20 – Portions de cartes de classification obtenues respectivement sans (a) et avec (b) post-traitement, permettant d'observer les effets (légers, somme toute) du lissage généré.

5.5.7 Pré-traitement

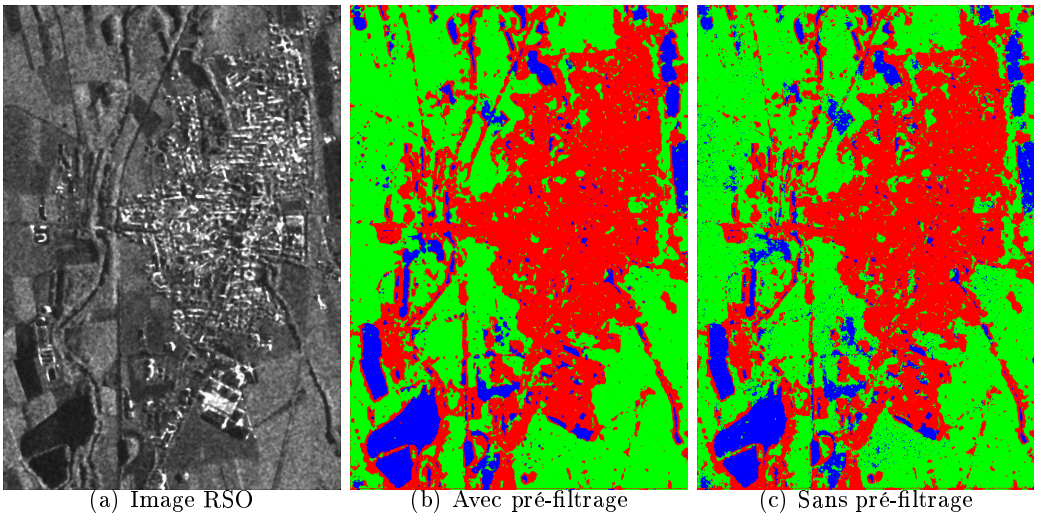
Au vu des effets du bruit de chatoiement sur la méthode hiérarchique, nous proposons comme expérience de débruiter légèrement l'image RSO initiale. Statistiquement et physiquement parlant, un tel débruitage est discutable puisque comme nous l'avons décrit dans le chapitre 2, le bruit comporte des informations sur l'image. En outre, la modélisation statistique est considérée comme encore valable, alors que mathématiquement, cela est tout aussi discutable. Cependant, nous avons été curieux d'observer comment évolue la classification avec un tel débruitage.

Nous présentons quelques filtres bien connus de la littérature. Le filtre sigma établit une moyenne de pixels dans une fenêtre glissante donnée, moyenne estimée au regard de différents types de noyaux. Les filtres de Lee et de Kuan sont généraux et couramment utilisés [Nicolas 2001], et visent à minimiser l'erreur quadratique moyenne. Le filtre Gamma est fondé sur la recherche du maximum a posteriori dans un cadre bayésien, sous hypothèse – restrictive – que la réflectivité du radar ainsi que le bruit suivent une distribution Gamma [Gagnon 1997]. Le filtre de Wiener, ou les versions adaptatives qui en découlent (par exemple, filtre de Frost), convoluent les valeurs de pixels dans une fenêtre de taille donnée avec une réponse impulsionnelle exponentielle [Gagnon 1997]. Le filtre de Kalman est un filtre $2D$ fondé sur le principe que les pixels sont markoviens, et leurs statistiques suivent un modèle causal auto-régressif [Gagnon 1997].

Les ondelettes ont aussi été utilisées [Gagnon 1997, Bijaoui 1996], mais nous n'avons pas pris en compte celles-ci pour des raisons de potentielle redondance avec la décomposition en ondelettes que nous utilisons pour les données.

L'utilisation de la transformée de Mellin dans le cadre spécifique du filtrage RSO a montré des résultats intéressants [Nicolas 2001] car la convolution de Mellin est plutôt bien adaptée au bruit multiplicatif. En outre, sa formulation peut prendre en compte diverses hypothèses sur le chatoiement.

Nous avons opté tout simplement pour un filtre médian, qui remplace la valeur d'un pixel par la valeur médiane de l'ensemble des pixels de son voisinage. Il est connu pour son applicabilité aux images en niveaux de gris, sa robustesse au bruit



(a) Image RSO (b) Avec pré-filtrage (c) Sans pré-filtrage

FIG. 5.21 – (a) : Image RSO de Cavallermaggiore après filtrage médian ; (b) : Carte de classification obtenue avec la méthode hiérarchique proposée, appliquée à l'image RSO filtrée ; (c) : Carte de classification obtenue avec la méthode hiérarchique proposée, appliquée à l'image RSO non filtrée (résultats de la sous-partie 5.5.4.1).

TAB. 5.6 – Effets d'un pré-traitement sur les taux de bonne classification.

	Zone d'eau	Urbain	Végétation	**Global**
Avec pré-filtrage	98,95%	94,00%	99,63%	**97,53%**
Sans pré-filtrage	98,20%	91,80%	98,35%	**96,12%**

de type "poivre et sel", et sa capacité à préserver les frontières [Arce 2005]. Celui-ci est choisi comme légèrement adaptatif. En effet, si nous traitons de l'image RSO de Cavallermaggiore monorésolution, nous appliquons, pour intégrer les données dans l'arbre, une transformée en ondelettes, que nous pouvons déjà considérer comme un filtre. Pour cette raison, nous avons choisi un filtrage médian assez fort en bas de l'arbre (meilleure résolution, donc la plus bruitée) et un filtrage très faible en haut de l'arbre (résolution la plus grossière, donc partiellement débruitée). Nous avons, ensuite, combiné ces images débruitées aux attributs de texture, puis classifié celles-ci en utilisant la méthode de classification proposée dans ce chapitre. Nous observons que le taux de classification global (Tab. 5.6) est meilleur. En outre, la carte de classification obtenue (Fig. 5.21(b)) semble plus robuste au bruit de chatoiement que lorsqu'aucun filtre n'est appliqué au préalable (Fig. 5.21(c)). Cependant, la classification obtenue semble un peu moins détaillée, ce qui est probablement dû aux effets de dilatation relatifs au filtre utilisé. Des modèles plus spécifiques de débruitage d'images RSO ont été récemment développés [Parrilli 2012], prenant en compte désormais la géométrie locale de l'acquisition, par exemple en utilisant les bonnes propriétés des champs de Markov combinés aux ondelettes [Zhang 2009b].

Cette étude n'est qu'une étude préliminaire, et nous avons pu entrevoir qu'un pré-traitement de l'image RSO initiale pourrait permettre d'améliorer la classification, sans toutefois pouvoir démontrer une quelconque justification théorique précise.

5.6 Conclusion

La méthode de classification supervisée proposée au cours de la thèse combine une modélisation statistique souple (mélanges finis et copules) avec un champ de Markov hiérarchique qui intègre une mise à jour de la probabilité a priori, améliorant de ce fait la robustesse du classifieur par rapport au bruit de chatoiement. Les étiquettes sont déterminées grâce à la recherche d'un critère MPM, ce qui se fait via un algorithme d'optimisation non itératif. Les résultats obtenus avec cette méthode sont un bon compromis entre les résultats assez lisses obtenus en utilisant la méthode des CM mono-échelles du chapitre 4, et les résultats assez bruités obtenus en utilisant la méthode proposée par Laferté et al. (sous-partie 5.3.3).

En outre, l'avantage principal de la méthode développée est qu'il s'agit d'un modèle hiérarchique qui permet de prendre en compte différents types de données, que ces données soient monorésolution ou multicapteur.

Une des limitations du MPM est, en général, les effets de bloc, que nous n'observons pas ici, grâce à la récursivité et la robustesse de la méthode proposée. Cependant, d'autres problématiques restent à explorer.

Une première problématique est liée à la limitation relative à la structure dyadique imposée sur les pixels, et qui réduit nos possibilités d'intégration d'observations dans les arbres, impliquant parfois un ré-échantillonnage des données. Dans [Yang 2006], les auteurs proposent une structure en quad-arbre tronquée, adaptée non pas aux sites, mais directement aux pixels en bordure des régions de façon à être plus flexibles. Un algorithme MPM est, ensuite, appliqué à cet arbre. Les pré-requis sont l'obtention d'une classification préalable à chaque niveau. Dans la même lignée, un arbre fondé sur une décomposition hiérarchique des régions par fusion des régions adjacentes a été considéré dans [Katartzis 2005]. Une fois encore, les auteurs font appel au MPM pour traiter ensuite les images. Ces diverses idées pourraient être intégrées dans notre méthode.

Une deuxième problématique, comme nous l'avons vu en toute fin de ce chapitre, tourne autour de l'utilisation de méthodes de pré-traitement et/ou de post-traitement, qui peuvent présenter certains avantages, comme, par exemple, l'amélioration des résultats par élimination de certains pixels isolés.

Quelques autres horizons restent à explorer, notamment au niveau de l'estimation des paramètres. Tout comme le paramètre β, le paramètre θ_n pourrait, dans un premier temps, être estimé par apprentissage sur des vérités de terrain exhaustives, afin de prendre en compte, au mieux, les probabilités de transition entre des pixels de classes différentes. Une telle configuration n'est, en général, possible qu'au niveau des frontières inter-classes. Ainsi, nous pourrions même envisager un paramètre θ_n spécifique à chacun des niveaux de l'arbre, et donc dépendant de la couche n. Puis,

par extension, l'estimation des paramètres β (a priori) et θ_n (transitions) pourrait devenir entièrement automatique, voire même semi-ou non-supervisée. Nous avions, par ailleurs, déjà discuté de ces possibilités d'automatisation dans le chapitre précédent pour le paramètre β (sous-partie 4.2.2.1).

Chapitre 6

Conclusion générale et Perspectives

Cette thèse, intitulée "Classification supervisée d'images d'observation de la Terre par utilisation de méthodes markoviennes", nous a permis de mettre en oeuvre divers aspects du traitement d'images. Les méthodes de classification développées, essentiellement statistiques, sont fondées sur la théorie bayésienne, et permettent de traiter différents types d'images. Dans notre cas, il s'agit d'images optiques et d'images radars, plus précisément radars à synthèse d'ouverture (RSO).

Ces méthodes nécessitent une étape d'apprentissage, qui se traduit par la modélisation statistique des niveaux de gris pour chacune des bandes en entrée de l'algorithme, et chacune des classes considérées pour la classification. Cette modélisation statistique est faite grâce au recours à des mélanges finis, qui sont des sommes pondérées de fonctions de densité de probabilité (FDP) bien choisies. Nous avons opté pour des mélanges de gaussiennes pour les images optiques, et des mélanges de distributions Gamma généralisées pour les images radars. Les paramètres des FDP sont automatiquement estimés grâce à des algorithmes stochastiques d'espérance-maximisation (SEM). Pour chacune des classes, les FDP des différentes bandes (par exemple, plusieurs polarisations RSO) sont combinées grâce à des copules d-variées, qui modélisent la dépendance entre les d bandes d'entrée.

Les probabilités jointes relatives aux différentes classes sont ensuite intégrées dans des modèles contextuels, qui visent à permettre une classification robuste au bruit, en particulier au bruit de chatoiement des images RSO.

Le premier modèle considéré est un champ de Markov, qui prend en compte un contexte spatial (voisinage des pixels), particulièrement adapté à des images mono-résolution. Nous avons étudié l'influence du choix du paramètre de ce champ sur la classification, et avons opté pour une valeur médiane, compromis entre une valeur trop faible qui génère une classification affectée par le bruit inhérent au capteur, et une valeur trop élevée, qui génère un sur-lissage. Nous avons ensuite comparé la méthode proposée à d'autres algorithmes de l'état de l'art, comme, par exemple, les séparateurs à vaste marge (SVM), et avons pu apprécier les bons résultats obtenus. En outre, la méthode est suffisamment générale pour traiter non seulement des images de télédétection (notre premier objectif), mais aussi des images biologiques (coupes histologiques). Il suffit juste d'avoir initialement une modélisation adéquate des statistiques des niveaux de gris des images.

Les images RSO qui nous ont été fournies sont, en pratique, monopolarisées (i.e.,

monobande). Nous avons donc opté pour l'introduction d'une information supplémentaire via l'extraction d'un attribut de texture bien choisi, qui met en relief les zones urbaines par rapport aux autres zones à classifier. Cela est, en outre, plus aisé avec des images RSO qu'avec des images optiques, car les ondes hyperfréquence sont davantage reflétées par les matériaux contenus dans les zones urbaines. Les images RSO et leurs attributs de texture sont ensuite combinés statistiquement grâce à des copules.

Le deuxième modèle considéré est un champ de Markov hiérarchique, qui vise à étendre l'étude précédente à des images multirésolution et/ou multicapteur recalées. Il est fondé sur un modèle mathématique en quad-arbre, qui offre de bonnes propriétés, notamment la causalité, qui permet d'appliquer des algorithmes non itératifs pour l'estimation des étiquettes. Cette estimation est obtenue par recherche du critère des modes des marginales a posteriori (MPM). Le contexte est cette fois-ci en échelle, et est donc entièrement déterminé par les probabilités de transition inter-niveaux. L'algorithme de classification hiérarchique proposé intègre une mise à jour de la probabilité a priori combinée à une troncature itérative du quad-arbre, afin de diminuer les effets du bruit des images traitées. Un des principaux avantages relatif à une telle structure hiérarchique est le fait de pouvoir utiliser des images à leur résolution initiale, sans nécessairement recourir à des techniques de fusion d'images multicapteur. Ce modèle hiérarchique permet aussi de classifier des images à simple résolution, en décomposant ces images grâce, par exemple, à des ondelettes, et d'intégrer à la fois l'image initiale et ses décompositions dans les différents niveaux d'observation de l'arbre. Nous avons étudié expérimentalement les influences des paramètres des probabilités a priori et de transition, ainsi que l'influence du choix des ondelettes, afin de choisir les paramètres optimaux. L'algorithme proposé a été validé expérimentalement sur divers jeux de données : monorésolution, multirésolution, et multicapteur. En revanche, la comparaison avec l'état de l'art s'est avérée ardue du fait du manque de méthodes de classification prenant en compte des données multirésolution à leur résolution initiale. Nous avons, cependant, pu constater que la méthode proposée permet de mettre en avant des détails sur la carte de classification, ce qui n'est pas pas forcément le cas avec d'autres méthodes (incluant notamment la méthode fondée sur des champs de Markov mono-échelle).

Nous avons, enfin, discuté de la considération potentielle de pré- et de post-traitements, qui pourraient améliorer la classification au détriment de certains détails malheureusement. Toute la philosophie de la classification des images, radars en particulier, est d'apprendre à faire des compromis entre une classification détaillée, mais légèrement bruitée, et une classification grossière, mais peu affectée par le bruit environnant.

Cette thèse nous a permis d'étudier divers aspects du traitement d'image, qui comprennent, bien entendu, la classification, mais aussi le filtrage et la fusion. Dans le cadre des risques naturels, la classification est importante, mais elle pourrait être complétée, par exemple, par de la détection de changements. Avec ces deux aspects, une grande partie des risques naturels pourraient être étudiés, dans toutes les phases du risque lui-même : prévention, secours, et reconstruction.

La détection de changements n'a été considérée que brièvement durant la thèse. La publication donnée en Annexe E.7 montre une des potentialités de la détection de changements, et les résultats encourageants obtenus montrent que l'analyse statistique a encore un bel avenir en traitement d'images.

Pour conclure ce travail de thèse, nous proposons de donner quelques perspectives possibles en vue d'améliorer le travail effectué, et de traiter des points encore délicats.

Perspective 1 : Validation graphique des copules

Nous avons utilisé le tracé du *chi-plot* dans le chapitre 5 comme tracé permettant de déterminer la dépendance des observations. Mais cette courbe peut aussi être utilisée pour en déduire la copule sous-jacente [Fisher 2001]. Dans un même objectif, des graphes de dépendance plus simples ont été proposés [Genest 2003], fondés sur une adaptation du tracé "quantile-quantile" : les *Kendall-plots*. Ceux-ci sont plus faciles à interpréter, et, par propriété, beaucoup plus faciles à relier aux copules. En outre, les *Kendall-plots* ont pour avantage d'être applicables à des contextes multivariés, ce qui nous intéresse dans le cadre de la sélection de la copule multivariée la plus adaptée.

L'idée principale serait donc de faire le choix de la meilleure copule non plus en effectuant un test statistique du χ^2, mais en utilisant cette méthode graphique, de façon à donner une validité à la fois théorique et visuelle.

Perspective 2 : Utilisation d'un voisinage adaptatif

Nous proposons, comme deuxième perspective, de prendre en compte un modèle de Potts un peu plus évolué que celui, non isotrope, présenté dans la partie 4.2.1 du chapitre 4. Deux modèles exploités dans la littérature seraient intéressants à considérer : le voisinage adaptatif [Smits 1997, Zhong 2007] et les champs de lignes [Zhang 1993], tous deux fondés sur la reformulation de la fonction énergie définissant la caractéristique locale du champ de Markov (voir 4.2.1 du chapitre 4). Le voisinage adaptatif nous semble relativement adapté et assez facile à exécuter. Il permet de prendre en compte la géométrie locale d'une classe donnée et donc d'introduire un aspect géométrique dans la classification. Cette géométrie est d'autant plus intéressante que nous travaillons sur des images à très haute résolution, pour lesquelles les aspects géométriques des rivières ou des routes, par exemple, peuvent être visibles. Le principe repose non plus sur le fait de prendre en compte un simple voisinage de 8 pixels autour d'un pixel donné, mais sur celui de posséder un "dictionnaire" de voisinages possibles, et de choisir celui qui donne la plus haute énergie localement au cours de l'étape de maximisation de l'a posteriori. Si nous considérons un cercle de rayon 2 pixels (Fig. 6.1(a)), nous aurons alors divers voisinages possibles, dont certains sont donnés dans la figure 6.1. Dans un tel cas, la fonction énergie peut être réécrite [Smits 1997] :

$$p(x_s) = \frac{1}{Z}\exp(-\frac{\beta}{N_s^\alpha} \times \sum_{s:\{s,t\}\in C} \delta_{x_s=x_t}) \quad \text{où } \delta_{x_s=x_t} = \begin{cases} 1, & \text{si } x_s = x_t \\ 0, & \text{sinon} \end{cases}. \qquad (6.1)$$

où nous rappelons que s désigne un site, x_s l'étiquette attribuée à ce site et x_t est une étiquette, telle que t et s appartiennent à la même clique dans une configuration de voisinage α donnée. N_s^α désigne le nombre de voisins du site s dans la configuration α donnée.

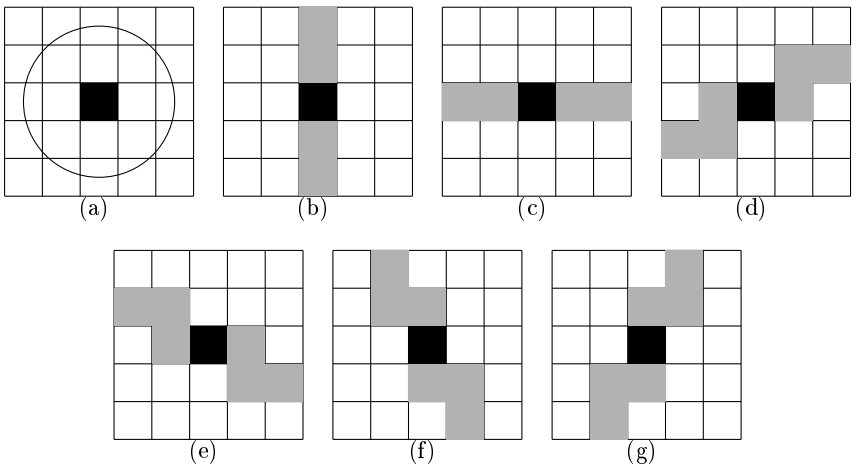


FIG. 6.1 – (a) Cercle de rayon 2 dans lequel sont inscrits les voisinages et (b)->(g) : exemples de voisinages possibles. Le pixel noir est le site s considéré, ses voisins sont représentés par du gris clair.

Nous avons étudié les influences de ces voisinages, et les résultats que nous avons obtenus sont encourageants, car ils montrent une augmentation du taux de classification global d'un peu plus de 1% en moyenne. Cependant, un travail important reste encore à faire sur l'étude des voisinages, ainsi que sur l'ordre des cliques à considérer. Cela permettrait globalement d'accéder à de meilleurs résultats de classification, plus détaillés. La considération des voisinages adaptatifs vise, comme nous l'avons mentionné précédemment, à prendre en compte un aspect géométrique local. L'ordre des cliques est, quant à lui, relatif au nombre de pixels à considérer dans le voisinage. Cet ordre est donc à étudier, d'une part pour que l'aspect géométrique soit suffisamment bien pris en compte, d'autre part pour que nous ne choisissions pas un nombre trop élevé de pixels, en vue d'économiser de la mémoire et du temps de calcul.

Perspective 3 : Optimisation du temps de calcul

Les résultats de classification avec la méthode hiérarchique (chapitre 5) sont obtenus en plus d'une trentaine de minutes de calculs en pratique. Nous avons pu entrevoir, dans le chapitre 4, les bénéfices apportés par les graph-cuts au niveau des temps de calcul. À ce jour, aucune méthode des graph-cuts n'a encore été introduite dans un algorithme de recherche du critère MPM, et serait adaptée dans le sens où nous travaillons avec un graphe (hiérarchique). Il serait donc intéressant d'explorer

une telle introduction, qui pourrait permettre de gagner un temps de calcul considérable, afin de nous rapprocher d'avantage d'un traitement en temps réel. Ce temps réel serait un véritable gain dans le cadre des risques naturels, car l'exploitation des données serait plus rapide et donc plus efficace. En outre, il serait intéressant d'étudier si une parallélisation de l'algorithme est possible, pour gagner du temps au niveau du calcul des probabilités.

L'optimisation du temps de calcul serait aussi intéressante dans le but de pouvoir traiter des images bien plus grandes (par exemple, 10000 pixels sur 10000 pixels), correspondant à des tailles typiques d'acquisition. Actuellement, le code implémenté ne prend pas en compte de telles tailles d'images, et si l'utilisateur souhaite obtenir une classification, il est préférable de subdiviser au préalable l'image originale et de traiter chaque bout indépendamment. Le traitement des images sans pré-découpage permettrait d'éviter une étape d'"assemblage" de divers bouts d'images, et pourrait permettre d'intégrer un plus grand nombre de classes dans la classification, ce qui est possible grâce à la flexibilité de la méthode statistique de l'apprentissage (mélanges finis).

Perspective 4 : Utilisation d'un quad-arbre légèrement différent

Cette perspective est, à notre sens, la plus importante et la plus intéressante à étudier, car elle apporterait une réelle innovation.

Un des principaux défauts du quad-arbre considéré dans le chapitre 5, est qu'il requiert une décomposition dyadique sur les résolutions des images (puisque l'on traite directement les pixels). Il est donc parfois nécessaire de ré-échantillonner les images initiales. Nous voudrions conserver le même modèle de quad-arbre, pour garder le même type d'algorithme, mais pouvoir l'appliquer sans recourir à un quelconque redimensionnement de pixels. Une des meilleures idées que nous ayons vues dans la littérature, et qui serait intéressante à exploiter, est l'utilisation du quad-arbre sur les régions [Yang 2006].

Une autre possibilité serait d'étudier la causalité d'un modèle d'arbre plus flexible, dans lequel les résolutions seraient les résolutions initiales, et non plus adaptées à une série dyadique. À un pixel ne correspondrait donc plus 4 pixels, mais un nombre possiblement non entier de pixels, ce qui implique l'introduction d'une notion de poids sur les descendants. Un tel modèle n'existe pas encore dans la littérature, il serait donc intéressant de l'étudier, pour créer une méthode de classification générale, applicable à tous les types d'images et toutes les résolutions.

Annexe A

Caractéristiques techniques des images traitées

Cette annexe vise à présenter de manière relativement détaillée les caractéristiques des images satellites utilisées comme illustrations dans cette thèse, afin d'éviter toute redondance, les images pouvant être utilisées dans différents chapitres. Les résultats sont présentés sous forme de tableau dans la page suivante.

Site	Capteur	Année	Polarisation	Mode d'acquisition (résolution)
Amiens (France)	CSK (©ASI)	2011	Simple (HH)	StripMap (2, 5 m)
Amiens (France)	CSK (©ASI)	2011	Dual	PingPong (5 m)
Cavallermaggiore (Italie)	CSK (©ASI)	2008	Simple (HH)	StripMap (2, 5 m)
Lombriasco (Italie)	CSK (©ASI)	2008	Simple (HH)	StripMap (2, 5 m)
Port-au-Prince (Haïti)	CSK (©ASI)	2010	Simple (HH)	StripMap (2, 5 m)
Port-au-Prince (Haïti)	GeoEye-1 (©GeoEye)	2010	4 bandes : RVB+PIR	résolution de 0, 6 m
Rosenheim (Allemagne)	TSX (©Infoterra)	2008	Simple (HH)	SpotLight (8, 2 m)

Site	Caractéristiques	Taille de l'image
Amiens (France)	geocoded, single-look	510×1200 pixels
Amiens (France)	geocoded	255×600 pixels
Cavallermaggiore (Italie)	geocoded, single-look	650×930 pixels
Lombriasco (Italie)	geocoded, single-look	950×700 pixels
Port-au-Prince (Haïti)	geocoded, single-look	320×400 pixels
Port-au-Prince (Haïti)	geocoded	1280×1600 pixels
Rosenheim (Allemagne)	ellipsoid corrected, 4-look	900×600 pixels

Annexe B

Expression générale de la densité de la copule de Clayton multivariée

Cette annexe présente les étapes de calcul aboutissant à l'expression générale de la densité de Clayton multivariée.

L'expression analytique de la copule multivariée est :

$$C(u_1, ..., u_d) = \left[\left(\sum_{i=1}^{d} u_i^{-\theta} \right) - d + 1 \right]^{-1/\theta}$$

Nous allons donc dériver cette expression variable par variable, et en déduire une expression générale pour la densité.

Dérivée par rapport à u_1 :

$$\frac{\partial C(u_1, ..., u_d)}{\partial u_1} = \frac{-1}{\theta}(-\theta u_1^{-\theta-1}) \left[\left(\sum_{i=1}^{d} u_i^{-\theta} \right) - d + 1 \right]^{\frac{-1-\theta}{\theta}} \tag{B.1}$$

$$= u_1^{-\theta-1} \left[\left(\sum_{i=1}^{d} u_i^{-\theta} \right) - d + 1 \right]^{\frac{-1-\theta}{\theta}} \tag{B.2}$$

Dérivée par rapport à u_2 :

$$\frac{\partial C(u_1, ..., u_d)}{\partial u_1 \partial u_2} = u_1^{-\theta-1} \frac{-1-\theta}{\theta}(-\theta) u_2^{-\theta-1} \left[\left(\sum_{i=1}^{d} u_i^{-\theta} \right) - d + 1 \right]^{\frac{-1-2\theta}{\theta}} \tag{B.3}$$

$$= u_1^{-\theta-1} u_2^{-\theta-1} (1+\theta) \left[\left(\sum_{i=1}^{d} u_i^{-\theta} \right) - d + 1 \right]^{\frac{-1-2\theta}{\theta}} \tag{B.4}$$

Dérivée par rapport à u_3 :

$$\frac{\partial C(u_1, ..., u_d)}{\partial u_1 \partial u_2 \partial u_3} = u_1^{-\theta-1} u_2^{-\theta-1} (1+\theta) \frac{(-1-2\theta)}{\theta}(-\theta) u_3^{-\theta-1} \left[\left(\sum_{i=1}^{d} u_i^{-\theta} \right) - d + 1 \right]^{\frac{-1-3\theta}{\theta}} \tag{B.5}$$

$$= u_1^{-\theta-1} u_2^{-\theta-1} u_3^{-\theta-1} (1+\theta)(1+2\theta) \left[\left(\sum_{i=1}^{d} u_i^{-\theta} \right) - d + 1 \right]^{\frac{-1-3\theta}{\theta}} \tag{B.6}$$

Soit par récursion :

$$\frac{\partial C(u_1,...,u_d)}{\partial u_1 ... \partial u_d} = \prod_{i=1}^{d} u_i^{-\theta-1} \prod_{n=0}^{d-1}(1+n\theta)\left[\left(\sum_{i=1}^{d} u_i^{-\theta}\right) - d + 1\right]^{\frac{-1-d\theta}{\theta}} \tag{B.7}$$

Annexe C

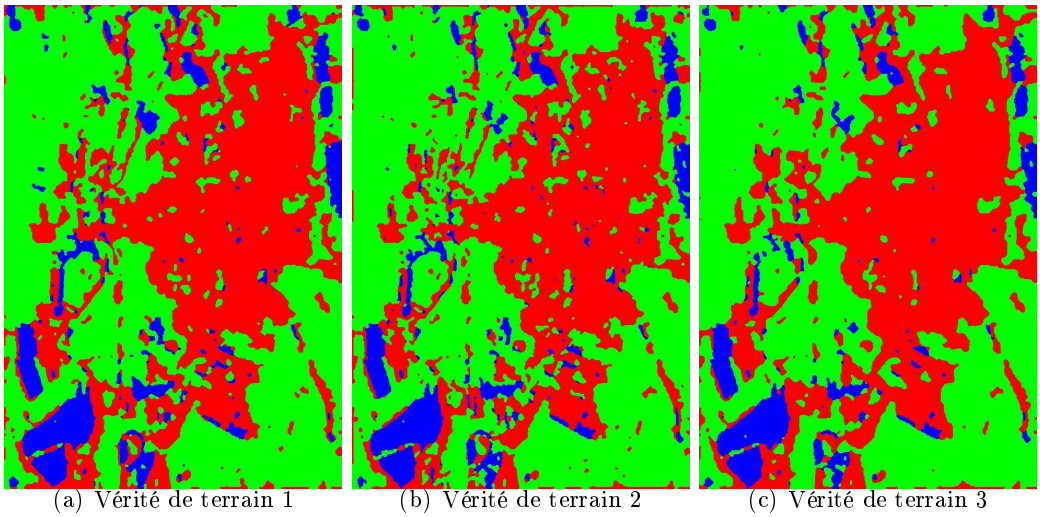
Étude de la robustesse de la classification par rapport à la base d'apprentissage

Dans cette annexe, nous utilisons l'acquisition COSMO-SkyMed de Cavallermaggiore en Italie afin d'étudier brièvement la robustesse de la méthode proposée dans le chapitre 4. Pour ce faire, nous avons sélectionné manuellement plusieurs bases d'apprentissage (Fig. C.1) et observé visuellement (Fig. C.2) et numériquement (Tab. C.1) les résultats de classification. Nous pouvons constater qu'aux niveaux qualitatif et quantitatif, les résultats obtenus sont très proches. La méthode proposée est donc robuste aux divers choix de vérités de terrain. En outre, ces vérités de terrain sont faciles à construire, ce qui permet à une personne non nécessairement experte de pouvoir établir ses propres bases d'apprentissage.

En revanche, nous n'avons porté d'étude sur la taille de la base d'apprentissage à partir de laquelle la performance du classifieur est dégradée, car cette étude est délicate dans le sens où notre expertise n'est pas suffisante pour juger si entre deux cartes de classification assez proches qualitativement et quantitativement, l'une est meilleure que l'autre.

(a) Vérité de terrain 1 (b) Vérité de terrain 2 (c) Vérité de terrain 3

Fig. C.1 – Diverses bases d'apprentissage.



(a) Vérité de terrain 1 (b) Vérité de terrain 2 (c) Vérité de terrain 3

FIG. C.2 – Résultats de classification obtenus avec diverses bases d'apprentissage.

TAB. C.1 – Résultats numériques de classification obtenus pour l'image de Cavallermaggiore pour chacune des vérités de terrain.

	Eau	Urbain	Végétation	**Global**
V.T.1	98,62%	98,42%	100%	**99,01%**
V.T.2	98,60%	94,50%	100%	**97,70%**
V.T.3	98,37%	98,24%	100%	**98,87%**

Annexe D

Formations, séminaires et transfert de logiciels

La thèse a été l'occasion de suivre diverses formations et de présenter le travail réalisé au cours de la thèse. En voici la liste chronologique :

- Juillet 2010 : École d'été Peyresq en traitement du signal et des images sur la thématique "Apprentissage en traitement du signal et des images".
- Octobre 2010 : Participation aux Doctoriales organisées conjointement par la DGA, l'école Polytechnique et ParisTech à Fréjus (France).
- Juillet 2011 : Auditrice des cours de l'école d'été SSIP (summer school in image processing) à Szeged (Hongrie) et présentation orale du travail de thèse.
- Décembre 2011 : Présentation orale du travail de thèse à Galderma R&D à Sophia Antipolis (France).
- Janvier 2012 : Présentation d'un poster lors des journées Pléiades à Toulouse (France) organisées par le CNES à l'occasion du lancement du satellite.
- Juin 2012 : Séminaire invité au laboratoire CRAN (Centre de Recherche en Automatique de Nancy).

La thèse a aussi été l'occasion d'assister à des séminaires du groupe de recherche (GdR) ISIS à Paris, ou des séminaires d'équipe à l'INRIA-SAM.

Le travail de thèse a abouti à la création de deux logiciels déposés à l'agence pour la protection des programmes (APP) :

- SCOMBO : "Supervised Classifier of Multi-Band Optical images", qui est l'implémentation pour les images multibande optiques de la méthode présentée dans la partie 4.2 du chapitre 4. Dépôt n°IDDN.FR.001.090004.000.S.P.2012.000.21000 pour la version 1.0. Dépôt n°IDDN.FR.001.090004.001.S.P.2012.000.21000 pour la version 1.1.
- H-SCOMBO : Hierarchical Supervised Classifier of Multi-Band Optical images, qui est l'implémentation pour les images multibande optiques de la méthode hiérarchique présentée dans la partie 5.3 du chapitre 5. Dépôt n°IDDN.FR.001.090006.000.S.P.2012.000.21000.

Le logiciel SCOMBO a été transféré au laboratoire Cutis, Galderma R&D, en mai 2012, ainsi qu'à l'unité mixte internationale Image and Pervasive Access Lab (IPAL) de Singapour en novembre 2012.

Annexe E

Publications - Conférences et Journaux

Nous listons dans cette annexe, l'intégralité des publications effectuées pendant la thèse, et donc relatives à ce manuscrit. Celles-ci sont disponibles en ligne sur le site internet de l'équipe Ayin, ainsi que sur la page personnelle d'Aurélie Voisin (*https ://team.inria.fr/ayin/aurelie-voisin/*).

Conférences internationales et nationales avec actes :

E.1 SPIE Remote Sensing 2010

A. Voisin, G. Moser, V. Krylov, S. B. Serpico, J. Zerubia. *Classification of very high resolution SAR images of urban areas by dictionary-based mixture models, copulas and Markov random fields using textural features*, SPIE Symposium on Remote Sensing 2010, Proc. of SPIE, volume 7830, 78300O, Toulouse (France), 20-23 septembre, 2010.

E.2 ICIP 2011

K. Kayabol, A. Voisin, J. Zerubia. *SAR image classification with non-stationary multinomial logistic mixture of amplitude and texture densities*, Proc. IEEE International Conference on Image Processing (ICIP), Bruxelles (Belgique), 11-14 septembre, 2011.

E.3 Gretsi 2011

A. Voisin, V. Krylov, J. Zerubia. *Classification bayésienne supervisée d'images RSO de zones urbaines à très haute résolution*, GRETSI "Symposium on Signal and Image Processing", Bordeaux (France), 5-8 septembre, 2011.

E.4 IS&T/SPIE 2012

A. Voisin, V. Krylov, G. Moser, S. B. Serpico, J. Zerubia. *Multichannel hierarchical image classification using multivariate copulas*, IS&T/SPIE Electronic Imaging 2012, Proc. of SPIE, volume 8296, 82960K, San Francisco (USA), 22-26 janvier, 2012.

E.5 GTTI 2012

A. Voisin, V. Krylov, G. Moser, S. B. Serpico, J. Zerubia. *Multiscale classification of very high resolution SAR images of urban areas by Markov random fields, copula functions, and texture extraction*, Réunion annuelle de l'association Gruppo nazionale Telecomunicazioni e Tecnologie dell'Informazione (GTTI), Cagliari (Italie), 25-27 juin, 2012.

E.6 EUSIPCO 2012

A. Voisin, V. Krylov, G. Moser, S. B. Serpico, J. Zerubia. *Classification of multisensor remote sensing images using an adaptive hierarchical Markovian model*, European Signal Processing Conference (EUSIPCO), Bucarest (Roumanie), 27-31 août, 2012.

E.7 ICIP 2012

V. Krylov, G. Moser, A. Voisin, S. B. Serpico, J. Zerubia. *Change detection with synthetic aperture radar images by Wilcoxon statistic likelihood ratio test*, Proc. IEEE International Conference on Image Processing (ICIP), Orlando (USA), 30 septembre-3 octobre, 2012.

Journaux internationaux :

E.8 GRSL 2012

A. Voisin, V. Krylov, G. Moser, S. B. Serpico, J. Zerubia. *Classification of very high resolution SAR images of urban areas using copulas and texture in a hierarchical Markov random field model*, IEEE Geoscience and Remote Sensing Letters, vol.10, no 1, pp 96-100, 2013.

E.9 TGRS 2012

A. Voisin, V. Krylov, G. Moser, S. B. Serpico, J. Zerubia. *Supervised classification of multi-sensor and multi-resolution remote sensing images with a hierarchical*

copula-based approach, IEEE Transactions on Geoscience and Remote Sensing, (soumis), 2012.

Bibliographie

[Achim 2002] A. Achim, A. Bezerianos et P. Tsakalides. *SAR image denoising : a multiscale robust statistical approach.* In 14th International Conference on Digital Signal Processing (DSP'02), volume 2, pages 1235–1238, 2002. (Cité en page 40.)

[Achim 2006] A. Achim, E. Kuruoglu et P. Tsakalides. *SAR image filtering based on the heavy-tailed Rayleigh model.* IEEE Trans. Image Process., vol. 15, no. 9, pages 2686–2693, 2006. (Cité en page 13.)

[Al-Ani 2011] A. Al-Ani et M. Deriche. *A new technique for combining multiple classifiers using the Dempster-Shafer theory of evidence.* Journal Of Artificial Intelligence Research, vol. 17, pages 333–361, 2011. (Cité en pages 85 et 86.)

[Amarsaikhan 2004] D. Amarsaikhan et T. Douglas. *Data fusion and multisource image classification.* International Journal of Remote Sensing, vol. 25, no. 17, pages 3529–3539, 2004. (Cité en page 87.)

[Arce 2005] G.R. Arce. Nonlinear signal processing : a statistical approach. Wiley, New Jersey, 2005. (Cité en page 119.)

[Arias-Castro 2009] E. Arias-Castro et D. L. Donoho. *Does median filtering truly preserve edges better than linear filtering ?* Annals of Statistics, vol. 37, no. 3, pages 1172–1206, 2009. (Cité en page 98.)

[Banjanin 1991] Z. B. Banjanin et D. S. Zrnic. *Clutter rejection for Doppler weather radars which use staggered pulses.* IEEE Trans. Geosci. Remote Sens., vol. 29, no. 4, pages 610–620, 1991. (Cité en page 15.)

[Barnard 1958] G. A. Barnard et T. Bayes. *Studies in the history of probability and statistics : IX. Thomas Bayes's essay towards solving a problem in the doctrine of chances.* Biometrika, vol. 45, no. 3/4, pages 293–315, 1958. (Cité en pages 24 et 60.)

[Bates 2000] P. D. Bates et A. P. J. De Roo. *A simple raster-based model for flood inundation simulation.* Journal of Hydrology, vol. 236, no. 1-2, pages 54–77, 2000. (Cité en page 56.)

[Baum 1970] L. E. Baum, T. Petrie, G. Soules et N. Weiss. *A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains.* IEEE Ann. Math. Stats, vol. 41, no. 1, pages 164–171, 1970. (Cité en page 93.)

[Benboudjema 2005] D. Benboudjema et W. Pieczynski. *Unsupervised image segmentation using triplet Markov fields.* Computer Vision and Image Understanding, vol. 99, no. 3, pages 476–498, 2005. (Cité en page 58.)

[Benediktsson 1992] J. A. Benediktsson et P. H. Swain. *Consensus theoretic classification methods.* IEEE Trans. Syst., Man, Cybern., vol. 22, no. 4, pages 688–704, 1992. (Cité en page 86.)

[Benediktsson 1997] J. A. Benediktsson, J. R. Sveinsson et P. H. Swain. *Hybrid consensus theoretic classification.* IEEE Trans. Geosci. Remote Sens., vol. 35, no. 4, pages 833–843, 1997. (Cité en page 86.)

[Benediktsson 1999] J. A. Benediktsson et I. Kanellopoulos. *Classification of multisource and hyperspectral data based on decision fusion.* IEEE Trans. Geosci. Remote Sens., vol. 37, no. 3, pages 1367–1377, 1999. (Cité en page 86.)

[Bertsekas 1996] D. P. Bertsekas. Constrained optimization and Lagrange multiplier methods. Athena Scientific, 1996. (Cité en page 70.)

[Besag 1974] J. Besag. *Spatial interaction and the statistical analysis of lattice systems.* Journal of the Royal Statistical Society, vol. 36, no. 2, pages 192–236, 1974. (Cité en page 59.)

[Besag 1986] J. Besag. *On the statistical analysis of dirty pictures.* Journal of the Royal Statistical Society, vol. 48, no. 3, pages 259–302, 1986. (Cité en pages 64 et 65.)

[Biernacki 2001] C. Biernacki, G. Celeux et G. Govaert. *Strategies for getting the highest likelihood in mixture models.* Research report 4255, INRIA, France, sep 2001. (Cité en page 32.)

[Biernacki 2003] C. Biernacki, G. Celeux et G. Govaert. *Choosing starting values for the EM algorithm for getting the highest likelihood in multivariate Gaussian mixture models.* Computational statistics and data analysis, vol. 41, no. 3-4, pages 561–575, 2003. (Cité en page 32.)

[Bijaoui 1996] A. Bijaoui, Y. Fang, Y. Bobichon et F. Rue. *The analysis of SAR images by multiscale methods.* In IEEE International Conference on Image Processing (ICIP'96), volume 3, pages 895–898, 1996. (Cité en page 118.)

[Bishop 1996] C. M. Bishop. Neural networks for pattern recognition. Oxford University Press, 1996. (Cité en page 25.)

[Bishop 2006] C. M. Bishop. Pattern recognition and machine learning. Springer-Verlag, New-York, 2006. (Cité en pages 68 et 70.)

[Bombrun 2011] L. Bombrun, G. Vasile, M. Gay et F. Totir. *Hierarchical segmentation of polarimetric SAR images using heterogeneous clutter models.* IEEE Trans. Geosci. Remote Sens., vol. 49, no. 2, pages 726–737, 2011. (Cité en page 46.)

[Boni 2007] G. Boni, F. Castelli, L. Ferraris, N. Pierdicca, S. Serpico et F. Siccardi. *High resolution COSMO-SkyMed SAR data analysis for civil protection from flooding events.* In IEEE International Geoscience and Remote Sensing Symposium (IGARSS'07), pages 6–9, Juillet 2007. (Cité en pages 16 et 54.)

[Bouman 1994] C. Bouman et M. Shapiro. *A multiscale random field model for Bayesian image segmentation.* IEEE Trans. Image Process., vol. 3, no. 2, pages 162–177, 1994. (Cité en pages 91 et 92.)

[Boykov 2001] Y. Boykov, O. Veksler et R. Zabih. *Efficient approximate energy minimization via graph cuts.* IEEE Trans. Pattern Anal. Mach. Intell., vol. 23, no. 11, pages 1222–1239, 2001. (Cité en page 66.)

[Boykov 2004] Y. Boykov et V. Kolmogorov. *An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision.* IEEE Trans. Pattern Anal. Mach. Intell., vol. 26, no. 9, pages 1124–1137, 2004. (Cité en page 66.)

[Breiman 1996] L. Breiman. *Bagging predictors.* Machine learning, vol. 24, no. 2, pages 123–140, 1996. (Cité en page 86.)

[Briem 2002] G. J. Briem, J. A. Benediktsson et J. R. Sveinsson. *Multiple classifiers applied to multisource remote sensing data.* IEEE Trans. Geosci. Remote Sens., vol. 40, no. 10, pages 2291–2299, 2002. (Cité en pages 85 et 86.)

[Brunel 2010] N. J.-B. Brunel, J. Lapuyade-Lahorgue et W. Pieczynski. *Modeling and unsupervised classification of multivariate hidden Markov chains with copulas.* IEEE Transactions on Automatic Control, vol. 55, no. 2, pages 338–349, 2010. (Cité en page 58.)

[Bruniquel 1997] J. Bruniquel et A. Lopes. *Multi-variate optimal speckle reduction in SAR imagery.* International Journal of Remote Sensing, vol. 18, no. 3, pages 603–627, 1997. (Cité en page 13.)

[Buades 2005] A. Buades, B. Coll et J. M. Morel. *A review of image denoising algorithms, with a new one.* Journal on Multiscale Modeling and Simulation, vol. 4, no. 2, pages 490–530, 2005. (Cité en page 40.)

[Campbell 2002] J. B. Campbell. Introduction to remote sensing. New York London : the Guilford Press, 2002. (Cité en page 7.)

[Carter 2000] M. Carter et B. Van Brunt. The Lebesgue-Stieltjes integral : a practical introduction. Springer, 2000. (Cité en page 50.)

[Celeux 1992] G. Celeux et G. Govaert. *A classification EM algorithm for clustering and two stochastic versions.* Comput. Statist. Data Anal., vol. 14, no. 3, pages 315–332, 1992. (Cité en page 68.)

[Celeux 1996] G. Celeux, D. Chauveau et J. Diebolt. *Stochastic versions of the EM algorithm : an experimental study in the mixture case.* Journal of Statistical Computation and Simulation, vol. 55, no. 4, pages 287–314, 1996. (Cité en pages 25 et 31.)

[Celeux 2003] G. Celeux, F. Forbes et N. Peyrard. *EM procedures using mean field-like approximations for Markov model-based image segmentation.* Pattern Recognition, vol. 36, no. 1, pages 131–144, 2003. (Cité en page 61.)

[Chabert 2011] M. Chabert et J.-Y. Tourneret. *Étude d'un système de Pearson bivarié pour des images hétérogènes d'observation de la Terre.* In Colloque GRETSI'11, numéro 258, Septembre 2011. (Cité en page 46.)

[Chamundeeswari 2007] V. V. Chamundeeswari, D. Singh et K. Singh. *Unsupervised land cover classification of SAR images by contour tracing.* In IEEE International Geoscience and Remote Sensing Symposium (IGARSS'07), pages 547–550, Juillet 2007. (Cité en page 57.)

[Chavez 1991] P. S. Chavez, S. C. Sides et J. A. Anderson. *Comparison of three different methods to merge multiresolution and multispectral data : Landsat TM and SPOT panchromatic.* Photogramm. Eng. Remote Sens., vol. 57, no. 3, pages 295–303, 1991. (Cité en page 87.)

[Chen 2004] Q. Chen et P. Gong. *Automatic variogram parameter extraction for textural classification of the panchromatic IKONOS imagery.* IEEE Trans. Geosci. Remote Sens., vol. 42, no. 5, pages 1106–1115, 2004. (Cité en pages 40 et 41.)

[Cheney 2009] M. Cheney et B. Borden. Fundamentals of radar imaging (CBMS-NSF regional conference series in applied mathematics). Society for Industrial Mathematics, 2009. (Cité en page 7.)

[Cho 1995] S.-B. Cho et J. H. Kim. *Combining multiple neural networks by fuzzy integral for robust classification.* IEEE Trans. Syst., Man, Cybern., vol. 25, no. 2, pages 380–384, 1995. (Cité en page 86.)

[Choi 2001] H. Choi et R. G. Baraniuk. *Multiscale image segmentation using wavelet-domain hidden Markov models.* IEEE Trans. Image Process., vol. 10, no. 9, pages 1309–1321, 2001. (Cité en page 56.)

[Cianci 2012] L. Cianci, G. Moser et S. B. Serpico. *Change detection from very high-resolution multisensor remote-sensing images by a Markovian approach.* In Proc. of IEEE GOLD (in press), 2012. (Cité en page 25.)

[CNE 2012] *Supplément CNES MAG* 52, Jan. 2012. www.cnes-multimedia.fr/cnesmag/interactif. (Cité en page 1.)

[Corbane 2009] C. Corbane, N. Baghdadi, X. Descombes, G. Wilson, N. Villeneuve et M. Petit. *Comparative study on the performance of multiparameter SAR data for operational urban areas extraction using textural features.* IEEE Geosci. Remote Sens. Lett., vol. 6, no. 4, pages 728–732, 2009. (Cité en page 40.)

[CSK 2007] *Guide de COSMO-SkyMed*, 2007. www.e-geos.it/products/pdf/csk-user_guide.pdf. (Cité en page 1.)

[Curran 1988] P. J. Curran. *The semivariogram in remote sensing : an introduction.* Remote Sensing of Environment, vol. 24, no. 3, pages 493–507, 1988. (Cité en page 41.)

[Datcu 2002] M. Datcu, F. Melgani, A. Piardi et S. B. Serpico. *Multisource data classification with dependence trees.* IEEE Trans. Geosci. Remote Sens., vol. 40, no. 3, pages 609–617, 2002. (Cité en page 46.)

[Daubechies 1988] I. Daubechies. *Orthonormal bases of compactly supported wavelets.* Communications on Pure and Applied Mathematics, vol. 41, no. 7, pages 909–996, 1988. (Cité en pages 56, 101 et 105.)

[Dekker 2003] R.J. Dekker. *Texture analysis and classification of ERS SAR images for map updating of urban areas in the Netherlands.* IEEE Trans. Geosci. Remote Sens., vol. 41, no. 9, pages 1950–1958, 2003. (Cité en page 37.)

[D'Elia 2003] C. D'Elia, G. Poggi et G. Scarpa. *A tree-structured Markov random field model for Bayesian image segmentation.* IEEE Trans. Image Process., vol. 12, no. 10, pages 1259–1273, 2003. (Cité en page 55.)

[Delignon 1997] Y. Delignon, A. Marzouki et W. Pieczynski. *Estimation of generalized mixtures and its application in image segmentation.* IEEE Trans. Image Process., vol. 6, no. 10, pages 1364–1375, 1997. (Cité en pages 33 et 57.)

[Dell'Acqua 2003] F. Dell'Acqua et P. Gamba. *Texture-based characterization of urban environments on satellite SAR images.* IEEE Trans. Geosci. Remote Sens., vol. 41, no. 5, pages 153–159, 2003. (Cité en page 13.)

[Dell'Acqua 2004] F. Dell'Acqua, P. Gamba, A. Ferrari, J. A. Palmason, J. A.Benediktsson et K. Arnason. *Exploiting spectral and spatial information in hyperspectral urban data with high resolution.* IEEE Geosci. Remote Sens. Lett., vol. 1, no. 4, pages 322–326, 2004. (Cité en page 84.)

[Dellepiane 1991] S. Dellepiane, D. D. Giusto, S. B. Serpico et G. Vernazza. *SAR image recognition by integration of intensity and textural information.* Int. J. Remote Sens., vol. 12, no. 9, pages 1915–1932, 1991. (Cité en page 40.)

[Demir 2011] B. Demir, C. Persello et L. Bruzzone. *Batch-mode active-learning methods for the interactive classification of remote sensing images.* IEEE Trans. Geosci. Remote Sens., vol. 49, no. 3, pages 1014–1031, 2011. (Cité en page 58.)

[Dempster 1977] A. P. Dempster, N. M. Laird et D. B. Rubin. *Maximum likelihood from incomplete data via the EM algorithm.* Journal of the Royal Statistical Society. Series B (Methodological), vol. 39, no. 1, pages 1–38, 1977. (Cité en page 31.)

[Derin 1990] H. Derin, P. A. Kelly, G. Vezina et S. G. Labitt. *Modeling and segmentation of speckled images using complex data.* IEEE Trans. Geosci. Remote Sens., vol. 28, no. 1, pages 76–87, 1990. (Cité en page 55.)

[Derrode 2007] S. Derrode et G. Mercier. *Unsupervised multiscale oil slick segmentation from SAR images using a vector HMC model.* Pattern Recognition, vol. 40, no. 3, pages 1135–1147, 2007. (Cité en page 57.)

[Descombes 1999] X. Descombes, R. D. Morris, J. Zerubia et M. Berthod. *Estimation of Markov random field prior parameters using Markov chain Monte Carlo maximum likelihood.* IEEE Trans. Image Process., vol. 8, no. 7, pages 954–963, 1999. (Cité en pages 61 et 63.)

[Devroye 1996] L. Devroye, L. Gyorfi et G. Lugosi. A probabilistic theory of pattern recognition. Springer, 1996. (Cité en page 73.)

[Dobson 1996] M. C. Dobson, L. E. Pierce et F. T. Ulaby. *Knowledge-based land-cover classification using ERS–1/JERS–1 SAR composites.* IEEE Trans. Geosci. Remote Sens., vol. 34, no. 1, pages 83–99, 1996. (Cité en pages 40 et 117.)

[Dong 1999] Y. Dong, B. C. Forester et A. K. Milne. *Segmentation of radar imagery using the Gaussian Markov random field model.* International Journal of Remote Sensing, vol. 20, no. 8, pages 1617–1639, 1999. (Cité en pages 40 et 55.)

[Dowman 2012] I. Dowman, K. Jacobsen, G. Konecny et R. Sandau. High resolution optical satellite imagery. Whittles Publishing, 2012. (Cité en page 6.)

[Du 2002] L. Du, M. R. Grunes et J. S. Lee. *Unsupervised segmentation of dual-polarization SAR images based on amplitude and texture characteristics.* International Journal of Remote Sensing, vol. 23, no. 20, pages 4383–4402, 2002. (Cité en page 40.)

[Dubes 1989] R. C. Dubes et A. K. Jain. *Random field models in image analysis.* Journal of Applied Statistics, vol. 16, no. 2, pages 131–164, 1989. (Cité en page 59.)

[Dubuisson 1990] B. Dubuisson. Diagnostic et reconnaissance des formes. Hermès, 1990. (Cité en page 67.)

[Duda 2000] R. O. Duda, P. E. Hart et D. G. Stork. Pattern classification. Wiley-Interscience, New-York, 2nd édition, 2000. (Cité en page 25.)

[Dupont 1994] S. Dupont et M. Berthod. *Interférométrie radar et déroulement de phase.* Research report 2344, INRIA, France, 1994. (Cité en page 7.)

[Efron 1994] B. Efron et R. J. Tibshirani. An introduction to the bootstrap. Chapman and Hall, 1st édition, 1994. (Cité en pages 57 et 72.)

[Eymard 1994] L Eymard, A. LeCornec et L. Tabary. *The ERS-1 microwave radiometer.* International Journal of Remote Sensing, vol. 15, no. 4, pages 845–857, 1994. (Cité en page 12.)

[Fabre 1994] E. Fabre. *Traitement du signal multi-résolution : conception de lisseurs rapides pour une famille de modèles.* Thèse de doctorat, Université de Rennes 1, France, 1994. (Cité en page 91.)

[Fauvel 2006] M. Fauvel, J. Chanussot et J. A. Benediktsson. *Decision fusion for the classification of urban remote sensing images.* IEEE Trans. Geosci. Remote Sens., vol. 44, no. 10, pages 2828–2838, 2006. (Cité en page 86.)

[Figueiredo 2002] M. A. T. Figueiredo et A. K. Jain. *Unsupervised learning of finite mixture models.* IEEE Trans. Pattern Anal. Mach. Intell., vol. 24, no. 3, pages 381–396, 2002. (Cité en pages 25, 30 et 35.)

[Fisher 2001] N. I. Fisher et P. Switzer. *Graphical assessment of dependence : is a picture worth* 100 *tests ?* The American Statistician, vol. 55, no. 3, pages 233–239, 2001. (Cité en pages 115 et 125.)

[Fjortoft 2003] R. Fjortoft, Y. Delignon, W. Pieczynski, M. Sigelle et F. Tupin. *Unsupervised classification of radar images using hidden Markov chains and hidden Markov random fields.* IEEE Trans. Geosci. Remote Sens., vol. 41, no. 3, pages 675–686, 2003. (Cité en pages 58 et 63.)

[Fogler 1994] R. J. Fogler, L. D. Hostetler et D. R. Hush. *SAR clutter suppression using probability density skewness.* IEEE Trans. on Aerospace and Electronic Systems, vol. 30, no. 2, pages 622–626, 1994. (Cité en page 15.)

[Fosgate 1997] C. H. Fosgate, H. Krim, W. W. Irving, W. C. Karl et A. S. Willsky. *Multiscale segmentation and anomaly enhancement of SAR imagery.* IEEE Trans. Image Process., vol. 6, no. 1, pages 7–20, 1997. (Cité en page 56.)

[Foucher 2001] S. Foucher, G. B. Benie et J.-M. Boucher. *Multiscale MAP filtering of SAR images.* IEEE Trans. Image Process., vol. 10, no. 1, pages 49–60, 2001. (Cité en page 13.)

[Foucher 2002] S. Foucher, M. Germain, J.-M. Boucher et G. B. Benie. *Multisource classification using ICM and Dempster-Shafer theory.* IEEE Trans. on Instrumentation and Measurement, vol. 51, no. 2, pages 277–281, 2002. (Cité en page 56.)

[Freeman 1992] A. Freeman. *SAR calibration : an overview.* IEEE Trans. Geosci. Remote Sens., vol. 30, no. 6, pages 1107–1121, 1992. (Cité en page 14.)

[Gagnon 1997] L. Gagnon et A. Jouan. *Speckle filtering of SAR images - A comparative study between complex-wavelet-based and standard filters.* In Proc. of SPIE, volume 3169, pages 80–91, 1997. (Cité en pages 13 et 118.)

[Galland 2009] F. Galland, J.-M. Nicolas, H. Sportouche, M. Roche, F. Tupin et P. Refregier. *Unsupervised synthetic aperture radar image segmentation using Fisher distributions.* IEEE Trans. Geosci. Remote Sens., vol. 47, no. 8, pages 2966–2972, 2009. (Cité en page 57.)

[Galloway 1975] M. M. Galloway. *Texture analysis using gray level run lengths.* Computer Graphics and Image Processing, vol. 4, no. 2, pages 172–179, 1975. (Cité en page 40.)

[Geman 1984] S. Geman et D. Geman. *Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images.* IEEE Trans. Pattern Anal. Mach. Intell., vol. 6, no. 6, pages 721–741, 1984. (Cité en pages 62, 64 et 65.)

[Genest 2003] C. Genest et J.-C. Boies. *Detecting dependence with Kendall plots.* The American Statistician, vol. 57, no. 4, pages 275–284, 2003. (Cité en pages 115 et 125.)

[Gidas 1989] B. Gidas. *A renormalization group approach to image processing problems.* IEEE Trans. Pattern Anal. Mach. Intell., vol. 11, no. 2, pages 164–180, 1989. (Cité en page 88.)

[Gomarasca 2009] M. A. Gomarasca. Basics of geomatics. Springer, 2009. (Cité en page 14.)

[Gonzalez 2008] R. C. Gonzalez et R. E. Woods. Digital image processing. Pearson Prentice Hall, 2008. (Cité en page 112.)

[Goshtasby 2005] A. A. Goshtasby. 2-D and 3-D image registration : for medical, remote sensing, and industrial applications. Wiley-Interscience, 2005. (Cité en page 98.)

[Graffigne 1995] C. Graffigne, F. Heitz, P. Perez, F. Preteux, M. Sigelle et J. Zerubia. *Hierarchical Markov random field models applied to image analysis : a review.* In Proc. SPIE Conf. on Neural, Morphological and Stochastic Methods in Image Proc., volume 2568, pages 2–17, 1995. (Cité en page 88.)

[Greco 2007] M. S. Greco et F. Gini. *Statistical analysis of high-resolution SAR ground clutter data.* IEEE Trans. Geosci. Remote Sens., vol. 45, no. 3, pages 566–575, 2007. (Cité en page 39.)

[Haralick 1973] R. M. Haralick, K. Shanmugam et I. Dinstein. *Textural features for image classification.* IEEE Trans. Syst., Man, Cybern., vol. 3, no. 6, pages 610–621, 1973. (Cité en pages 41 et 42.)

[Hardie 2004] R. C. Hardie, M. T. Eismann et G. L. Wilson. *MAP estimation for hyperspectral image resolution enhancement using an auxiliary sensor.* IEEE Trans. Image Process., vol. 13, no. 9, pages 1174–1184, 2004. (Cité en page 46.)

[Hastings 1970] W. K. Hastings. *Monte Carlo sampling methods using Markov chains and their applications.* Biometrika, vol. 57, no. 1, pages 97–109, 1970. (Cité en pages 62 et 66.)

[Heitz 1994] F. Heitz, P. Perez et P. Bouthemy. *Multiscale minimization of global energy functions in some visual recovery problems.* CVGIP : image understanding, vol. 59, no. 1, pages 125–134, 1994. (Cité en page 88.)

[Herbrich 2002] R. Herbrich. Learning kernel classifiers. MIT Press, 2002. (Cité en page 70.)

[Hu 1999] Y. H. Hu, H. B. Lee et F. L. Scarpace. *Optimal linear spectral unmixing.* IEEE Trans. Geosci. Remote Sens., vol. 37, no. 1, pages 639–644, 1999. (Cité en page 88.)

[Huard 2006] D. Huard, G. Evin et A.-C. Fabre. *Bayesian copula selection.* Computational statistics and data analysis, vol. 51, no. 2, pages 809–822, 2006. (Cité en page 48.)

[Jackson 2001] Q. Jackson et D. A. Landgrebe. *An adaptive classifier design for high-dimensional data analysis with a limited training data set.* IEEE Trans. Geosci. Remote Sens., vol. 39, no. 12, pages 2664–2679, 2001. (Cité en page 63.)

[Jacob 2002] A. M. Jacob, E. M. Hemmerly et D. Fernandes. *SAR image classification using a neural classifier based on Fisher criterion.* In Proc. of the VII Brazilian Symposium on Neural Networks Microelectromechanical Systems (SBRN'02), pages 168–172, 2002. (Cité en page 55.)

[Jain 1988] A. K. Jain. Fundamentals of digital image processing. Prentice Hall, 1988. (Cité en page 112.)

[Jain 1991] A. K. Jain et F. Farrokhnia. *Unsupervised texture segmentation using Gabor filters.* Pattern Recognition, vol. 24, no. 12, pages 1167–1186, 1991. (Cité en page 40.)

[Jain 2000] A. K. Jain, R. P. W. Duin et J. Mao. *Statistical pattern recognition : a review*. IEEE Trans. Pattern Anal. Mach. Intell., vol. 22, no. 1, pages 4–37, 2000. (Cité en page 73.)

[Jakeman 1976] E. Jakeman et P. Pusey. *A model for non-Rayleigh sea echo*. IEEE Trans. on Antennas and Propagation, vol. 24, no. 6, pages 806–814, 1976. (Cité en page 34.)

[Joe 1990] H. Joe. *Multivariate concordance*. Journal of multivariate analysis, vol. 35, no. 1, pages 12–30, 1990. (Cité en pages 49 et 50.)

[Katartzis 2005] A. Katartzis, I. Vanhamel et H. Sahli. *A hierarchical Markovian model for multiscale region-based classification of vector-valued images*. IEEE Trans. Geosci. Remote Sens., vol. 43, no. 3, pages 548–558, 2005. (Cité en page 120.)

[Kato 1992] Z. Kato, J. Zerubia et M. Berthod. *Satellite image classification using a modified Metropolis dynamics*. In IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP'92), volume 3, pages 573–576, 1992. (Cité en page 65.)

[Kayabol 2010] K. Kayabol, E. E. Kuruoglu, J. L. Sanz, B. Sankur, E. Salerno, et D. Herranz. *Adaptive Langevin sampler for separation of t-distribution modeled astrophysical maps*. IEEE Trans. Image Process., vol. 19, no. 9, pages 2357–2368, 2010. (Cité en page 67.)

[Ketelaar 2010] V. B. H. (Gini) Ketelaar. Satellite radar interferometry : subsidence monitoring techniques (remote sensing and digital image processing). Springer, 1st édition, 2010. (Cité en page 12.)

[Kirkpatrick 1983] S. Kirkpatrick, C. D. Gelatt et M. P. Vecchi. *Optimization by simulated annealing*. Science, vol. 220, no. 4598, pages 671–680, 1983. (Cité en page 64.)

[Kittler 1998] J. Kittler, M. Hatef, R. P. W. Duin et J. Matas. *On combining classifiers*. IEEE Trans. Pattern Anal. Mach. Intell., vol. 20, no. 3, pages 226–239, 1998. (Cité en page 86.)

[Klauder 1960] J. R. Klauder, Price A. C, S. Darlington et W. J. Albersheim. *The theory and design of chirp radars*. Bell System Technical Journal, vol. 39, no. 4, pages 745–808, 1960. (Cité en page 9.)

[Kohavi 1995] R. Kohavi. *A study of cross-validation and bootstrap for accuracy estimation and model selection*. In International Joint Conference on Artificial Intelligence (IJCAI'95), pages 1137–1143, 1995. (Cité en pages 67 et 72.)

[Kojadinovic 2010] I. Kojadinovic et J. Yan. *Comparison of three semiparametric methods for estimating dependence parameters in copula models*. Insurance : Mathematics and Economics, vol. 47, no. 1, pages 52–63, 2010. (Cité en page 49.)

[Kolmogorov 2004] V. Kolmogorov et R. Zabih. *What energy functions can be minimized via graph cuts ?* IEEE Trans. Pattern Anal. Mach. Intell., vol. 26, no. 2, pages 147–159, 2004. (Cité en page 66.)

[Krishnapuram 2005] B. Krishnapuram, L. Carin, M. A. T. Figueiredo et A. J. Hartemink. *Sparse multinomial logistic regression : fast algorithms and generalization bounds.* IEEE Trans. Pattern Anal. Mach. Intell., vol. 27, no. 6, pages 957–968, 2005. (Cité en page 68.)

[Krylov 2010] V. Krylov et J. Zerubia. *Generalized gamma mixtures for supervised SAR image classification.* In Proc. of International Conference on Computer Graphics and Vision (Graphicon'2010), pages 107–110, Saint Petersburg, Russia, Septembre 2010. (Cité en page 34.)

[Krylov 2011a] V. Krylov, G. Moser, S. B. Serpico et J. Zerubia. *Enhanced dictionary-based SAR amplitude distribution estimation and its validation with very high-resolution data.* IEEE Geosci. Remote Sens. Lett., vol. 8, no. 1, pages 148–152, 2011. (Cité en page 34.)

[Krylov 2011b] V. Krylov, G. Moser, S. B. Serpico et J. Zerubia. *On the method of logarithmic cumulants for parametric probability density function estimation.* Research report 7666, INRIA, France, jul 2011. (Cité en page 34.)

[Krylov 2011c] V. Krylov, G. Moser, S. B. Serpico et J. Zerubia. *Supervised high resolution dual polarization SAR image classification by finite mixtures and copulas.* IEEE J. Sel. Top. Signal Proc., vol. 5, no. 3, pages 554–566, 2011. (Cité en pages 48 et 50.)

[Kumar 2010] P. Kumar. *Probability distributions and estimation of Ali-Mikhail-Haq copula.* Applied Mathematical Sciences, vol. 4, no. 14, pages 657–666, 2010. (Cité en page 48.)

[Kuruoglu 2003] E. Kuruoglu et J. Zerubia. *Skewed α-stable distributions for modeling textures.* Pattern Recognition Letters, vol. 24, no. 1-3, pages 339–348, 2003. (Cité en page 27.)

[Kuruoglu 2004] E. E. Kuruoglu et J. Zerubia. *Modeling SAR images with a generalization of the Rayleigh distribution.* IEEE Trans. Image Process., vol. 13, no. 4, pages 527–533, 2004. (Cité en page 27.)

[Laferté 1995] J.-M. Laferté, F. Heitz, P. Perez et E. Fabre. *Hierarchical statistical models for the fusion of multiresolution image data.* In Proceedings of the 5th International Conference on Computer Vision (ICCV'95), pages 908–913, Cambridge, USA, Juin 1995. (Cité en pages 90 et 91.)

[Laferté 1996] J.-M. Laferté. *Contribution à l'analyse d'images par modèles markoviens sur des graphes hiérarchiques. Application à la fusion de données multirésolutions.* Thèse de doctorat, Université de Rennes 1, 1996. (Cité en page 91.)

[Laferté 2000] J.-M. Laferté, P. Perez et F. Heitz. *Discrete Markov modeling and inference on the quad-tree.* IEEE Trans. Image Process., vol. 9, no. 3, pages 390–404, 2000. (Cité en pages iv, 83, 91, 93, 94, 105 et 106.)

[Lakshmanan 1989] S. Lakshmanan et H. Derin. *Simultaneous parameter estimation and segmentation of Gibbs random fields using simulated annealing.* IEEE

Trans. Pattern Anal. Mach. Intell., vol. 11, no. 8, pages 799–813, 1989. (Cité en page 63.)

[Lee 1980] J.-S. Lee. *Digital image enhancement and noise filtering by use of local statistics.* IEEE Trans. Pattern Anal. Mach. Intell., vol. 2, no. 2, pages 165–168, 1980. (Cité en page 40.)

[Lee 1989] J.-S. Lee et I. Jurkevich. *Segmentation of SAR images.* IEEE Trans. Geosci. Remote Sens., vol. 27, no. 6, pages 674–680, 1989. (Cité en page 26.)

[Lee 1994] J.-S. Lee, L. Jurkevich, P. Dewaele, P. Wambacq et A. Oosterlinck. *Speckle filtering of synthetic aperture radar images : A review.* Remote Sensing Reviews, vol. 8, no. 4, pages 313–340, 1994. (Cité en page 13.)

[Lee 2009] J.-S. Lee et E. Pottier. Polarimetric radar imaging : from basics to applications. CRC Press, 1st édition, 2009. (Cité en page 11.)

[Lehmann 2007] E. Lehmann et J. Romano. Testing statistical hypotheses. Springer, 3rd revised édition, 2007. (Cité en page 50.)

[LeMoigne 2011] J. LeMoigne, N. S. Netanyahu et R. D. Eastman. Image registration for remote sensing. Cambridge University Press, 2011. (Cité en page 98.)

[Li 1995] S. Z. Li. Markov random field modeling in computer vision. Springer, 1995. (Cité en page 60.)

[Li 2000] J. Li, R. M. Gray et R. A. Olshen. *Multiresolution image classification by hierarchical modeling with two-dimensional hidden Markov models.* IEEE Transactions on Information Theory, vol. 46, no. 5, pages 1826–1841, 2000. (Cité en page 85.)

[Li 2011] H.-C. Li, W. Hong, Y.-R. Wu et P.-Z. Fan. *On the empirical-statistical modeling of SAR images with generalized gamma distribution.* IEEE J. Sel. Top. Signal Process., vol. 5, no. 3, pages 386–397, 2011. (Cité en pages 28 et 29.)

[Liu 2009] G. Liu, Q. Qin, T. Mei, W. Xie et L. Wang. *Supervised image segmentation based on tree-structured MRF model in wavelet domain.* IEEE Geosci. Remote Sens. Lett., vol. 6, no. 4, pages 850–854, 2009. (Cité en page 56.)

[Lombardo 2003] P. Lombardo, C. J. Oliver, T. Macri Pellizzeri et M. Meloni. *A new maximum-likelihood joint segmentation technique for multitemporal SAR and multiband optical images.* IEEE Trans. Geosci. Remote Sens., vol. 41, no. 11, pages 2500–2518, 2003. (Cité en page 46.)

[Lu 2007] D. Lu et Q. Weng. *A survey of image classification methods and techniques for improving classification performance.* International Journal of Remote Sensing, vol. 28, no. 5, pages 823–870, 2007. (Cité en page 54.)

[Luettgen 1994] M. R. Luettgen, W. Karl et A. S. Willsky. *Efficient multiscale regularization with applications to the computation of optical flow.* IEEE Trans. Image Process., vol. 3, no. 1, pages 41–64, 1994. (Cité en page 91.)

[Madhok 2002] V. Madhok et D. A. Landgrebe. *A process model for remote sensing data analysis.* IEEE Trans. Geosci. Remote Sens., vol. 40, no. 3, pages 680–686, 2002. (Cité en page 80.)

[Madsen 2001] S. N. Madsen, W. Edelstein, L. D. Di Domenico et J. La Brecque. *A geosynchronous synthetic aperture radar ; for tectonic mapping, disaster management and measurements of vegetation and soil moisture.* In IEEE International Geoscience and Remote Sensing Symposium (IGARSS'01), volume 1, pages 447–449, 2001. (Cité en page 12.)

[Mallat 2008] S. G. Mallat. A wavelet tour of signal processing. Academic Press, 3rd édition, 2008. (Cité en pages 56, 98 et 105.)

[Mantero 2003] P. Mantero, G. Moser et S. B. Serpico. *Partially supervised classification of remote sensing images using SVM-based probability density estimation.* In IEEE Workshop on Advances in Techniques for Analysis of Remotely Sensed Data, pages 327–336, Octobre 2003. (Cité en page 25.)

[Marroquin 1987] J. Marroquin, S. Mitter et T. Poggio. *Probabilistic solution of ill-posed problems in computational vision.* Journal of the American Statistical Association, vol. 82, no. 397, pages 76–89, 1987. (Cité en page 91.)

[Martins 2006] C. I. O. Martins, F. N. S. Medeiros, E. A. Carvalho et F. N. Bezerra. *Combining watershed and statistical analysis for SAR image segmentation.* In IEEE Conference on Image Radar, pages 823–828, 2006. (Cité en page 55.)

[Massonnet 2008] D. Massonnet et J.-C. Souyris. Imaging with synthetic aperture radar. Efpl Press, 2008. (Cité en page 10.)

[Mather 2011] P. M. Mather et M. Koch. Computer processing of remotely-sensed images : An introduction. Wiley, 4th édition, 2011. (Cité en pages 87 et 101.)

[McLachlan 2000] G. McLachlan et D. Peel. Finite mixture models. New York : John Wiley & Sons, 2000. (Cité en page 35.)

[Meischner 2010] P. Meischner. Weather radar : principles and advanced applications. Springer, 2010. (Cité en page 12.)

[Mercier 2008] G. Mercier, G. Moser et S. B. Serpico. *Conditional copulas for change detection in heterogeneous remote sensing images.* IEEE Trans. Geosci. Remote Sens., vol. 46, no. 5, pages 1428–1441, 2008. (Cité en page 48.)

[Mercier 2009] G. Mercier et P.-L. Frison. *Statistical characterization of the Sinclair matrix : application to polarimetric image segmentation.* In IEEE International Geoscience and Remote Sensing Symposium (IGARSS'09), volume 3, pages 717–720, 2009. (Cité en page 50.)

[Metwalli 2010] M. R. Metwalli, A. H . Nasr, O. S. Farag Allah, S. El-Rahaie et F. E. Abd El-Samie. *Satellite image fusion based on principal component analysis and high-pass filtering.* Journal of the Optical Society of America, vol. 27, no. 6, pages 1385–1394, 2010. (Cité en page 87.)

[Morales 2008] D. I. Morales, M. Moctezuma et F. Parmiggiani. *Detection of oil slicks in SAR images using hierarchical MRF.* In IEEE International Geoscience and Remote Sensing Symposium (IGARSS'08), volume 3, pages 1390–1393, 2008. (Cité en page 57.)

[Morrison 2006] R. L. Morrison et N. Do Minh. *Multichannel autofocus algorithm for synthetic aperture radar.* In IEEE International Conference on Image Processing (ICIP'06), pages 2341–2344, 2006. (Cité en page 15.)

[Moser 2006a] G. Moser, S. B. Serpico et J. Zerubia. *Dictionary-based Stochastic Expectation Maximization for SAR amplitude probability density function estimation.* IEEE Trans. Geosci. Remote Sens., vol. 44, no. 1, pages 188–199, 2006. (Cité en page 33.)

[Moser 2006b] G. Moser, S. B. Serpico et J. Zerubia. *SAR amplitude probability density function estimation based on a generalized Gaussian model.* IEEE Trans. Image Process., vol. 15, no. 6, pages 1429–1442, 2006. (Cité en pages 28 et 34.)

[Moser 2010] G. Moser et S. B. Serpico. *Contextual remote-sensing image classification by support vector machines and Markov random fields.* In IEEE International Geoscience and Remote Sensing Symposium (IGARSS'10), pages 3728–3731, 2010. (Cité en page 71.)

[Nelsen 2006] R. B. Nelsen. An introduction to copulas. Springer, New York, 2nd édition, 2006. (Cité en pages 24, 47, 48, 49, 50, 112 et 115.)

[Nicolas 2000] J.-M. Nicolas et A. Maruani. *Lower-order statistics : a new approach for probability density functions defined on R^+.* In Proceedings of the 10th European Signal Processing Conference (EUSIPCO'00), pages 805–808, 2000. (Cité en page 33.)

[Nicolas 2001] J.-M. Nicolas. *Filtrage homomorphique optimal d'images RSO.* In Acte de conférence du Groupe d'Études du Traitement du Signal et des Images (Gretsi'01), 2001. (Cité en page 118.)

[Noferini 2007] L. Noferini, M. Pieraccini, D. Mecatti, G. Macaluso, G. Luzi et C. Atzeni. *DEM by ground-based SAR interferometry.* IEEE Geosci. Remote Sens. Lett., vol. 4, no. 4, pages 659–663, 2007. (Cité en page 12.)

[Oliver 2004] C. Oliver et S. Quegan. Understanding Synthetic Aperture Radar images. SciTech Publishing, 2004. (Cité en pages 6, 7, 10, 13, 26, 27, 28 et 30.)

[Palubinskas 2011] G. Palubinskas et P. Reinartz. *A brief introduction to boosting.* In Joint Urban Remote Sensing Event (JURSE'11), pages 313–316, 2011. (Cité en page 87.)

[Parrilli 2012] S. Parrilli, M. Poderico, C. V. Angelino et L. Verdoliva. *A nonlocal SAR image denoising algorithm based on LLMMSE wavelet shrinkage.* IEEE Trans. Geosci. Remote Sens., vol. 50, no. 2, pages 606–616, 2012. (Cité en page 119.)

[Parzen 1962] E. Parzen. *On estimation of a probability density function and mode.* The Annals of Mathematical Statistics, vol. 33, no. 3, pages 1065–1076, 1962. (Cité en page 25.)

[Pellizzeri 2002] T. Macri Pellizzeri, C. J. Oliver, P. Lombardo et T. Bucciarelli. *Improved classification of SAR images by segmentation and fusion with optical images.* International Radar Conference, vol. 2002, no. CP490, pages 158–161, 2002. (Cité en page 46.)

[Peng 1995] A. Peng et W. Pieczynski. *Adaptive mixture estimation and unsupervised local bayesian image segmentation.* Graphical Models and Image Processing, vol. 57, no. 5, pages 389–399, 1995. (Cité en page 57.)

[Pentland 1984] A. P. Pentland. *Fractal-based description of natural scenes.* IEEE Trans. Pattern Anal. Mach. Intell., vol. 6, no. 6, pages 661–674, 1984. (Cité en page 40.)

[Pieczynski 2000] W. Pieczynski et A.-N. Tebbache. *Pairwise Markov random fields and segmentation of textured images.* Mach. Graph. Vision, vol. 9, no. 3, pages 705–718, 2000. (Cité en page 58.)

[Poggi 2005] G. Poggi, G. Scarpa et J. Zerubia. *Supervised segmentation of remote sensing images based on a tree-structured MRF model.* IEEE Trans. Geosci. Remote Sens., vol. 43, no. 8, pages 1901–1911, 2005. (Cité en page 56.)

[Pohl 1999] A. Pohl. *Tools and methods for fusion of images of different spatial resolution.* International Archives of Photogrammetry and Remote Sensing, vol. 32, 1999. (Cité en page 87.)

[Polidori 1998] L. Polidori. Cartographie radar. Macmillan Education, Australia, 1998. (Cité en pages 7 et 10.)

[Prata 1990] A. J. F. Prata, R. P. Cechet, I. J. Barton et D. T. Llewellyn-Jones. *The along track scanning radiometer for ERS-1-scan geometry and data simulation.* IEEE Trans. Geosci. Remote Sens., vol. 28, no. 1, pages 3–13, 1990. (Cité en page 12.)

[Qing 2004] X. Qing, Y. Jie et D. Siyi. Unsupervised multiscale image segmentation using wavelet domain hidden Markov tree. Lecture Notes in Computer Science. Springer Berlin - Heidelberg, 4th édition, 2004. (Cité en page 56.)

[Quelle 1993] H.-C. Quelle, Y. Delignon et A. Marzouki. *Unsupervised Bayesian segmentation of SAR images using the Pearson system distributions.* In IEEE International Geoscience and Remote Sensing Symposium (IGARSS'93), volume 3, pages 1538–1540, 1993. (Cité en page 29.)

[Raney 1971] R.K. Raney. *Synthetic aperture imaging radar and moving target.* IEEE Transactions on Aerospace and Electronic Systems, vol. AES-7, no. 3, pages 499–505, 1971. (Cité en page 15.)

[Reigber 2000] A. Reigber et A. Moreira. *First demonstration of airborne SAR tomography using multibaseline L-band data.* IEEE Trans. Geosci. Remote Sens., vol. 38, no. 5, pages 2142–2152, 2000. (Cité en page 12.)

[Richards 2006] J. A. Richards et X. Jia. Remote sensing digital image analysis : an introduction. Springer-Verlag, Berlin, 4th édition, 2006. (Cité en pages 7, 54 et 98.)

[Rignot 1991] E. Rignot et R. Chellappa. *Segmentation of synthetic-aperture-radar complex data.* J. Opt. Soc. Am. A, vol. 8, no. 9, pages 1499–1509, 1991. (Cité en page 55.)

[Rissanen 1978] J. Rissanen. *Modeling by shortest data description.* Automatica, vol. 14, no. 5, pages 465–471, 1978. (Cité en page 35.)

[Savu 2008] C. Savu et M. Trede. *Goodness-of-fit tests for parametric families of Archimedean copulas.* Quantitative Finance, vol. 8, no. 2, pages 109–116, 2008. (Cité en page 48.)

[Schapire 1999] R. E. Schapire. *A brief introduction to boosting.* In Proceedings of the 16th International Joint Conference on Artificial Intelligence, volume 16, pages 1401–1406, 1999. (Cité en page 86.)

[Schwarz 1978] G. Schwarz. *Estimating the dimension of a model.* Annals of Statistics, vol. 6, no. 2, pages 461–464, 1978. (Cité en page 35.)

[Serpico 2006] S. B. Serpico et G. Moser. *Weight parameter optimization by the Ho-Kashyap algorithm in MRF models for supervised image classification.* IEEE Trans. Geosci. Remote Sens., vol. 44, no. 12, pages 3695–3705, 2006. (Cité en pages 61 et 64.)

[Serpico 2012] S. B. Serpico, L. Bruzzone, G. Corsini, W. Emery, P. Gamba, A. Garzelli, G. Mercier, J. Zerubia, N. Acito, B. Aiazzi, F. Bovolo, F. Dell'Acqua, M. De Martino, M. Diani, V. Krylov, G. Lisini, C. Marin, G. Moser, A. Voisin et C. Zoppetti. *Development and validation of multitemporal image analysis methodologies for multirisk monitoring of critical structures and infrastructures.* In IEEE International Geoscience and Remote Sensing Symposium (IGARSS'12), pages 5506–5509, Munich, 2012. (Cité en page 25.)

[Shakhnarovish 2005] G. Shakhnarovish, T. Darrell et P. Indyk. Nearest-neighbor methods in learning and vision. MIT Press, 2005. (Cité en page 67.)

[Shao 1995] J. Shao et D. Tu. The jackknife and bootstrap. Springer-Verlag, 1995. (Cité en page 72.)

[Silveira 2009] M. Silveira et S. Heleno. *Separation between water and land in SAR images using region-based level sets.* IEEE Geosci. Remote Sens. Lett., vol. 6, no. 3, pages 471–475, 2009. (Cité en page 56.)

[Smits 1997] P. C. Smits et S. G. Dellepiane. *Synthetic aperture radar image segmentation by a detail preserving Markov random field approach.* IEEE Trans. Geosci. Remote Sens., vol. 35, no. 4, pages 844–857, 1997. (Cité en page 125.)

[Sneddon 1972] I. Sneddon. The use of integral transforms. McGraw-Hill, New York, 1972. (Cité en pages ix, 27, 28, 33 et 34.)

[Soergel 2010] U. Soergel. Radar remote sensing of urban areas (remote sensing and digital image processing). Springer, 1st édition, 2010. (Cité en page 12.)

[Storvik 2003] B. Storvik, G. Storvik et R. Fjortoft. *Joint distributions for correlated radar images.* In IEEE International Geoscience and Remote Sensing Symposium (IGARSS'03), volume 3, pages 2011–2013, 2003. (Cité en page 46.)

[Storvik 2005] B. Storvik, R. Fjortoft et A. H. S. Solberg. *A Bayesian approach to classification of multiresolution remote sensing data.* IEEE Trans. Geosci. Remote Sens., vol. 43, no. 3, pages 539–547, 2005. (Cité en page 85.)

[Storvik 2009] B. Storvik, G. Storvik et R. Fjortoft. *On the combination of multisensor data using meta-Gaussian distributions.* IEEE Trans. Geosci. Remote Sens., vol. 47, no. 7, pages 2372–2379, 2009. (Cité en pages 47 et 88.)

[Stratton 1941] A. Stratton. Electromagnetic theory. McGraw-Hill, New York, 1941. (Cité en page 9.)

[Strutt 1899] J. W. Strutt. *On the transmission of light through an atmosphere containing small particles in suspension, and on the origin of the blue of the sky.* Philosophical Magazine, vol. 47, no. 5, pages 375–394, 1899. (Cité en page 9.)

[Su 2004] F. Su, L. Ni, D. Li et H. Sun. *Classification of SAR image based on gray cooccurrence matrix and support vector machine.* In Proc. of the 7th International Conference on Signal Processing (ICSP'04), volume 2, pages 1385–1388, 2004. (Cité en page 55.)

[Sun 2007] Y. Sun, X. Li, H. Gong, W. Zhao et Z. Gong. *A study on optical and SAR data fusion for extracting flooded area.* In IEEE International Geoscience and Remote Sensing Symposium (IGARSS'07), pages 3086–3089, Juillet 2007. (Cité en page 88.)

[Theodoridis 2008] S. Theodoridis et K. Koutroumbas. Pattern recognition. Academic Press, 4th édition, 2008. (Cité en page 73.)

[Tison 2003a] C. Tison, J.-M. Nicolas et F. Tupin. *Accuracy of Fisher distributions and log-moment estimation to describe amplitude distributions of high resolution SAR images over urban areas.* In IEEE International Geoscience and Remote Sensing Symposium (IGARSS'03), volume 3, pages 1999–2001, Juillet 2003. (Cité en page 27.)

[Tison 2003b] C. Tison, F. Tupin et H. Maitre. *Retrieval of building shapes from shadows in high resolution SAR interferometric images.* In IEEE International Geoscience and Remote Sensing Symposium (IGARSS'04), volume 3, pages 1788–1791, 2003. (Cité en page 15.)

[Tison 2004] C. Tison, J.-M. Nicolas, F. Tupin et H. Maitre. *A new statistical model for Markovian classification of urban areas in high-resolution SAR images.* IEEE Trans. Geosci. Remote Sens., vol. 42, no. 10, pages 2046–2057, 2004. (Cité en pages 28, 29 et 33.)

[Tu 2004] T.-M. Tu, P. S. Huang, C.-L. Hung et C.-P. Chang. *A fast intensity-hue-saturation fusion technique with spectral adjustment for IKONOS imagery.* IEEE Geosci. Remote Sens. Lett., vol. 1, no. 4, pages 309–312, 2004. (Cité en page 101.)

[Tuia 2009] D. Tuia, F. Ratle, F. Pacifici, M. Kanevski et W. J. Emery. *Active learning methods for remote sensing image classification.* IEEE Trans. Geosci. Remote Sens., vol. 47, no. 7, pages 2218–2232, 2009. (Cité en page 58.)

[Tuia 2012] D. Tuia, J. Munoz-Mari et G. Camps-Valls. *Remote sensing image segmentation by active queries.* Pattern Recognition, vol. 45, pages 2180–2192, 2012. (Cité en page 58.)

[Ulaby 1989] F. T. Ulaby et C. Dobson. Handbook of radar scattering statistics for terrain. Artech House, 1989. (Cité en page 26.)

[Unal 2009] C. Unal. *Spectral polarimetric radar clutter suppression to enhance atmospheric echoes.* J. Atmos. Oceanic Technol., vol. 26, no. 9, pages 1781–1797, 2009. (Cité en page 15.)

[Vapnik 2000] V. Vapnik. The nature of statistical learning theory. N-Y : Springer-Verlag, 2nd édition, 2000. (Cité en pages 55, 68 et 105.)

[Vetterli 2010] M. Vetterli, J. Kovacevic et V. K. Goyal. Fourier and wavelet signal processing. Protected by copyright under the Attribution-NonCommercial-NoDerivs 3.0 Unported License from Creative Commons, 2010. http ://www.fourierandwavelets.org. (Cité en page 105.)

[Wahl 1996] D. E. Wahl, P. H. Eichel, D. C. Ghiglia, P. A. Thompson et C. V. Jakowatz. Spotlight-mode synthetic aperture radar : a signal processing approach. Springer, 1996. (Cité en page 17.)

[Wang 2005] Z. Wang, D. Ziou, C. Armenakis, D. Li et Q. Li. *A comparative analysis of image fusion methods.* IEEE Trans. Geosci. Remote Sens., vol. 43, no. 6, pages 1391–1402, 2005. (Cité en page 87.)

[Waske 2008] B. Waske et S. van der Linden. *Classifying multilevel imagery from SAR and optical sensors by decision fusion.* IEEE Trans. Geosci. Remote Sens., vol. 46, no. 5, pages 1457–1466, 2008. (Cité en page 86.)

[Weizman 2008] L. Weizman et J. Goldberger. *Detection of urban zones in satellite images using visual words.* In IEEE International Geoscience and Remote Sensing Symposium (IGARSS'08), pages 160–163, Juillet 2008. (Cité en page 56.)

[Wen 2009] C. Wen, Y. Zhang et K. Deng. *Urban area classification in high resolution SAR based on texture features.* In International Conference on Geospatial Solutions for Emergency Management, volume 28, pages 281–285, Beijing, China, 2009. (Cité en pages 37, 41 et 57.)

[Whittaker 1990] J. Whittaker. Graphical models in applied multivariate statistics. Wiley, 1990. (Cité en page 91.)

[Wolpert 1992] D. H. Wolpert. *Stacked generalization.* Neural Networks, vol. 5, no. 2, pages 241–259, 1992. (Cité en page 86.)

[Xia 1997] Z.-G. Xia et F. M. Henderson. *Understanding the relationships between radar response patterns and the bio- and geophysical parameters of urban areas.* IEEE Trans. Geosci. Remote Sens., vol. 35, no. 1, pages 93–101, 1997. (Cité en page 13.)

[Xu 1992] L. Xu, A. Krzyzak et C. Y. Suen. *Methods of combining multiple classifiers and their applications to handwriting recognition.* IEEE Trans. Syst., Man, Cybern., vol. 22, no. 3, pages 418–435, 1992. (Cité en page 86.)

[Xue 2003] X. R. Xue, Y. N. Zhang, R. C. Zhao, F. Duan et Y. Chen. *A new method of SAR image segmentation based on neural network.* In Proc. of the fifth International Conference on Computational Intelligence and Multimedia Applications (ICCIMA'03), pages 149–153, 2003. (Cité en page 55.)

[Yang 2006] Y. Yang, H. Sun et C. He. *Supervised SAR image MPM segmentation based on region-based hierarchical model.* IEEE Geosci. Remote Sens. Lett., vol. 3, no. 4, pages 517–521, 2006. (Cité en pages 120 et 127.)

[Yang 2009] W. Yang, D. Dai, B. Triggs et G.-S. Xia. *Semantic labeling of SAR images with hierarchical Markov aspect models.* Rapport de recherche HAL 00433600, INRIA, France, nov 2009. (Cité en page 56.)

[Yu 2003] Y. Yu et Q. Cheng. *MRF parameter estimation by an accelerated method.* Pattern Recognition Letters, vol. 24, no. 9-10, pages 1251–1259, 2003. (Cité en page 63.)

[Zerubia 1994] J. Zerubia, Z. Kato et M. Berthod. *Multi-temperature annealing : a new approach for the energy-minimization of hierarchical Markov random field models.* In Proceedings of the 12th International Conference on Pattern Recognition (IAPR'94), volume 1, pages 520–522, Octobre 1994. (Cité en page 91.)

[Zhang 1993] J. Zhang. *The mean field theory in EM procedures for blind Markov random field image restoration.* IEEE Trans. Image Process., vol. 2, no. 1, pages 27–40, 1993. (Cité en page 125.)

[Zhang 2005] H. Zhang, J. E. Fritts et S. A. Goldman. *A fast texture feature extraction method for region-based image segmentation.* In 16th Annual Symposium on Image and Video Communication and Processing, SPIE, 2005. (Cité en page 40.)

[Zhang 2009a] J.-G. Zhang, X.-B. Wen, X. Jiao et L. Wang. *Multiscale Markov random field method for SAR image segmentation.* In 2nd International Congress on Image and Signal Processing (CISP'09), pages 1–5, Octobre 2009. (Cité en page 57.)

[Zhang 2009b] Y. Zhang, F. Duan et R. Cui. *An edge-preserving wavelet denoising method based on MRF.* In 5th International Conference on Image and Graphics (ICIG'09), pages 67–71, 2009. (Cité en page 119.)

[Zhong 2007] P. Zhong, F. Liu et R. Wang. *A New MRF framework with dual adaptive contexts for image segmentation.* In International Conference on Computational Intelligence and Security, pages 351–355, Décembre 2007. (Cité en page 125.)

[Zhou 1998] J. Zhou, D. L. Civco et J. A. Silander. *A wavelet transform method to merge Landsat TM and SPOT panchromatic data.* Int. J. Remote Sens., vol. 19, no. 4, pages 743–757, 1998. (Cité en page 87.)

[Zhukov 1999] B. Zhukov, D. Oertel, F. Lanzl et G. Reinhackel. *Unmixing-based multisensor multiresolution image fusion.* IEEE Trans. Geosci. Remote Sens., vol. 37, no. 3, pages 1212–1226, 1999. (Cité en page 87.)

Résumé : La classification d'images de télédétection incluant des zones urbaines permet d'établir des cartes d'utilisation du sol et/ou de couverture du sol, ou de zones endommagées par des phénomènes naturels (tremblements de terre, inondations...). Les méthodes de classification développées au cours de cette thèse sont des méthodes supervisées fondées sur des modèles markoviens.
Une première approche a porté sur la classification d'images d'amplitudes issues de capteurs RSO (radar à synthèse d'ouverture) à simple polarisation et mono-résolution. La méthode choisie consiste à modéliser les statistiques de chacune des classes par des modèles de mélanges finis, puis à intégrer cette modélisation dans un champ de Markov. Afin d'améliorer la classification au niveau des zones urbaines, non seulement affectées par le bruit de chatoiement, mais aussi par l'hétérogénéité des matériaux qui s'y trouvent, nous avons extrait de l'image RSO un attribut de texture qui met en valeur les zones urbaines (typiquement, variance d'Haralick). Les statistiques de cette information texturelle sont combinées à celles de l'image initiale via des copules bivariées.
Par la suite, nous avons cherché à améliorer la méthode de classification par l'utilisation d'un modèle de Markov hiérarchique sur quad-arbre. Nous avons intégré, dans ce modèle, une mise à jour de l'a priori qui permet, en pratique, d'aboutir à des résultats moins sensibles bruit de chatoiement. Les données mono-résolution sont décomposées hiérarchiquement en ayant recours à des ondelettes. Le principal avantage d'un tel modèle est de pouvoir utiliser des images multi-résolution et/ou multi-capteur et de pouvoir les intégrer directement dans l'arbre. En particulier, nous avons travaillé sur des données optiques (type GeoEye) et RSO (type COSMO-SkyMed) recalées. Les statistiques à chacun des niveaux de l'arbre sont modélisées par des mélanges finis de lois normales pour les images optiques et de lois gamma généralisées pour les images RSO. Ces statistiques sont ensuite combinées via des copules multivariées et intégrées dans le modèle hiérarchique.
Les méthodes ont été testées et validées sur divers jeux de données mono-/multi-résolution RSO et/ou optiques.

Mots clés : Classification supervisée, Champ de Markov hiérarchique, Radar à synthèse d'ouverture, Modélisation statistique, Données multi-capteur

Supervised classification of high-resolution remote sensing images including urban areas by using Markovian models

Abstract : The classification of remote sensing images including urban areas is relevant in the context of the management of natural disasters (earthquakes, floodings...), and allows to determine land-use and establish land cover maps, or to localise damaged areas. The supervised classification methods developed during this PhD thesis are essentially based on Markovian models.
The first part of the study deals with the classification of single-polarized, mono-resolution synthetic aperture radar (SAR) amplitude images. The selected method consists in modeling the class-conditional statistics by resorting to finite mixture models, and then to plug these statistics into a Markov random field (MRF). To improve the classification results in urban areas, affected not only by speckle noise, but particularly difficult to process given their heterogeneity, we extracted a textural feature from the initial image which aims at discriminating the urban areas (e.g. Haralick's variance). The textural feature statistics are combined with those of SAR amplitude image by using bivariate copulas.
Next, we extended the single-scale model to a hierarchical Markov random field integrated in a quad-tree structure. This model includes a prior update that experimentally leads to results less affected by speckle. The mono-resolution data are hierarchically decomposed using a wavelet transform. The main advantage of this approach is that multi-resolution and/or multi-sensor acquisitions can directly be integrated in the tree. We applied the proposed hierarchical classification to SAR COSMO-SkyMed multi-resolution acquisitions, and to coregistered optical/SAR data. At each tree level, the statistics are independently modeled by finite mixtures of Gaussian distributions when considering optical images, and by finite mixtures of generalized Gamma distributions when considering the SAR amplitude data. Such statistics are then combined by using multivariate copulas, and plugged in the hierarchical model.
We tested and validated these methods on real SAR and optical data.

Keywords : Supervised classification, Hierarchical Markov random fields, Synthetic aperture radar, Statistical modeling, Multi-sensor data